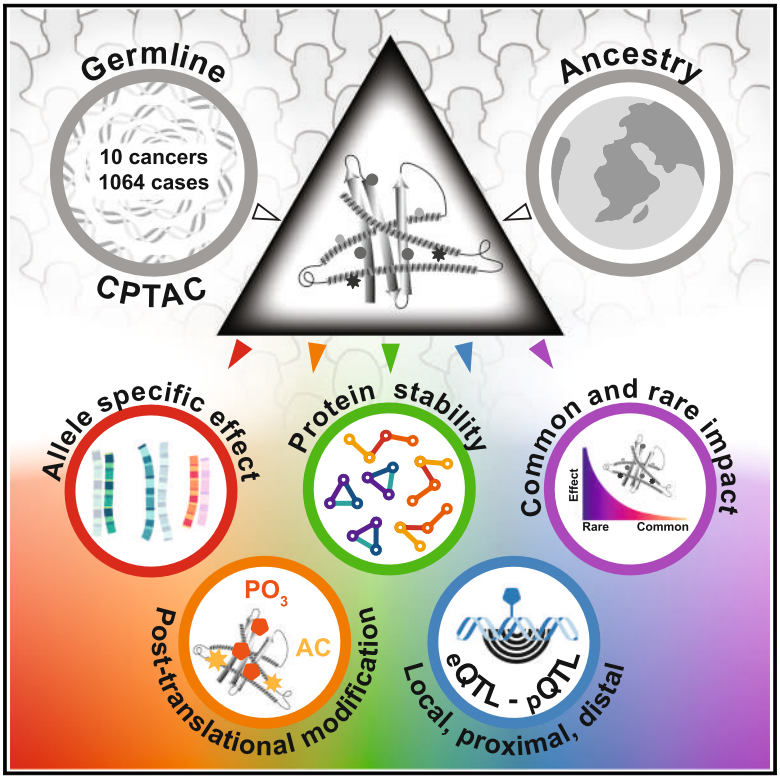
Cell

Precision proteogenomics reveals pan-cancer impact of germline variants

Graphical abstract



Highlights

- Precision peptidomics reveals germline variants' impact on cancer patient's proteomes
- Germline variants impact PTMs and protein stability and have allele-specific effects
- QTL analyses reveal genes and proteins under germline genetic control in cancer patients
- PRSs reveal cumulative impact of common germline variants on cancer proteomes

Authors

Fernanda Martins Rodrigues, Nadezhda V. Terekhanova, Kathleen J. Imbach, ..., Eduard Porta-Pardo, Li Ding, Clinical Proteomic Tumor Analysis Consortium

Correspondence

zeynep.gumus@mssm.edu (Z.H.G.), matthew.bailey@byu.edu (M.H.B.), gadgetz@broadinstitute.org (G.G.), eporta@carrerasresearch.org (E.P.-P.), lding@wustl.edu (L.D.)

In brief

Precision proteogenomics analysis of 1,064 cancer patients across ten cancer types unveils the ways in which rare and common germline variants shape the cancer proteome; the findings highlight the contribution of germline genetics in tumor heterogeneity and oncogenesis.

Martins Rodrigues et al., 2025, Cell *188*, 1–24
May 1, 2025
https://doi.org/10.1016/j.cell.2025.03.026

CellPress

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

Cell

CellPress
OPEN ACCESS

Article

Precision proteogenomics reveals pan-cancer impact of germline variants

Fernanda Martins Rodrigues,[1,2,3,31] Nadezhda V. Terekhanova,[1,2,3,31] Kathleen J. Imbach,[4,5,31] Karl R. Clauser,[6,31] Myvizhi Esai Selvan,[7,8,9,31] Isabel Mendizabal,[10,11,12,32] Yifat Geffen,[6,13,32] Yo Akiyama,[6,32] Myranda Maynard,[6] Tomer M. Yaron,[14] Yize Li,[1,2,3] Song Cao,[1,2,3] Erik P. Storrs,[1,2,3] Olivia S. Gonda,[15] Adrian Gaite-Reguero,[10] Akshay Govindan,[1,2,3] Emily A. Kawaler,[16] Matthew A. Wyczalkowski,[1,2,3] Robert J. Klein,[7] Berk Turhan,[7]

(Author list continued on next page)

[1]Department of Medicine, Washington University in St. Louis, Saint Louis, MO, USA
[2]McDonnell Genome Institute, Washington University in St. Louis, Saint Louis, MO, USA
[3]Department of Genetics, Washington University in St. Louis, St. Louis, MO 63110, USA
[4]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Spain
[5]Universitat Autonoma de Barcelona, Barcelona, Spain
[6]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[7]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[8]Center for Thoracic Oncology, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[9]Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[10]Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Derio, Spain
[11]Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[12]Translational Prostate Cancer Research Lab, CIC bioGUNE-Basurto, Biocruces Bizkaia Health Research Institute, Derio, Spain
[13]Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
[14]Meyer Cancer Center, Department of Medicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA
[15]Department of Biology, Brigham Young University, Salt Lake City, UT, USA
[16]Applied Bioinformatics Laboratories, New York University Langone Health, New York City, NY, USA
[17]Department of Pathology, University of Michigan, Ann Arbor, MI, USA
[18]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[19]Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY, USA
[20]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, USA

(Affiliations continued on next page)

SUMMARY

We investigate the impact of germline variants on cancer patients' proteomes, encompassing 1,064 individuals across 10 cancer types. We introduced an approach, "precision peptidomics," mapping 337,469 coding germline variants onto peptides from patients' mass spectrometry data, revealing their potential impact on post-translational modifications, protein stability, allele-specific expression, and protein structure by leveraging the relevant protein databases. We identified rare pathogenic and common germline variants in cancer genes potentially affecting proteomic features, including variants altering protein abundance and structure and variants in kinases (*ERBB2* and *MAP2K2*) impacting phosphorylation. Precision peptidome analysis predicted destabilizing events in signal-regulatory protein alpha (SIRPA) and glial fibrillary acid protein (GFAP), relevant to immunomodulation and glioblastoma diagnostics, respectively. Genome-wide association studies identified quantitative trait loci for gene expression and protein levels, spanning millions of SNPs and thousands of proteins. Polygenic risk scores correlated with distal effects from risk variants. Our findings emphasize the contribution of germline genetics to cancer heterogeneity and high-throughput precision peptidomics.

INTRODUCTION

The germline genome of each individual person has a unique combination of millions of genetic variants that influence virtually all biological processes throughout life, including cancer evolution. Many studies have demonstrated the critical importance of germline genomics, from cancer risk assessment to the development of tailored treatments.[1] The earliest germline genomics

CellPress
OPEN ACCESS

Cell
Article

Karsten Krug,[6] D.R. Mani,[6] Felipe da Veiga Leprevost,[17] Alexey I. Nesvizhskii,[17,18] Steven A. Carr,[6] David Fenyö,[19] Michael A. Gillette,[6] Antonio Colaprico,[20,21] Antonio Iavarone,[21,22] Ana I. Robles,[23] Kuan-lin Huang,[7,24] Chandan Kumar-Sinha,[17,25] François Aguet,[6] Alexander J. Lazar,[26] Lewis C. Cantley,[27] Urko M. Marigorta,[10,11] Zeynep H. Gümüş,[7,8,9,*] Matthew H. Bailey,[15,*] Gad Getz,[6,13,28,*] Eduard Porta-Pardo,[4,29,*] Li Ding,[1,2,3,30,33,*] and Clinical Proteomic Tumor Analysis Consortium

[21]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, USA
[22]Department of Neurological Surgery, Department of Biochemistry and Molecular Biology, University of Miami, Miller School of Medicine, Miami, FL, USA
[23]Office of Cancer Clinical Proteomics Research, National Cancer Institute, Rockville, MD, USA
[24]Center for Transformative Disease Modeling, Tisch Cancer Institute, Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[25]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA
[26]Departments of Pathology and Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[27]Dana Farber Cancer Institute, Boston, MA, USA
[28]Harvard Medical School, Boston, MA, USA
[29]Barcelona Supercomputing Center (BSC), Barcelona, Spain
[30]Siteman Cancer Center, Washington University in St. Louis, Saint Louis, MO, USA
[31]These authors contributed equally
[32]These authors contributed equally
[33]Lead contact
*Correspondence: zeynep.gumus@mssm.edu (Z.H.G.), matthew.bailey@byu.edu (M.H.B.), gadgetz@broadinstitute.org (G.G.), eporta@carrerasresearch.org (E.P.-P.), lding@wustl.edu (L.D.)
https://doi.org/10.1016/j.cell.2025.03.026

studies of cancer-prone families identified highly penetrant risk genes.[2–6] These targeted-gene and linkage studies were followed by array-based genome-wide association studies (GWASs), which identified many common variants (minor allele frequency [MAF] $\geq$ 1%) associated with tissue-specific[7,8] or pan-cancer risk.[9] While common risk variants typically have small effect sizes when seen individually, they discriminate individuals at high risk when combined as polygenic risk scores (PRSs).[10,11] Furthermore, many common germline variants regulate proximal and distal expression of genes in specific tissues and tumors with potentially additive effects.[12,13]

With advances in sequencing technologies, it has become feasible to identify rare and low-frequency variants (MAF < 1%) with moderate to high penetrance, associated with tissue-specific[14,15] or overall cancer risk,[16–18] mechanisms and pathways in tumor development,[19] tumor immune microenvironment,[20,21] mutational burden,[22,23] mutational signatures,[24,25] loss of heterozygosity (LOH),[19] and clinical variables such as age of cancer onset[16,22] and survival.[26] However, the impact of germline variants on the cancer proteome and post-translational modification (PTM) landscapes is poorly understood, specifically on oncogenic signaling pathways and their impact on cancer formation and evolution.

We analyzed the pan-cancer Clinical Proteomic Tumor Analysis Consortium (CPTAC) datasets from genomic, transcriptomic, proteomic, acetylomic, and phosphoproteomic analytes to generate precision proteogenomic profiles. These datasets provide a unique resource to study the impact of germline genomics on molecular oncogenic processes. Integrative multi-omic analyses revealed new putative pathogenic (P) rare germline variants in cancer predisposition genes (CPGs). Furthermore, common variants in these genes were associated with reduced levels of tumor suppressors in both primary tumors and normal adjacent tissues (NATs). Additionally, common germline variants at specific protein phosphorylation and acetylation sites influenced phosphorylation and acetylation levels or resulted in the emergence of new PTM sites. Our precision peptidomics data also identified allele-specific protein and PTM (ASP) effects and germline indels associated with protein stabilization, destabilization, or alternative products. Finally, whole-genome sequencing (WGS) and quantitative trait loci (QTL) analyses identified common variants affecting protein expression levels in normal and tumor tissues, impacting cancer-associated pathways. Our results highlight the power of integrative multi-omic approaches to illuminate the impact of germline variants across cancer phenotypes, revealing important biological insights into the role of germline genomics. These findings suggest that precision proteogenomics could inform patient risk stratification and prevention and interception approaches.

RESULTS

Precision peptidomic and PTM analysis of coding germline variants

CPTAC provides a proteogenomic dataset that includes common and rare germline variants across 10 cancer types. We processed and analyzed proteogenomic, clinical, and demographic data from 1,064 prospectively collected tumor and matching blood samples, including: whole-exome sequencing (WES), RNA sequencing (RNA-seq), proteome, and phosphoproteome data from all 10 cancer types; WGS from seven cancer types; and, acetylome data from six (Figure 1A). CPTAC also includes proteogenomic data from paired NATs from eight cancer types (n = 548/1,064 cases). All 1,064 matching blood samples passed WES quality control criteria and were used for germline variant calling, with average coverage ranging between 105× and 357× across target regions, and overall coverage of 25×–280× across a prioritized list of 160 CPGs (STAR Methods; Figures S1A and S1B; Table S1).

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

CellPress
OPEN ACCESS

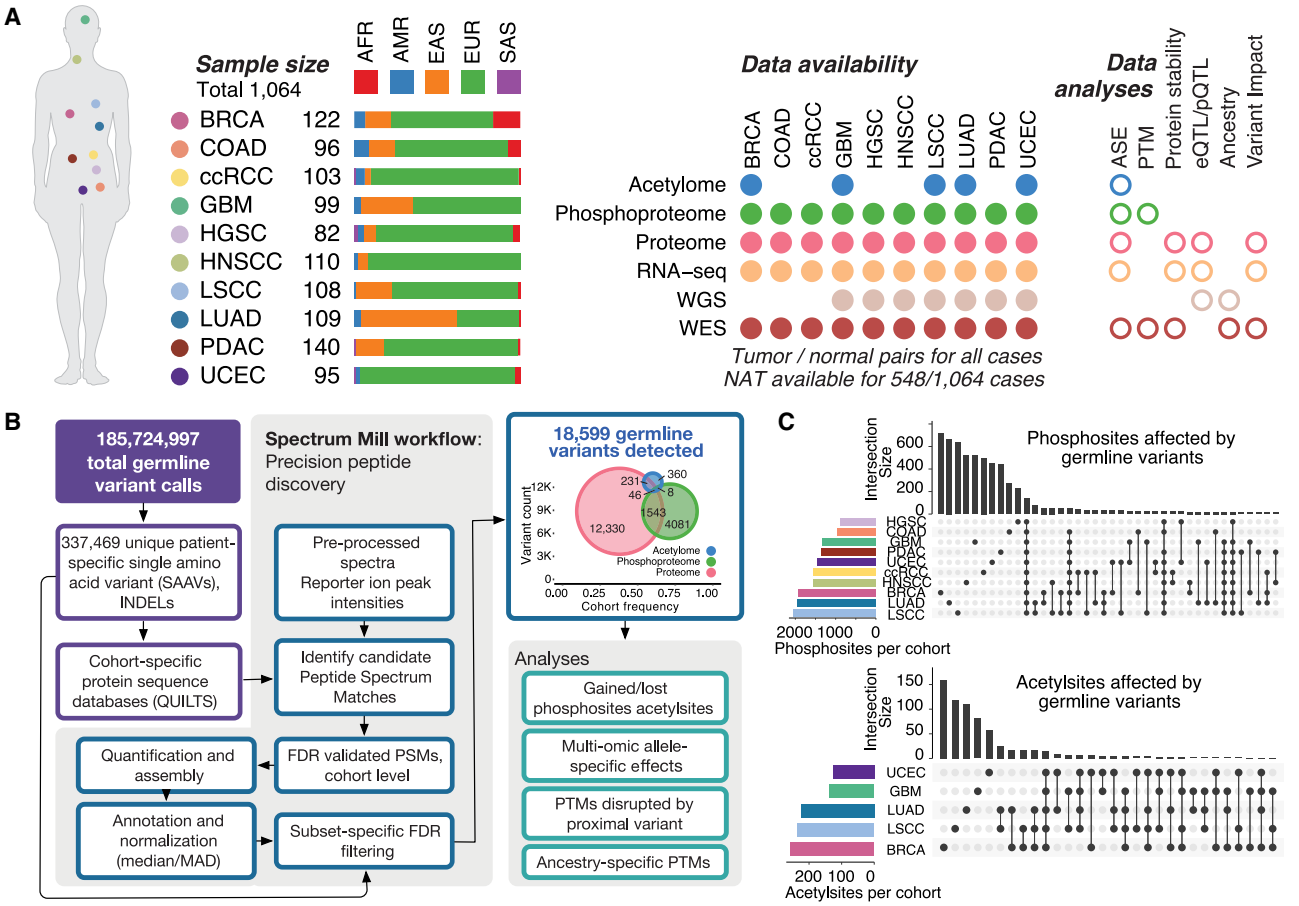


Figure 1. CPTAC dataset overview and precision peptidomics workflow

(A) The CPTAC cohort of 1,064 individuals of different genetic ancestries across 10 cancer types and available data types. Colors in top distribution represent genetic ancestry: African (AFR); admixed American (AMR); East Asian (EAS); European (EUR); South Asian (SAS).

(B) Our precision peptidomics workflow, representing the implementation of the Spectrum Mill workflow on the LC-MS/MS datasets to yield peptide spectrum matches (PSMs) that detected 18,599 germline variants in the proteome, phosphoproteome, and acetylome datasets.

(C) Overview of phosphorylation (upper) and acetylation (lower) sites affected by germline variants across cancer types based on the precision peptide data. Variants occur nearby or directly at the site, with 78% of phosphosites and 84% of acetylsites having germline variants located at 10 or fewer amino acids from the PTM site.

See also Figure S1.

A total of 185,724,997 germline variants were called from WES data (STAR Methods). Variants were filtered and annotated, resulting in 27,104,152 germline variant calls (563,036 unique variants) in exonic regions (~25,474 variants per sample). Individuals of African genetic ancestry (AFR) showed the highest average number of exonic germline variants per individual (30,510), with the lowest being for those of European (EUR) ancestry (25,205; Figure S1C). The germline exomes exhibited an average transition-transversion (TiTv) ratio of 2.74% and >99% concordance with dbSNP (Figure S1D). We derived ancestral (ANC) status information for 27,104,152 exonic germline variants (STAR Methods). Throughout this manuscript, we refer to individual variants in terms of ANC or derived (DER) alleles instead of major and minor alleles, respectively, according to their ANC status.

We also characterized the impact of non-coding variants on gene expression and protein abundance in 779 CPTAC samples from seven cancer types for which WGS data were available (STAR Methods). Given the low-pass nature of our WGS dataset, we phased and imputed the genotypes using GLIMPSE[27] using a set of high-quality variants from 2,504 unrelated samples from Phase 3 of the 1,000 Genomes Project, which were resequenced to high coverage by the New York Genome Center (NYGC).[28] For quality control, variants called from WGS and WES were compared for the seven cancer types (STAR Methods). Overall, 94.6% of variants overlapping between WES and WGS had the same genotypes in the same samples (Figure S1E; Table S1D).

WGS data was also used to confirm genetic ancestry predictions obtained from WES. Ancestry was first predicted from WES using a random forest classifier for all individuals (STAR Methods), while WGS was used to refine ancestry for 9 individuals of Slavic origin in the glioblastoma (GBM), head and neck squamous cell carcinoma (HNSCC), lung squamous cell carcinoma (LSCC), pancreatic ductal adenocarcinoma (PDAC), and

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

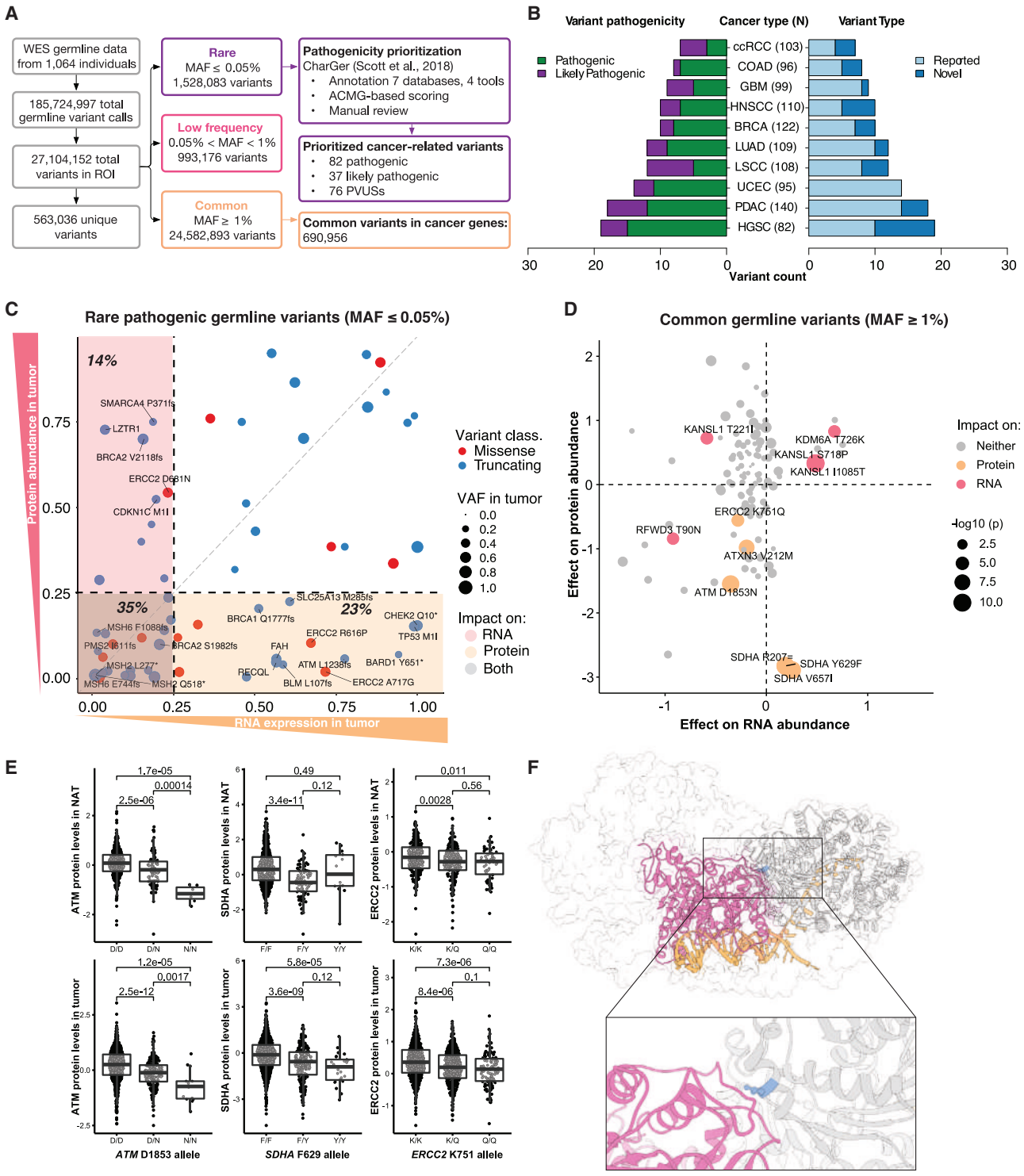


Figure 2. Impact of rare pathogenic and common germline variants on gene and protein expression

(A) Schematic of filtering and classification of germline variants. Purple boxes describe the prioritization procedure for rare variants; yellow boxes show processing for common variants.

(B) (Left) Distribution of rare pathogenic/likely pathogenic (P/LP) variants across 10 cancer types. (Right) Distribution of variants previously reported in any of the TCGA, gnomAD, and UKBB datasets (light blue) or novel to this study (dark blue).

(C) Gene expression (x axis) and protein abundance (y axis) quantiles of proteins from P/LP variant carriers. Yellow and pink shading indicates variants with impact on protein levels and gene expression, respectively; gray denotes variants with effect in both.

(legend continued on next page)

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

uterine corpus endometrial carcinoma (UCEC) cohorts which were misclassified in the WES-based predictions, but correctly classified as EUR using WGS (Figure S1F; STAR Methods).

Next, we combined proteomics and genomics datasets to create protein sequence databases for each individual using the proteogenomic integration tool Quantitative Integrated Library of Translated SNPs/Splicing (QUILTS)[29] (STAR Methods). From the total of 185,724,997 germline variants from WES, we incorporated 337,469 unique patient-specific coding variants that mapped to Gencode v34 reference protein sequences (Figure 1B). Using these databases for each cancer cohort, the proteomics liquid chromatography-tandem mass spectrometry (LC-MS/MS) datasets were searched with the Spectrum Mill workflow (STAR Methods) to yield peptide spectrum matches (PSMs) that matched peptides from the reference proteome, germline variants, or somatic mutations. We detected peptides for 18,599 unique coding germline variants in the proteome, phosphoproteome, and acetylome, with the majority having low frequencies at the cohort level. Among the variants detected, 1,828 were in more than one dataset, while the majority were in only one: 12,330 in the proteome, 4,081 in the phosphoproteome, and 360 in the acetylome (Figure 1B). Scrutiny of the location of these variants revealed 8,046 PTM sites (7,353 phosphosites and 693 acetylation sites) affected by germline variants, 150 of which were detected across all cancers and 5,459 detected in a single cancer (Figure 1C). The pattern of high cancer-type specificity of PTM sites is consistent with our previous study[30] and suggests a role for PTMs in tissue/cell-specific regulation and signaling.

Looking at the peptide-length distribution for reference and alternative alleles (Figure S1G), there is a tendency for peptides carrying the alternative allele to be longer than all reference proteome-derived peptides, regardless of the specific proteomics dataset (protein, phosphorylation, or acetylation). While the higher minimum-score thresholds employed in the subset specific false discovery rate (ssFDR) filtering of the proteome dataset to maintain suitable false discovery rate (FDR) levels will bias against shorter peptides, a random variant in a protein is more likely to occur in a peptide that spans a longer proportion of that protein.

Proteogenomic modeling of rare pathogenic and common germline variants

Germline variants associated with cancer likely have different P mechanisms depending on allele frequency (AF): rare P variants are oftentimes more damaging to protein function than common variants.[31,32] Here, we investigated the landscape of rare P and common germline variants in the CPTAC cohort, leveraging multi-omics information from tumor and matching NAT samples.

From 27,104,152 total exonic germline variants, a minority of them were rare (1,528,083 variants; gnomAD AF $\leq$ 0.05%), followed by low frequency (993,176; 0.05% < gnomAD AF < 1%), and common variants (24,582,893, gnomAD AF $\geq$ 1%; Figure 2A). These proportions are similar to other large-scale databases of population genomics, such as UK Biobank (UKBB).

Considering that rare P germline variants play important roles in cancer susceptibility,[16,18] we aimed to identify such events using CharGer[33] (STAR Methods; Figures 2A and S2). We identified 119 P and likely pathogenic (LP) variants across CPGs (Table S1) affecting 115 individuals (10.8% of the cohort; Figure 2B). The majority of P/LP variants likely represent loss-of-function events (i.e., nonsense, frameshift, start-loss, and splice-site variants; 75%, n = 89), with the remaining being missense variants predicted to be deleterious (n = 30; Table S2). These variants were also observed in other cohorts (The Cancer Genome Atlas [TCGA],[16] gnomAD, and UKBB) at extremely low frequencies (mean gnomAD AF = 0.0001, and mean UKBB AF = 0.0002). Furthermore, 34 variants (29%) were private to the CPTAC cohort (Figure 2B). We also observed that carriers were younger at diagnosis compared with non-carriers for the breast cancer (BRCA), colorectal adenocarcinoma (COAD), and clear cell renal cell carcinoma (ccRCC) cohorts (Figure S2A).

To evaluate the impact of germline variants in a somatic context, we investigated LOH events using allele fractions from tumor-normal data to identify variants positively selected in the tumor based on the two-hit hypothesis[17,34,35] (STAR Methods). From 119 P/LP variants, we observed 21 (17.6%) and 11 (9.2%) variants undergoing significant (FDR $\leq$ 0.05) and suggestive (0.05 < FDR $\leq$ 0.15) LOH in the tumor, respectively (Figure S2C). For 15 of 21 (71%) significant LOH, we observed deletion of the respective gene detected by the tool Genomic Identification of Significant Targets in Cancer, version 2 (GISTIC2).[36] Also, 6 (5%) of 119 P/LP variants co-occurred with non-silent somatic mutations in the same gene.

Next, we explored the molecular consequences of these 119 P/LP variants using protein and RNA expression data, focusing on 65 P/LP variants for which both RNA and protein levels were available. Consistent with a loss-of-function phenotype, P/LP variant carriers displayed lower RNA expression and protein levels (within-cancer-type quantile means of 0.36 and 0.29, respectively, compared with 0.5 for the entire cohort; Figures 2C and S2D). This was observed for variants affecting members of the mismatch repair (MMR) pathway (*PMS2*, *MSH2*, and *MSH6*) associated with low RNA expression and protein levels (expression quantiles < 0.25). We observed that 4 of 5 carriers of P/LP variants in those genes (*MSH2* L277*, *MSH6* E744fs, *MSH2* Q518*, and *PMS2* I611fs) were also identified as microsatellite instability (MSI)-high samples (Table S1 from Li et al.[37]), consistent with the fact that carriers of P/LP variants

(D) Effects of common germline variants in cancer genes in their protein abundance (y axis) and RNA expression (x axis). Effect is calculated as the slope in the regression model. Dot size reflects the $-\log_{10}$ of the FDR adjusted p values from the regression model.
(E) Protein levels (y axis) in NAT or tumor samples for ATM, SDHA, and ERCC2 according to genotype (x axis). p values from pairwise Wilcoxon tests between genotype groups are provided, and data are represented as median and interquartile range.
(F) Mapping of the ERCC2 K751 position on the PDB: 6RO4. Blue represents the residue, gray denotes ERCC2, pink represents ERCC3, orange highlights the DNA molecule.
See also Figure S2.

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

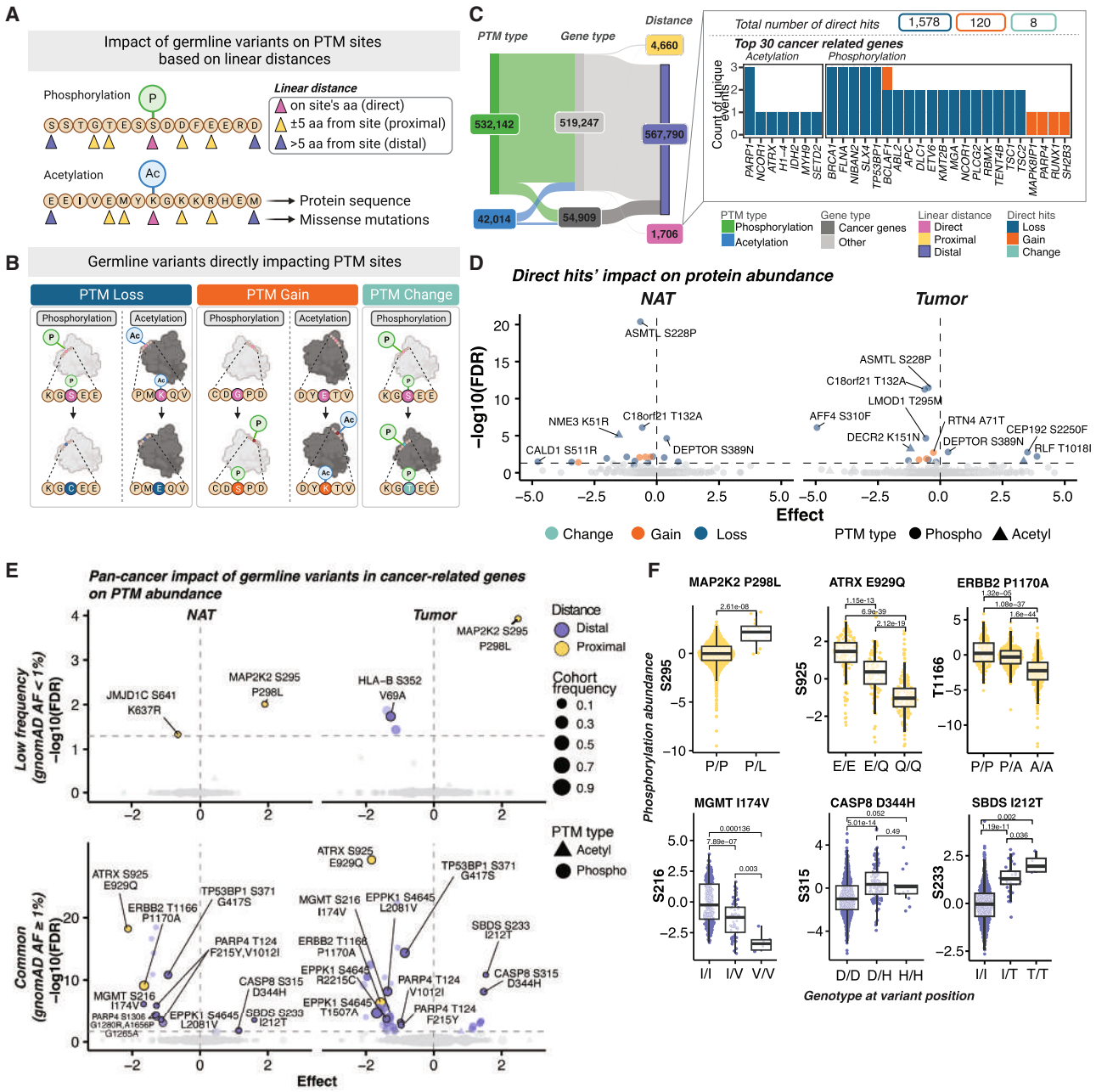


Figure 3. Impact of missense germline variants on PTM sites based on linear distances

(A) Depiction of how missense variants may impact PTM sites based on linear distances: direct hit (colocalizes with PTM site); proximal (within 5 amino acids); or distal (located beyond 5 amino acids). Created in Biorender.

(B) Direct hits are classified based on their consequence: loss, change, and gain. Created in Biorender.

(C) Distribution of direct, proximal, and distal events detected in CPTAC, beside a bar plot summarizing the distribution of direct hits across the top 30 cancer-related genes.

(D) Significance (y axis) and effect (x axis) of direct-hit events on global protein levels in NATs (left) and tumor samples (right) from linear model results. Points are colored by variant consequence and shaped according to PTM type. Effect (x axis) is calculated as the slope in the regression model, and y axis reflects the $-\log_{10}$ of the FDR adjusted *p* values from model.

(E) Significance (y axis) and effect (x axis) of proximal or distal variants in cancer-related genes on phosphorylation and acetylation levels of their corresponding PTM sites in NATs (left) and tumors (right). (Top and bottom) Results from rare/low frequency (gnomAD AF < 1%) and common variants (gnomAD AF ≥ 1%), respectively. Colors represent variant distance to the PTM sites; shapes represent PTM type; and sizes represent the frequency of the event in the CPTAC cohort (pan-cancer level). Only events for which protein abundance differences were not observed are labeled. Events in *HLA-A* and *HLA-B* were removed from common

(legend continued on next page)

in core MMR pathway genes tend to develop an MSI cancer phenotype.[38] Most variants had comparable quantiles of both gene and protein expression (rho = 0.49, $p = 3.08 \times 10^{-5}$; Figure 2C). Interestingly, we also observed outliers, including *TP53* M1I, *ERCC2* A717G, and *ATM* L1283fs, which were associated with high RNA expression but low protein abundance of the respective genes, highlighting the importance of proteomics to assess the functional impact of variants.

Next, we explored the potential effects of common germline variants (gnomAD AF $\geq$ 1%) in our list of 160 CPGs and 299 cancer driver genes[37,39] (Figure 2D). We observed variants in *ATM*, *SDHA*, and *ERCC2* with no detectable effect on RNA expression, but lower protein levels in carriers of the DER alleles in tumor and matched NAT samples (Figure 2E). *ERCC2* K751Q has been associated with lower DNA-repair activity *in vitro* and better outcomes in patients treated with chemotherapy,[40,41] consistent with the DER allele lowering DNA-repair efficiency. A structural alignment of the AlphaFold2 model for ERCC2 (Protein Data Bank [PDB]: 6RO4) suggests that K751 could sit at the binding interface between ERCC2 and ERCC3 (Figure 2F). This, together with previous *in vitro* and clinical data, and lower protein levels, suggests that the DER allele may damage the stability of the complex. Further experiments are needed to validate this hypothesis. In conclusion, the overall lower protein levels for core proteins of the DNA-repair machinery suggest that, even if these are common variants and with no detectable effects at the RNA level, they could potentially have important clinical impacts.

Direct, proximal, and distal effects of germline variants on PTM sites

Germline variants may mediate cancer risk through dysregulation of signaling pathways.[42,43] For example, variants might change a PTM site to abrogate its ability to become phosphorylated or acetylated[44–48] or alter the motifs that make it recognizable by enzymes, making it more or less likely to become modified. We explored the impact of rare/low frequency (gnomAD AF < 1%) and common (gnomAD AF $\geq$ 1%) missense variants co-localizing, proximal (within 5 amino acids), or distal (beyond 5 amino acids) to PTM sites at the linear distance (Figure 3A). For germline variants directly overlapping a PTM site, three scenarios were assessed: (1) loss of a PTM site; (2) creation of a new site; or, in the case of phosphorylations, (3) change of the substrate, e.g., serine to threonine (Figure 3B). To focus on protein-coding variants, we evaluated missense germline variants from WES to identify reference peptides in the (phopsho/acetyl)proteomics datasets with the matching amino acids for both alleles across the entire cohort (STAR Methods). We observed 532,142 proximal, distal, and direct-hit events involving single phosphorylation sites and 42,014 events involving acetylation sites. Of these, 1,706 variants directly overlapped a site, 4,660 were proximal, and 567,790 were distal to a site on the same protein (Figure 3C; Table S3). Most PTM-related genetic variants (92.6%) were associated with phosphorylation rather than acetylation sites, reflecting the higher abundance of phosphorylation PTMs in our dataset (Figures 1A and 3C).

Regarding variants overlapping a PTM site, PTM losses were the most frequent events: 1,578 losses detected across all proteins, compared to 120 gain and 8 changes (Figure 3C). Of these, we observe 115 loss and 5 gain events across the lists of 160 CPGs (Table S1), 299 cancer driver genes,[37,39] and 624 other cancer genes[17] including *ATRX*, *BRCA1*, *TP53BP1*, and *PARP4* (Figure 3C; Table S3). Samples with variants affecting PTMs in these proteins displayed differences in protein abundance compared with those with reference alleles (STAR Methods). Specifically, 16 proteins with variants located at a PTM site exhibited significant dysregulation in NATs, of which 14 were also observed in tumors (generalized linear model [GLM] FDR $\leq$ 0.05; Figure 3D; Table S3). For example, we noted a small but statistically significant increase in the level of DEP containing MTOR interacting protein (DEPTOR) in the presence of the S389N phosphosite loss allele. DEPTOR is associated with suppression of the mechanistic target of rapamycyin kinase (mTOR) complexes 1/2 (mTORC1/2),[49] and the S389N variant (rs4871827, gnomAD AF = 0.33) is at the interface between DEPTOR and mTOR.[50] To understand whether this variant has broader mTOR pathway effects, we tested for changes in proteins or phosphoproteins in pathway members between variant carriers and non-carriers (STAR Methods), as even modest changes in protein abundance may elicit downstream effects (Table S3). We found a slight decrease in MAP2K2 T25 phosphorylation levels in HNSCC (GLM FDR = 0.0163; Wilcoxon FDR = 0.00097 between non-carriers and heterozygous (HET) individuals; Figure S3A). In PDAC, EIF4EBP1 showed decreased phosphorylation at S83/S101 and T36/T37 (GLM FDR = 0.02 and 0.027, respectively). Moreover, patients homozygous for the DER allele of *DEPTOR* S389N showed the lowest phosphorylation levels at both EIF4EBP1 sites (Wilcoxon FDR = 0.018 and 0.036, respectively; Figures S3B and S3C; Table S3). The T37 site in EIF4EBP1 is involved in hyperphosphorylation-dependent disruption of eIF4E binding.[51] Beyond *DEPTOR*, several other PTM-overlapping variants showed associations with phosphorylation levels of pathway members, including *ERBB2* P1170A on PAK1 S220s/T225t phosphorylation in the pan-cancer cohort (GLM FDR = 0.043; Figure S3D), *HLA-B* V69A on HSP90AA1 S763s phosphorylation in BRCA (GLM FDR = 0.005; Figure S3E), and *CASP8* D344H on SEPTIN4 S605s phosphorylation in GBM (GLM FDR = 0.048; Figure S3F).

Next, we quantified the association of proximal or distal variants with phosphorylation/acetylation abundance differences on reference peptides at the pan-cancer level (STAR Methods). For rare/low-frequency variants, to increase statistical power, we collapsed all individuals harboring a proximal or distal variant into a single variable at the gene level (STAR Methods). We identified 46 variants associated with phosphorylation and

variant results (bottom) (see Table S3D for complete list of tested events and Table S3E for protein abundance differences results). Effect (x axis) is calculated as the slope in the regression model, and y axis reflects the $-\log_{10}$ of the FDR adjusted *p* values from the model.
(F) PTM levels according to patient genotype status for variants proximal (top) and distal (bottom) to the sites. FDR adjusted *p* values from pairwise Wilcoxon tests between genotype groups are provided, and data are represented as median and interquartile range.
See also Figure S3.

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

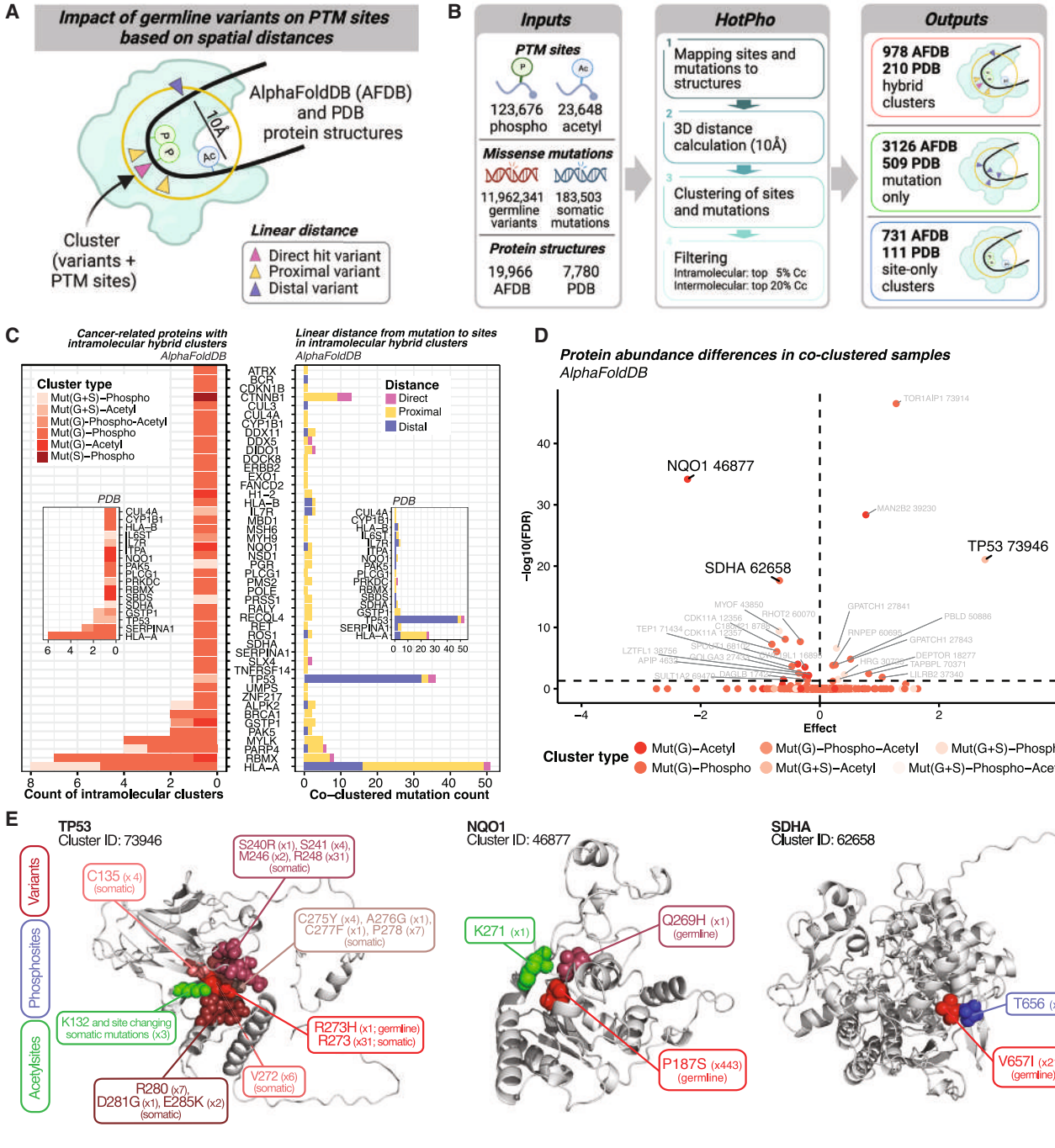


Figure 4. Spatially interacting missense germline variants, somatic mutations, and PTM sites

(A) Depiction of how missense germline variants and somatic mutations may interact with PTM sites based on spatial distances, showing an overview of HotPho analyses, which map input mutations and PTM sites onto protein structures. Created in Biorender.

(B) HotPho pipeline. Created in Biorender.

(C) (Left) Number of intramolecular hybrid clusters in cancer-related proteins detected in AFDB and PDB (inner). (Right) Number of germline variants and somatic mutations in each hybrid cluster that are directly overlapping a PTM site in the same cluster at a linear distance (direct), within 5 amino acids (proximal), or beyond 5 amino acids (distal).

(D) Protein level differences in samples involved in hybrid clusters detected in AlphaFoldDB structures vs. not. Dots represent a cluster, where color depicts its type based on involved events. AFDB cluster ID is shown beside protein names. Effect (x axis) is calculated as the slope in the regression model, and y axis reflects the $-\log_{10}$ of the FDR adjusted p values from the model.

(legend continued on next page)

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

acetylation changes at proximal or distal sites, of which 9 are in cancer genes (Figure 3E; Table S3D), including *MAP2K2* P298L (rs200371894, gnomAD AF = 9.2e−4) associated with higher phosphorylation at the S295 phosphosite in tumor and NATs (Figures 3E and 3F).

For common germline variants, we then analyzed each variant individually, detecting 815 common proximal or distal variants associated with phosphorylation and acetylation-level differences in carriers vs. non-carriers, with 78 events in cancer genes (Figure 3E; Table S3D). One notable proximal event is variant *ATRX* E929Q (rs3088074, gnomAD AF = 0.3747), which is associated with lower phosphorylation levels at the S925 site in both tumor and NAT samples at the pan-cancer level (Figure 3F). This event was also significant in ccRCC, GBM, PDAC, and UCEC (Table S3D). This variant was recently reported as the most frequent mutation in a Lebanese GBM cohort.[52] A similar effect was observed for the T1166 phosphosite in ERBB2 proximal to the P1170A variant (rs1058808, gnomAD AF = 0.60; Figure 3F). This variant has had conflicting interpretations, with some studies reporting a lack of association with increased cancer risk,[53–58] but others reporting associations with increased expression of HER2 protein in BRCA,[59] risk of lung cancer in a Korean population,[60] gastric cancers,[61,62] epithelial ovarian cancer,[63] and of cervical cancer.[64]

Focusing on distal events in cancer genes, we observe the *MGMT* I174V variant associated with decreased phosphorylation at S216 (Figure 3F). *MGMT* plays a role in DNA damage repair. Epigenetic alterations of this gene are reported in GBM, colorectal, gastric cancers,[65] and its loss is associated with increased melanoma risk.[66] This variant, however, has been suggested as benign in melanoma.[67] Additionally, the role of S216 phosphosite is unknown. The degree of S315 phosphorylation in CASP8 exhibited a similar pattern relative to the distal D344H variant (rs1045485, gnomAD AF = 0.091; Figure 3F), which has been associated with a reduced risk for breast and prostate cancer.[68,69] In contrast, patients with the *SBDS* I212T variant (rs79344818, gnomAD AF = 0.025) demonstrated higher phosphorylation at S233 compared with those with the ANC allele (Figure 3F). No significant acetylation differences due to proximal or distal events were observed in cancer genes. While no global protein abundance changes were associated with PTM-proximal events in cancer-related genes, 52 distal events indicated a significant protein abundance disparity including variant D1853N in *ATM*, variants Y629F and V657I in *SDHA*, and P187S in *NQO1* (Table S3E).

Finally, we identified several cancer-related pathways enriched in proteins harboring PTM-proximal events (STAR Methods), including oxidative phosphorylation and p53 pathways (Figure S3G), DNA damage-repair pathways, and pathways specific to GBM, endometrial cancer, BRCA, and head and neck cancer (Figure S3H). For PTM-distal events, we see similar enrichment for DNA damage-repair pathways, Wnt beta-catenin, and the phosphatidylinositol 3-kinase (PI3K)/Akt/mTOR signaling[70] (Figure S3H).

Our findings suggest that germline variants are not solely associated with protein levels but may mediate PTM on particular protein residues due to their changes in amino acid context, affecting oncogenic signaling pathways.

Spatially interacting germline variants, somatic mutations, and PTM sites

Protein structures provide insights into the functional consequences of genetic changes, as variants in close proximity in a protein's three-dimensional (3D) space tend to be associated with similar phenotypes.[71–74] Structure prediction algorithms, such as AlphaFold2[75,76] provide 3D models of all human proteins, allowing us to explore the clustering of somatic mutations, germline variants, and PTM sites in 3D space across the entire proteome.

We used HotSpot3D[72] and HotPho[73] to evaluate co-clustering events using 7,780 experimental human protein structures from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB)[77,78] and 19,966 human protein models from the AlphaFold Protein Structure Database (AFDB) v4,[75,76,79] as well as a total of 123,676 phosphosites, 23,648 acetylsites, 183,503 missense somatic mutations, and 11,962,341 missense germline variants (Figure 4A; STAR Methods). We focused on intramolecular clusters among the top 5% cluster closeness (Cc) score, a metric that evaluates the enrichment in studied features (i.e., genetic variants and PTM sites). We found 210 hybrid, 509 mutation-only, and 111 site-only clusters from PDB and 978 hybrid, 3,126 mutation-only, and 731 site-only clusters from AFDB (Figures 4B and S4A; Table S4; STAR Methods).

Overall, results from PDB and AFDB structures are in agreement (Figures S4B and S4C), with 56.9% of clusters identified in PDB also detected in AFDB (9.1% of all clusters detected in AFDB and PDB combined). Given the larger number of AFDB structures, more clusters were observed for that database. The proteins with the most clusters are highly enriched in human leukocyte antigen (HLA) molecules in AFDB and PDB (HLA-DRB5, HLA-DRB1, HLA-A, HLA-B, HLA-DQA1, and HLA-C; Figure S4B), likely due to the high germline and somatic variability of HLA genes.[80–83] We observed an enrichment of clusters in cancer-related genes (Table S4) at the top 5% Cc score compared with other genes (Figure S4A, bottom). Most hybrid clusters in these proteins involve at least one missense germline variant and phosphosite, with a few involving acetylsites (Figure 4C), including clusters in TP53, RET, BRCA1, PMS2, POLE, SDHA, IL7R, RBMX, ERBB2, and CTNNB1. CTNNB1 has a cluster associated with endometrial carcinoma involving two phosphosites (S33 and S29) and a few recurrent somatic mutations at positions D32 and G34.[84]

(E) AlphaFoldDB protein structures depicting detected clusters, rendered with Pymol. From left to right: TP53 cluster 73,946 (AFDB: AF-P04637-F1); NQO1 cluster 46,877 (AFDB: AF-P15559-F1); SDHA cluster 62,658 (AFDB: AF-P31040-F1). Green spheres represent acetylsites; blue represents phosphosites; 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 notes one individual).
See also Figure S4.

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

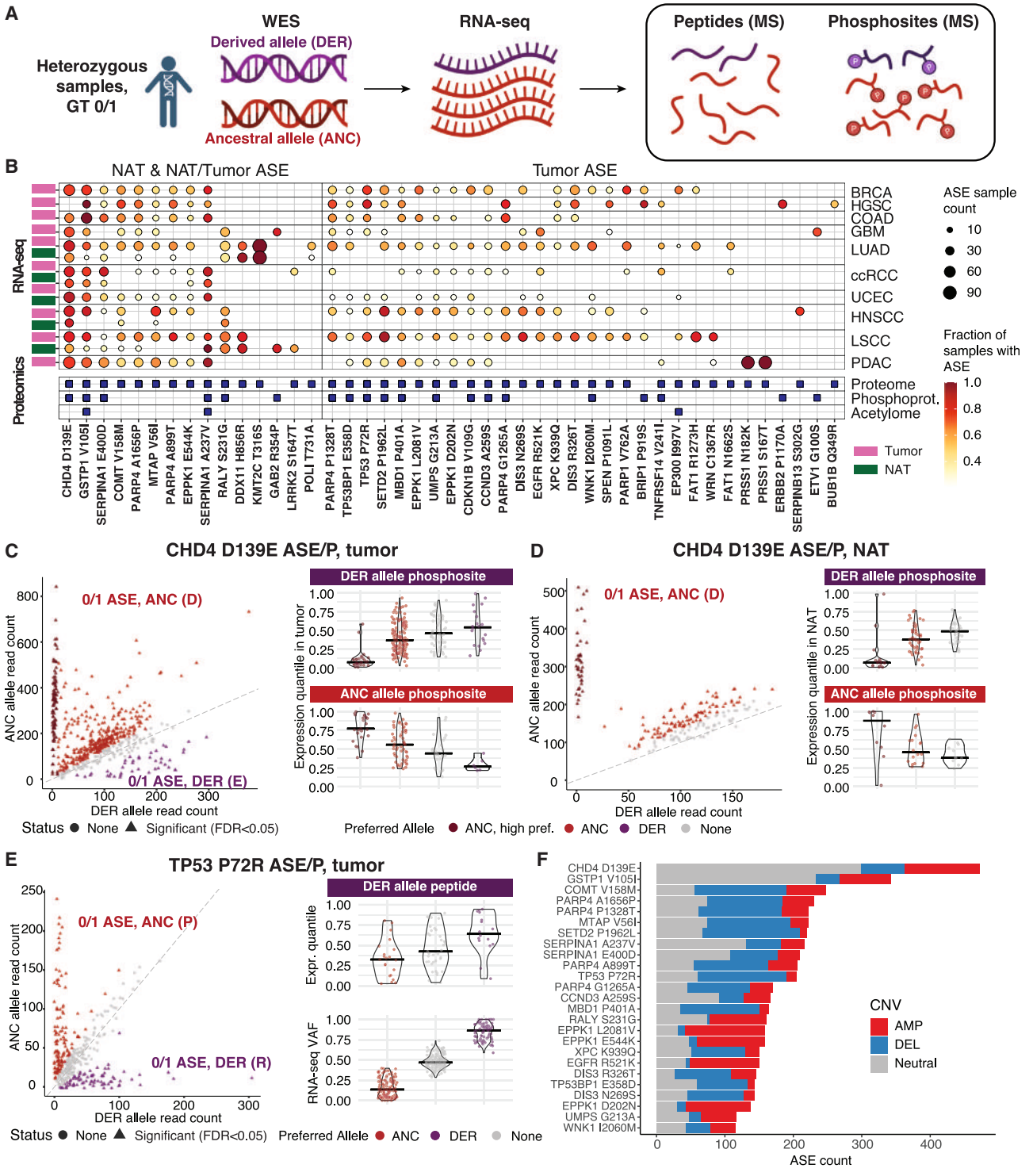


Figure 5. Allele-specific effects on gene, protein, and PTM expression

(A) Schematic of data acquisition from HET patients for allele-specific expression (ASE) analyses (GT, genotype; WES, whole-exome sequencing; MS, mass spectrometry).

(B) Top ASE events in tumor and NAT samples across 10 cohorts with frequency >50%. Dot color denotes frequency, while size represents the number of ASE events in the respective cohort. The y axis represents data type (RNA or proteomics), and in the case of RNA, the cancer and sample type (tumor or NAT) from which the results were derived.

(legend continued on next page)

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

Most hybrid clusters involve germline variants linearly proximal to PTM sites. However, our analysis also identified variants distal from the sites, but close in the 3D protein structure (Figure 4C). We found 155 germline variants overlapping PTM sites within clusters, 1,136 proximal variants, and 458 co-clustered with distal sites. Of these, a few distal examples are found in the top clusters of cancer proteins (Figures 4C and S4D; Table S4), including a TP53 cluster (Figure 4E) involving the K132 acetylsite, some somatic mutations (K132N, K132R) at this site that lead to its loss, and distal germline variants like the rare P R273H that affects TP53-DNA binding, promoting cancer cell survival.[85–92] This gain-of-function variant also increases TP53 levels in samples with co-clustering events at the pan-cancer level (FDR = 8.75e−22; Figure 4D; Table S4C). Higher TP53 levels are also observed in BRCA, COAD, and GBM samples with variants in this cluster (Figure S4E; Table S4D).

Another example is a NQO1 cluster (Figure 4E) involving a common P germline variant (P187S), and the acetylsite K271. NQO1 is involved in detoxification of carcinogens whose dysregulation is associated with many cancers.[93–103] The presence of P187S can significantly reduce NQO1 activity by lowering flavin adenin dinucleotide (FAD) affinity and impairing protein stability, increasing cancer risk.[104,105] NQO1 stability is linked to its C-terminal domain, where K271 is located. P187S accelerates ubiquitin-dependent proteasomal degradation driven by K271 and other sites, affecting both polymorphic and wild-type forms of the protein.[106–111] K271 acetylation also disrupts FAD binding,[112–114] reducing protein activity. Our analyses revealed lower NQO1 abundance in co-clustering samples at the pan-cancer level (FDR = 6.73e−35; Figure 4D), and in samples with variants in this cluster compared to non-carriers in BRCA, ccRCC, COAD, GBM, HNSCC, lung adenocarcinoma (LUAD), and PDAC (Figure S4E; Table S4D).

We also identify a SDHA cluster involving the germline variant V657I and proximal to the phosphosite T656 (Figure 4E). Although V657I is classified as benign in ClinVar, studies suggest its potential pathogenicity in pheochromocytoma/paraganglioma, renal cell carcinoma, and gastrointestinal tumors.[115–117] Supporting this, we observed significantly lower SDHA levels among co-clustered samples at the pan-cancer level (Figures 4D and S4E; Table S4C). We also observe a cluster involving a germline variant overlapping the S389 phosphosite of DEPTOR (S389N) also detected in our analyses of direct-hit events (Figures 3D and S4F). This cluster prioritizes a nearby phosphosite (Y385) and somatic mutations (R386W, V388M, and L393V) and seems to be associated with a small but statistically significant increase of DEPTOR protein levels in co-clustering samples at the pan-cancer level (Figure 4D; Table S4C). Functional studies would be needed to further investigate this variant's impact. Collectively, our 3D spatial clustering analyses can help prioritize variants and PTM sites that may influence cancer susceptibility.

ASP effects revealed by precision peptidomics

Genetic variants can lead to preferential expression of one allele, a phenomenon called allele-specific expression (ASE)[118,119] associated with cancer and other diseases.[120–122] Although allele-specific effects on proteins have been explored,[123–126] ASE of germline variants has not been extensively explored using proteomics at the pan-cancer level. We evaluated ASP expression, focusing on HET carriers of variants in cancer genes (Figure 5A).

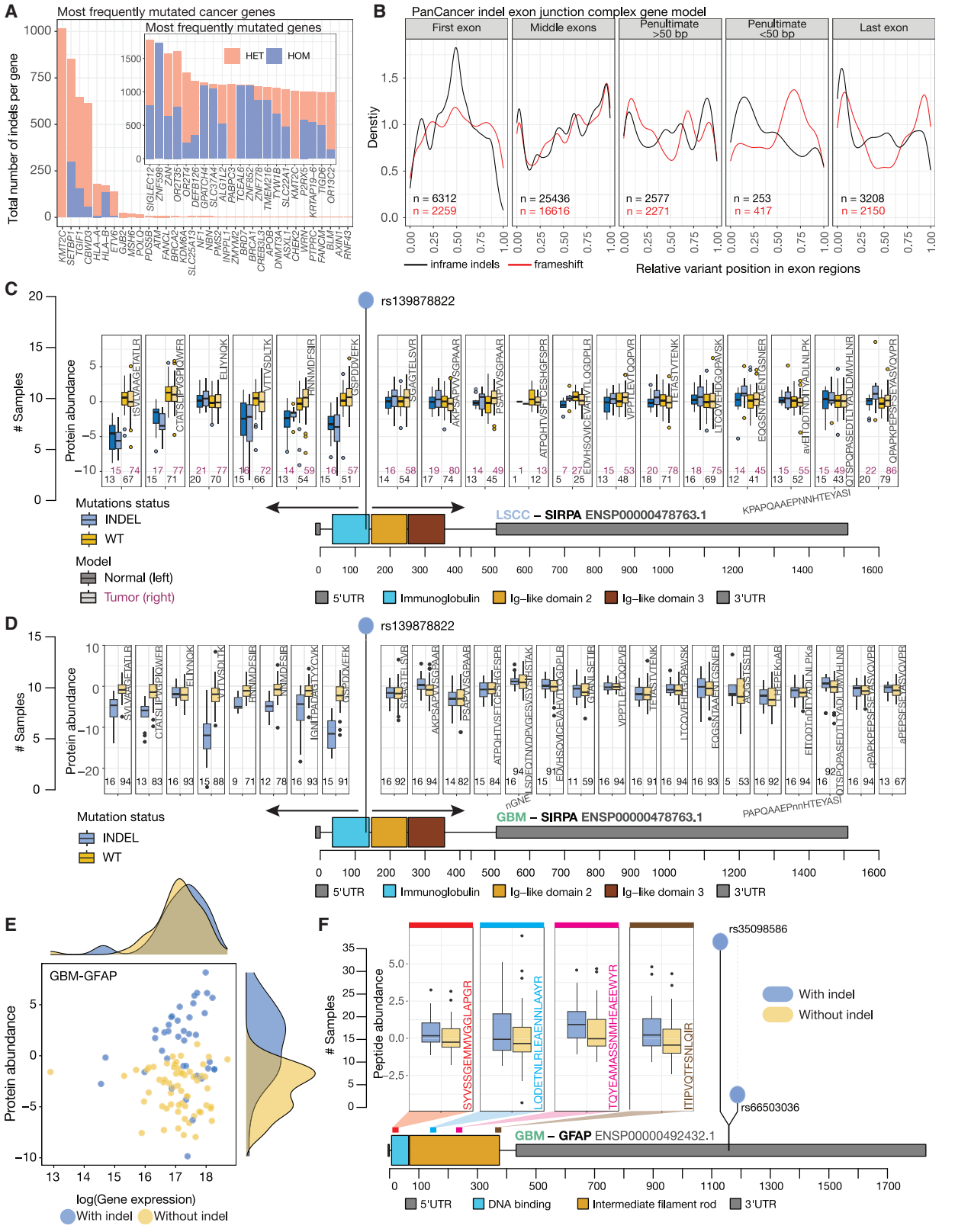
We used a classic approach for ASE detection by calculating read counts for the ANC and DER alleles in tumor and NAT RNA expression data of all HET individuals in CPTAC,[119,122] focusing on variants in 624 cancer-related genes[17] detected in the proteome, phosphoproteome, or acetylome datasets (STAR Methods; Figure 1B). We detected 17,971 ASE instances in tumor and 4,057 in NAT samples (FDR < 0.05; Table S5) affecting 184 and 101 genes, respectively (Figure 5B). To evaluate the ASE impact on the proteome, we compared protein and PTM abundances between samples with preference for the ANC vs. the DER allele for each ASE event. From 45 ASE instances frequently observed across samples in different cancer types (Figure 5B), 23 (51%) showed significant and suggestive differences (FDR < 0.2) in protein and/or PTM abundances (Figures S5A and S5B; Table S5E). To explore the impact of ASE of germline variants on protein interactions, we compared protein and PTM abundances of interacting partners of proteins with ASE germline variants between samples with preference for the ANC vs. the DER allele. In total, we observed 7 PTMs and 7 proteins associated with those ASE events (FDR < 0.2; Table S5).

We observed a higher number of tumor ASE events compared with NAT, consistent with previous findings.[121,127] The higher prevalence of ASE in tumors (after accounting for having fewer NAT samples; Figure S5C) is likely due to genetic factors, including copy-number variations (CNVs), while NAT ASEs should be enriched with epigenetically regulated events.[128] Tumor-only ASE events included variants in genes associated with DNA damage response and cell cycle, such as *AURKB* T299M, *MLH1* I219V, variants in *PARP4*, and the *BRIP1* S919P variant in BRCA and ovarian high-grade serous carcinoma (HGSC) patients. *BRIP1* is involved in the homologous recombination pathway, which is frequently altered in these two cancers.[129]

Common germline variants can also lead to ASE. This is the case of D139E in *CHD4* (rs1639122, gnomAD AF = 0.4), a chromatin-remodeling enzyme that regulates DNA damage

(C–E) Evaluation of ASE on gene, protein, and PTM expression. (Left) Read counts for ancestral (ANC) and derived (DER) alleles of *CHD4* D139E in tumor (C) and NAT (D), and *TP53* P72R in tumor (E). Dots represent individual samples, and color denotes significant (FDR ≤ 0.05) preference for the ANC (red) or DER (purple) allele. Samples with high preference for the ANC allele of *CHD4* D139E are colored in dark red. (Right) Peptides for: DER allele CHD4 E139 (phosphopeptide sequence: RKEEEEEDDDDDsKEPK), ANC allele CHD4 D139 (phosphopeptide sequence: RKEEEEEDDDDDDsKEPK), and DER allele TP53 R72 (peptide sequences: MPEAAPR and (R)VAPAPAAPTPAAPAPAPSWPLSSSVPSQK, that are possible only with R72).
(E) (Bottom right) RNA-seq VAFs for *TP53* R72 allele.
(F) Distribution of CNVs affecting respective genes across samples with ASE.
See also Figure S5.

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026



(legend on next page)

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

response[130] (Figures 5C and 5D, left; Figures S5D–S5F). D139E affects the high mobility group (HMG) box-like-domain of CHD4 essential for DNA binding and nucleosome remodeling.[131] To dissect its ASP effects, we subdivided HET individuals into groups with preference for either the ANC or DER alleles, and those showing no allele preference based on gene expression data (STAR Methods). We defined individuals with $\leq$10 reads covering the DER allele as the group with high preference for the ANC allele (Figures 5C and 5D, dark red). Our results revealed that CHD4 S145 peptides containing the DER allele are more abundant in tumor samples with preference for the DER allele compared with samples with preference for the ANC allele at the phosphosite level ($|\log_2(\text{fold change})| = 1.36$, Wilcoxon rank sum test $p = 0.001$; Figures 5C and 5D, right; Table S5E). An even higher difference was observed between samples with high preference for the ANC allele vs. the DER allele ($|\log_2(\text{fold change})| = 4.23$, Wilcoxon rank sum test $p = 1.91 \times 10^{-10}$; Figures 5C and 5D, right).

In the tumor-only ASEs, we observed the *TP53* P72R variant (rs1042522, gnomAD AF = 0.67; Figure 5E), widely studied in cancer,[132–134] and shown to impact apoptosis, cell cycle arrest, and DNA damage repair.[134,135] We identified this variant at peptide and phosphosite levels (Figure 5B) and noticed frequent ASE for either ANC (33%) or DER (20%) alleles. We also identified several phosphosites of TP53 targets that showed significant (FDR < 0.05) and suggestive (FDR < 0.1) differences between samples with preference for the P72 allele vs. the R72 allele, including MAP4 N86nS94sT101t, WRAP53 S54s, and ARFGEF2 S227s (Figure S5G; Table S5F). Finally, we evaluated how many ASEs were driven by CNV events in our cohort (Figures S5H and S5I) and observed that many samples with ASE were also harboring a CNV in the same gene (Figure 5F), which could explain a higher fraction of ASE events in tumors compared with NAT samples, as previously reported.[136] Our findings highlight the power of precision proteomics in evaluating the impacts of allele-specific events.

The influence of germline indels on protein stability

Precision proteogenomics enables the detection of protein-destabilizing and -stabilizing germline insertions and deletions (indels), an unexplored germline effect in cancer. We identified 103,428 "high-impact" germline indels in CPTAC (STAR Methods), with the most frequent indels occurring in non-cancer genes (e.g., *SIGLEC12*, *ZNF598*, *ZAN*, and *OR2T35*; Figure 6A). Among cancer genes, we observed a high fraction of indels in *KMT2C*, a methyltransferase implicated in many cancer types, oftentimes through somatic truncating single-nucleotide variants (SNVs) or indels[137] (Figure 6A). All *KMT2C* indels were HET, suggesting haploinsufficiency. Other frequent indels in cancer-related genes occurred in *SETBP1*, *TGIF1*, and *CBWD3*, consistent with known germline indel effects.[138,139]

Premature stop codons >50 base pairs (bps) upstream of the last exon junction (EJ) likely trigger higher nonsense mediated decay (NMD) via EJ complex (EJC) than those within 50 bps of the last EJ.[140–142] Consistent with this model, we observed a dramatic shift in inframe and frameshift mutation abundance depending on the distance to the last EJ across cancer types (Figures 6B and S6). Interestingly, no cancer-related genes exhibited frameshift events within the 50 bps before the last EJ, whereas 34 cancer-related genes had inframe indels within the last 50 bps of the last EJ. Among the latter, *NCOR2* (30 samples) had frequent inframe deletions. *NCOR2*, a predicted tumor suppressor gene, recruits HDAC3 to promote histone deacetylation. The absence of germline frameshift events in cancer genes near the penultimate EJC suggests a low tolerance of these variants in cancer genes.

To assess intra-genetic variability at the peptide level, we developed methods to detect proteins with variable stability. We compared aggregated peptide expression upstream and downstream of common indels, identifying 45 proteins with significant differences in peptide abundances at indel sites (STAR Methods; Table S6). For instance, indel rs139878822 in the signal regulatory protein alpha (*SIRPA*; encoding the CD47 receptor) altered peptide abundance in LSCC and GBM (Figures 6C and 6D). Of note, the ELIYNQK peptide, upstream of the indel, has a 100% protein sequence identity to SIRPβ and SIRPγ likely contributing to the higher upstream abundances for that single peptide. SIRPα is involved in the negative regulation of the mitogen-activated protein kinase (MAPK) cascade[143] and CD47-mediated "*don't eat me*" immune signaling cascade.[144] Our results suggest that carriers of indel rs139878822 likely acquire an alternative start site that truncates SIRPA's first immunoglobulin domain.

Using a multi-omic linear discriminant analysis (moLDA), we discovered mutations associated with strong deviations from the expected relationships between protein abundance and RNA expression (STAR Methods). This revealed 45 genes with exceptional moLDA scores across cancer types, e.g., *CPNE1,*

Figure 6. Impact of germline insertion and deletion variants on protein expression

(A) Total number of indels per gene, highlighting the most frequently mutated cancer genes (main plot) and the top 21 genes from all possible genes (inner plot). Blue and orange colors indicate homozygous (HOM) and heterozygous (HET) mutation counts, respectively.

(B) Two density plot lines illustrate the relative location (according to exon) and abundance of germline indels across all genes. The penultimate exon junction (EJ) plots indicate exonic mutations in the second-to-last exon relative to the last EJs: >50 and <50 bps, respectively. Annotated inframe indels are indicated in black, and frameshift indels are shown in red.

(C) Lolliplot and faceted peptide comparisons of samples carrying the *SIRPA* indel compared with non-carriers. Boxplots to the left and right of the lolliplot correspond to peptides upstream and downstream of the mutations, respectively. Data are represented as median and interquartile range. Only LSCC samples are displayed.

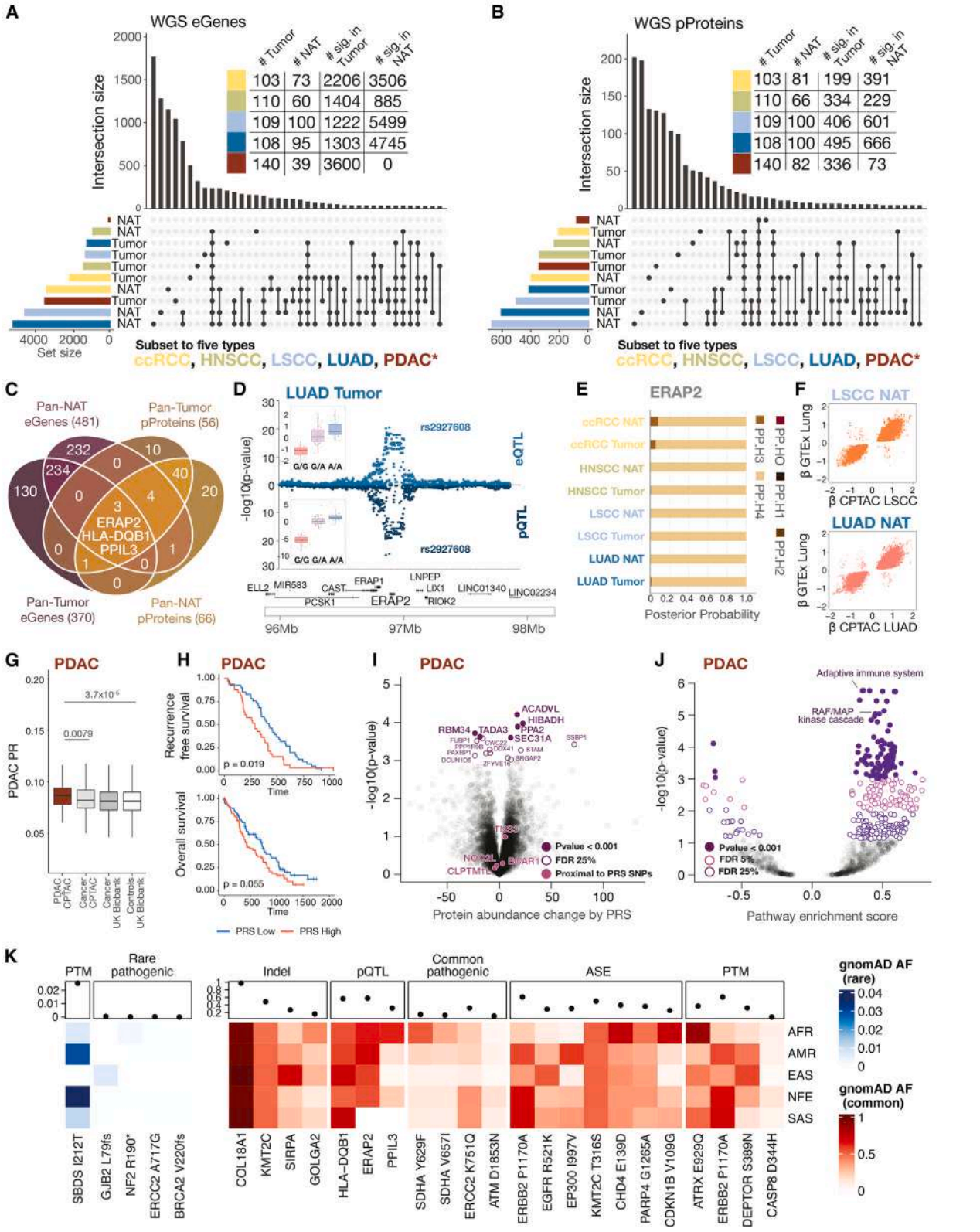
(D) Similar to (C) but highlights mutations in GBM patients.

(E) Gene expression and protein abundance for each sample in the GBM cohort. Dots represent samples colored based on indel mutation status. Margins display data density for protein abundance and gene expression.

(F) Similar to (C) and (D) but for *GFAP* indels in the GBM cohort.

See also Figure S6.

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026



(legend on next page)

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

OAS1, CARD8, CASP7, and *ITIH1* (Figure S6). We highlight a novel association between a common variant in the 3′UTR and protein expression of the glial fibrillary acid protein (GFAP) in GBM tumors (Figure 6E, top x axis). No association is detectable between indel carriers and non-carriers at RNA expression level, but there is a drastic shift in protein abundance (Figure 6E, right y axis, Welch's t test p value = 1.323×10^{-8}). This association found in the UTR of *GFAP* has not been previously reported, probably because of its lack of influence on RNA levels and scarce proteomics data in this disease. GFAP is a critical GBM biomarker,[145] and a "promising therapeutic target"[146]. A breakdown of carriers and non-carriers at the peptide level indicated an increase in the abundance and stability of the entire protein (Figure 6F). Furthermore, miRWalk,[147] an miRNA binding prediction tool, suggests strong binding of miR-137 at this indel site. Collectively, these analyses underscore the utility of multi-omic integration in linking genomic, expression, and proteomic changes to cancer mechanisms.

Omics-wide association of common germline variants and ANC variants with proteomics impacts

Most germline variations occur in non-coding regions of the genome, which regulate cellular processes. To characterize their regulatory impact on gene expression and protein abundance, we performed QTL analyses. Germline variant calling was performed on blood-derived WGS samples followed by imputation using the NYGC 1,000 Genomes Project genome.[28] QTLs affecting transcript (eQTL) and protein (pQTL) abundance were mapped in NAT and tumors across ccRCC, HNSCC, LSCC, LUAD, and PDAC (Figures 7A and 7B; Table S7A). We observed that the expression levels of ~5% and ~10% of the total tested genes (eGenes) and the abundance levels of ~4% and ~5% total tested proteins (pProteins) were associated with WGS germline variants in tumor and NAT, respectively (Table S7). Furthermore, ~12% and 15% of pQTLs were also eQTLs in tumor and NAT, respectively.

Pan-cancer analysis identified 237 eGenes and 47 pProteins that were shared across all NATs and cancers we studied, suggesting cross-tissue QTLs (Figure 7C). Interestingly, *ERAP2*, *HLA-DQB1*, and *PPIL3* were under germline genetic control across all NATs and cancers at both gene expression and protein levels. To determine whether the causal genetic variant was the same for transcript expression and protein abundance for each of these three genes, we conducted a Bayesian test for colocalization of all eQTL-pQTL *cis*-pairs. We discovered that for *ERAP2* (Figure 7D), the same variant drives both eQTLs and pQTLs (Figure 7E). We also show the effect of the lowest p value *cis*-SNP (rs2927608) on *ERAP2* in Figure 7D. Similarly, in *HLA-DQB1* and *PPIL3*, we observed that eQTLs and pQTLs shared the same causal variants in most of the NAT and tumor tissues (Table S7F). As a positive control, we compared the *cis*-eQTLs of NAT and tumor tissues of LSCC and LUAD with normal lung eQTL data from the Genotype-Tissue Expression (GTEx) Consortium (Figure 7F; Table S7G), showing that ~60% of eQTLs (~65% eGenes) and ~50% of eQTLs (~60% eGenes) in NAT and tumors, respectively, were also identified in GTEx at 1% FDR. Furthermore, >95% of common *cis*-eQTLs had the same allelic effects (beta direction) in both lung GTEx and our lung NAT.

Given their prevalence across tissues and -omics datasets, and their role in disease risk in other immune related diseases, we tested whether the expressions of *ERAP2*, *HLA-DQB1*, and *PPIL3* correlated with patient survival. Indeed, the expression of *ERAP2* and *HLA-DQB1* was positively associated with overall survival in HNSCC. Note that 109 of 110 HNSCC individuals were HPV-negative. Furthermore, we observed the same trend in the TCGA HNSCC cohort (Figure S7A).

We calculated PRSs using variants discovered through prior GWAS to evaluate the global impact of personal risk in CPTAC participants (Table S7H). For GBM, LSCC, and PDAC, PRSs were associated with cancer diagnosis as compared with other cancer types in CPTAC and healthy controls from UKBB (Figures 7G and S7B). PRSs also stratified patients by disease aggressiveness, as indicated by disease recurrence and overall survival rates in PDAC (Figure 7H; same patterns observed for LSCC). Considering the potential of PRSs, and that most risk variants from GWAS are non-coding, we characterized their regulatory impact on the tumor proteome. We modeled the effect of PRSs on protein abundance while controlling for clinical, demographic, and molecular covariates. We observed few proteins

Figure 7. eQTL, pQTL, and polygenic risk assessment of samples with tumor and normal WGS

(A) Shared number of eGenes (genes with significant eQTLs).

(B) pProteins (proteins with significant pQTLs) across NAT and tumor tissues of different cancer types indicated by an UpSetR plot[148] (top 40). Asterisks indicate that eQTL analysis was not performed for normals due to the limited number of samples.

(C) Intersection of eGenes and pProteins from a Pan-CPTAC (CCRCC, HNSCC, LSCC, and LUAD) comparison across NAT and tumors.

(D) All p values of *cis*-eQTLs and -pQTLs associated with *ERAP2* in LUAD tumor samples. Plot insets highlight the effect of rs2927608 alleles on *ERAP2* RNA expression (top) and protein abundance (bottom).

(E) Colocalization results of eQTLs and pQTLs in *ERAP2* across NAT and tumors of different cancer types (PP: posterior probabilities supporting each hypothesis; H0: no causal variant; H1: causal variant for RNA expression only; H2: causal variant for protein abundance only; H3: distinct causal variants; H4: common causal variants for eQTL and pQTL).

(F) Comparison of beta coefficients of common *cis*-eQTLs at 1% FDR between GTEx lung and CPTAC LSCC NAT (top) and LUAD NAT (bottom).

(G) Literature-based polygenic risk scores (PRSs) calculated on CPTAC PDAC samples and compared with orthogonal datasets. Data are represented as median and interquartile range. p values for statistical significance for the comparison against "Cancer CPTAC" and "Controls UK Biobank" are provided (t test).

(H) Kaplan-Myer plots estimating recurrence free survival (top) and overall survival (bottom) for samples with high and low PRSs.

(I) Protein abundance changes that correlate with PRS, highlighting that proximal genes (magenta) change less than distal genes.

(J) GSEA shows that high PRS samples are enriched for genes in the adaptive immune system and the RAF/MAPK cascades.

(K) The gnomAD ANC allele frequency (AF) for our top findings, separated by the section in which they are described. Top annotations show overall gnomAD AF separated by rare and common germline variants (left: gnomAD AF $\leq$ 0.05%, right: gnomAD AF > 0.05%). y axis displays the ANC population.

See also Figure S7.

Please cite this article in press as: Martins Rodrigues et al., Precision proteogenomics reveals pan-cancer impact of germline variants, Cell (2025), https://doi.org/10.1016/j.cell.2025.03.026

associated with PRS (Figure 7I), implying limited impact at the single-protein level in CPTAC. However, a pathway-based approximation of these results with gene-set enrichment analyses (GSEAs) showed significant overrepresentation of several biological processes (Figure 7J; Table S7I), suggesting that genetic risk has a cumulative impact that converges in certain biological processes rather than large alterations in specific proteins. Antigen presentation was among the top pathways associated with common risk for PDAC, consistent with its high heritability estimated by pan-cancer immunity studies,[20] in addition to platelet function[149] and L1 cell adhesion molecule (L1CAM) related neural microenvironment remodeling.[150] Common variants also impacted protein levels of the RAS/MAPK pathway, which is mutated in 96% of pancreatic ductal tumors.[151]

We also examined whether variants in this study vary in prevalence across genetic ancestries. While our analyses accounted for ancestry as a covariate (STAR Methods), we recognize that some variants may differ in frequency among individuals from different genetic backgrounds. To explore this, we selected ~150 statistically significant variants from our analyses and compared their ancestry-specific AF using gnomAD for the groups relevant to CPTAC: admixed-American (AMR), East Asian (EAS), non-Finnish European (NFE), and South Asian (SAS). We observed some variants with varying AFs among the five ancestry groups, while others showed consistent AF across all groups (Figure 7K). For instance, the truncating *SIRPA* indel is more common in EAS individuals, while the *CHD4* E139D variant exhibiting strong ASE is more frequent in AFR individuals. In contrast, variants like the top SNP from QTL analysis for *HLA-DQB1* (rs9273472) and *CASP8* D344H, which influenced a distal phosphorylation site, showed similar AFs across all ancestries in gnomAD.

DISCUSSION

While most cancer genomics studies have focused on the role of somatic mutations, the number of germline variants greatly exceed that of somatic mutations in a cancer cell. The composition of these variants is unique, and their effects in oncogenic processes and cancer evolution remain poorly understood. We have leveraged the CPTAC cohort with multiple cancer types to explore the impact of germline variations on cancer-relevant genes through multiple-omics layers: from DNA to RNA, protein abundance, and PTM.

To assess the effects of coding variants and their association with cognate proteins (and PTMs), we used precision peptidomics, i.e., the quantification of peptides carrying genetic variants from individual patients. Integrating bulk proteomic and transcriptomic data with germline variants, we derived mechanistic inferences on the effects of coding variants. Point mutations at or near phosphosites altering downstream biological processes were noted in both tumor and NAT samples. Similar regulatory mechanisms are seen for mutations far from phosphosites in linear distance. We have highlighted examples where a distal linear effect is likely caused by the genetic variants and the PTM sites being close in 3D, benefiting from predicted 3D models by AlphaFold2. We are mindful that those models are imperfect, particularly regarding the relative spatial arrangement of different domains within the same protein.[152] Finally, we also show that germline indels can shape peptide and protein abundance through effects that cannot be discerned at the RNA level.

We explored the impact of non-coding variants on both gene expression and protein abundance (QTL analyses), reporting genes and proteins under germline genetic control across different NATs and tumors (https://immuneregulation.mssm.edu). Comparison of our lung NAT eQTLs with lung eQTLs from GTEx showed an extensive overlap, validating our approach. Beyond highlighted genes from colocalization and survival analyses, there are additional tissue-specific or multi-cancer eGenes and pProteins that merit further investigation. In recent years, large consortia like GTEx have generated genome-wide catalogs of regulatory effects that were critical in understanding the molecular consequences of germline loci identified by GWAS.[153] Here, we provide a pan-tissue catalog of matched gene expression and protein abundance in tumors and NATs that expands such efforts. We also observed that the collective effect of known GWAS risk variants in PDAC, measured as PRS, correlated better with protein levels within oncogenic pathways that are distal to the loci that are part of the PRS. These results suggest that, on top of their local impact in *cis*, GWAS loci can collectively alter global proteomic regulation in *trans*. Despite the case-control design of the cancer discovery GWASs performed to date, our results confirm that a PRS can stratify patients according to disease aggressiveness and overall survival rates.[10] These findings underscore the value of proteogenomics in interpreting germline variant effects on cancer phenotypes and clinical outcomes.

Finally, genetic ancestry might influence the effects of germline variants.[154,155] While diverse, spanning five key genetic ancestries—EUR (n = 786), AFR (n = 40), EAS (n = 194), SAS (n = 5), and AMR (n = 39)—the CPTAC cohort remains underpowered for discovery of novel contributors to cancer phenotypes for specific genetic ancestries other than EUR. Also, our cohort is relatively small compared with larger genomic studies.[39,156–158] Despite this limitation, we uncovered ancestry-independent associations of proteomic, phospho-proteomic, and transcriptomic variations by accounting for genetic ancestry in our analyses.

In conclusion, the germline genome is the fundamental arena where the drama of cancer unfolds and is depicted. Amid mutational chaos, the germline plays a critical role that can enable or constrain the evolution of cancer, dictating the odds of many clinically relevant phenomena: from cancer driver mutations[11,159] to immune responses against cancer cells.[20] A deeper understanding, afforded by proteomics, illuminates this complexity, unveiling altered protein function as pivotal in carcinogenesis.

Limitations of the study

While our dataset is one of the largest multi-omic resources available, we remain underpowered due to sample size. Our cohort included patients predominantly of EUR genetic ancestry, with smaller subsets of other ancestries. Future proteogenomic studies need to include more diverse populations. All -omics datasets were from bulk analytes, limiting our ability to resolve impacts of germline variants on specific cell types. We only used the common variants imputed from the 1,000 Genomes

Project,[28] as we did not have high-coverage WGS data. Current proteomic pipelines rely on generic peptide references to quantify peptide abundance. We addressed this limitation by identifying personalized peptides, but single peptides reflect diverse allele frequencies from populations and our cancer-specific cohort. While protein and gene-level quantification are mitigated by aggregating many peptides, we remain conservative when addressing the impact of single peptides. AlphaFoldDB expanded our structural analysis to all human proteins, but its models are not experimentally validated. Finally, validation of our findings is challenging due to the limited availability of comparable comprehensive datasets, so some of our results will likely evolve as more samples are analyzed.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Dr. Li Ding (lding@wustl.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw and processed proteomics as well as open-access genomic data, can be obtained via Proteomic Data Commons (PDC) at https://pdc.cancer.gov/pdc/cptac-pancancer. Raw genomic and transcriptomic data files can be accessed via the Genomic Data Commons (GDC) Data Portal at https://portal.gdc.cancer.gov with dbGaP Study Accession: phs001287.v17.p6. Complete CPTAC Pan-Cancer controlled and processed data, including the precision proteogenomics data generated in this manuscript, can be accessed via the Cancer Data Service (CDS). The CPTAC Pan-Cancer data hosted in CDS is controlled data and can be accessed through the NCI DAC approved, dbGaP compiled whitelists. Users can access the data for analysis through the Seven Bridges Cancer Genomics Cloud (SB-CGC) which is one of the NCI-funded Cloud Resource/platform for compute intensive analysis. Instructions to access data are as follows: (1) create an account on CGC, Seven Bridges (https://cgc-accounts.sbgenomics.com/auth/register; (2) get approval from dbGaP to access the controlled study (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001287.v17.p6); (3) log into CGC to access Cancer Data Service (CDS) File Explore; (4) copy data into your own space and start analysis and exploration; (5) visit the CDS page to see what studies are available and instructions and guides to use the resources (https://dataservice.datacommons.cancer.gov/#/data).

Data used in this publication were generated by CPTAC, accessible through dbGaP accession numbers phs000892.v6.p1 (“CPTAC Proteogenomic Confirmatory Study”) and phs001287.v17.p6 (“CPTAC Proteogenomic Study”).

We focused on the CPTAC samples with both genomic and proteomic data available to investigate the Pan-Cancer proteogenomic impacts of oncogenic drivers. DOIs are listed in the key resources table. Any additional information and code required to reanalyze the data reported in this paper is available from the lead contact upon request.

CONSORTIA

The members of the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium are Eunkyung An, Meenakshi Anurag, Jasmin Bavarva, Chet Birger, Michael J. Birrer, Anna P. Calinawan, Michele Ceccarelli, Daniel W. Chan, Arul M. Chinnaiyan, Hanbyul Cho, Shrabanti Chowdhury, Marcin P. Cieslik, Daniel Cui Zhou, Corbin Day, Marcin J. Domagalski, Yongchao Dou, Brian J. Druker, Nathan Edwards, Matthew J. Ellis, Steven M. Foltz, Alicia Francis, Tania J. Gonzalez Robles, Sara J.C. Gosline, Runyu Hong, Galen Hostetter, Yingwei Hu, Tara Hiltke, Chen Huang, Emily Huntsman, Eric J. Jaehnig, Scott D. Jewell, Jiayi Ji, Wen Jiang, Lizabeth Katsnelson, Karen A. Ketchum, Iga Kolodziejczak, Jonathan T. Lei, Yuxing Liao, Caleb M. Lindgren, Tao Liu, Weiping Ma, Wilson McKerrow, Chelsea J. Newton, Robert Oldroyd, Gilbert S. Omenn, Amanda G. Paulovich, Francesca Petralia, Boris Reva, Karin D. Rodland, Henry Rodriguez, Kelly V. Ruggles, Dmitry Rykunov, Sara R. Savage, Eric E. Schadt, Michael Schnaubelt, Tobias Schraink, Zhiao Shi, Richard D. Smith, Xiaoyu Song, Yizhe Song, Jimin Tan, Ratna R. Thangudu, Nicole Tignor, Joshua M. Wang, Pei Wang, Ying Wang, Bo Wen, Maciej Wiznerowicz, Xinpei Yi, Bing Zhang, Hui Zhang, Xu Zhang, Zhen Zhang, David I. Heiman, Jared L. Johnson, Liang-Bo Wang, Lijun Yao, Mathangi Thiagarajan, Mehdi Mesri, Özgün Babur, Pietro Pugliese, Qing Zhang, Samuel H. Payne, Saravana M. Dhanasekaran, Shankara Anand, Shankha Satpathy, Stephan Schürer, Vasileios Stathias, Wen-Wei Liang, Wenke Liu, and Yige Wu.

ACKNOWLEDGMENTS

We would like to thank the participants and investigators from the National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium (CPTAC). This work was supported by NCI-CPTAC under award numbers U24CA210955, U24CA210985, U24CA210986, U24CA210954, U24CA210967, U24CA210972, U24CA210979, U24CA210993, U01CA214114, U01CA214116, and U01CA214125 as well as U24CA210972 (D.F., and L.D.), U24CA210979 (G.G.), U24CA270823 (M.A.G.), and contract number GR0012005 (L.D.). This work was also supported by NCI U24CA211006 and R01HG009711 to L.D. The Spanish Ministry of Science supports E.P.-P. and K.J.I. (RYC2019-026415-I and PID2019-107043RA-I00) and U.M.M. (RYC2020-030632-I and PID2019-108244RA-I00). I.M. is supported by Fundación Cris Contra el Cáncer (PR_TPD_2020-19). This research was conducted using the UK Biobank Resource under application numbers 54343 and 74382 (to E.P.-P. and U.M.M., respectively).

This project is funded in part with federal funds from the NCI, National Institutes of Health, under contract no. HHSN261201500003I, Task Order no. HHSN26100064. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

AUTHOR CONTRIBUTIONS

Study conception and design, Z.H.G., E.P.-P., L.D., M.H.B., and G.G.; performed experiments or data collection, F.M.R., N.V.T., Y.L., Y.A., A.I.R., Y.G., F.d.V.L., and A.I.N.; multi-omic & statistical analyses, F.M.R., N.V.T., K.J.I., K.R.C., M.M., K.K., M.E.S., I.M., Y.G., Y.A., T.M.Y., S.C., E.P.S., Y.L., O.S.G., A.G., E.A.K., U.M.M., Z.H.G., M.H.B., E.P.-P., B.T., and R.J.K.; data interpretation & biological analysis, F.M.R., N.V.T., K.R.C., A.C., K.-I.H., C.K.-S., F.A., A.J.L., L.C.C., U.M.M., Z.H.G., M.H.B., G.G., E.P.-P., and L.D.; writing, F.M.R., N.V.T., K.J.I., K.R.C., M.E.S., I.M., Y.G., Y.A., C.K.-S., A.J.L., U.M.M., Z.H.G., D.F., M.A.W., M.H.B, G.G., E.P.-P., and L.D.; supervision, D.R.M., M.A.G., D.F., S.A.C., Z.H.G., M.H.B., G.G., E.P.-P., and L.D.; administration, G.G., A.I.R., and L.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human subjects
  - Clinical data annotation
- METHOD DETAILS
  - Harmonized genome alignment
  - Germline variant calling and filtering from WES
  - Somatic mutation and copy number variant calling from WES
  - Germline variant calling and filtering from WGS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.03.026.

Received: October 9, 2023
Revised: April 29, 2024
Accepted: March 13, 2025
Published: April 14, 2025

REFERENCES

1. Robson, M., Im, S.-A., Senkus, E., Xu, B., Domchek, S.M., Masuda, N., Delaloge, S., Li, W., Tung, N., Armstrong, A., et al. (2017). Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. N. Engl. J. Med. *377*, 523–533. https://doi.org/10.1056/NEJMoa1706450.
2. Villani, A., Tabori, U., Schiffman, J., Shlien, A., Beyene, J., Druker, H., Novokmet, A., Finlay, J., and Malkin, D. (2011). Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study. Lancet Oncol. *12*, 559–567. https://doi.org/10.1016/S1470-2045(11)70119-X.
3. Robson, M., and Offit, K. (2007). Clinical practice. Management of an inherited predisposition to breast cancer. N. Engl. J. Med. *357*, 154–162. https://doi.org/10.1056/NEJMcp071286.
4. Galiatsatos, P., and Foulkes, W.D. (2006). Familial adenomatous polyposis. Am. J. Gastroenterol. *101*, 385–398. https://doi.org/10.1111/j.1572-0241.2006.00375.x.
5. Rahman, N. (2014). Realizing the promise of cancer predisposition genes. Nature *505*, 302–308. https://doi.org/10.1038/nature12981.
6. Lindor, N.M., Petersen, G.M., Hadley, D.W., Kinney, A.Y., Miesfeldt, S., Lu, K.H., Lynch, P., Burke, W., and Press, N. (2006). Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. JAMA *296*, 1507–1517. https://doi.org/10.1001/jama.296.12.1507.
7. Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., et al. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature *452*, 633–637. https://doi.org/10.1038/nature06885.
8. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D.P., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*, 1087–1093. https://doi.org/10.1038/nature05887.
9. Rashkin, S.R., Graff, R.E., Kachuri, L., Thai, K.K., Alexeeff, S.E., Blatchins, M.A., Cavazos, T.B., Corley, D.A., Emami, N.C., Hoffman, J.D., et al. (2020). Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. Nat. Commun. *11*, 4423. https://doi.org/10.1038/s41467-020-18246-6.
10. Klein, R.J., and Gümüş, Z.H. (2022). Are polygenic risk scores ready for the cancer clinic?-a perspective. Transl. Lung Cancer Res. *11*, 910–919. https://doi.org/10.21037/tlcr-21-698.
11. Porta-Pardo, E., Sayaman, R., Ziv, E., and Valencia, A. (2020). The landscape of interactions between cancer polygenic risk scores and somatic alterations in cancer cells. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.28.316851.
12. Bicak, M., Wang, X., Gao, X., Xu, X., Väänänen, R.-M., Taimen, P., Lilja, H., Pettersson, K., and Klein, R.J. (2020). Prostate cancer risk SNP rs10993994 is a trans-eQTL for SNHG11 mediated through MSMB. Hum. Mol. Genet. *29*, 1581–1591. https://doi.org/10.1093/hmg/ddaa026.
13. Li, Q., Stram, A., Chen, C., Kar, S., Gayther, S., Pharoah, P., Haiman, C., Stranger, B., Kraft, P., and Freedman, M.L. (2014). Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. Hum. Mol. Genet. *23*, 5294–5302. https://doi.org/10.1093/hmg/ddu228.
14. Esai Selvan, M., Zauderer, M.G., Rudin, C.M., Jones, S., Mukherjee, S., Offit, K., Onel, K., Rennert, G., Velculescu, V.E., Lipkin, S.M., et al. (2020). Inherited Rare, Deleterious Variants in ATM Increase Lung Adenocarcinoma Risk. J. Thorac. Oncol. *15*, 1871–1879. https://doi.org/10.1016/j.jtho.2020.08.017.
15. Esai Selvan, M., Klein, R.J., and Gümüş, Z.H. (2019). Rare, Pathogenic Germline Variants in Fanconi Anemia Genes Increase Risk for Squamous Lung Cancer. Clin. Cancer Res. *25*, 1517–1525. https://doi.org/10.1158/1078-0432.CCR-18-2660.
16. Huang, K.-L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. Cell *173*, 355–370.e14. https://doi.org/10.1016/j.cell.2018.03.039.
17. Lu, C., Xie, M., Wendl, M.C., Wang, J., McLellan, M.D., Leiserson, M.D.M., Huang, K.-L., Wyczalkowski, M.A., Jayasinghe, R., Banerjee, T., et al. (2015). Patterns and functional implications of rare germline variants across 12 cancer types. Nat. Commun. *6*, 10086. https://doi.org/10.1038/ncomms10086.
18. Esai Selvan, M., Onel, K., Gnjatic, S., Klein, R.J., and Gümüş, Z.H. (2023). Germline rare deleterious variant load alters cancer risk, age of onset and tumor characteristics. NPJ Precis. Oncol. *7*, 13. https://doi.org/10.1038/s41698-023-00354-3.
19. Kanchi, K.L., Johnson, K.J., Lu, C., McLellan, M.D., Leiserson, M.D.M., Wendl, M.C., Zhang, Q., Koboldt, D.C., Xie, M., Kandoth, C., et al. (2014). Integrated analysis of germline and somatic variants in ovarian cancer. Nat. Commun. *5*, 3156. https://doi.org/10.1038/ncomms4156.
20. Sayaman, R.W., Saad, M., Thorsson, V., Hu, D., Hendrickx, W., Roelands, J., Porta-Pardo, E., Mokrab, Y., Farshidfar, F., Kirchhoff, T., et al. (2021). Germline genetic contribution to the immune landscape of cancer. Immunity *54*, 367–386.e8. https://doi.org/10.1016/j.immuni.2021.01.011.
21. Shahamatdar, S., He, M.X., Reyna, M.A., Gusev, A., AlDubayan, S.H., Van Allen, E.M., and Ramachandran, S. (2020). Germline Features Associated with Immune Infiltration in Solid Tumors. Cell Rep. *30*, 2900–2908.e4. https://doi.org/10.1016/j.celrep.2020.02.039.
22. Qing, T., Mohsen, H., Marczyk, M., Ye, Y., O'Meara, T., Zhao, H., Townsend, J.P., Gerstein, M., Hatzis, C., Kluger, Y., et al. (2020). Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. Nat. Commun. *11*, 2438. https://doi.org/10.1038/s41467-020-16293-7.
23. Sun, X., Xue, A., Qi, T., Chen, D., Shi, D., Wu, Y., Zheng, Z., Zeng, J., and Yang, J. (2021). Tumor Mutational Burden Is Polygenic and Genetically

Associated with Complex Traits and Diseases. Cancer Res. *81*, 1230–1239. https://doi.org/10.1158/0008-5472.CAN-20-3459.

24. Liu, Y., Gusev, A., Heng, Y.J., Alexandrov, L.B., and Kraft, P. (2022). Somatic mutational profiles and germline polygenic risk scores in human cancer. Genome Med. *14*, 14. https://doi.org/10.1186/s13073-022-01016-y.

25. Wang, S., Pitt, J.J., Zheng, Y., Yoshimatsu, T.F., Gao, G., Sanni, A., Oluwasola, O., Ajani, M., Fitzgerald, D., Odetunde, A., et al. (2019). Germline variants and somatic mutation signatures of breast cancer across populations of African and European ancestry in the US and Nigeria. Int. J. Cancer *145*, 3321–3333. https://doi.org/10.1002/ijc.32498.

26. Musa, J., Cidre-Aranaz, F., Aynaud, M.-M., Orth, M.F., Knott, M.M.L., Mirabeau, O., Mazor, G., Varon, M., Hölting, T.L.B., Grossetête, S., et al. (2019). Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes. Nat. Commun. *10*, 4128. https://doi.org/10.1038/s41467-019-12071-2.

27. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nat. Genet. *53*, 120–126. https://doi.org/10.1038/s41588-020-00756-0.

28. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19. https://doi.org/10.1016/j.cell.2022.08.004.

29. Ruggles, K.V., Tang, Z., Wang, X., Grover, H., Askenazi, M., Teubl, J., Cao, S., McLellan, M.D., Clauser, K.R., Tabb, D.L., et al. (2016). An Analysis of the Sensitivity of Proteogenomic Mapping of Somatic Mutations and Novel Splicing Events in Cancer. Mol. Cell. Proteomics *15*, 1060–1071. https://doi.org/10.1074/mcp.M115.056226.

30. Geffen, Y., Anand, S., Akiyama, Y., Yaron, T.M., Song, Y., Johnson, J.L., Govindan, A., Babur, Ö., Li, Y., Huntsman, E., et al. (2023). Pan-cancer analysis of post-translational modifications reveals shared patterns of protein regulation. Cell *186*, 3945–3967.e26. https://doi.org/10.1016/j.cell.2023.07.013.

31. Momozawa, Y., and Mizukami, K. (2021). Unique roles of rare variants in the genetics of complex diseases in humans. J. Hum. Genet. *66*, 11–23. https://doi.org/10.1038/s10038-020-00845-2.

32. Park, J.-H., Gail, M.H., Weinberg, C.R., Carroll, R.J., Chung, C.C., Wang, Z., Chanock, S.J., Fraumeni, J.F., and Chatterjee, N. (2011). Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc. Natl. Acad. Sci. USA *108*, 18026–18031. https://doi.org/10.1073/pnas.1114759108.

33. Scott, A.D., Huang, K.-L., Weerasinghe, A., Mashl, R.J., Gao, Q., Martins Rodrigues, F., Wyczalkowski, M.A., and Ding, L. (2019). CharGer: clinical Characterization of germline variants. Bioinformatics *35*, 865–867. https://doi.org/10.1093/bioinformatics/bty649.

34. Knudson, A.G. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA *68*, 820–823. https://doi.org/10.1073/pnas.68.4.820.

35. Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. Nat. Rev. Cancer *1*, 157–162. https://doi.org/10.1038/35101031.

36. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

37. Li, Y., Porta-Pardo, E., Tokheim, C., Bailey, M.H., Yaron, T.M., Stathias, V., Geffen, Y., Imbach, K.J., Cao, S., Anand, S., et al. (2023). Pan-cancer proteogenomics connects oncogenic drivers to functional states. Cell *186*, 3921–3944.e25. https://doi.org/10.1016/j.cell.2023.07.014.

38. Peltomäki, P., Nyström, M., Mecklin, J.-P., and Seppälä, T.T. (2023). Lynch Syndrome Genetics and Clinical Implications. Gastroenterology *164*, 783–799. https://doi.org/10.1053/j.gastro.2022.08.058.

39. Bailey, M.H., Tokheim, C., Porta-Pardo, E., SenGupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18. https://doi.org/10.1016/j.cell.2018.02.060.

40. Xiao, S., Cui, S., Lu, X., Guan, Y., Li, D., Liu, Q., Cai, Y., Jin, C., Yang, J., Wu, S., et al. (2016). The ERCC2/XPD Lys751Gln polymorphism affects DNA repair of benzo[a]pyrene induced damage, tested in an in vitro model. Toxicol. In Vitro *34*, 300–308. https://doi.org/10.1016/j.tiv.2016.04.015.

41. Zhang, G., Guan, Y., Zhao, Y., van der Straaten, T., Xiao, S., Xue, P., Zhu, G., Liu, Q., Cai, Y., Jin, C., et al. (2017). ERCC2/XPD Lys751Gln alter DNA repair efficiency of platinum-induced DNA damage through P53 pathway. Chem. Biol. Interact. *263*, 55–65. https://doi.org/10.1016/j.cbi.2016.12.015.

42. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

43. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discov. *12*, 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059.

44. Krassowski, M., Paczkowska, M., Cullion, K., Huang, T., Dzneladze, I., Ouellette, B.F.F., Yamada, J.T., Fradet-Turcotte, A., and Reimand, J. (2018). ActiveDriverDB: human disease mutations and genome variation in post-translational modification sites of proteins. Nucleic Acids Res. *46*, D901–D910. https://doi.org/10.1093/nar/gkx973.

45. Kim, Y., Kang, C., Min, B., and Yi, G.-S. (2015). Detection and analysis of disease-associated single nucleotide polymorphism influencing post-translational modification. BMC Med. Genomics *8* (*Suppl 2*), S7. https://doi.org/10.1186/1755-8794-8-S2-S7.

46. Yang, Y., Peng, X., Ying, P., Tian, J., Li, J., Ke, J., Zhu, Y., Gong, Y., Zou, D., Yang, N., et al. (2019). Awesome: a database of SNPs that affect protein post-translational modifications. Nucleic Acids Res. *47*, D874–D880. https://doi.org/10.1093/nar/gky821.

47. Patrick, R., Kobe, B., Lê Cao, K.-A., and Bodén, M. (2017). PhosphoPICK-SNP: quantifying the effect of amino acid variants on protein phosphorylation. Bioinformatics *33*, 1773–1781. https://doi.org/10.1093/bioinformatics/btx072.

48. Huang, K.-Y., Lee, T.-Y., Kao, H.-J., Ma, C.-T., Lee, C.-C., Lin, T.-H., Chang, W.-C., and Huang, H.-D. (2019). dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. Nucleic Acids Res. *47*, D298–D308. https://doi.org/10.1093/nar/gky1074.

49. Wang, Z., Zhong, J., Inuzuka, H., Gao, D., Shaik, S., Sarkar, F.H., and Wei, W. (2012). An evolving role for DEPTOR in tumor development and progression. Neoplasia *14*, 368–375. https://doi.org/10.1593/neo.12542.

50. Klen, J., Goričar, K., Horvat, S., Stojan, J., and Dolžan, V. (2019). DEPTOR polymorphisms influence late complications in type 2 diabetes patients. Pharmacogenomics *20*, 879–890. https://doi.org/10.2217/pgs-2019-0058.

51. Gingras, A.C., Raught, B., Gygi, S.P., Niedzwiecka, A., Miron, M., Burley, S.K., Polakiewicz, R.D., Wyslouch-Cieszynska, A., Aebersold, R., and Sonenberg, N. (2001). Hierarchical phosphorylation of the translation inhibitor 4E-BP1. Genes Dev. *15*, 2852–2864. https://doi.org/10.1101/gad.912401.

52. Saadeh, F.S., Morsi, R.Z., El-Kurdi, A., Nemer, G., Mahfouz, R., Charafeddine, M., Khoury, J., Najjar, M.W., Khoueiry, P., and Assi, H.I. (2020). Correlation of genetic alterations by whole-exome sequencing with clinical outcomes of glioblastoma patients from the Lebanese population. PLoS One *15*, e0242793. https://doi.org/10.1371/journal.pone.0242793.

53. von der Heyde, S., Wagner, S., Czerny, A., Nietert, M., Ludewig, F., Salinas-Riester, G., Arlt, D., and Beißbarth, T. (2015). mRNA Profiling Reveals Determinants of Trastuzumab Efficiency in HER2-Positive Breast Cancer. PLoS One *10*, e0117818. https://doi.org/10.1371/journal.pone.0117818.

54. Tong, S.Y., Ha, S.Y., Ki, K.D., Lee, J.M., Lee, S.K., Lee, K.B., Kim, M.K., Cho, C.H., and Kwon, S.Y. (2009). The effects of obesity and HER-2 polymorphisms as risk factors for endometrial cancer in Korean women. BJOG *116*, 1046–1052. https://doi.org/10.1111/j.1471-0528.2009.02186.x.

55. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. *39*, 870–874. https://doi.org/10.1038/ng2075.

56. Breyer, J.P., Sanders, M.E., Airey, D.C., Cai, Q., Yaspan, B.L., Schuyler, P.A., Dai, Q., Boulos, F., Olivares, M.G., Gao, Y.-T., et al. (2009). Heritable Variation of ERBB2 and Breast Cancer Risk. Cancer Epidemiol. Biomarkers Prev. *18*, 1252–1258. https://doi.org/10.1158/1055-9965.EPI-08-1202.

57. Benusiglio, P.R., Lesueur, F., Luccarini, C., Conroy, D.M., Shah, M., Easton, D.F., Day, N.E., Dunning, A.M., Pharoah, P.D., and Ponder, B.A. (2005). Common ERBB2 polymorphisms and risk of breast cancer in a white British population: a case–control study. Breast Cancer Res. *7*, R204–R209. https://doi.org/10.1186/bcr982.

58. Poole, E.M., Curtin, K., Hsu, L., Kulmacz, R.J., Duggan, D.J., Makar, K.W., Xiao, L., Carlson, C.S., Slattery, M.L., Caan, B.J., et al. (2011). Genetic variability in EGFR, Src and HER2 and risk of colorectal adenoma and cancer. Int. J. Mol. Epidemiol. Genet. *2*, 300–315.

59. Su, Y., Jiang, Y., Sun, S., Yin, H., Shan, M., Tao, W., Ge, X., and Pang, D. (2015). Effects of HER2 genetic polymorphisms on its protein expression in breast cancer. Cancer Epidemiol. *39*, 1123–1127. https://doi.org/10.1016/j.canep.2015.08.011.

60. Jo, U.H., Han, S.G.L., Seo, J.H., Park, K.H., Lee, J.W., Lee, H.J., Ryu, J.S., and Kim, Y.H. (2008). The genetic polymorphisms of HER-2 and the risk of lung cancer in a Korean population. BMC Cancer *8*, 359. https://doi.org/10.1186/1471-2407-8-359.

61. Vázquez-Ibarra, K.C., Bustos-Carpinteyro, A.R., García-Ruvalcaba, A., Magaña-Torres, M.T., Gutiérrez-Aguilar, R., Marín-Contreras, M.E., Santiago-Luna, E., and Sánchez-López, J.Y. (2019). The ERBB2 gene polymorphisms rs2643194, rs2934971, and rs1058808 are associated with increased risk of gastric cancer. Braz. J. Med. Biol. Res. *52*, e8379. https://doi.org/10.1590/1414-431X20198379.

62. Alanazi, I.O., Shaik, J.P., Parine, N.R., Azzam, N.A., Alharbi, O., Hawsawi, Y.M., Oyouni, A.A.A., Al-Amer, O.M., Alzahrani, F., Almadi, M.A., et al. (2021). Association of HER1 and HER2 Gene Variants in the Predisposition of Colorectal Cancer. J. Oncol. *2021*, 6180337. https://doi.org/10.1155/2021/6180337.

63. Chen, H., Zhai, Z., Xie, Q., Lai, Y., and Chen, G. (2021). Correlation between SNPs of PIK3CA, ERBB2 3′UTR, and their interactions with environmental factors and the risk of epithelial ovarian cancer. J. Assist. Reprod. Genet. *38*, 2631–2639. https://doi.org/10.1007/s10815-021-02177-2.

64. Gao, Y., Tang, X., Cao, J., Rong, R., Yu, Z., Liu, Y., Lu, Y., Liu, X., Han, L., Liu, J., et al. (2019). The Effect of HER2 Single Nucleotide Polymorphisms on Cervical Cancer Susceptibility and Survival in a Chinese Population. J. Cancer *10*, 378–387. https://doi.org/10.7150/jca.27976.

65. Paska, A.V., and Hudler, P. (2015). Aberrant methylation patterns in cancer: a clinical view. Biochem. Med. *25*, 161–176. https://doi.org/10.11613/BM.2015.017.

66. Appelqvist, F., Yhr, M., Erlandson, A., Martinsson, T., and Enerbäck, C. (2014). Deletion of the MGMT gene in familial melanoma. Genes Chromosomes Cancer *53*, 703–711. https://doi.org/10.1002/gcc.22180.

67. De Summa, S.D., Guida, M., Tommasi, S., Strippoli, S., Pellegrini, C., Fargnoli, M.C., Pilato, B., Natalicchio, I., Guida, G., and Pinto, R. (2017). Genetic profiling of a rare condition: co-occurrence of albinism and multiple primary melanoma in a Caucasian family. Oncotarget *8*, 29751–29759. https://doi.org/10.18632/oncotarget.12777.

68. Lubahn, J., Berndt, S.I., Jin, C.H., Klim, A., Luly, J., Wu, W.S., Isaacs, S., Wiley, K., Isaacs, W.B., Suarez, B.K., et al. (2010). Association of CASP8 D302H polymorphism with reduced risk of aggressive prostate carcinoma. Prostate *70*, 646–653. https://doi.org/10.1002/pros.21098.

69. Vahednia, E., Shandiz, F.H., Bagherabad, M.B., Moezzi, A., Afzaljavan, F., Tajbakhsh, A., Kooshyar, M.M., and Pasdar, A. (2019). The Impact of CASP8 rs10931936 and rs1045485 Polymorphisms as well as the Haplotypes on Breast Cancer Risk: A Case-Control Study. Clin. Breast Cancer *19*, e563–e577. https://doi.org/10.1016/j.clbc.2019.02.011.

70. Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell *173*, 321–337.e10. https://doi.org/10.1016/j.cell.2018.03.035.

71. Porta-Pardo, E., Garcia-Alonso, L., Hrabe, T., Dopazo, J., and Godzik, A. (2015). A Pan-Cancer Catalogue of Cancer Driver Protein Interaction Interfaces. PLoS Comput. Biol. *11*, e1004518. https://doi.org/10.1371/journal.pcbi.1004518.

72. Niu, B., Scott, A.D., SenGupta, S., Bailey, M.H., Batra, P., Ning, J., Wyczalkowski, M.A., Liang, W.-W., Zhang, Q., McLellan, M.D., et al. (2016). Protein-structure-guided discovery of functional mutations across 19 cancer types. Nat. Genet. *48*, 827–837. https://doi.org/10.1038/ng.3586.

73. Huang, K.-L., Scott, A.D., Zhou, D.C., Wang, L.-B., Weerasinghe, A., Elmas, A., Liu, R., Wu, Y., Wendl, M.C., Wyczalkowski, M.A., et al. (2021). Spatially interacting phosphorylation sites and mutations in cancer. Nat. Commun. *12*, 2313. https://doi.org/10.1038/s41467-021-22481-w.

74. Kamburov, A., Lawrence, M.S., Polak, P., Leshchiner, I., Lage, K., Golub, T.R., Lander, E.S., and Getz, G. (2015). Comprehensive assessment of cancer missense mutation clustering in protein structures. Proc. Natl. Acad. Sci. USA *112*, E5486–E5495. https://doi.org/10.1073/pnas.1516373112.

75. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

76. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

77. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242. https://doi.org/10.1093/nar/28.1.235.

78. Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nat. Struct. Biol. *10*, 980. https://doi.org/10.1038/nsb1203-980.

79. The UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

80. Sidney, J., Peters, B., Frahm, N., Brander, C., and Sette, A. (2008). HLA class I supertypes: a revised and updated classification. BMC Immunol. *9*, 1. https://doi.org/10.1186/1471-2172-9-1.

81. Al Naqbi, H., Mawart, A., Alshamsi, J., Al Safar, H., and Tay, G.K. (2021). Major histocompatibility complex (MHC) associations with diseases in ethnic groups of the Arabian Peninsula. Immunogenetics *73*, 131–152. https://doi.org/10.1007/s00251-021-01204-x.

82. Medhasi, S., and Chantratita, N. (2022). Human Leukocyte Antigen (HLA) System: Genetics and Association with Bacterial and Viral Infections. J. Immunol. Res. *2022*, 9710376. https://doi.org/10.1155/2022/9710376.

83. Janeway, C.A., Travers, P., Walport, M., and Shlomchik, M. (2001). Immunobiology: the Immune System in Health and Disease, Fifth Edition (Garland Publishing).

84. Dou, Y., Kawaler, E.A., Cui Zhou, D., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al. (2020). Proteogenomic Characterization of Endometrial Carcinoma. Cell *180*, 729–748.e26. https://doi.org/10.1016/j.cell.2020.01.026.

85. Dong, P., Tada, M., Hamada, J.-I., Nakamura, A., Moriuchi, T., and Sakuragi, N. (2007). p53 dominant-negative mutant R273H promotes invasion and migration of human endometrial cancer HHUA cells. Clin. Exp. Metastasis *24*, 471–483. https://doi.org/10.1007/s10585-007-9084-8.

86. Li, J., Yang, L., Gaur, S., Zhang, K., Wu, X., Yuan, Y.-C., Li, H., Hu, S., Weng, Y., and Yen, Y. (2014). Mutants TP53 p.R273H and p.R273C but not p.R273G Enhance Cancer Cell Malignancy. Hum. Mutat. *35*, 575–584. https://doi.org/10.1002/humu.22528.

87. Xiao, G., Lundine, D., Annor, G.K., Canar, J., Ellison, V., Polotskaia, A., Donabedian, P.L., Reiner, T., Khramtsova, G.F., Olopade, O.I., et al. (2020). Gain-of-Function Mutant p53 R273H Interacts with Replicating DNA and PARP1 in Breast Cancer. Cancer Res. *80*, 394–405. https://doi.org/10.1158/0008-5472.CAN-19-1036.

88. Tan, B.S., Tiong, K.H., Choo, H.L., Chung, F.F.-L., Hii, L.W., Tan, S.H., Yap, I.K., Pani, S., Khor, N.T., Wong, S.F., et al. (2015). Mutant p53-R273H mediates cancer cell survival and anoikis resistance through AKT-dependent suppression of BCL2-modifying factor (BMF). Cell Death Dis. *6*, e1826. https://doi.org/10.1038/cddis.2015.191.

89. Lang, G.A., Iwakuma, T., Suh, Y.-A., Liu, G., Rao, V.A., Parant, J.M., Valentin-Vega, Y.A., Terzian, T., Caldwell, L.C., Strong, L.C., et al. (2004). Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. Cell *119*, 861–872. https://doi.org/10.1016/j.cell.2004.11.006.

90. Goh, A.M., Coffill, C.R., and Lane, D.P. (2011). The role of mutant p53 in human cancer. J. Pathol. *223*, 116–126. https://doi.org/10.1002/path.2784.

91. Solomon, H., Madar, S., and Rotter, V. (2011). Mutant p53 gain of function is interwoven into the hallmarks of cancer. J. Pathol. *225*, 475–478. https://doi.org/10.1002/path.2988.

92. Wang, H., Guo, M., Wei, H., and Chen, Y. (2023). Targeting p53 pathways: mechanisms, structures, and advances in therapy. Signal Transduct. Target. Ther. *8*, 1–35. https://doi.org/10.1038/s41392-023-01347-1.

93. Belinsky, M., and Jaiswal, A.K. (1993). NAD(P)H:Quinone oxidoreductase1 (DT-diaphorase) expression in normal and tumor tissues. Cancer Metastasis Rev. *12*, 103–117. https://doi.org/10.1007/BF00689804.

94. Fagerholm, R., Hofstetter, B., Tommiska, J., Aaltonen, K., Vrtel, R., Syrjäkoski, K., Kallioniemi, A., Kilpivaara, O., Mannermaa, A., Kosma, V.-M., et al. (2008). NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. Nat. Genet. *40*, 844–853. https://doi.org/10.1038/ng.155.

95. Tossetta, G., Fantone, S., Goteri, G., Giannubilo, S.R., Ciavattini, A., and Marzioni, D. (2023). The Role of NQO1 in Ovarian Cancer. Int. J. Mol. Sci. *24*, 7839. https://doi.org/10.3390/ijms24097839.

96. Prawan, A., Buranrat, B., Kukongviriyapan, U., Sripa, B., and Kukongviriyapan, V. (2009). Inflammatory cytokines suppress NAD(P)H:quinone oxidoreductase-1 and induce oxidative stress in cholangiocarcinoma cells. J. Cancer Res. Clin. Oncol. *135*, 515–522. https://doi.org/10.1007/s00432-008-0483-2.

97. Kolesar, J.M., Pritchard, S.C., Kerr, K.M., Kim, K., Nicolson, M.C., and McLeod, H. (2002). Evaluation of NQO1 gene expression and variant allele in human NSCLC tumors and matched normal lung tissue. Int. J. Oncol. *21*, 1119–1124. https://doi.org/10.3892/ijo.21.5.1119.

98. Cresteil, T., and Jaiswal, A.K. (1991). High levels of expression of the NAD(P)H:quinone oxidoreductase (NQO1) gene in tumor cells compared to normal cells of the same origin. Biochem. Pharmacol. *42*, 1021–1027. https://doi.org/10.1016/0006-2952(91)90284-c.

99. Tossetta, G., and Marzioni, D. (2023). Targeting the NRF2/KEAP1 pathway in cervical and endometrial cancers. Eur. J. Pharmacol. *941*, 175503. https://doi.org/10.1016/j.ejphar.2023.175503.

100. Siegel, D., Franklin, W.A., and Ross, D. (1998). Immunohistochemical detection of NAD(P)H:quinone oxidoreductase in human lung and lung tumors. Clin. Cancer Res. *4*, 2065–2070.

101. Siegel, D., and Ross, D. (2000). Immunodetection of NAD(P)H:quinone oxidoreductase 1 (NQO1) in human tissues. Free Radic. Biol. Med. *29*, 246–253. https://doi.org/10.1016/s0891-5849(00)00310-5.

102. Awadallah, N.S., Dehn, D., Shah, R.J., Russell Nash, S., Chen, Y.K., Ross, D., Bentz, J.S., and Shroyer, K.R. (2008). NQO1 expression in pancreatic cancer and its potential use as a biomarker. Appl. Immunohistochem. Mol. Morphol. *16*, 24–31. https://doi.org/10.1097/PAI.0b013e31802e91d0.

103. Tossetta, G., Fantone, S., Montanari, E., Marzioni, D., and Goteri, G. (2022). Role of NRF2 in Ovarian Cancer. Antioxidants (Basel) *11*, 663. https://doi.org/10.3390/antiox11040663.

104. Beaver, S.K., Mesa-Torres, N., Pey, A.L., and Timson, D.J. (2019). NQO1: A target for the treatment of cancer and neurological diseases, and a model to understand loss of function disease mechanisms. Biochim. Biophys. Acta Proteins Proteom. *1867*, 663–676. https://doi.org/10.1016/j.bbapap.2019.05.002.

105. Lajin, B., and Alachkar, A. (2013). The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: a comprehensive meta-analysis. Br. J. Cancer *109*, 1325–1337. https://doi.org/10.1038/bjc.2013.357.

106. Siegel, D., Anwar, A., Winski, S.L., Kepa, J.K., Zolman, K.L., and Ross, D. (2001). Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H:quinone oxidoreductase 1. Mol. Pharmacol. *59*, 263–268. https://doi.org/10.1124/mol.59.2.263.

107. Medina-Carmona, E., Neira, J.L., Salido, E., Fuchs, J.E., Palomino-Morales, R., Timson, D.J., and Pey, A.L. (2017). Site-to-site interdomain communication may mediate different loss-of-function mechanisms in a cancer-associated NQO1 polymorphism. Sci. Rep. *7*, 44532. https://doi.org/10.1038/srep44532.

108. Medina-Carmona, E., Betancor-Fernández, I., Santos, J., Mesa-Torres, N., Grottelli, S., Batlle, C., Naganathan, A.N., Oppici, E., Cellini, B., Ventura, S., et al. (2019). Insight into the specificity and severity of pathogenic mechanisms associated with missense mutations through experimental and structural perturbation analyses. Hum. Mol. Genet. *28*, 1–15. https://doi.org/10.1093/hmg/ddy323.

109. Martínez-Limón, A., Alriquet, M., Lang, W.-H., Calloni, G., Wittig, I., and Vabulas, R.M. (2016). Recognition of enzymes lacking bound cofactor by protein quality control. Proc. Natl. Acad. Sci. USA *113*, 12156–12161. https://doi.org/10.1073/pnas.1611994113.

110. Pacheco-Garcia, J.L., Anoz-Carbonell, E., Vankova, P., Kannan, A., Palomino-Morales, R., Mesa-Torres, N., Salido, E., Man, P., Medina, M., Naganathan, A.N., et al. (2021). Structural basis of the pleiotropic and specific phenotypic consequences of missense mutations in the multifunctional NAD(P)H:quinone oxidoreductase 1 and their pharmacological rescue. Redox Biol. *46*, 102112. https://doi.org/10.1016/j.redox.2021.102112.

111. Muñoz, I.G., Morel, B., Medina-Carmona, E., and Pey, A.L. (2017). A mechanism for cancer-associated inactivation of NQO1 due to P187S and its reactivation by the consensus mutation H80R. FEBS Lett. *591*, 2826–2835. https://doi.org/10.1002/1873-3468.12772.

112. Medina-Carmona, E., Rizzuti, B., Martín-Escolano, R., Pacheco-García, J.L., Mesa-Torres, N., Neira, J.L., Guzzi, R., and Pey, A.L. (2019). Phosphorylation compromises FAD binding and intracellular stability of wild-type and cancer-associated NQO1: Insights into flavo-proteome stability. Int. J. Biol. Macromol. *125*, 1275–1288. https://doi.org/10.1016/j.ijbiomac.2018.09.108.

113. Medina-Carmona, E., Fuchs, J.E., Gavira, J.A., Mesa-Torres, N., Neira, J.L., Salido, E., Palomino-Morales, R., Burgos, M., Timson, D.J., and Pey, A.L. (2017). Enhanced vulnerability of human proteins towards disease-associated inactivation through divergent evolution. Hum. Mol. Genet. *26*, 3531–3544. https://doi.org/10.1093/hmg/ddx238.

114. Pacheco-Garcia, J.L., Loginov, D.S., Anoz-Carbonell, E., Vankova, P., Palomino-Morales, R., Salido, E., Man, P., Medina, M., Naganathan, A.N., and Pey, A.L. (2022). Allosteric Communication in the Multifunctional and Redox NQO1 Protein Studied by Cavity-Making Mutations. Antioxidants (Basel) *11*, 1110. https://doi.org/10.3390/antiox11061110.

115. Dwight, T., Benn, D.E., Clarkson, A., Vilain, R., Lipton, L., Robinson, B.G., Clifton-Bligh, R.J., and Gill, A.J. (2013). Loss of SDHA Expression Identifies SDHA Mutations in Succinate Dehydrogenase–deficient Gastrointestinal Stromal Tumors. Am. J. Surg. Pathol. *37*, 226–233. https://doi.org/10.1097/PAS.0b013e3182671155.

116. Burnichon, N., Brière, J.-J., Libé, R., Vescovo, L., Rivière, J., Tissier, F., Jouanno, E., Jeunemaitre, X., Bénit, P., Tzagoloff, A., et al. (2010). SDHA is a tumor suppressor gene causing paraganglioma. Hum. Mol. Genet. *19*, 3011–3020. https://doi.org/10.1093/hmg/ddq206.

117. Kamai, T., Higashi, S., Murakami, S., Arai, K., Namatame, T., Kijima, T., Abe, H., Jamiyan, T., Ishida, K., Shirataki, H., et al. (2021). Single nucleotide variants of succinate dehydrogenase A gene in renal cell carcinoma. Cancer Sci. *112*, 3375–3387. https://doi.org/10.1111/cas.14977.

118. Knight, J.C. (2004). Allele-specific gene expression uncovered. Trends Genet. *20*, 113–116. https://doi.org/10.1016/j.tig.2004.01.001.

119. Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. Genome Biol. *16*, 195. https://doi.org/10.1186/s13059-015-0762-6.

120. van Beek, D., Verdonschot, J., Derks, K., Brunner, H., de Kok, T.M., Arts, I.C.W., Heymans, S., Kutmon, M., and Adriaens, M. (2023). Allele-specific expression analysis for complex genetic phenotypes applied to a unique dilated cardiomyopathy cohort. Sci. Rep. *13*, 564. https://doi.org/10.1038/s41598-023-27591-7.

121. Robles-Espinoza, C.D., Mohammadi, P., Bonilla, X., and Gutierrez-Arcelus, M. (2021). Allele-specific expression: applications in cancer and technical considerations. Curr. Opin. Genet. Dev. *66*, 10–19. https://doi.org/10.1016/j.gde.2020.10.007.

122. Wang, Z., Fan, X., Shen, Y., Pagadala, M.S., Signer, R., Cygan, K.J., Fairbrother, W.G., Carter, H., Chung, W.K., and Huang, K.-L. (2021). Non-cancer-related pathogenic germline variants and expression consequences in ten-thousand cancer genomes. Genome Med. *13*, 147. https://doi.org/10.1186/s13073-021-00964-1.

123. Wingo, T.S., Duong, D.M., Zhou, M., Dammer, E.B., Wu, H., Cutler, D.J., Lah, J.J., Levey, A.I., and Seyfried, N.T. (2017). Integrating Next-Generation Genomic Sequencing and Mass Spectrometry to Estimate Allele-Specific Protein Abundance in Human Brain. J. Proteome Res. *16*, 3336–3347. https://doi.org/10.1021/acs.jproteome.7b00324.

124. Johansson, H.J., Socciarelli, F., Vacanti, N.M., Haugen, M.H., Zhu, Y., Siavelis, I., Fernandez-Woodbridge, A., Aure, M.R., Sennblad, B., Vesterlund, M., et al. (2019). Breast cancer quantitative proteome and proteogenomic landscape. Nat. Commun. *10*, 1600. https://doi.org/10.1038/s41467-019-09018-y.

125. Wu, L., and Snyder, M. (2015). Impact of allele-specific peptides in proteome quantification. Proteomics Clin. Appl. *9*, 432–436. https://doi.org/10.1002/prca.201400126.

126. Shi, J., Wang, X., Zhu, H., Jiang, H., Wang, D., Nesvizhskii, A., and Zhu, H.-J. (2018). Determining Allele-Specific Protein Expression (ASPE) Using a Novel Quantitative Concatamer-Based Proteomics Method. J. Proteome Res. *17*, 3606–3612. https://doi.org/10.1021/acs.jproteome.8b00620.

127. Liu, Z., Dong, X., and Li, Y. (2018). A Genome-Wide Study of Allele-Specific Expression in Colorectal Cancer. Front. Genet. *9*, 570. https://doi.org/10.3389/fgene.2018.00570.

128. Bell, C.G., and Beck, S. (2009). Advances in the identification and analysis of allele-specific expression. Genome Med. *1*, 56. https://doi.org/10.1186/gm56.

129. Polak, P., Kim, J., Braunstein, L.Z., Karlic, R., Haradhavala, N.J., Tiao, G., Rosebrock, D., Livitz, D., Kübler, K., Mouw, K.W., et al. (2017). A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat. Genet. *49*, 1476–1486. https://doi.org/10.1038/ng.3934.

130. Larsen, D.H., Poinsignon, C., Gudjonsson, T., Dinant, C., Payne, M.R., Hari, F.J., Rendtlew Danielsen, J.M., Menard, P., Sand, J.C., Stucki, M., et al. (2010). The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. J. Cell Biol. *190*, 731–740. https://doi.org/10.1083/jcb.200912135.

131. Silva, A.P.G., Ryan, D.P., Galanty, Y., Low, J.K.K., Vandevenne, M., Jackson, S.P., and Mackay, J.P. (2016). The N-terminal Region of Chromodomain Helicase DNA-binding Protein 4 (CHD4) Is Essential for Activity and Contains a High Mobility Group (HMG) Box-like-domain That Can Bind Poly(ADP-ribose). J. Biol. Chem. *291*, 924–938. https://doi.org/10.1074/jbc.M115.683227.

132. De Souza, C., Madden, J., Koestler, D.C., Minn, D., Montoya, D.J., Minn, K., Raetz, A.G., Zhu, Z., Xiao, W.-W., Tahmassebi, N., et al. (2021). Effect of the p53 P72R Polymorphism on Mutant TP53 Allele Selection in Human Cancer. J. Natl. Cancer Inst. *113*, 1246–1257. https://doi.org/10.1093/jnci/djab019.

133. Bojesen, S.E., and Nordestgaard, B.G. (2008). The common germline Arg72Pro polymorphism of p53 and increased longevity in humans. Cell Cycle *7*, 158–163. https://doi.org/10.4161/cc.7.2.5249.

134. Marin, M.C., Jost, C.A., Brooks, L.A., Irwin, M.S., O'Nions, J., Tidy, J.A., James, N., McGregor, J.M., Harwood, C.A., Yulug, I.G., et al. (2000). A common polymorphism acts as an intragenic modifier of mutant p53 behaviour. Nat. Genet. *25*, 47–54. https://doi.org/10.1038/75586.

135. Siddique, M., and Sabapathy, K. (2006). Trp53-dependent DNA-repair is affected by the codon 72 polymorphism. Oncogene *25*, 3489–3500. https://doi.org/10.1038/sj.onc.1209405.

136. Tuch, B.B., Laborde, R.R., Xu, X., Gu, J., Chung, C.B., Monighetti, C.K., Stanley, S.J., Olsen, K.D., Kasperbauer, J.L., Moore, E.J., et al. (2010). Tumor Transcriptome Sequencing Reveals Allelic Expression Imbalances Associated with Copy Number Alterations. PLoS One *5*, e9317. https://doi.org/10.1371/journal.pone.0009317.

137. Fagan, R.J., and Dingwall, A.K. (2019). COMPASS Ascending: Emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. Cancer Lett. *458*, 56–65. https://doi.org/10.1016/j.canlet.2019.05.024.

138. Fang, F., Liu, C., Li, Q., Xu, R., Zhang, T., and Shen, X. (2022). The Role of SETBP1 in Gastric Cancer: Friend or Foe. Front. Oncol. *12*, 908943. https://doi.org/10.3389/fonc.2022.908943.

139. Wang, H., Gao, Y., Qin, L., Zhang, M., Shi, W., Feng, Z., Guo, L., Zhu, B., and Liao, S. (2023). Identification of a novel de novo mutation of SETBP1 and new findings of SETBP1 in tumorgenesis. Orphanet J. Rare Dis. *18*, 107. https://doi.org/10.1186/s13023-023-02705-6.

140. Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. Mol. Cell. Biol. *18*, 5272–5283. https://doi.org/10.1128/MCB.18.9.5272.

141. Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. EMBO J. *17*, 3484–3494. https://doi.org/10.1093/emboj/17.12.3484.

142. Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat. Genet. *48*, 1112–1118. https://doi.org/10.1038/ng.3664.

143. Kapoor, G.S., and O'Rourke, D.M. (2023). Editorial Expression of Concern: SIRPa1 receptors interfere with the EGFRvIII signalosome to inhibit glioblastoma cell transformation and migration. Oncogene *42*, 2154. https://doi.org/10.1038/s41388-023-02740-4.

144. Gholamin, S., Mitra, S.S., Feroze, A.H., Liu, J., Kahn, S.A., Zhang, M., Esparza, R., Richard, C., Ramaswamy, V., Remke, M., et al. (2017). Disrupting the CD47-SIRPα anti-phagocytic axis by a humanized anti-CD47 antibody is an efficacious treatment for malignant pediatric brain tumors. Sci. Transl. Med. *9*, eaaf2968. https://doi.org/10.1126/scitranslmed.aaf2968.

145. Jung, C.S., Foerch, C., Schänzer, A., Heck, A., Plate, K.H., Seifert, V., Steinmetz, H., Raabe, A., and Sitzer, M. (2007). Serum GFAP is a diagnostic marker for glioblastoma multiforme. Brain *130*, 3336–3341. https://doi.org/10.1093/brain/awm263.

146. Radu, R., Petrescu, G.E.D., Gorgan, R.M., and Brehar, F.M. (2022). GFAPδ: A Promising Biomarker and Therapeutic Target in Glioblastoma. Front. Oncol. *12*, 859247. https://doi.org/10.3389/fonc.2022.859247.

147. Sticht, C., De La Torre, C., Parveen, A., and Gretz, N. (2018). miRWalk: An online resource for prediction of microRNA binding sites. PLoS One *13*, e0206239. https://doi.org/10.1371/journal.pone.0206239.

148. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.

149. Mai, S., and Inkielewicz-Stepniak, I. (2021). Pancreatic Cancer and Platelets Crosstalk: A Potential Biomarker and Target. Front. Cell Dev. Biol. *9*, 749689. https://doi.org/10.3389/fcell.2021.749689.

150. Na'ara, S., Amit, M., and Gil, Z. (2019). L1CAM induces perineural invasion of pancreas cancer cells by upregulation of metalloproteinase expression. Oncogene *38*, 596–608. https://doi.org/10.1038/s41388-018-0458-y.

151. Drosten, M., and Barbacid, M. (2020). Targeting the MAPK Pathway in KRAS-Driven Tumors. Cancer Cell *37*, 543–550. https://doi.org/10.1016/j.ccell.2020.03.013.

152. Akdel, M., Pires, D.E.V., Pardo, E.P., Jänes, J., Zalevsky, A.O., Mészáros, B., Bryant, P., Good, L.L., Laskowski, R.A., Pozzati, G., et al. (2022). A structural biology community assessment of AlphaFold2 applications. Nat. Struct. Mol. Biol. *29*, 1056–1067. https://doi.org/10.1038/s41594-022-00849-w.

153. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660. https://doi.org/10.1126/science.1262110.

154. Abdellaoui, A., Yengo, L., Verweij, K.J.H., and Visscher, P.M. (2023). 15 years of GWAS discovery: Realizing the promise. Am. J. Hum. Genet. *110*, 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011.

155. Shi, H., Burch, K.S., Johnson, R., Freund, M.K., Kichaev, G., Mancuso, N., Manuel, A.M., Dong, N., and Pasaniuc, B. (2020). Localizing Components of Shared Transethnic Genetic Architecture of Complex Traits from GWAS Summary Data. Am. J. Hum. Genet. *106*, 805–817. https://doi.org/10.1016/j.ajhg.2020.04.012.

156. Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., et al. (2020). A compendium of mutational cancer driver genes. Nat. Rev. Cancer *20*, 555–572. https://doi.org/10.1038/s41568-020-0290-x.

157. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93. https://doi.org/10.1038/s41586-020-1969-6.

158. Ding, L., Bailey, M.H., Porta-Pardo, E., Thorsson, V., Colaprico, A., Bertrand, D., Gibbs, D.L., Weerasinghe, A., Huang, K.L., Tokheim, C., et al. (2018). Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. Cell *173*, 305–320.e10. https://doi.org/10.1016/j.cell.2018.03.033.

159. Carrot-Zhang, J., Soca-Chafre, G., Patterson, N., Thorner, A.R., Nag, A., Watson, J., Genovese, G., Rodriguez, J., Gelbard, M.K., Corrales-Rodriguez, L., et al. (2021). Genetic Ancestry Contributes to Somatic Mutations in Lung Cancers from Admixed Latin American Populations. Cancer Discov. *11*, 591–598. https://doi.org/10.1158/2159-8290.CD-20-1165.

160. Li, Y., Dou, Y., Da Veiga Leprevost, F., Geffen, Y., Calinawan, A.P., Aguet, F., Akiyama, Y., Anand, S., Birger, C., Cao, S., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. Cancer Cell *41*, 1397–1406. https://doi.org/10.1016/j.ccell.2023.06.009.

161. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv.

162. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

163. Freed, D., Aldana, R., Weber, J.A., and Edwards, J.S. (2017). The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. Preprint at bioRxiv. https://doi.org/10.1101/115717.

164. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

165. Andrews, S. (2010). FastQC: A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

166. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

167. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

168. Kalayci, S., Selvan, M.E., Ramos, I., Cotsapas, C., Harris, E., Kim, E.-Y., Montgomery, R.R., Poland, G., Pulendran, B., Tsang, J.S., et al. (2019). ImmuneRegulation: a web-based tool for identifying human immune regulatory elements. Nucleic Acids Res. *47*, W142–W150. https://doi.org/10.1093/nar/gkz450.

169. Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics *34*, 867–868. https://doi.org/10.1093/bioinformatics/btx699.

170. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

171. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865–2871. https://doi.org/10.1093/bioinformatics/btp394.

172. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594.

173. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613. https://doi.org/10.1093/nar/gky1131.

174. Kassambara, A., Kosinski, M., and Biecek, P. (2021). survminer: Drawing Survival Curves using 'ggplot2'. (R package version 0.4.9). https://rpkgs.datanovia.com/survminer/index.html.

175. Zhou, W., Chen, T., Chong, Z., Rohrdanz, M.A., Melott, J.M., Wakefield, C., Zeng, J., Weinstein, J.N., Meric-Bernstam, F., Mills, G.B., et al. (2015).

TransVar: a multilevel variant annotator for precision genomics. Nat. Methods *12*, 1002–1003. https://doi.org/10.1038/nmeth.3622.

176. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576. https://doi.org/10.1101/gr.129684.111.
177. Kandoth, C., Gao, J., qwangmsk, Mattioni, M., Struck, A., Boursin, Y., Penson, A., and Chavan, S. (2018). mskcc/vcf2maf: vcf2maf v1.6.16 (Zenodo) https://doi.org/10.5281/zenodo.1185418.
178. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
179. Paten, B., Herrero, J., Fitzgerald, S., Beal, K., Flicek, P., Holmes, I., and Birney, E. (2008). Genome-wide nucleotide-level mammalian ancestor reconstruction. Genome Res. *18*, 1829–1843. https://doi.org/10.1101/gr.076521.108.
180. Paten, B., Herrero, J., Beal, K., Fitzgerald, S., and Birney, E. (2008). Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. Genome Res. *18*, 1814–1828. https://doi.org/10.1101/gr.076554.108.
181. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424. https://doi.org/10.1038/gim.2015.30.
182. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
183. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081. https://doi.org/10.1038/nprot.2009.86.
184. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *Chapter 7*, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76.
185. Basu, S., and Pan, W. (2011). Comparison of statistical tests for disease association with rare variants. Genet. Epidemiol. *35*, 606–619. https://doi.org/10.1002/gepi.20609.
186. Ouspenskaia, T., Law, T., Clauser, K.R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B.A., Le, P.M., et al. (2022). Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. Nat. Biotechnol. *40*, 209–217. https://doi.org/10.1038/s41587-021-01021-3.
187. Niu, L., Stinson, S.E., Holm, L.A., Lund, M.A.V., Fonvig, C.E., Cobuccio, L., Meisner, J., Juel, H.B., Thiele, M., Krag, A., et al. (2025). Plasma Proteome Variation and its Genetic Determinants in Children and Adolescents. Nat. Genet. *57*, 635–646. https://doi.org/10.1038/s41588-025-02089-2.
188. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.
189. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
190. Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S.L., Digles, D., et al. (2018). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. *46*, D661–D667. https://academic.oup.com/nar/article/46/D1/D661/4612963.
191. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. Nucleic Acids Res. *50*, D988–D995. https://doi.org/10.1093/nar/gkab1049.
192. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.
193. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507. https://doi.org/10.1038/nprot.2011.457.
194. Győrffy, B. (2023). Discovery and ranking of the most robust prognostic biomarkers in serous ovarian cancer. GeroScience *45*, 1889–1898. https://doi.org/10.1007/s11357-023-00742-4.
195. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.
196. Strimmer, K. (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics *24*, 1461–1462. https://doi.org/10.1093/bioinformatics/btn209.
197. Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst. *12*, 477–479. https://doi.org/10.1039/C5MB00663E.