Mechanistic Insights into Tumorigenesis from Serum Proteins

Valur Emilsson[1,2,*], Valborg Gudmundsdottir[1,2], Sean Bankier[3], Elisabet A. Frick[2], Thorarinn Jonmundsson[2], Kari Arnarsson[2], Hulda K Ingvarsdottir[2], Heida Bjarnadottir[2], Joseph Loureiro[4], Eva Jacobsen[2], Thor Aspelund[1], Eirikur Briem[5], Lenore J Launer[6], Esther Bastiaannet[7], Tom Michoel[3], Sigurdis Haraldsdottir[2,8], Sigridur K Bodvarsdottir[9], Thorarinn Gudjonsson[10], Nancy Finkel[4], Anthony P Orth[11], Lori L Jennings[4], John R Lamb[12] and Vilmundur Gudnason[1,2,*]

[1]Icelandic Heart Association, Holtasmari 1, IS-201 Kopavogur, Iceland.
[2]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.
[3]Computational Biology Unit, Department of Informatics, University of Bergen, 5020 Bergen, Norway
[4]Novartis Biomedical Research, 22 Windsor Street, Cambridge, MA 02139, USA.
[5]Department of Genetics and Molecular Medicine, University Hospital, Reykjavík, Iceland
[6]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, MD, USA
[7]Cancer Epidemiology, University of Zurich, Epidemiology, Biostatistics and Prevention Institute, CH-8001 Zurich, Switzerland
[8]Department of Medical Oncology, University Hospital, Reykjavik, Iceland
[9]Biomedical Center, School of Health Sciences, University of Iceland, Reykjavík, Iceland
[10]Biomedical Center, School of Health Sciences, University of Iceland and Department of Hematology, University Hospital, Reykjavík, Iceland
[11]Novartis Biomedical Research, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA
[12]Monoceros Biosystems, 12636 High Bluff Drive, Suite 400, San Diego, CA. 92130, USA

*Correspondence: valur@hjarta.is and v.gudnason@hjarta.is

Abstract

Improving early cancer detection would have a transformative effect on patient survival and associated societal costs. Ideally, this would involve tests that are minimally invasive, cancer-type specific and provide mechanistic insights. To address this need, we analyzed associations between 7,523 human serum proteins and 13 cancer types in 5,376 participants from the prospective, population-based AGES Reykjavik cohort. The study included 1,235 cancer cases spanning the digestive, genitourinary, respiratory, and female reproductive systems, as well as skin cancer. The analysis was conducted both longitudinally and cross-sectionally, with adjustments made for various well-established cancer risk factors. After accounting for age, sex, clinical, and lifestyle factors, 526 serum proteins were significantly associated with either prevalent (diagnosed prior to blood draw) or incident (diagnosed after blood draw) clinical presentation of the various types of cancer. Additionally, 776 circulating proteins were influenced by known genetic risk loci for various cancers, including 114 of the 526 mentioned above. Some serum protein associations were shared across cancer types, both prevalent and incident, as well as with genetic susceptibility loci. To contextualize these findings, we integrated our results with both internal and external datasets, including known cancer genes, germline genetic risk loci, tumor- and tissue-specific expression profiles, oncogenes and tumor suppressor genes, and circulating protein networks. This integrative analysis highlights distinct functional categories of protein involvement and reveals the complex and specific etiology of cancer. These findings support the potential for population-level surveillance, early cancer detection, and molecular insights into tumorigenesis.

Introduction

The process of tumorigenesis is one where individual cells gain through a series of incremental steps, the ability to grow and survive independent of central control. Tumorigenesis is frequently a decades-long process marked by the accumulation of genetic alterations that lead to genomic instability and clonal expansion[1,2], which are among the hallmarks of cancer. Environmental factors and genetic susceptibility can expedite this process[3,4]. Although many aspects of tumorigenesis are known[2,5], significant gaps remain, emphasizing the importance of elucidating the underlying mechanisms to advance early detection and preventive strategies. Breast, lung, and colorectal cancers are the most common cancers in women, accounting for 51% of all new cases[6], while prostate, lung, and colorectal cancers account for 48% of all incident cases in males[6]. Although the global burden of cancer-related morbidity continues to rise[7], largely driven by population aging, cancer-related mortality has declined since peaking in 1991, likely due to healthier lifestyle choices, advances in therapies, and improved detection methods[6]. However, emerging evidence suggests that early-onset cancer among younger individuals is on the rise[8]. Despite the decline in mortality, cancer remains the second leading cause of death worldvide[9], surpassed only by heart disease.

Cancer etiology is multifactorial, with genetic and environmental factors contributing to the disease[10]. Germline (inherited), somatic, and epigenetic factors all play a role in the development of cancer[11]. Both rare high penetrance (i.e., *BRCA1* and *BRCA2* genes) and common low penetrance germline variants confer a risk of developing breast and ovarian cancers[12,13]. Recently, it has been shown that common genetic variants contribute to the risk of common malignancies[14], with hundreds of these risk loci identified to date. Cancers also arise from somatic mutations in specific driver genes[15,16], typically ranging from 1 to 10 driver mutations

per genome, depending on the cancer type, which can be sufficient to transform a normal cell into advanced cancer[15]. In addition, widespread copy number alterations, rearrangements, and epigenetic changes are frequently observed, particularly in later-stage cancer[16,17]. Importantly, cancer is now increasingly recognized as a complex systemic disease involving extensive interactions with non-tumor cells and other factors locally and systemically, indicating the need for a comprehensive systems-level understanding of its etiology[10].

Current clinical practice includes screening diagnostics like mammography for breast cancer, Pap smear for cervical cancer, and chest imaging scans for lung cancer though they have suffered from high false positive rates and low compliance[18]. Studies have employed orthogonal mass spectrometry to discover peptides in serum and plasma that are specific for certain established tumors[19]. Only a few such biomarkers have been identified, including the prostate-specific antigen (PSA, aka KLK3)[20], the alpha-fetoprotein (AFP) for liver cancer[21], the carcinoembryonic antigen (CEA) for colorectal cancer[22], and cancer antigen 125 (CA-125, aka MUC16) for ovarian cancer[23]. Recent studies identifying somatic mutations or alterations in DNA methylation within cell-free circulating tumor-derived DNA have been designed to detect both asymptomatic and full-blown cancers[24,25], presenting promising opportunities for early-stage cancer detection[26]. Less research has focused on large-scale, unbiased proteome analysis of proteins associated with different types of cancer. However, a recent study utilizing a multiplex proteomics platform to quantify 1,463 plasma proteins identified several associations with various cancer types in the UK Biobank data[27], highlighting the connection between the plasma proteome and cancer development.

Given these early results, a more comprehensive prospective population-based survey of circulating proteins and their associations to prevalent (prior diagnosis) and future cancers would

be a valuable advancement to the field. Indeed, proteins are essential players in all biological processes, directly influencing the development and progression of diseases like cancer. They also connect genotype variation to disease-related outcomes, thereby revealing disease-specific mechanisms. Although the specific functional roles of most circulating proteins are not yet fully understood, our previous findings suggest that they reflect highly coordinated inter-individual variations in protein levels, indicating the involvement of systemic regulatory pathways or global homeostasis[28,29]. Plasma and serum proteins have been associated with a broad spectrum of diseases of different etiologies[28-38]. While many proteins in biofluids such as serum and plasma have been associated with various diseases and exhibit coordinated regulation, their potential links to cancer remain largely unexplored, highlighting a critical yet understudied area of research.

The current study details the application of the highly sensitive aptamer-based technology to assess the relationship of 7,523 serum proteins to 13 different cancer types in 5,376 older adults from the prospective population-based Age, Gene/Environment Susceptibility Reykjavik Study (AGES) cohort[39]. We investigated the association between the circulating proteome and cancer diagnosis prior to the time of blood draw or diagnosed within the subsequent 13.6 years. Additionally, we investigated sex-specific effects for cancers that are not inherently sex-specific, accounted for established clinical and lifestyle risk factors, and examined the relationship between the serum proteome and genetic risk factors across different cancer types. This study offers a detailed understanding of the serum proteomic signatures associated with various cancer types, providing mechanistic insights into potential biomarkers for early detection and personalized cancer risk assessments.

Results

Study population and study overview

The present study builds on the population-based, prospective AGES cohort[39] of older adults (N = 5,764, mean age 76.6 ± 5.6 years, age range 66-96 years, 57% female), which is extensively annotated for disease risk factors, disease endpoints, comorbidities, genotype and deep serum proteomics data and includes real-time follow-up information. Table 1 summarizes selected baseline characteristics of the prospective, population-based AGES study (n = 5,376), in which 7,523 circulating serum proteins were measured. The table presents sex-stratified demographic, biochemical, clinical, physiological, and anthropometric data, with cancer types grouped by shared organ systems. Data are shown separately for individuals with a prior cancer diagnosis (prevalent cases) and those at future risk (incident cases). These encompass 13 different cancer types, including those of the digestive system (esophagus, stomach, colon, rectum, and pancreas), genitourinary system (kidney, prostate, and bladder), the female reproductive system (breast, ovary, and corpus uteri), respiratory system (lung and bronchus), and skin (melanoma). Incident cancer refers to newly diagnosed cases during the follow-up period. In this study, follow-up for incident cancer cases lasted up to 13.6 years from baseline, with person-years calculated from the participant's first AGES study visit until the earliest of cancer diagnosis, death, or end of follow-up. Prevalent cancer cases at the AGES baseline visit refer to patients who had already been diagnosed with cancer before enrollment in the study, many of whom had likely undergone cancer-related treatment. The number of patients diagnosed with each of the 13 cancer types examined in this study is detailed in Supplementary Table S1, while an overview of the study is provided in Figure 1.

Across 13 distinct cancer types, 1,235 individuals were identified as either prevalent or incident cases (Table 1 and Supplementary Table S2). There were 1,414 diagnoses overall (Figure 2A), with 179 individuals diagnosed with more than one type of cancer. We will use the abbreviations for each cancer type as presented in Figure 2A from this point onward. In the AGES study, 62% of cases (n = 770) were diagnosed with new-onset cancer during the 13.6 year follow-up period. The exceptions to this were cancer cases of the female reproductive system, especially breast cancer (BRC) which had high prevalence of breast cancer patients at the baseline visit (Table 1 and Supplementary Table S1). Supplementary Tables S2-S7 provide descriptive statistics stratified by all 13 cancers combined or categorized by cancer types within the same body system, with prevalent and incident cases considered together. Potential variations in the distribution of known risk factors are indicated. Among the many modifiable risk factors, tobacco smoking, alcohol consumption and obesity continue to be the most important risk factors for the onset of many different types of cancer[40,41]. For example, patients with cancers of the digestive and female reproductive systems exhibit significantly different distributions of BMI categories (Supplementary Tables S3 and S6). Additionally, smoking status showed significant associations with all different cancer types, apart from skin cancer (Supplementary Tables S2-S7). Furthermore, individuals with genitourinary malignancies consumed significantly more alcohol than those without these cancers (Supplementary Table S4). In addition to the 13 cancer types described above, there were 684 individual cases of other distinct cancer diagnoses in the AGES study. Unlike the strong associations usually seen between the 13 prevalent cancers and different lifestyle factors, these 684 patients showed no such links to common risk factors (Supplementary Table S8).

To evaluate the association between serum proteins and incident cancers diagnosed after the blood draw, a time-to-event Cox proportional-hazards model[42] was applied to the log2-transformed serum proteomics data. For examining the associations of serum proteins with prevalent cancers diagnosed before blood draw, logistic regression analysis was used. Unless otherwise noted, all regression analyses were adjusted for age, sex, and estimated glomerular infiltration rate (eGFR), which we will refer to as the "standard covariates". To account for multiple hypothesis testing, the false discovery rate (FDR) was adjusted using the Benjamini-Hochberg method[43]. Leveraging the opportunity to compare different patient groups within the AGES study, providing a form of internal validation for findings beyond a single cancer type, and to maximize discovery across the sample, we report results using both FDR thresholds of <0.05 and <0.10 as noted.

In this study, we explored the relationships between thousands of proteins and 13 types of cancer, as well as a broader category encompassing any type of cancer. Given the extensive scope of the analysis, we have highlighted select findings in the Results section to illustrate the study's breadth and significance. To underline key cancer-related biological processes and mechanisms represented in the findings, we attempt to categorize the highlighted cancer-associated proteins into four functional groups based on their potential roles in cancer: tumor-specific (oncogenes and/or tumor suppressor genes), tissue-specific (reflecting the tumor's tissue of origin), genetic susceptibility (common low-penetrance germline variants), and a broad category of tumor–host interaction proteins, encompassing external factors (e.g., lifestyle, systemic influences), tumor microenvironment (TME), and epithelial–mesenchymal transition (EMT). A comprehensive and detailed discussion of each of the different cancer types is

provided in the Supplementary Text (of Supplementary Information). To facilitate readability, each tumor type in the Supplement is organized based on its tissue of origin.

Observational study linking the serum proteome to incident cancers

In the Cox regression analyses, 328 proteins were associated with at least one of the 13 incident cancers at FDR < 0.05 (Supplementary Table S9, Table 2, Figure 2B, and Supplementary Text), and 535 proteins at FDR < 0.10 (Supplementary Table S9), using various covariate adjustments, as well as sex-specific analyses for cancers that are not inherently linked to a particular sex. The overlap in findings across different cancer types was limited to only a few serum proteins, highlighting the origin-specificity of these associations. For instance, at FDR < 0.05, SIGLEC6 was associated with incident bladder cancer (BLC) and pancreatic cancer (PAC), while HAVCR1 was linked to incident kidney cancer (KIC) and lung cancer (LUC) (Supplementary Table S9). Additionally, 11 other proteins, including MPP2, TRAT1, TMEM106A, UBE2E3, TREM2, CREBBP, ITIH1, and RAB22A, were associated with multiple types of incident cancer at FDR < 0.05 or < 0.10 (Supplementary Table S9). Interestingly, CREBBP is a tumor-specific protein encoded by well-established tumor suppressor gene[44,45], with an enriched mutation frequency across several cancer types (https://depmap.org/portal)[46]. The other proteins may play diverse roles in tumor–host interactions, including immune modulation (SIGLEC6, MPP2, TRAT1)[47-49], EMT and/or metastasis (TMEM106A, RAB22A)[50,51], TME remodeling (TREM2)[52], and cellular senescence (UBE2E3)[53]. The kidney-specific expression of HAVCR1 (https://www.proteinatlas.org/), together with its association with incident KIC, suggests it may reflect tumor burden and could potentially contribute to organ-specific mechanisms of tumorigenesis.

Several genes, gene expression traits, and proteins corresponding to serum proteins associated with new-onset cancers in the AGES study have been previously linked to their respective cancer types. For instance, *CXCL8* (aka *IL8*), is a known pro-tumor effector in colon cancer (COC) involved in resistance to immune checkpoint inhibitors and TME modulation[54,55]. Consistently, our study showed a positive association between CXCL8 and COC (Supplementary Table S9, Figure 3A). Furthermore, KLK3 (aka PSA), a glycoprotein enzyme primarily produced by the prostate gland and one of the most recognized prostate cancer (PRC) biomarkers linked to various PRC-related clinical outcomes[56], was positively associated with incident PRC in the AGES study (Supplementary Table S9, Figure 3A). Another example is acid phosphatase ACP3 (aka PAP), a protein produced specifically by the prostate gland and strongly associated with incident PRC (Supplementary Table S9, Figure 3A). Elevated serum levels of ACP3 have been associated with accelerated progression of PRC[57], and the protein is a key target for cellular immunotherapy, which has been shown to improve survival in men with metastatic prostate cancer[58]. The observed associations of the prostate-specific proteins KLK3 and ACP3 with incident prostate cancer are anticipated, given their established use as clinical biomarkers for the disease. WNT10B is an additional example, associated with incident BRC in the AGES study (Supplementary Table S9, Figure 3A), and previously linked to BRC[59,60], particularly aggressive subtypes like triple-negative BRC. This association is thought to be driven by activation of the Wnt/β-catenin pathway, which is crucial for development and tissue homeostasis but also contributes to poor prognosis and increased metastasis in BRC patients[59,60]. Several other proteins or associated genes have previously been linked to various cancer types. Examples include CTNNB1[61], BCL2L14[62], WNT7A[63-65], PTPN6[66], PRKCZ[67], WFDC2 (aka HE4)[68-70] and EPOR[71], all of which were associated with one or more incident cancer type at an FDR < 0.05 in

the AGES study (Supplementary Table S9). While *CTNNB1* is a well-established canonical oncogene[44,45], *EPOR* has more recently been implicated as an oncogene[44,45], with truncated rearrangements identified in certain forms of leukemia[72,73]. Cancer-related roles of the remaining proteins are likely mediated through tumor–host interactions. These include inflammation and the mucin network (BCL2L14)[74], the TME and metastasis (WNT7A)[75], as well as various immune-related functions (PTPN6, PRKCZ, and WFDC2)[67,76,77].

Observational study linking the serum proteome to prevalent cancers

Logistic regression of prevalent cancers with sufficient case numbers ($n \geq 10$) identified 223 associated serum proteins at FDR < 0.05 (Supplementary Table S10, Table 2, Figure 2B, and Supplementary Text), and 350 proteins associated at FDR < 0.10. (Supplementary Table S10). Consistent with the results from the analysis of incident cancer types, the overlap of proteins associated with different prevalent cancer types was confined to just a few, notably the trefoil factors TFF2 and TFF3, which were linked to both stomach cancer (STC) and BRC at FDR < 0.05 (Supplementary Table S10, Figure 3B). TFF2 and TFF3 are specifically expressed in mucin-secreting epithelial tissues of the gastric mucosa and gastrointestinal tract[78], respectively, where they contribute to maintaining epithelial integrity[79]. While their altered expression in STC suggests a context-dependent role in tumor biology, their involvement in BRC remains unclear. An additional four proteins, C7, OMD, LRRC15, and CLIC4, were associated with more than one prevalent cancer type at FDR < 0.05 or <0.10, primarily STC, BRC, PRC, and rectal cancer (REC) (Supplementary Table S10). For instance, the serum protein CLIC4 was positively associated with prevalent STC at FDR < 0.05 (Supplementary Table S10, Figure 3B), and with prevalent BRC and REC at FDR < 0.10 (Supplementary Table S10). Notably, CLIC4, a protein involved in ion transport and present in both soluble and membrane-bound forms, has been

linked to various cancer types, primarily due to its role in the TME and regulation of oxidative stress[80], with elevated expression serving as a negative prognostic factor for patient survival[81]. The serum proteins C7, OMD, and LRRC15 have also been linked to pathways involved in tumor-host interactions[82-84]. Well-established cancer-associated genes encoding proteins like GKN2 and MSMB, which are known to be linked to stomach and prostate cancers[85,86], respectively, were specifically associated with the prevalent forms of these cancer types in the AGES study, with no connection to other prevalent cancers (Supplementary Table S10, Figure 3B, and Supplementary Text). Notably, the stomach-specific protein CBLIF, whose role in cancer remains unclear, also showed specificity for prevalent STC in this analysis (Supplementary Tables S10-S11, Figure 3B). These proteins are primarily expressed in the organs where their respective tumors arise: GKN2 and CBLIF in the stomach, and MSMB in the prostate. All three proteins exhibit a negative association with their respective prevalent cancer types in the AGES study (Supplementary Tables S10, Figure 3B). The inverse relationship of these proteins with the respective cancers diagnosed before protein measurements may be attributed to some patients in the cohort having undergone partial or total gastrectomy or prostatectomy. Supporting the negative association results in the AGES study, however, both GKN2 and MSMB have previously been shown to be secreted and downregulated in their respective cancers[87,88]. Therefore, treatment alone is unlikely to fully account for the observed reduction in these protein levels.

In many cases, the proteins associated with cancer are specifically expressed in the tissue of tumor origin. The proteins linked to prevalent STC are particularly noteworthy due to their enrichment in the stomach and the broader gastrointestinal tract, as illustrated in Figure 4. More to the point, with 25 cases diagnosed with STC prior to study enrollment and protein

measurements (Figure 4A), a total of 33 serum proteins were significantly associated with prevalent STC at FDR < 0.05 (Figure 4B). Many of the proteins associated with prevalent STC, notable for their previous links to gastric cancer as mentioned above, are exclusively expressed in regions of the digestive system (Figure 4C-D, and Supplementary Table S11). Supplementary Figures S1–S5 display volcano plots of serum proteins associated with various incident and/or prevalent cancer types, organized by the body system corresponding to the tumor's site of diagnosis.

The impact of additional covariate adjustments and sex-specific analyses

The most common modifiable risk factors for cancer include tobacco use, alcohol consumption, and excess body weight (overweight and obesity). In addition to adjusting for age, sex (where applicable), and eGFR, we accounted for these factors when supported by epidemiological evidence linking them to specific types of cancer. Considering the established epidemiological associations between body height and the risk of PRC[89], and blood pressure and KIC[6], we included adjustments for these factors in our analysis (Supplementary Tables S9-S10, Supplementary Text). Some cancer types were more influenced than others by the inclusion of additional adjustments for well-established risk factors specific to each cancer type, and in some cases additional proteins were detected (Supplementary Text, Supplementary Figure S6A-N). Finally, for many cancers that are not inherently sex-specific, new serum protein associations were identified when the data was analyzed separately for each sex (Supplementary Tables S9-S10, Supplementary Text).

The cancer type most affected by the inclusion of additional covariate adjustment beyond the standard covariates, was LUC (Supplementary Figure S6I). This is not surprising, considering the high proportion of current smokers in the combined LUC patient group (38.8% vs. 11.4%)

and the low proportion of never smokers (7.6% vs. 43.6%) (Supplementary Table S5). In this study, 216 serum proteins were significantly associated with incident LUC (Supplementary Table S9), and six proteins with prevalent LUC (Supplementary Table S10), at FDR of < 0.05 when using standard covariates only. After adjustment for smoking, 10 proteins remained significant for the incident LUC, while no proteins remained significant for prevalent LUC. The proteins still significantly associated with incident LUC included WFDC2, SCGB3A1, CLEC3B, CNTN3, and GDF11 (Supplementary Table S9). Notably, WFDC2, SCGB3A1, and CLEC3B are differentially regulated in tumors of LUC or lung squamous cell carcinoma (Supplementary Table S11). For instance, SCGB3A1 is specifically expressed in the lung (Supplementary Table S11), and lower levels of this protein have been reported in non-small cell lung cancer, suggesting a role in modulating tumor progression[90,91]. Moreover, SCGB3A1 and WFDC2 are markers of early secretory cells in the lung conducting airway epithelium and may thus help identify distinct cell populations within tumors[92,93]. The serum proteins CLEC3B and GDF11 have been linked to tumor-host interactions, through their involvement in EMT[94] and the TME[95], respectively. Further, GDF11, a negative regulator of muscle growth and a member of the transforming growth factor-beta (TGF-β) superfamily, has been associated with muscle wasting in lung cancer-related cachexia in mouse models[96]. Given the strong influence of smoking on lung cancer, proteins associated with the disease may reflect interactions with smoking status, potentially capturing biological responses to tobacco exposure.

Since we adjusted for eGFR as part of the standard covariates in all our analyses, we anticipated that this would most likely influence the observational findings related to KIC, as tumors affecting the kidneys are likely to cause renal dysfunction. In fact, when adjusting for age and sex alone, 624 serum proteins were associated with prevalent KIC at an FDR < 0.05, and these

proteins were also independently linked to eGFR (data not shown). When eGFR was included as a covariate, three proteins were found to be associated with prevalent KIC (Supplementary Table S10). Excluding eGFR from the incident KIC adjustment model yielded three significant protein associations, including HAVCR1, which also appeared among the four proteins identified when eGFR was included (Supplementary Text). This may indicate that renal dysfunction was already present in KIC patients at the baseline visit but had not yet manifested during the prodromal phase of the disease. The proteins associated with either incident or prevalent KIC using the standard covariate adjustment remained significantly linked to KIC after further adjustment for additional covariates (Supplementary Tables S9-S10, Supplementary Text).

In certain types of cancer, including incident esophageal cancer (ESC), prevalent STC, prevalent REC, prevalent BRC, incident and prevalent corpus uteri cancer (CUC), incident ovarian cancer (OVC), and prevalent cutaneous melanoma (CMC), full adjustment led to a reduction in the total number of protein associations (Supplementary Figure S6A-B, D, J-L, M, Supplementary Text), though many associations remain. In contrast, for other cancers, full adjustment had little effect or led to the identification of additional protein associations, as seen in incident STC, incident and prevalent KIC and PRC, incident and prevalent BLC (Supplementary Figure S6B, F-H, and Supplementary Text). Among proteins highlighted above, several remained robust to additional covariate adjustments beyond the standard set, including KLK3 and ACP3 for incident PRC, MSMB for prevalent PRC, and WFDC2 and SCGB3A1 for incident LUC, among others (Supplementary Tables S9–S10). Proteins that became significant only after full adjustment included, for example, APOM for incident PRC, TNFSF4 (aka OX40L) for incident STC, POLR1C and CHST12 for incident PAC, TSTA3 for incident KIC, and PAGR1 and UBE2E3 for incident CUC (Supplementary Table S9 and Supplementary Text). Although the roles of

POLR1C and TSTA3 in cancer remain unclear, the other proteins have been implicated in various aspects of tumor-host interactions[97-99]. For example, immunotherapy may increase OX40L levels, a stimulatory checkpoint on dendritic cells, while concurrently decreasing the expression of the inhibitory checkpoint PD-L1 on tumor cells[100]. This dual effect enhances the effectiveness of antitumor immunotherapy in gastrointestinal cancer[100]. Both POLR1C and CHST12 are differentially regulated in pancreatic tumors compared to normal samples (Supplementary Table S11). High expression levels of TSTA3 in kidney tumors are associated with poor survival in KIC patients[101]. In summary, presenting results with both standard and full adjustment offers a more refined and comprehensive understanding of the findings, ensuring that potential confounding factors are adequately controlled.

Many cancers are inherently sex-specific due to the presence of organs unique to either males or females, such as the prostate, ovaries, and corpus uteri, as well as most cancers in the breast, while others affect both sexes. In the AGES study, some cancers that are not inherently sex-specific still showed a skewed sex distribution. For instance, we found that males outnumber females in cancers like ESC, BLC, and KIC, while other cancers have a more balanced male-to-female ratio (Supplementary Text). Analyzing the sexes separately for these cancers is important for fully understanding the distinct biological and molecular characteristics of cancer in males and females. Thus, the sexes for these cancers were analyzed both together and separately, resulting in identification of several protein associations in one sex (Supplementary Tables S9-S10). In cancers such as COC, STC, REC, BLC, KIC, and CMC, new protein associations were identified when the sexes were analyzed separately (Supplementary Tables S9-S10, Supplementary Text). These include proteins such as BCL2L14, noted above, which is positively associated with incident COC in males only and has been previously linked to colon

tumorigenesis[74]. In females only, CUL1 is positively linked to incident STC, but elevated CUL1 levels have also been associated with poor prognosis in gastric cancer patients[102]. Other examples include CSF1, which is positively correlated with incident BLC in males only, and high CSF1 levels have been linked to poorer overall survival in BLC patients[103]. The proteins CUL1 and CSF1 may influence tumor–host interactions through distinct mechanisms, CUL1 by modulating the DNA damage response and apoptosis[102], and CSF1 through immune regulation of the TME[104]. ASMTL is another protein positively associated with incident CMC in females only (Supplementary Table S9) and may contribute to melatonin biosynthesis in a sex-specific manner[105], a pathway linked to circadian regulation and tumorigenesis[106]. In the male-only analysis of ESC, additional protein associations emerged upon excluding females, suggesting a male-dominant effect in protein associations within ESC in the AGES study (Supplementary Text). Notably, significant protein associations were only observed for incident PAC when the sexes were analyzed separately, with no overlap between the findings in males and females (Supplementary Table S9). While analyzing sexes separately may reduce statistical power due to smaller sample sizes, sex-specific biological differences such as hormonal, immune, or genetic factors, can uncover distinct molecular signatures, as demonstrated in our study. Pooling sexes together may in some cases obscure these differences, making separate analyses more informative despite the reduced power.

Proteins associated with both incident and prevalent cancer states

All prevalent cases were excluded from the analysis of new-onset cancer cases to ensure no overlap between the two patient groups. The groups of prevalent and incident cancer cases are therefore independent of each other in terms of both analysis and results. A total of 25 protein associations overlapped between incident and prevalent cancer types at FDR < 0.05 (Figure 5A).

The examples include four proteins, i.e., WFDC2, CLEC3B, CRTAC1, and KLK3, that showed overlap for the same cancer type, while others, including trefoil factors TFF1-3, MET (aka c-MET), GDF15, EGFR, and MSMB, were associated with different cancer types (Supplementary Tables S9-S10). Interestingly, both MET and EGFR are considered canonical oncogenes[44,45], playing pivotal roles in tumorigenesis. Both CRTAC1 and GDF15 are implicated in tumor-host interactions through the EMT process[107,108], while the roles of the other overlapping proteins have been discussed above.

While the direction of effect was consistent across the disease states (prevalent or incident) for most of these proteins, some proteins, including KLK3, MSMB and the trefoil factors TFF1 and TFF2, exhibited opposite directions of effect depending on the disease state and/or cancer type (Supplementary Tables S9-S10). For instance, KLK3, which was directly associated with an incident PRC, is inversely related to prevalent PRC (Figure 5B). The negative correlation between KLK3 and prevalent PRC may reflect the impact of prostatectomy or androgen deprivation therapy, treatments likely undergone by many individuals diagnosed before enrollment in the AGES study, which are known to lower PSA levels[109], the product of the *KLK3* gene. Figure 5C displays the associations of the 25 overlapping proteins with various cancer types and conditions.

Analysis of a combined group of patients with a history of any cancer type

While many findings are cancer-type specific, previous research, especially pan-cancer genetic studies, has revealed shared genetic associations across multiple cancer types[110,111]. Since some serum proteins may serve as pan-cancer markers reflecting fundamental cancer biology, and to enhance the power to detect new associations with the 7,523 serum proteins, we pooled cases with a diagnosis of any cancer type (ATC), including the 1,235 previously analyzed cases and an

additional 684 cases with cancers other than the above 13 types, such as cancers of the larynx, brain, testis, liver, small intestine, Hodgkin lymphoma, multiple myeloma, and several other cancer types, most represented by fewer than 10 patients. Specifically, the associations between the 7,523 serum proteins and a group of 1,916 individuals with ATC were assessed. Overall, the ATC study group included 835 cases of prevalent cancers (499 females) and 1,081 cases of incident cancers (535 females).

A total of 291 proteins were associated with incident ATC at an FDR < 0.05 (Supplementary Table S12, Figure 6A), while 55 proteins were linked to prevalent ATC at the same threshold (Supplementary Table S13, Figure 6B). These associations were identified through both combined and sex-specific analyses. Additionally, considering the diversity of combined cancer types, we employed fully adjusted models that accounted for age, sex, eGFR, BMI, alcohol use, and smoking status: key independent risk factors across various cancer types, as described above and in the Supplementary Text. While numerous new protein associations were identified, i.e. 189 for new-onset ATC and 28 for prevalent ATC, many proteins were also found in the analyses of individual prevalent or incident cancer types (Figure 7A). Of the 291 proteins linked to incident ATC, 104 (36%) overlapped with those identified through analyses of individual cancer types (Figure 7A). Notable serum proteins, such as WFDC2 and GDF15, which are components of the Cancer Seek blood test panel for early detection of various cancers[25], were associated with incident ATC in our study as well as with incident LUC (Supplementary Tables S9 and S12). Similarly, 25 (44%) of the proteins associated with prevalent ATC were also identified in individual cancer analyses (Figure 7A). Two proteins, KLK3 and TFF3, were identified across all four study groups (Figure 7A). The prostate-specific protein KLK3, as expected, was exclusively linked to males due to its specific association with PRC. Notably,

KLK3 was no longer significantly associated with incident or prevalent ATC when PRC cases were excluded from the ATC group (data not shown). This also demonstrates that the aptamer-based method effectively differentiates tissue-specific expression. In line with the differing effects seen in incident versus prevalent PRC (Figure 5B), KLK3 serum quintiles showed opposing associations with incident and prevalent ATC (Supplementary Figure S7A), although the effect sizes were attenuated in the combined ATC group relative to the individual cancer types (Supplementary Figure S7B). In contrast, TFF3 was positively associated with several individual cancer types, including incident LUC and prevalent BRC, STC, and PRC (Supplementary Tables S9-S10, Figure 5C). These findings align with previous research that associated TFF3 with these and other cancer types[112]. Finally, a recent study examined the plasma proteome in mice with transplanted human lung, breast, colon, or ovarian tumors to determine the timing of protein detection in plasma[113]. Notably, we observed a significant enrichment of serum proteins associated with ATC across various human tumor xenograft models (Supplementary Figure S8A-D, Supplementary Text). Proteins such as CXCL8, the top hit for incident COC (Supplementary Table S9), and the proto-oncogene MET were consistently identified across all cancer xenograft models. In conclusion, pooling cancer cases is valuable as it produces results that somewhat mirror those of individual cancer analyses, albeit with smaller effect sizes, while also uncovering numerous new protein associations, likely due to increased statistical power from a larger sample size and new cases included.

Interestingly, among the genes encoding serum proteins associated with incident or prevalent cancers, 24 were classified as oncogenes and 11 as tumor suppressor genes[44,45]. These included for instance well-known oncogenes such as *EGFR*, *MET*, *RET*, *FGFR1*, and *CTNNB1*, as well as tumor suppressors like *CREBBP*, *ARID1A*, *TP53*, and *CDH1* (Supplementary Tables S9–S10,

Supplementary Tables S12–S13). Pathway and tissue enrichment analyses (GSEA or ORA; see Supplementary Text) of proteins associated with individual or any cancer types often revealed enrichment in relevant tissues and oncology-related pathways (Supplementary Tables S14A–B and S15A–B), though results varied with the number of cancer-associated proteins.

Associations between serum proteins and germline genetic cancer risk

The variability in levels of most circulating proteins can be partly attributed to inherited (germline) genetic variation, offering additional support for the observational associations of proteins to cancer and enabling the identification of proteins causally linked to specific cancer types. We explored the relationships between the 7,523 serum proteins that represent over 30% of all annotated human protein-coding genes, and the latest meta-analyzed GWAS for different cancer types (Supplementary Text). Additionally, we investigated pan-cancer GWAS that encompassed a wide range of cancers[111,114], including those analyzed here and others not covered in this study.

A total of 300 independent GWAS lead SNPs for 13 different cancer types examined in this study were associated ($P < 0.00001$) with 737 proteins, represented by 800 aptamers (Supplementary Table S16). In contrast to the distinct cancer types identified in the observational analysis (see above), relationships between serum proteins and genetic risk factors of cancer were more widely shared across multiple cancers, with 210 proteins (28.5%) associated with more than one cancer type (Supplementary Table S16). For instance, 23 proteins regulated by cancer GWAS risk loci, either in *cis* or *trans*, were linked to four or more distinct cancer types (Supplementary Table S16). Additionally, protein hotspots such as those at chr. 3p21.1 (*ITIH1* and *ITIH3* loci), chr. 6p21.3 (*MHC* locus), chr. 9q34.2 (*ABO* locus), chr. 12q24.1 (*SH2B3* locus), chr. 14q32.1 (*SERPINA1* locus), chr. 14q32.3 (*AKT1* locus), and chr. 19q13.3 (*FUT2* locus)

explained many of the associations between genetic variants and serum protein levels linked to both individual cancer types and across different types of cancer (Supplementary Table S16). This suggests a notable pleiotropic effect in the molecular etiology of different cancer types. Moreover, these hotspots suggest phenotyping pleiotropy, as they have been implicated in numerous diseases beyond cancer[29,30]. Notably, *AKT1* at the hotspot locus 14q32.3 is a canonical oncogene with gain-of-function mutations implicated in various cancer types[44,45], while *SH2B3*, located at the 12q24.1 hotspot locus, functions as a tumor suppressor gene with loss-of-function mutations linked to lymphoblastic leukemia[44,45,115]. However, the expression levels of AKT1 and SH2B3 were not linked to the corresponding lead cancer-associated SNPs. Some proteins regulated in *cis* by these established cancer risk loci include for example chymotrypsinogen B2 (CTRB2) for PAC and alpha-2-glycoprotein 1, zinc-binding (AZGP1) for colorectal cancer (Supplementary Table S16). Interestingly, CTRB2 is specifically expressed in the pancreas[116], is downregulated in pancreatic tumors[117], and low CTRB2 levels are linked to higher PAC risk (Supplementary Table S16), suggesting that reduced CTRB2 may play a causal role in PAC. As another highlighted protein associated with genetic cancer susceptibility, AZGP1 negatively regulates angiogenesis and may play roles in both TME and EMT[118,119].

Of the 737 genetically affected proteins, 109 were also identified in the association analysis of various incident and/or prevalent cancer types (Supplementary Tables S9-S10, Supplementary Table S16, Figure 7B) representing a significant enrichment (FET P-value = 4.6×10^{-15}). Among the overlapping proteins, 14 were associated with the same cancer type in both the observational and genetic analyses. Eight proteins, i.e., CD163, CDH5, GOLM1, MET, MSMB, OMD, SERPIND1, and TFF1, were shared across all three study groups (Figure 7B). The roles of MET, MSMB, OMD, and TFF1 in cancer, including tumor specificity, genetic associations, tissue-

specific expression, and tumor-host interactions, have been discussed above. The serum proteins CD163, CDH5, GOLM1, OMD, and SERPIND1 have been linked to various aspects of tumor-host interactions, including immune checkpoint regulation, EMT, TME, and metastasis[83,120-123]. Several of these overlapping proteins are strongly regulated by *cis*-acting genetic variants (Supplementary Figure S9).

Integrating multiple cancer types into a single pan-cancer study not only reinforces associations identified in individual cancer cohort studies but also highlights shared molecular etiology across multiple cancer types. Unlike the ATC, which combines cases from different cancer types, the pan-cancer genomics analysis simultaneously tests genotype associations with different individual cancer types. Pan-cancer GWAS have identified numerous risk loci shared across multiple malignancies[111,124]. One study examining 18 cancer types in 408,786 individuals of European ancestry from two large independent cohorts, revealed a significant number of common genetic risk factors among different types of cancer, with 25 genomic regions containing 136 independent SNPs that were associated with at least two cancer types[111]. We identified 34 of these SNPs as controlling one or more serum proteins, encompassing 158 proteins in total (Supplementary Table S17). These findings introduce 39 additional serum protein associations with genetic risk factors for individual cancers, increasing the total to 776 proteins influenced by known genetic risk factors for cancer, 114 of which were also identified in the observational analysis. Notably, several independent variants on chromosome 6 within the MHC region influenced the same proteins, sometimes in opposite directions, as detailed in Supplementary Table S17. According to Rashkin et al.[111], these variants are associated with both distinct and overlapping sets of cancers. Several proteins linked to genetic variants in the pan-cancer study were also found among serum proteins affected by genetic variants associated with

specific cancer types mentioned earlier (Supplementary Tables S16-S17). The collective evidence presented here points to a substantial shared molecular basis across diverse cancer types.

It is noteworthy that 38 genes encoding serum proteins linked to germline genetic susceptibility loci for cancer are classified as either oncogenes and/or tumor suppressors[44,45], including the well-known oncogene *MET* and the prototypical tumor suppressor gene *TP53* (Supplementary Tables S16-S17). What these tumor-specific proteins have in common is that they are all regulated by distal *trans*-acting genetic variants, which exert weak or modest effects on protein levels (Supplementary Tables S16-S17).

Integrating the cancer-associated proteins with internal and external data sources

Given the limited number of studies that match the depth of the proteome analyzed in this study, we instead focused on the representation of genes encoding the cancer-associated proteins within the broader set of publicly known cancer-related genes. For this, we utilized the Geneshot search engine[125] with the term "cancer gene" which produced 9,952 entries (~50% of all human protein encoding genes) ranked by the number of associated publications, with each entry having at least one publication. We performed a hypergeometric test to assess the enrichment of genes encoding proteins linked to incident and/or prevalent cancers, within the top 20% of the ranked cancer genes. Our findings show a significant enrichment of cancer-associated protein-coding genes among the most frequently cited cancer genes: for proteins associated with incident cancer types ($P = 3\times10^{\square\square}$), prevalent cancer ($P = 4\times10^{\square 11}$), incident ATC ($P = 3\times10^{\square 11}$), and prevalent ATC ($P = 0.00008$) (Figure 8A). This analysis included well-established cancer genes such as the oncogenes *FGFR1*, *MET*, *EGFR*, and *CTNNB1*, as well as *CXCL8* and *KLK3* highlighted above, and *NCAM1* and *CCK* which have been linked to tumor–host interactions within the TME[126,127].

The proteins encoded by these genes were associated with future cancer risk in the present study. Additionally, several highly cited genes including the oncogenes *RET* and *HSP90AA1*[44,45], the tumor suppressor *TP53*[44,45], and others such as *CDH17*[128], *HSP90AB1*[129], and *PIK3C3*[130], which play diverse roles in tumor-host interactions, encode proteins that were associated with prevalent cancer in the AGES study. In contrast to the enrichment observed among the most highly cited cancer genes, a similar analysis of the bottom 20% of least-cited cancer-related papers revealed no evidence of enrichment, with only 0 to 4 overlapping genes (P = 0.967).

A similar hypergeometric test revealed a highly significant enrichment of genes encoding proteins controlled by genetic cancer risk loci in the AGES study among highly cited cancer genes ($P = 9\times10^{-26}$) (Figure 8B), surpassing the enrichment found for proteins identified in the observational analysis. These included, for example, the oncogene *MET* and the tumor suppressor genes *TP53* and *CDH1*[44,45], along with cancer-related genes *CDH17*, *EGF*, and *TGFB1*, all of which are known to play roles in tumor–host interactions[128,131,132]. EGF and TGFB1 exhibit *cis*-acting associations with their respective cancer types, whereas the others are *trans*-acting (Supplementary Table S16). No significant enrichment was found when compared with the bottom 20% of the least-cited cancer genes ($P = 0.693$). In summary, these findings suggest that many of the serum proteins identified in our study have well-established mechanistic roles in cancer biology.

Given the intricate and systemic nature of cancer, utilizing biological networks provides insights into the mechanistic relationships involved in cancer onset and progression. We previously reconstructed the first circulating serum protein networks in humans, encompassing both the undirected co-regulatory network[29] and, more recently, the circulating causal protein network (CPN)[133], which was mapped using causal inference analysis. Although individual proteins

associated with incident or prevalent cancer show minimal overlap (see above), proteins from these two groups share co-regulatory network modules (Supplementary Table S18, Supplementary Text, and Supplementary Figure S10), suggesting they may be interconnected and involved in related biological processes. The co-regulatory modules they share have demonstrated strong associations with cardiometabolic and cardiovascular diseases, as well as overall and disease-specific survival[29].

The CPNs offer further insights into causal relationships between proteins that are not apparent in co-regulatory networks, as they distinguish between cause and correlation. We examined the enrichment of cancer-associated proteins within specific CPN subnetworks, which include both genetic regulators and target proteins[133]. To provide a more comprehensive analysis, we incorporated protein findings from both observational and genetic studies. Notably, many cancer-associated proteins were highly enriched in the CPN networks, with a particularly strong enrichment observed for proteins linked to cancer risk loci (Supplementary Table S19). Overall, 18 CPN subnetworks were enriched for proteins associated with new-onset cancer, four for prevalent cancers, and 42 for proteins linked to cancer risk loci (Supplementary Table S19). Considering the predictive power of biological networks[29], we highlight the CPN networks enriched for various types of incident cancers along with their interconnectedness (Supplementary Figure S11A-B). Moreover, within solid tissues, a subset of these networks adopts a cascade-like configuration and is strongly enriched for physical protein–protein interactions (PPIs) (6 expected edges vs. 29 observed edges; $P = 8 \times 10^{-11}$) (Figure 9A-B). For example, several network regulators converge on PTPN11 (aka SHP2) (Figure 9A), a protein encoded by a well-established canonical oncogene[44,45], which regulates 83 serum proteins including the oncogenic protein AKT1[133], and is associated with prevalent BRC in this study

(Supplementary Table S10). Although these network regulators linked to PTPN11 have not been directly implicated in this study, they represent potential candidates in cancer biology. The PPI network revealed seven distinct functional clusters encompassing proteins involved in complement activation, protein folding, proteasome core complex, platelet development, NK-cell lectin-like receptor binding, allograft rejection, and vault protein inter-alpha-trypsin domain (Figure 9B). Interestingly, the platelet-related PPI network includes PTPN11, while many other proteins across the various PPIs are associated with genetic susceptibility to different cancers. Several of these risk loci, for example, ITIH1, ITIH3, CFB, and MICA, also exhibit strong *cis*-acting regulation in the context of their respective cancers (Figure 9B, Supplementary Table S16). The accumulated data are in line with our previous observation that these networks are strongly represented in functional and physical PPIs within solid tissues as well as serum protein networks that also span tissue boundaries[28,29].

In the main text, we have highlighted selected cancer-associated proteins by categorizing their roles into four key areas: tumor-specific functions, tissue specificity, genetic susceptibility, and a broad category related to tumor–host interactions. The latter includes external factors such as lifestyle and systemic influences (e.g., metabolism, inflammation), as well as tumor-intrinsic processes like the TME, EMT and metastatic processes. This evidence-based and hypothesis-driven protein classification is illustrated in Figure 10 with representative examples that underscore key cancer-related biological mechanisms reflected in our findings. Supplementary Table S20 also summarizes all discussed proteins along with relevant information about their classification. In summary, this framework promotes a more holistic view of cancer etiology and may aid in prioritizing proteins with potential actionable or novel roles for further study.

Discussion

Cancer is a complex and heterogeneous disease marked by uncontrolled cell growth, invasion to surrounding tissues, and metastasis. It evolves over an extended period through the accumulation of advantageous genetic and epigenetic alterations that impair normal cellular functions and promote malignant transformation. Despite significant advances in treatment, cancer remains the second leading cause of death globally, highlighting the urgent need for better strategies in early detection and prevention. The research presented here has focused on identifying serum proteins associated with both current and future cancers across 13 distinct types. Among the many cancers examined, this study includes some of the most aggressive and treatment-resistant types, such as esophageal, gastric, pancreatic, and ovarian cancers, which often lack effective early diagnostic tools and are diagnosed at advanced stages, contributing to high mortality rates. The purpose of this work was to facilitate early detection of different cancers and to provide insight into the molecular mechanisms driving tumorigenesis.

Serum proteins linked to past diagnoses and future cancer risk

In the prospective, population-based AGES study, serum levels of 7,523 proteins were measured in 5,376 individuals. More than half (53%) of all cancers is diagnosed in individuals aged 65 or older, the entry age for the AGES study, and this burden is projected to grow significantly as the global population ages[9,134]. More to the point, the contributing factors to cancer in older adults may differ from those in younger individuals, due to multiple age-related changes, including, for instance, impaired DNA repair, chronic inflammation, cellular senescence, and cumulative exposure to carcinogens[135]. We examined associations between serum proteins and 13 different cancer types (each with ≥10 cases), including both cancers diagnosed prior to blood collection (prevalent) and those diagnosed within 13.6 years after sampling (incident). The analysis

revealed 526 protein-cancer associations at FDR < 0.05. In general, the associated proteins were specific to individual cancer types, and those linked to prevalent versus incident cancers showed minimal overlap. Nevertheless, several proteins were associated with two or more cancer types or timing, suggesting shared molecular pathways across different cancers.

It is important to note that the prevalent cancer cases in our study reflect individuals with a prior diagnosis of cancer. As such, this group likely includes a heterogeneous mix of clinical states, ranging from patients who may have undergone successful treatment with no current tumor burden, to those with metastatic disease or experiencing relapse. This diversity in disease stage, treatment history, and tumor activity at the time of sampling introduces biological variability that should be considered when interpreting the associations between protein levels and prevalent cancer.

Proteins associated with genetic susceptibility loci

Large-scale GWAS meta-analyses have identified numerous genetic loci associated with elevated cancer risk. In our dataset, 776 serum proteins were regulated in *cis* or *trans* by these susceptibility loci, including 114 that overlapped with proteins associated with prevalent or incident cancers (OR = 2.64, enrichment P-value = 1×10^{-15}). These genomic loci are considered causal drivers of cancer onset, though their individual effect sizes are typically weak or modest. The finding that susceptibility loci regulate serum protein levels suggests that some of these proteins could play a causal role in tumorigenesis. Particular attention should be given to proteins associated with future cancers and those already implicated as oncogenes or tumor suppressors in literature. Notably, several proteins influenced by susceptibility loci include well-known oncogenes such as *MET* and tumor suppressors like *TP53*. The finding that subtle impact of expression changes in oncogenes or tumor suppressors is associated with cancer risk supports

their functional relevance. This contrasts with the more familiar model of tumorigenesis driven by somatic mutations with large functional impacts. In earlier work, we proposed that an individual's non-cancer state can be represented as a point in high-dimensional gene expression space, with tumorigenesis conceptualized as movement through this space toward a tumor state[136]. In this framework, the Euclidean distance between normal and tumor states reflects the likelihood of tumor development, with shorter distances being more probable. The present findings, linking susceptibility loci to expression shifts in key proteins, support this model, suggesting these changes may collectively modulate the probability of cancer progression.

Presentation of mechanistically involved cancer genes

The cancer-associated serum proteins identified in this study were significantly enriched in those previously implicated in tumorigenesis, as evidenced by numerous studies linking them to cancer. This list included, for example, well-known tumor suppressor genes such as *TP53*, *CREBBP*, and *CDH1*, as well as oncogenes like *MET*, *EGFR*, *RET*, and *CTNNB1*, which have extensive experimental support from cell and mouse models, as well as human samples, highlighting their key roles as drivers of cancer. This suggests that proteins predicting incident tumors could serve as candidate biomarkers and may also provide insights into the molecular drivers of disease before clinical symptoms appear. This implies that certain proteins and pathways may form the foundation for pre-clinical interventions, especially in individuals at high risk of developing cancer. To better understand the proteins associated with cancer, it is useful to categorize them into distinct groups for individual consideration. These categories included: genetic susceptibility proteins, which are linked to cancer risk loci and may directly influence the likelihood of tumorigenesis; tumor-host interaction proteins, which are associated with behavioral risk factors (e.g., smoking) or other conditions affecting cancer risk (e.g., obesity,

age); tissue-specific proteins, whose serum levels are altered by cancer development in specific tissues; and tumor-specific proteins, which include both oncogenes and tumor suppressors and likely result from the growing tumor itself.

The identification of serum proteins associated with cancer that correspond to known oncogenes or tumor suppressor genes offers a biologically meaningful link between systemic factors and intracellular tumor biology. While these genes are well-characterized within tumor tissues, their detection at the protein level in circulation is understudied and may reflect tumor cell shedding, secretion, or broader tumor-host interactions. Such proteins may serve as mechanistically informed biomarkers that are reflective of underlying oncogenic processes. Importantly, their presence in preclinical cancer suggests that proteomic changes can precede clinical manifestation of cancer, supporting the utility of serum proteomics for early detection. This convergence of genetic, tissue-level, and circulating evidence strengthens the translational potential of these proteins, offering opportunities for more specific and biologically interpretable cancer surveillance strategies. We note in passing that circulating proteins are regulated in a coordinated manner across tissue boundaries[29,137], reflecting both tissue-intrinsic processes and inter-tissue communication, and thus provide a unique window into the systemic complexity of diseases such as cancer.

Early detection and targeting of mechanisms driving pre-clinical tumorigenesis

The identification of dysregulated proteins years before the cancer diagnosis suggests that early detection is achievable. While current methods like circulating tumor DNA (ctDNA) analysis show promise, they are limited by low abundance in early-stage disease[138]. In contrast, proteomic biomarkers detected through the aptamer-based affinity method offering superior precision and sensitivity[139], potentially surpassing existing techniques by providing deeper insights into early

tumor biology. The presentation of cancer-associated proteins within key oncogenic pathways further indicates the potential for developing early intervention strategies.

Proteins associated with both current and future cancers not only inform us about the nature of these tumors but also suggest potential drivers of tumorigenesis, raising the possibility of intervention before clinical presentation. An intriguing question is whether these same proteins, linked to prevalent or incident cancers, could provide valuable information for patient's post-surgery or treatment. For instance, could early indicators of relapses be detected? Although our current data does not address this, it presents an interesting avenue for future investigation.

Many cancer-associated proteins are linked to germline risk variants and integrating genetic and proteomic profiling could enable long-term risk stratification. Combining polygenic risk scores with proteomic monitoring could help identify individuals who would benefit from more intensive surveillance or preventive therapies. A multi-omics approach that includes susceptibility markers, ctDNA, and serum proteins should be explored to enhance the identification of at-risk individuals before clinical presentation. This approach could open the door to more intensive monitoring and early interventions, such as surgery or therapy.

Serum protein biomarkers present a powerful tool for early cancer detection, risk prediction, and understanding the mechanisms underlying tumorigenesis. By leveraging susceptibility loci, environmental factors, and tissue-specific signatures, these biomarkers have the potential to revolutionize cancer screening and prevention. Future research should focus on validating these proteins in prospective cohorts, integrating intensive imaging and monitoring, and developing targeted interventions to intercept cancer at its earliest stages. In conclusion, our findings highlight the potential of circulating proteins as biomarkers of tumor dynamics, which may reflect changes in tumor growth, the microenvironment, and systemic mediators, although this

has yet to be directly proven. These insights may lay the foundation for formulating hypotheses and directing future research, with significant implications for early cancer detection and future risk assessment.

Material and Methods

Study population

Cohort participants aged 66 through 96 years at the time of blood collection were from the AGES study[39], a single-center, prospective, population-based study of older adults (N = 5,764, mean age 76.6±6 years). The AGES study was formed between 2002 and 2006, and its participants were randomly selected from the surviving members of the established 40-year-long population-based prospective Reykjavik study[140,141], with a 72% recruiting rate. The Reykjavik Study, a prospective cardiovascular survey, recruited a random sample of 30,795 adults born between 1907 and 1935 who lived in the greater Reykjavik area in 1967, that were examined in six phases from 1967 to 1996[140,141]. Measurements in the AGES study, including, for example, brain and vascular imaging, are designed to assess four biologic systems: vascular, neurocognitive (including sensory), musculoskeletal, and body composition/metabolism.[39] All participants are of European ancestry. A decade-long collaboration with large genetic and epidemiology consortia of multiple disease-related phenotypes revealed no discernible difference between the Icelandic population and other European ancestry cohorts[142-144]. This study was approved (approval number VSN-00-063) by the National Bioethics Committee in Iceland, which serves as the Icelandic Heart Association's institutional review board in accordance with the Helsinki Declaration, and by the US National Institutes of Health, National Institute on Aging Intramural Institutional Review Board, with all participants providing informed consent.

Prevalent cancer cases were those with a history of cancer at the baseline visit, while the incident malignancies were diagnosed after the first visit with a 12-year follow-up based on hospital records and cancer registries. The diagnosis of different cancers was based on cancer in the 10th revision of the WHO International Statistical Classification of Diseases (ICD-10): esophagus (C15), stomach (C16), colon (C18), rectum (C20), pancreas (C25), prostate (C61), kidney (C64), bladder (C67), breast (C50), corpus uteri (C54), ovaries (C56), lung and bronchus (C34), and malignant melanoma (C43). Systolic and diastolic blood pressure were measured twice with subjects in a supine position using a Mercury sphygmomanometer. We categorized smoking status as never smoked, former smoker, or current smoker, while alcohol consumption was determined as units per week. Height was measured in meters, while body mass index (BMI) was expressed in kg/m^2. The estimated glomerular filtration rate (eGFR) was estimated according to the Chronic Kidney Disease Epidemiology Collaboration equation[145]. Finally, the present study included only individuals who had their serum proteome measured (see below), which amounted to 5,376 AGES participants.

Proteomics profiling assay

Blood samples were collected at the AGES baseline visit after an overnight fast, and serum samples prepared using a standardized protocol and stored in 0.5mL aliquots at -80°C. Serum samples collected from the inception period, i.e., from 2002 to 2006, were used to generate proteomics data. Before the protein measurements, all serum samples from this period went through their first freeze-thaw cycle. Serum protein levels from 5,376 AGES study participants were quantified using the multiplex SomaScan v4.1 proteomic platform[146], which uses modified DNA aptamers designed to bind target proteins with high affinity and specificity. Here, 7,523 aptamers mapping to 6,586 UniProt IDs were measured in total of 8,592 samples (two time

points). Thus, some proteins were targeted by more than one aptamer. In such cases, individual aptamers had distinct binding sites (epitopes) or binding affinity[29]. Examples include duplicate aptamers targeting single-pass transmembrane proteins (one binding the extracellular domain and another the intracellular loop), aptamers targeting multimers (e.g., interleukins), and duplicate aptamers produced in distinct expression systems. Of the 7,523 aptamers, 233 aptamers were derived from mouse-human chimeras, intended to target proteins from both species. The SOMAmer-based platform measures proteins with femtomole (fM) detection limits and a broad detection range (>8-log dynamic range) of concentration[147]. To avoid batch or time of processing biases, the order of sample collection and separate sample processing for protein measurements were randomized, and all samples run as a single set at SomaLogic Inc. (Boulder, CO, US). All aptamers that passed quality control exhibited median intra-assay and inter-assay coefficients of variation (CV) below 4% at both time points, measured five years apart. Hybridization controls were used to correct systematic variability in detection and calibrator samples of three dilution sets (20% (1:5), 0.5% (1:200), and 0.005% (1:20,000)) were included so that the degree of fluorescence was a quantitative reflection of protein concentration. The adaptive normalization by maximum likelihood (ANML) method was employed to normalize QC replicates and samples using point and variance estimations from a normal U.S. population. Consistent target specificity of aptamers was indicated by direct (through mass spectrometry) and/or indirect validation[29].

Genotype data and the identification of pQTLs

The genotype data includes assayed and imputed genotype data for 5,636 AGES participants[30]. The genotyping arrays used were Illumina Hu370CNV and Illumina GSA BeadChip, which were quality controlled by eliminating variants with call rates <95% and HWE P-value < 1×10^{-6}. The arrays were imputed against the Haplotype Reference Consortium imputation panel r1.1 and

post-imputation quality control was performed separately for each platform. Variants with imputation quality R2 < 0.7, MAF < 0.01, as well as monomorphic and multiallelic variants, were removed before merging to generate a dataset with 7,506,463 variants for 5,656 AGES individuals as previously described[30]. These variants were associated to each of the aptamers on the v4.1-7k serum protein panel to identify *cis* (proximal) and *trans* (distal) acting protein quantitative trait loci (pQTLs), in the same way as previously described[30].

Statistical analysis

Before the analyses, protein data were transformed using a log2 scale, and extreme outlier values excluded, defined as values above the 99.5th percentile of the distribution of 99th percentile cutoffs across all proteins. The relationship between serum protein levels and prevalent cancer was examined cross-sectionally using logistic regression analysis, while the associations of serum proteins with incident cancer were assessed longitudinally *via* the Cox proportional-hazards model. Functional enrichment analysis has been described elsewhere. Briefly though: it was performed using Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) with the R packages ClusterProfiler[148] and fgsea[149], the background set comprised all proteins tested. To account for multiple hypothesis testing, we applied the Benjamini-Hochberg correction with a threshold of FDR < 0.05 to determine the statistical significance of the associations between serum proteins and cancers, as well as for ORA inclusion. Analyses were conducted using R version 4.2.1.

Acknowledgement

The authors acknowledge the contribution of the Icelandic Heart Association (IHA) staff to the AGES study, as well as the involvement of all study participants. The proteomics work was

carried out in collaboration with Novartis Biomedical Research (NIBR). We thank the Icelandic Cancer Society for supporting this study through funding for V.E. The National Institute on Aging (NIA) contracts N01-AG-12100 and HHSN271201200022C for V. G. financed the study. V. G. received funding from the NIA (1R01AG065596-01A1), and IHA received a grant from the Icelandic Parliament. T.M. acknowledges support from the Research Council of Norway (project number 312045), the European Union Horizon Europe (European Innovation Council) (grant agreement number 101115381), and the L. Meltzers Hoyskolefond.

Figure Legends

Figure 1. The figure depicts the study procedures, as well as the materials and study cohort.

Figure 2. (**A**) The numbers represent the number of individuals diagnosed with any of 13 distinct cancer types or those with a history of any type of cancer (ATC). The total does not distinguish between incident and prevalent diagnoses, as we separate these groups in the study. Different cancer types are grouped based on the location of malignancy within the corresponding organ system. Red numbers indicate prevalent cases, while blue numbers represent incident cases. (**B**) The number of serum proteins associated with prevalent (black columns) or incident (red columns) cancers across various types, including ATC, adjusted for standard covariates (age, sex, and eGFR).

Figure 3. (**A**) A curated set of serum proteins associated with future risk of cancer and (**B**) a similar set of proteins linked to prevalent cancer.

Figure 4. (**A**) Digestive system cancers examined in the current study. (**B**) A volcano plot of proteins associated with prevalent stomach cancer (STC), and (**C**) their tissue distribution. (**D**) Association plot of selected serum proteins linked to STC.

Figure 5. (**A**) The figure shows the number of proteins shared between incident and prevalent cancers, highlighting examples of proteins in the overlap. (**B**) The plot illustrates the contrasting effect of KLK3 (aka PSA) in incident versus prevalent prostate cancer (PRC), with the effect represented as a hazard ratio for incident PRC and an odds ratio for prevalent PRC. (**C**) A heatmap displaying 25 serum proteins associated with both incident and prevalent cancer types.

Figure 6. The figure presents a volcano plot of proteins associated with (**A**) new-onset cancers of any type (ATC) and (**B**) prevalent ATC. Proteins directly associated with ATC are shown in red,

while those downregulated are shown in blue. The plot highlights a selected set of proteins. The regression analysis incorporated the standard covariate adjustment.

Figure 7. (**A**) The figure illustrates the overlap between proteins associated with different patient groups, including incident cancers (combined from proteins linked to any of the 13 different incident cancer types), incident ATC, prevalent cancers (combined from proteins linked to any of the 13 distinct prevalent cancer types), and prevalent ATC. We note that in some cases, where multiple aptamers target the same protein, proteins with the same annotation may appear in different overlap groups. (**B**) The figure illustrates the overlap between proteins identified in the observational study of incident and prevalent cancers and those detected through proteogenomic analysis of genetic cancer risk.

Figure 8. (**A**) The plot illustrates the enrichment of cancer-associated proteins in the current study among highly cited cancer genes from the literature, as detailed in the main text. (**B**) The plot shows the enrichment of proteins from the observational analysis and those linked to genetic cancer risk among highly cited cancer genes from the literature, as described in the main text.

Figure 9. (**A**) Hierarchical representation of interacting network regulators for CPNs enriched with cancer-associated proteins. (**B**) Physical protein-protein interactions among network regulators of CPN subnetworks enriched for cancer-associated proteins across multiple cancer types, as identified by the STRING database. These edges represent physical interactions, and unconnected network regulators are excluded from visualization.

Figure 10. The figure illustrates the categorization of cancer-associated proteins highlighted in this study. The main groups include tumor-specific proteins (encoded by oncogenes or tumor suppressor genes), tissue-specific proteins (reflecting changes related to the tumor's tissue of

origin), genetic susceptibility proteins (linked to common low-penetrance germline variants), and tumor-host interaction proteins. The latter is a broad category encompassing proteins that do not fit neatly into the other groups and instead reflect the complex interactions between the tumor and its host environment. This includes proteins influenced by lifestyle factors (e.g., smoking) and systemic conditions (such as inflammation or metabolic dysfunction), as well as those involved in tumor-intrinsic processes like the tumor microenvironment (TME) originating from stromal cells, immune cells, or the extracellular matrix, and epithelial-mesenchymal transition (EMT), which connects tumor cell behavior with microenvironmental remodeling. Additionally, this category covers proteins associated with metastatic progression. Examples of cancer-associated proteins are highlighted within each category.

References

1. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78-81 (2015).
2. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
3. Frank, S.A. Genetic predisposition to cancer - insights from population genetics. *Nat Rev Genet* **5**, 764-772 (2004).
4. Wu, S., Powers, S., Zhu, W. & Hannun, Y.A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529**, 43-47 (2016).
5. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31-46 (2022).
6. Siegel, R.L., Miller, K.D., Fuchs, H.E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J Clin* **72**, 7-33 (2022).
7. Fitzmaurice, C., *et al.* Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. *JAMA oncology* **4**, 1553-1568 (2018).
8. Koh, B., *et al.* Patterns in Cancer Incidence Among People Younger Than 50 Years in the US, 2010 to 2019. *JAMA Network Open* **6**, e2328171-e2328171 (2023).
9. Bray, F., *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229-263 (2024).
10. Swanton, C., *et al.* Embracing cancer complexity: Hallmarks of systemic disease. *Cell* **187**, 1589-1616 (2024).
11. Qing, T., *et al.* Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. *Nat Commun* **11**, 2438 (2020).
12. Antoniou, A., *et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* **72**, 1117-1130 (2003).
13. Gudmundsson, J., *et al.* Frequent occurrence of BRCA2 linkage in Icelandic breast cancer families and segregation of a common BRCA2 haplotype. *Am J Hum Genet* **58**, 749-756 (1996).
14. Turnbull, C., Sud, A. & Houlston, R.S. Cancer genetics, precision prevention and a call to action. *Nat Genet* **50**, 1212-1218 (2018).
15. Martincorena, I., *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041.e1021 (2017).
16. Vogelstein, B., *et al.* Cancer genome landscapes. *Science* **339**, 1546-1558 (2013).
17. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-734 (2011).
18. Shieh, Y., *et al.* Population-based screening for cancer: hope and hype. *Nature reviews. Clinical oncology* **13**, 550-565 (2016).
19. Villanueva, J., *et al.* Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* **116**, 271-284 (2006).
20. Catalona, W.J., *et al.* Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N Engl J Med* **324**, 1156-1161 (1991).
21. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *Journal of hepatology* **69**, 182-236 (2018).
22. Gold, P. & Freedman, S.O. Specific carcinoembryonic antigens of the human digestive system. *The Journal of experimental medicine* **122**, 467-481 (1965).
23. Drescher, C.W., *et al.* Longitudinal screening algorithm that incorporates change over time in CA125 levels identifies ovarian cancer earlier than a single-threshold rule. *J Clin Oncol* **31**, 387-392 (2013).

24. Chen, X., *et al.* Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat Commun* **11**, 3475 (2020).
25. Cohen, J.D., *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926-930 (2018).
26. Bruhm, D.C., *et al.* Genomic and fragmentomic landscapes of cell-free DNA for early cancer detection. *Nat Rev Cancer* (2025).
27. Papier, K., *et al.* Identifying proteomic risk factors for cancer using prospective and exome analyses of 1463 circulating proteins and risk of 19 cancers in the UK Biobank. *Nat Commun* **15**, 4010 (2024).
28. Lamb, J.R., Jennings, L.L., Gudmundsdottir, V., Gudnason, V. & Emilsson, V. It's in Our Blood: A Glimpse of Personalized Medicine. *Trends Mol Med* (2020).
29. Emilsson, V., *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769-773 (2018).
30. Gudjonsson, A., *et al.* A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nature Communications* **13**, 1-13 (2022).
31. Emilsson, V., *et al.* A proteogenomic signature of age-related macular degeneration in blood. *Nat Commun* **13**, 3401 (2022).
32. Emilsson, V., *et al.* Coding and regulatory variants are associated with serum protein levels and disease. *Nature Communications* **13**, 1-11 (2022).
33. Emilsson, V., Gudnason, V. & Jennings, L.L. Predicting health and life span with the deep plasma proteome. *Nature medicine* **25**, 1815-1816 (2019).
34. Williams, S.A., *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat Med* **25**, 1851-1857 (2019).
35. Lehallier, B., *et al.* Undulating changes in human plasma proteome profiles across the lifespan. *Nat Med* **25**, 1843-1850 (2019).
36. Axelsson, G.T., *et al.* The Proteomic Profile of Interstitial Lung Abnormalities. *American journal of respiratory and critical care medicine* **206**, 337-346 (2022).
37. Gudmundsdottir, V., *et al.* Circulating protein signatures and causal candidates for type 2 diabetes. *Diabetes* **69**, 1843-1853 (2020).
38. Emilsson, V., *et al.* Proteomic prediction of incident heart failure and its main subtypes. *Eur J Heart Fail* **26**, 87-102 (2024).
39. Harris, T.B., *et al.* Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* **165**, 1076-1087 (2007).
40. Brown, K.F., *et al.* The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *British journal of cancer* **118**, 1130-1141 (2018).
41. Gapstur, S.M., *et al.* Alcohol and Cancer: Existing Knowledge and Evidence Gaps across the Cancer Continuum. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **31**, 5-10 (2022).
42. Cox, D.R. Regression Models and Life-Tables. **34**, 187-202 (1972).
43. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* **Vol.57**, No. 1: 289-300 (1995).
44. Suehnholz, S.P., *et al.* Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. *Cancer Discov* **14**, 49-65 (2024).
45. Chakravarty, D., *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **17**, 16 (2017).
46. Tsherniak, A., *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e516 (2017).
47. Nunes, J., *et al.* Siglec-6 as a therapeutic target for cell migration and adhesion in chronic lymphocytic leukemia. *Nat Commun* **15**, 5180 (2024).

48. Yang, Z., *et al.* Effect of MPP2 and its DNA methylation levels on prognosis of colorectal cancer patients. *World journal of surgical oncology* **22**, 290 (2024).
49. Lu, C., *et al.* Analysis of Circulating Immune Subsets in Primary Colorectal Cancer. *Cancers* **14**(2022).
50. Shi, S., *et al.* TMEM106A transcriptionally regulated by promoter methylation is involved in invasion and metastasis of hepatocellular carcinoma. *Acta biochimica et biophysica Sinica* **54**, 1008-1020 (2022).
51. Fan, B., Wang, L. & Wang, J. RAB22A as a predictor of exosome secretion in the progression and relapse of multiple myeloma. *Aging* **16**, 4169-4190 (2024).
52. Pierro, E.W., *et al.* Trem2 deficiency attenuates breast cancer tumor growth in lean, but not obese or weight loss, mice and is associated with alterations of clonal T cell populations. *bioRxiv* (2024).
53. Schroder, A.K., *et al.* miR-379-5p affects breast cancer cell behavior by targeting UBE2E3 ubiquitin conjugating enzyme. *PLoS One* **19**, e0310315 (2024).
54. Fisher, R.C., *et al.* Disrupting Inflammation-Associated CXCL8-CXCR1 Signaling Inhibits Tumorigenicity Initiated by Sporadic- and Colitis-Colon Cancer Stem Cells. *Neoplasia (New York, N.Y.)* **21**, 269-281 (2019).
55. Fang, S., *et al.* CXCL8 Up-Regulated LSECtin through AKT Signal and Correlates with the Immune Microenvironment Modulation in Colon Cancer. *Cancers* **14**(2022).
56. Srinivasan, S., *et al.* A PSA SNP associates with cellular function and clinical outcome in men with prostate cancer. *Nat Commun* **15**, 9587 (2024).
57. Veeramani, S., *et al.* Cellular prostatic acid phosphatase: a protein tyrosine phosphatase involved in androgen-independent proliferation of prostate cancer. *Endocrine-related cancer* **12**, 805-822 (2005).
58. Kantoff, P.W., *et al.* Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med* **363**, 411-422 (2010).
59. El Ayachi, I., *et al.* The WNT10B Network Is Associated with Survival and Metastases in Chemoresistant Triple-Negative Breast Cancer. *Cancer Res* **79**, 982-993 (2019).
60. Wend, P., *et al.* WNT10B/β-catenin signalling induces HMGA2 and proliferation in metastatic triple-negative breast cancer. *EMBO Mol Med* **5**, 264-279 (2013).
61. Chan, J.J., *et al.* Pan-cancer pervasive upregulation of 3' UTR splicing drives tumourigenesis. *Nature cell biology* **24**, 928-939 (2022).
62. Hartman, M.L. & Czyz, M. BCL-G: 20 years of research on a non-typical protein from the BCL-2 family. *Cell death and differentiation* **30**, 1437-1446 (2023).
63. Huang, X., *et al.* Wnt7a activates canonical Wnt signaling, promotes bladder cancer cell invasion, and is suppressed by miR-370-3p. *J Biol Chem* **293**, 6693-6706 (2018).
64. MacLean, J.A., 2nd, King, M.L., Okuda, H. & Hayashi, K. WNT7A Regulation by miR-15b in Ovarian Cancer. *PLoS One* **11**, e0156109 (2016).
65. Kondratov, A.G., *et al.* Alterations of the WNT7A gene in clear cell renal cell carcinomas. *PLoS One* **7**, e47012 (2012).
66. Cui, P., *et al.* Pan-cancer analysis of the prognostic and immunological roles of SHP-1/ptpn6. *Sci Rep* **14**, 23083 (2024).
67. Abdelatty, A., *et al.* Pan-Cancer Study on Protein Kinase C Family as a Potential Biomarker for the Tumors Immune Landscape and the Response to Immunotherapy. *Frontiers in cell and developmental biology* **9**, 798319 (2021).
68. Xiong, Y., *et al.* WFDC2 suppresses prostate cancer metastasis by modulating EGFR signaling inactivation. *Cell death & disease* **11**, 537 (2020).
69. Chen, Y., *et al.* WAP four-disulfide core domain protein 2 gene(WFDC2) is a target of estrogen in ovarian cancer cells. *Journal of ovarian research* **9**, 10 (2016).
70. Min, B. & Wang, Y. WFDC2 is a potential prognostic and immunotherapy biomarker in lung adenocarcinoma. *The Journal of international medical research* **52**, 3000605241258893 (2024).

71. Zhang, Y., Wang, S., Han, S. & Feng, Y. Pan-Cancer Analysis Based on EPOR Expression With Potential Value in Prognosis and Tumor Immunity in 33 Tumors. *Frontiers in oncology* **12**, 844794 (2022).
72. Roberts, K.G., *et al.* Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer cell* **22**, 153-166 (2012).
73. Iacobucci, I., *et al.* Truncating Erythropoietin Receptor Rearrangements in Acute Lymphoblastic Leukemia. *Cancer cell* **29**, 186-200 (2016).
74. Nguyen, P.M., *et al.* Loss of Bcl-G, a Bcl-2 family member, augments the development of inflammation-associated colorectal cancer. *Cell death and differentiation* **27**, 742-757 (2020).
75. Avgustinova, A., *et al.* Tumour cell-derived Wnt7a recruits and activates fibroblasts to promote tumour aggressiveness. *Nat Commun* **7**, 10305 (2016).
76. Matous, J.G., *et al.* Shp-1 regulates the activity of low-affinity T cells specific to endogenous self-antigen during melanoma tumor growth and drives resistance to immune checkpoint inhibition. *Journal for immunotherapy of cancer* **13**(2025).
77. Bingle, L., *et al.* WFDC2 (HE4): a potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung. *Respiratory research* **7**, 61 (2006).
78. Uhlén, M., *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
79. Taupin, D. & Podolsky, D.K. Trefoil factors: initiators of mucosal healing. *Nature reviews. Molecular cell biology* **4**, 721-732 (2003).
80. Suh, K.S., *et al.* Reciprocal modifications of CLIC4 in tumor epithelium and stroma mark malignant progression of multiple human cancers. *Clin Cancer Res* **13**, 121-131 (2007).
81. Singha, B., *et al.* CLIC1 and CLIC4 complement CA125 as a diagnostic biomarker panel for all subtypes of epithelial ovarian cancer. *Sci Rep* **8**, 14725 (2018).
82. Suryawanshi, S., *et al.* Complement pathway is frequently altered in endometriosis and endometriosis-associated ovarian cancer. *Clin Cancer Res* **20**, 6163-6174 (2014).
83. Papadaki, V., *et al.* Two Secreted Proteoglycans, Activators of Urothelial Cell-Cell Adhesion, Negatively Contribute to Bladder Cancer Initiation and Progression. *Cancers* **12**(2020).
84. Yang, Z., Yu, B., Hu, J., Jiang, L. & Jian, M. LRRC25 Is a Potential Biomarker for Predicting Immunotherapy Response in Patients with Gastric Cancer. *Digestive diseases and sciences* **70**, 1395-1410 (2025).
85. Liu, F. & Wu, H. Prognostic Value of Gastrokine-2 (GKN2) and Its Correlation with Tumor-Infiltrating Immune Cells in Lung Cancer and Gastric Cancers. *Journal of inflammation research* **13**, 933-944 (2020).
86. Guy, M., *et al.* Identification of new genetic risk factors for prostate cancer. *Asian journal of andrology* **11**, 49-55 (2009).
87. Chung Nien Chin, S., *et al.* Coordinate expression loss of GKN1 and GKN2 in gastric cancer via impairment of a glucocorticoid-responsive enhancer. *American journal of physiology. Gastrointestinal and liver physiology* **319**, G175-g188 (2020).
88. Bergström, S.H., Järemo, H., Nilsson, M., Adamo, H.H. & Bergh, A. Prostate tumors downregulate microseminoprotein-beta (MSMB) in the surrounding benign prostate epithelium and this response is associated with tumor aggressiveness. *Prostate* **78**, 257-265 (2018).
89. Ali, M.W., *et al.* A risk variant for Barrett's esophagus and esophageal adenocarcinoma at chr8p23.1 affects enhancer activity and implicates multiple gene targets. *Hum Mol Genet* **31**, 3975-3986 (2022).
90. Mazumdar, J., *et al.* HIF-2alpha deletion promotes Kras-driven lung tumor development. *Proc Natl Acad Sci U S A* **107**, 14182-14187 (2010).
91. Krop, I., *et al.* HIN-1, an inhibitor of cell growth, invasion, and AKT activation. *Cancer Res* **65**, 9659-9669 (2005).

92. Reynolds, S.D., Reynolds, P.R., Pryhuber, G.S., Finder, J.D. & Stripp, B.R. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. *American journal of respiratory and critical care medicine* **166**, 1498-1509 (2002).
93. Goldfarbmuren, K.C., *et al.* Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nat Commun* **11**, 2485 (2020).
94. Lu, X., Shen, J., Huang, S., Wang, H. & Liu, D. Down-regulation of CLEC3B facilitates epithelial-mesenchymal transition, migration and invasion of lung adenocarcinoma cells. *Tissue & cell* **76**, 101802 (2022).
95. Ungaro, F., *et al.* Lymphatic endothelium contributes to colorectal cancer growth via the soluble matrisome component GDF11. *International journal of cancer* **145**, 1913-1920 (2019).
96. Ojima, C., *et al.* Peptide-2 from mouse myostatin precursor protein alleviates muscle wasting in cancer-associated cachexia. *Cancer science* **111**, 2954-2964 (2020).
97. Xu, T., *et al.* ApoM suppresses kidney renal clear cell carcinoma growth and metastasis via the Hippo-YAP signaling pathway. *Archives of biochemistry and biophysics* **743**, 109642 (2023).
98. Thapa, B., *et al.* OX40/OX40 ligand and its role in precision immune oncology. *Cancer metastasis reviews* **43**, 1001-1013 (2024).
99. Liu, K., Li, L. & Han, G. CHST12: a potential prognostic biomarker related to the immunotherapy response in pancreatic adenocarcinoma. *Frontiers in endocrinology* **14**, 1226547 (2023).
100. Li, Z., *et al.* Immune checkpoint reprogramming via sequential nucleic acid delivery strategy optimizes systemic immune responses for gastrointestinal cancer immunotherapy. *Cancer letters* **599**, 217152 (2024).
101. Chandrashekar, D.S., *et al.* UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia (New York, N.Y.)* **25**, 18-27 (2022).
102. Bai, J., *et al.* Overexpression of Cullin1 is associated with poor prognosis of patients with gastric cancer. *Human pathology* **42**, 375-383 (2011).
103. Leblond, M.M., Zdimerova, H., Desponds, E. & Verdeil, G. Tumor-Associated Macrophages in Bladder Cancer: Biological Role, Impact on Therapeutic Response and Perspectives for Immunotherapy. *Cancers* **13**(2021).
104. Maldonado, M.D.M., Schlom, J. & Hamilton, D.H. Blockade of tumor-derived colony-stimulating factor 1 (CSF1) promotes an immune-permissive tumor microenvironment. *Cancer immunology, immunotherapy : CII* **72**, 3349-3362 (2023).
105. Tan, D.X. & Hardeland, R. The Reserve/Maximum Capacity of Melatonin's Synthetic Function for the Potential Dimorphism of Melatonin Production and Its Biological Significance in Mammals. *Molecules (Basel, Switzerland)* **26**(2021).
106. Masri, S. & Sassone-Corsi, P. The emerging link between cancer, metabolism, and circadian rhythms. *Nat Med* **24**, 1795-1803 (2018).
107. Yang, J., Fan, L., Liao, X., Cui, G. & Hu, H. CRTAC1 (Cartilage acidic protein 1) inhibits cell proliferation, migration, invasion and epithelial-mesenchymal transition (EMT) process in bladder cancer by downregulating Yin Yang 1 (YY1) to inactivate the TGF-β pathway. *Bioengineered* **12**, 9377-9389 (2021).
108. Peake, B.F., Eze, S.M., Yang, L., Castellino, R.C. & Nahta, R. Growth differentiation factor 15 mediates epithelial mesenchymal transition and invasion of breast cancers through IGF-1R-FoxM1 signaling. *Oncotarget* **8**, 94393-94406 (2017).
109. de Crevoisier, R., *et al.* Early PSA decrease is an independent predictive factor of clinical failure and specific survival in patients with localized prostate cancer treated by radiotherapy with or without androgen deprivation therapy. *Annals of oncology : official journal of the European Society for Medical Oncology* **21**, 808-814 (2010).
110. Jiang, X., *et al.* Shared heritability and functional enrichment across six solid cancers. *Nat Commun* **10**, 431 (2019).

111. Rashkin, S.R., *et al.* Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun* **11**, 4423 (2020).
112. Ishibashi, Y., *et al.* Serum TFF1 and TFF3 but not TFF2 are higher in women with breast cancer than in women without breast cancer. *Sci Rep* **7**, 4846 (2017).
113. Gupta, S., *et al.* Plasma proteome of growing tumors. *Sci Rep* **13**, 12195 (2023).
114. Byun, J., *et al.* Cross-ancestry genome-wide meta-analysis of 61,047 cases and 947,237 controls identifies new susceptibility loci contributing to lung cancer. *Nat Genet* **54**, 1167-1177 (2022).
115. Perez-Garcia, A., *et al.* Genetic loss of SH2B3 in acute lymphoblastic leukemia. *Blood* **122**, 2425-2432 (2013).
116. Schwenk, J.M., *et al.* The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J Proteome Res* **16**, 4299-4310 (2017).
117. Tang, Z., *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* **45**, W98-w102 (2017).
118. Wen, R.M., *et al.* AZGP1 deficiency promotes angiogenesis in prostate cancer. *Journal of translational medicine* **22**, 383 (2024).
119. Deng, L., *et al.* AZGP1 activation by lenvatinib suppresses intrahepatic cholangiocarcinoma epithelial-mesenchymal transition through the TGF-β1/Smad3 pathway. *Cell death & disease* **14**, 590 (2023).
120. Lauwers, Y., *et al.* Imaging of tumor-associated macrophage dynamics during immunotherapy using a CD163-specific nanobody-based immunotracer. *Proc Natl Acad Sci U S A* **121**, e2409668121 (2024).
121. Li, Y., Wu, Q., Lv, J. & Gu, J. A comprehensive pan-cancer analysis of CDH5 in immunological response. *Front Immunol* **14**, 1239875 (2023).
122. Feng, P., *et al.* Golgi protein 73: the driver of inflammation in the immune and tumor microenvironment. *Front Immunol* **15**, 1508034 (2024).
123. Guo, Q., *et al.* SERPIND1 Affects the Malignant Biological Behavior of Epithelial Ovarian Cancer via the PI3K/AKT Pathway: A Mechanistic Study. *Frontiers in oncology* **9**, 954 (2019).
124. Lindström, S., *et al.* Genome-wide analyses characterize shared heritability among cancers and identify novel cancer susceptibility regions. *Journal of the National Cancer Institute* **115**, 712-732 (2023).
125. Lachmann, A., *et al.* Geneshot: search engine for ranking genes from arbitrary text queries. *Nucleic Acids Res* **47**, W571-w577 (2019).
126. Thinyakul, C., *et al.* Hippo pathway in cancer cells induces NCAM1(+)αSMA(+) fibroblasts to modulate tumor microenvironment. *Commun Biol* **7**, 1343 (2024).
127. Smith, J.P. & Solomon, T.E. Cholecystokinin and pancreatic cancer: the chicken or the egg? *American journal of physiology. Gastrointestinal and liver physiology* **306**, G91-g101 (2014).
128. Jacobsen, F., *et al.* Cadherin-17 (CDH17) expression in human cancer: A tissue microarray study on 18,131 tumors. *Pathology, research and practice* **256**, 155175 (2024).
129. Wang, D., *et al.* HSP90AB1 as the Druggable Target of Maggot Extract Reverses Cisplatin Resistance in Ovarian Cancer. *Oxidative medicine and cellular longevity* **2023**, 9335440 (2023).
130. Chu, C.A., *et al.* The Role of Phosphatidylinositol 3-Kinase Catalytic Subunit Type 3 in the Pathogenesis of Human Cancer. *International journal of molecular sciences* **22**(2021).
131. Kim, J., Kong, J., Chang, H., Kim, H. & Kim, A. EGF induces epithelial-mesenchymal transition through phospho-Smad2/3-Snail signaling pathway in breast cancer cells. *Oncotarget* **7**, 85021-85032 (2016).
132. Hua, W., Ten Dijke, P., Kostidis, S., Giera, M. & Hornsveld, M. TGFβ-induced metabolic reprogramming during epithelial-to-mesenchymal transition in cancer. *Cellular and molecular life sciences : CMLS* **77**, 2103-2123 (2020).
133. Bankier, S., *et al.* Circulating causal protein networks linked to future risk of myocardial infarction. *medRxiv* (2025).

134. Pal, S.K., Katheria, V. & Hurria, A. Evaluating the older patient with cancer: understanding frailty and the geriatric assessment. *CA Cancer J Clin* **60**, 120-132 (2010).
135. Laconi, E., Marongiu, F. & DeGregori, J. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *British journal of cancer* **122**, 943-952 (2020).
136. Lamb, J.R., *et al.* Predictive genes in adjacent normal tissue are preferentially altered by sCNV during tumorigenesis in liver cancer and may rate limiting. *PLoS One* **6**, e20090 (2011).
137. Lamb, J.R., Jennings, L.L., Gudmundsdottir, V., Gudnason, V. & Emilsson, V. It's in our blood: a glimpse of personalized medicine. *Trends in molecular medicine* **27**, 20-30 (2021).
138. Tan, A.C., *et al.* Detection of circulating tumor DNA with ultradeep sequencing of plasma cell-free DNA for monitoring minimal residual disease and early detection of recurrence in early-stage lung cancer. *Cancer* **130**, 1758-1765 (2024).
139. Rooney, M.R., *et al.* Plasma proteomic comparisons change as coverage expands for SomaLogic and Olink. *medRxiv* (2024).
140. Sigurdsson, E., Thorgeirsson, G., Sigvaldason, H. & Sigfusson, N. Unrecognized myocardial infarction: epidemiology, clinical characteristics, and the prognostic role of angina pectoris. The Reykjavik Study. *Ann Intern Med* **122**, 96-102 (1995).
141. Merkler, A.E., *et al.* Association Between Unrecognized Myocardial Infarction and Cerebral Infarction on Magnetic Resonance Imaging. *JAMA Neurol* **76**, 956-961 (2019).
142. Psaty, B.M., *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* **2**, 73-80 (2009).
143. Igap & Schellenberg, G.D. International Genomics of Alzheimer's Disease Project (IGAP) genome-wide association study. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **8**, P101 (2012).
144. Emerging Risk Factors, C., *et al.* The Emerging Risk Factors Collaboration: analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. *Eur J Epidemiol* **22**, 839-869 (2007).
145. Inker, L.A., *et al.* Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med* **367**, 20-29 (2012).
146. Cespedes Feliciano, E.M., *et al.* Association of Prediagnostic Frailty, Change in Frailty Status, and Mortality After Cancer Diagnosis in the Women's Health Initiative. *JAMA network open* **3**, e2016747 (2020).
147. Corey, K.E., *et al.* ADAMTSL2 protein and a soluble biomarker signature identify at-risk non-alcoholic steatohepatitis and fibrosis in adults with NAFLD. *Journal of hepatology* **76**, 25-33 (2022).
148. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-287 (2012).
149. Korotkevich, G., *et al.* Fast gene set enrichment analysis. 060012 (2021).

Table 1. Baseline characteristics of AGES participants by sex, cancer group (body system), and disease status, with 7,523 serum proteins measured.

Characteristic	Variable*	Females	Males	P-value	Total
Demographics					
	Numbers	3077 (57.2)	2299 (42.8)	N/A	5,376
	AGE (years)	76.7 (5.4)	76.5 (5.7)	0.232	76.6 (5.6)
Anthropometry					
	BMI (kg/m^2)	27.2 (4.8)	26.8 (3.8)	0.004	27.0 (4.4)
	BMI category			<0.001	
	BMI<25 kg/m^2	1050 (34.2)	747 (32.5)		1797 (33.5)
	BMI =20-30 kg/m^2	1264 (41.1)	1115 (48.5)		2379 (44.3)
	BMI ≥ 30 kg/m^2	758 (24.7)	435 (18.9)		1193 (22.2)
Lifestyle					
	Smoking status			<0.001	
	Never	1577 (52.7)	645 (28.8)		2222 (42.4)
	Former	1031 (34.5)	1337 (59.6)		2368 (45.2)
	Current	384 (12.8)	261 (11.6)		645 (12.3)
	Alcohol use	9.3 (21.1)	22.0 (41.3)	<0.001	14.7 (32.0)
Physiological					
	eGFR (ml/min/1.73m^2)	63.4 (15.6)	64.9 (15.1)	<0.001	64.0 (15.4)
	SBP (mmHg)	142.1 (20.9)	143.2 (20.4)	0.073	142.6 (20.7)
	DBP (mmHg)	72.2 (9.5)	76.2 (9.6)	<0.001	73.9 (9.7)
Other					
	Follow-up period (any cancer)	10.4 [6.2, 11.6]	8.6 [4.5, 11.1]	<0.001	10.0 [5.3, 11.4]
All 13 cancers examined in current study, combined					
	All cases	597 (19.6)	638 (28.0)	<0.001	1235 (23.2)
	Prevalent cases	237 (7.8)	228 (10.0)	0.006	465 (8.7)
	Incident cases	360 (11.7)	410 (17.8)	<0.001	770 (14.3)
Digestive system cancer					
	All cases	166 (5.5)	152 (6.7)	0.075	318 (6.0)
	Prevalent cases	45 (1.5)	49 (2.1)	0.084	94 (1.8)
	Incident cases	121 (3.9)	103 (4.5)	0.402	224 (4.2)
Genitourinary system cancer					
	All cases	65 (2.1)	420 (18.4)	<0.001	485 (9.1)
	Prevalent cases	31 (1.0)	170 (7.5)	<0.001	201 (3.8)
	Incident cases	34 (1.1)	250 (10.9)	<0.001	284 (5.3)
Respiratory system cancers					
	All cases	97 (3.2)	79 (3.5)	0.663	176 (3.3)
	Prevalent cases	11 (0.4)	9 (0.4)	1.000	20 (0.4)
	Incident cases	86 (2.8)	70 (3.1)	0.663	156 (2.9)
Cancers of the female reproductive system					
	All cases	281 (9.2)	6 (0.3)	<0.001	287 (5.4)
	Prevalent cases	154 (5.1)	2 (0.1)	<0.001	156 (2.9)
	Incident cases	127 (4.1)	4 (0.2)	<0.001	131 (2.4)
Skin cancer					
	All cases	50 (1.6)	43 (1.9)	0.576	93 (1.7)
	Prevalent cases	8 (0.3)	6 (0.3)	1.000	14 (0.3)
	Incident cases	42 (1.4)	37 (1.6)	0.544	79 (1.5)

*Numbers are mean (SD) for continuous-, N (%) for categorical- and median [IQR] for skewed variables. The reported P-values are two-sided. Abbreviations: BMI, body mass index; eGFR, estimated glomerular filtration rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; N/A, not applicable. The 13 cancer types encompass those affecting the digestive system (esophagus, stomach, colon, rectum, and pancreas); the genitourinary system (kidney, prostate, and bladder); the female reproductive system (breast, ovary, and corpus uteri); the respiratory system (lung and bronchus); and skin melanoma.

Table 2. Serum protein associations with different cancer types in the AGES study (FDR < 0.05).

Body system	Tumor site (ICD-10 code)	Condition	Number of linked proteins*	Number of linked proteins**	Two examples
Digestive system	Esophagus (C15)	Incident	16	9	CTNNB1, RAET1L
		Prevalent	NA	NA	NA
	Stomach (C16)	Incident	2	4	LECT2, CUL1
		Prevalent	33	14	GKN2, TFF1
	Colon (C18)	Incident	9	6	CXCL8, BCL2L14
		Prevalent	0	NA	NA
	Rectum (C20)	Incident	13	7	HNRNPA1, WNT7A
		Prevalent	17	9	COLEC12, COL6A3
	Pancreas (C25)	Incident	6	4	PTPN6, CHST12
		Prevalent	NA	NA	NA
Genitourinary system	Kidney (C64)	Incident	4	5	HAVCR1, GIMAP4
		Prevalent	3	5	EPHB2, GRP
	Prostate (C61)	Incident	5	6	KLK3, ACP3
		Prevalent	7	7	KLK3, MSMB
	Bladder (C67)	Incident	25	29	MPP2, PRKCZ
		Prevalent	5	5	NR3C2, PRPS1
Respiratory system	Lung (C34)	Incident	216	10	WFDC2, CLEC3B
		Prevalent	6	0	WFDC2, TP53
Female reproductive system	Breast (C50)	Incident	2	2	TXLNA, WNT10B
		Prevalent	133	100	MET, RET
	Corpus uteri (C54)	Incident	11	11	CRLF2, RAB32
		Prevalent	3	1	OSCAR, DCN
	Ovary (C56)	Incident	2	1	EPOR, EVA1B
		Prevalent	NA	NA	NA
Skin	Melanocytes (C43)	Incident	10	9	PCSK1N, GHRH
		Prevalent	6	2	FHIT, MICALL1
Any	Any site	Incident	278	163	WFDC2, GDF15
		Prevalent	43	35	FETUB, IGFBP4

*Standard covariate adjustment includes age, sex, and eGFR. Analyses were performed both jointly and separately by sex for cancers that are not inherently sex specific. **Additional covariate adjustments were tailored to known risk factors for specific cancers, which may include BMI, smoking status, alcohol consumption, height, or blood pressure (refer to Supplementary Text of Supplementary Material for further details). In some cases, additional proteins were identified beyond those detected with the standard covariate adjustment.

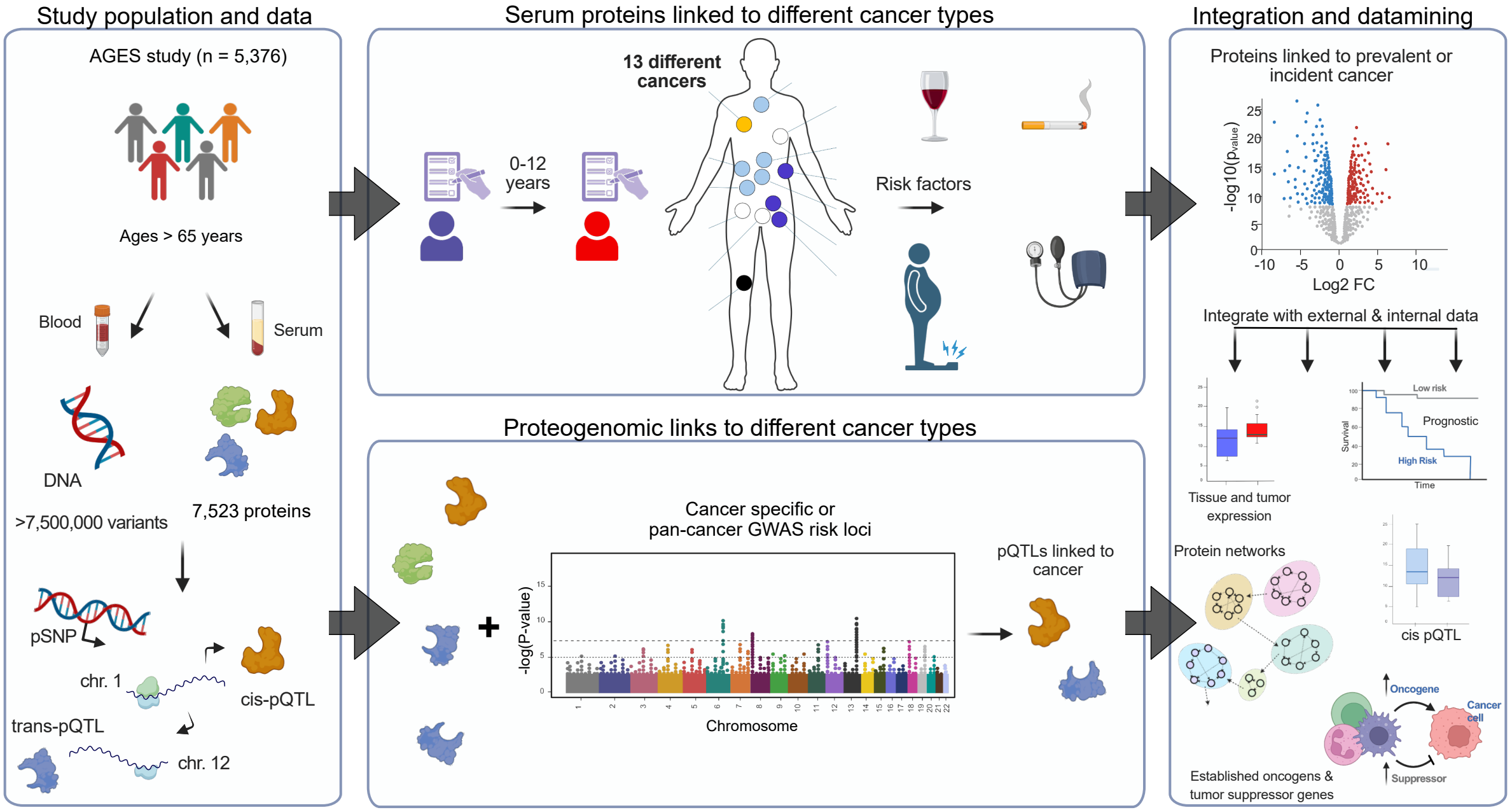
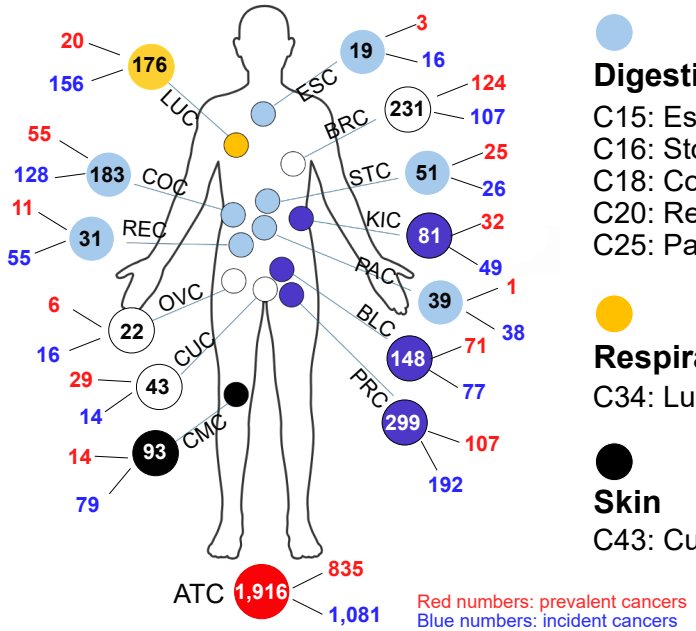

Study population and data
AGES study (n = 5,376)
Ages > 65 years
Blood
Serum
DNA
>7,500,000 variants
7,523 proteins
pSNP
chr. 1
cis-pQTL
trans-pQTL
chr. 12
Serum proteins linked to different cancer types
13 different cancers
0-12 years
Risk factors
Proteogenomic links to different cancer types
Cancer specific or pan-cancer GWAS risk loci
-log(P-value)
Chromosome
pQTLs linked to cancer
Integration and datamining
Proteins linked to prevalent or incident cancer
-log10(pvalue)
Log2 FC
Integrate with external & internal data
Tissue and tumor expression
Low risk
Prognostic
High Risk
Survival
Time
Protein networks
cis pQTL
Oncogene
Cancer cell
Suppressor
Established oncogens & tumor suppressor genes

Figure 1

A

Cancer case count in the current study



Digestive system
C15: Esophageal cancer (ESC)
C16: Stomach cancer (STC)
C18: Colon cancer (COC)
C20: Rectal cancer (REC)
C25: Pancreatic cancer (PAC)

Respiratory system
C34: Lung cancer (LUC)

Skin
C43: Cutaneous melanoma (CMC)

Genitourinary system
C64: Kidney cancer (KIC)
C61: Prostate cancer (PRC)
C67: Bladder cancer (BLC)

Female reproductive system
C50: Breast cancer (BRC)
C54: Cancer of corpus uteri (CUC)
C56: Ovarian cancer (OVC)

Any type of cancer (ATC)

B

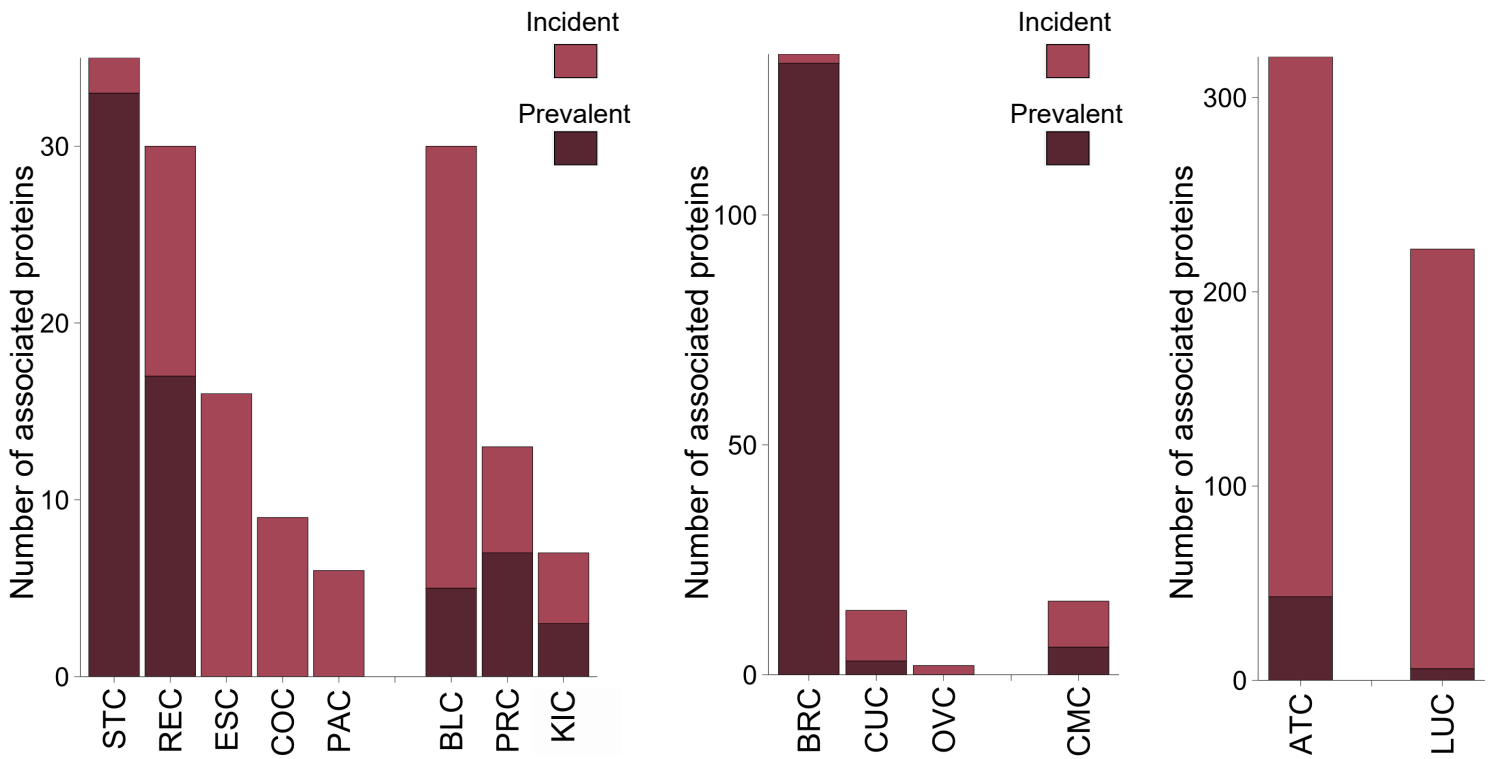


Figure 2

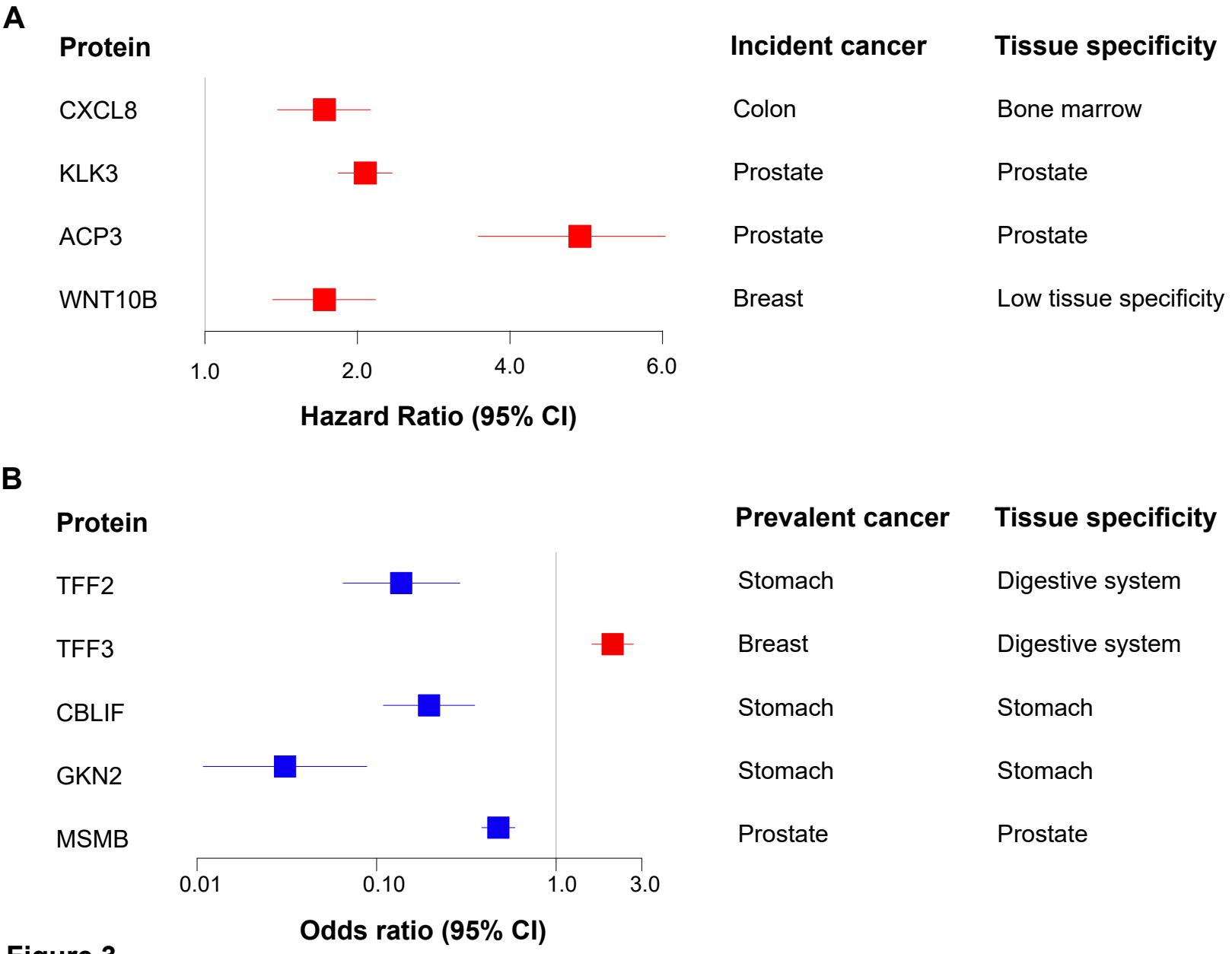

A
Protein
Incident cancer
Tissue specificity
CXCL8
Colon
Bone marrow
KLK3
Prostate
Prostate
ACP3
Prostate
Prostate
WNT10B
Breast
Low tissue specificity
1.0
2.0
4.0
6.0
Hazard Ratio (95% CI)
B
Protein
Prevalent cancer
Tissue specificity
TFF2
Stomach
Digestive system
TFF3
Breast
Digestive system
CBLIF
Stomach
Stomach
GKN2
Stomach
Stomach
MSMB
Prostate
Prostate
0.01
0.10
1.0
3.0
Odds ratio (95% CI)

Figure 3

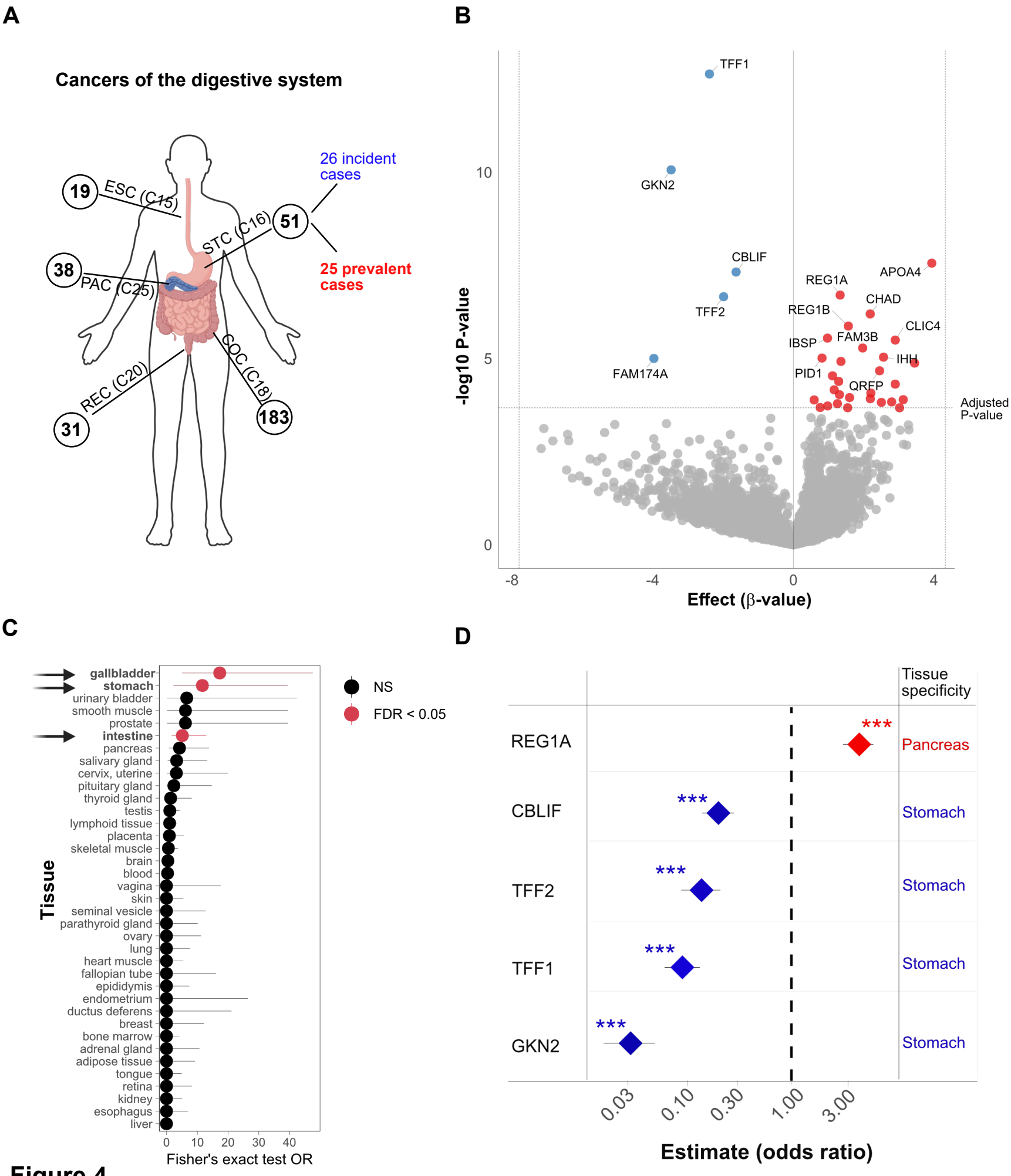

A
Cancers of the digestive system
19
ESC (C15)
38
PAC (C25)
STC (C16)
51
26 incident cases
25 prevalent cases
REC (C20)
31
COC (C18)
183
B
TFF1
GKN2
CBLIF
TFF2
FAM174A
REG1A
APOA4
CHAD
REG1B
CLIC4
IBSP
FAM3B
IHH
PID1
QRFP
Adjusted P-value
-log10 P-value
Effect (β-value)
C
gallbladder
stomach
urinary bladder
smooth muscle
prostate
intestine
pancreas
salivary gland
cervix, uterine
pituitary gland
thyroid gland
testis
lymphoid tissue
placenta
skeletal muscle
brain
blood
vagina
skin
seminal vesicle
parathyroid gland
ovary
lung
heart muscle
fallopian tube
epididymis
endometrium
ductus deferens
breast
bone marrow
adrenal gland
adipose tissue
tongue
retina
kidney
esophagus
liver
Tissue
NS
FDR < 0.05
Fisher's exact test OR
D
Tissue specificity
REG1A
Pancreas
CBLIF
Stomach
TFF2
Stomach
TFF1
Stomach
GKN2
Stomach
***
Estimate (odds ratio)

Figure 4

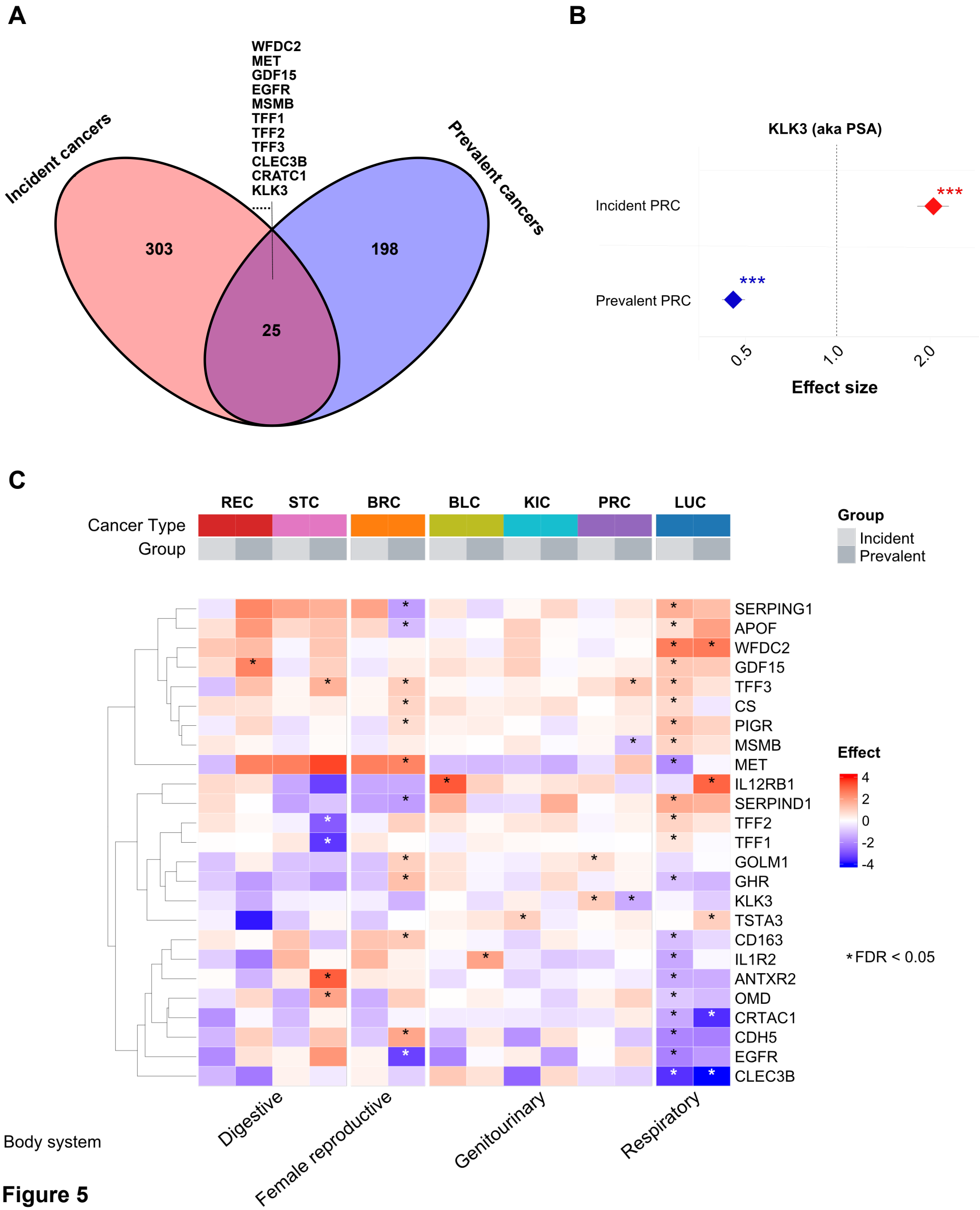

A
WFDC2
MET
GDF15
EGFR
MSMB
TFF1
TFF2
TFF3
CLEC3B
CRATC1
KLK3
.....
Incident cancers
Prevalent cancers
303
198
25
B
KLK3 (aka PSA)
Incident PRC
Prevalent PRC
***
***
0.5
1.0
2.0
Effect size
C
REC
STC
BRC
BLC
KIC
PRC
LUC
Cancer Type
Group
Group
Incident
Prevalent
SERPING1
APOF
WFDC2
GDF15
TFF3
CS
PIGR
MSMB
MET
IL12RB1
SERPIND1
TFF2
TFF1
GOLM1
GHR
KLK3
TSTA3
CD163
IL1R2
ANTXR2
OMD
CRTAC1
CDH5
EGFR
CLEC3B
Effect
4
2
0
-2
-4
* FDR < 0.05
Body system
Digestive
Female reproductive
Genitourinary
Respiratory

Figure 5

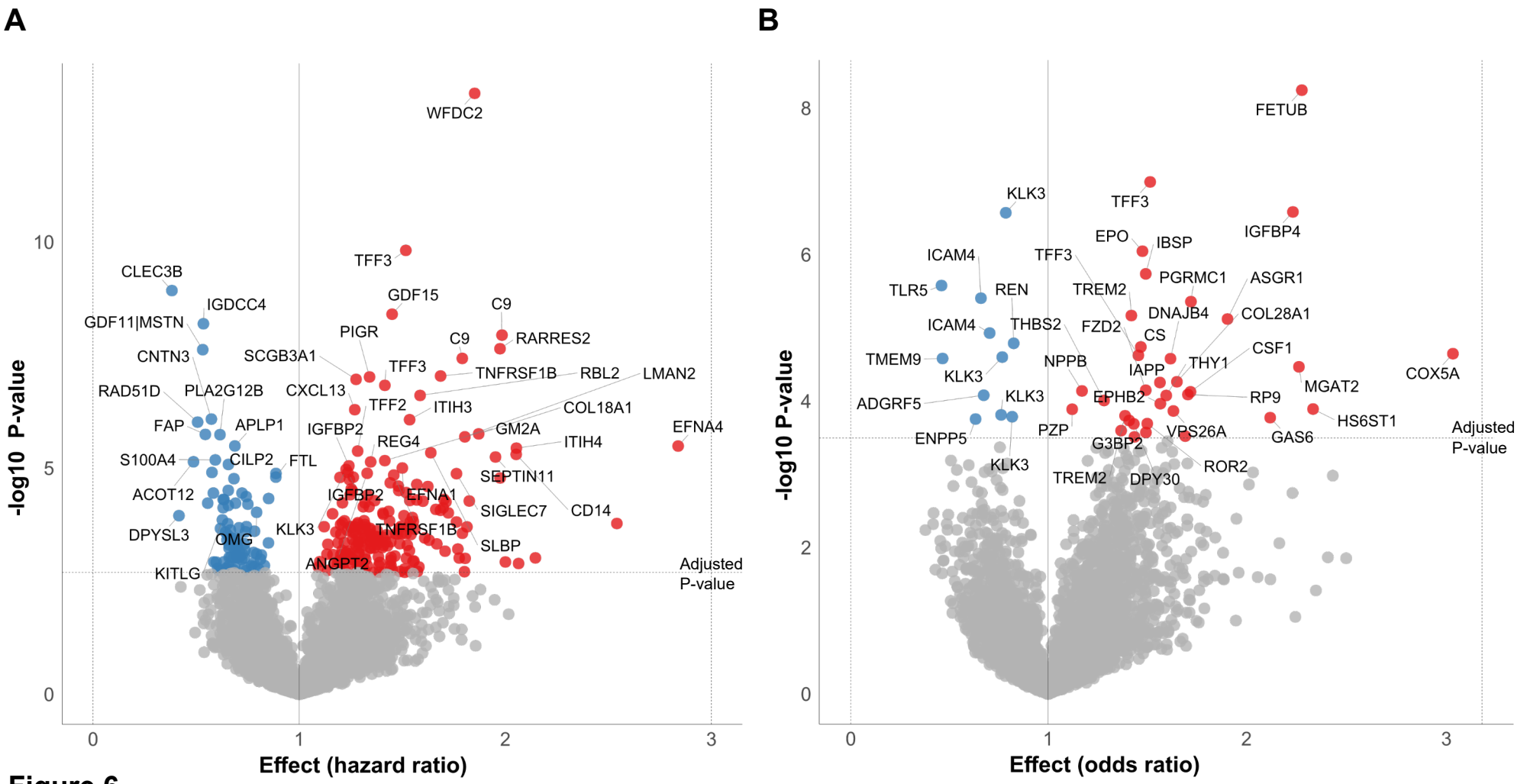

A
WFDC2
TFF3
CLEC3B
GDF15
C9
IGDCC4
GDF11|MSTN
PIGR
C9
RARRES2
CNTN3
SCGB3A1
TFF3
TNFRSF1B
RBL2
LMAN2
RAD51D
PLA2G12B
CXCL13
TFF2
ITIH3
COL18A1
FAP
APLP1
IGFBP2
GM2A
EFNA4
REG4
ITIH4
S100A4
CILP2
FTL
SEPTIN11
ACOT12
IGFBP2
EFNA1
SIGLEC7
CD14
DPYSL3
OMG
KLK3
TNFRSF1B
SLBP
KITLG
ANGPT2
Adjusted
P-value
-log10 P-value
Effect (hazard ratio)
B
FETUB
KLK3
TFF3
IGFBP4
EPO
IBSP
ICAM4
TFF3
PGRMC1
ASGR1
TLR5
REN
TREM2
DNAJB4
COL28A1
ICAM4
THBS2
FZD2
CS
CSF1
TMEM9
NPPB
THY1
COX5A
KLK3
IAPP
MGAT2
ADGRF5
KLK3
EPHB2
RP9
HS6ST1
ENPP5
PZP
VPS26A
GAS6
Adjusted
P-value
G3BP2
KLK3
ROR2
TREM2
DPY30
-log10 P-value
Effect (odds ratio)

Figure 6

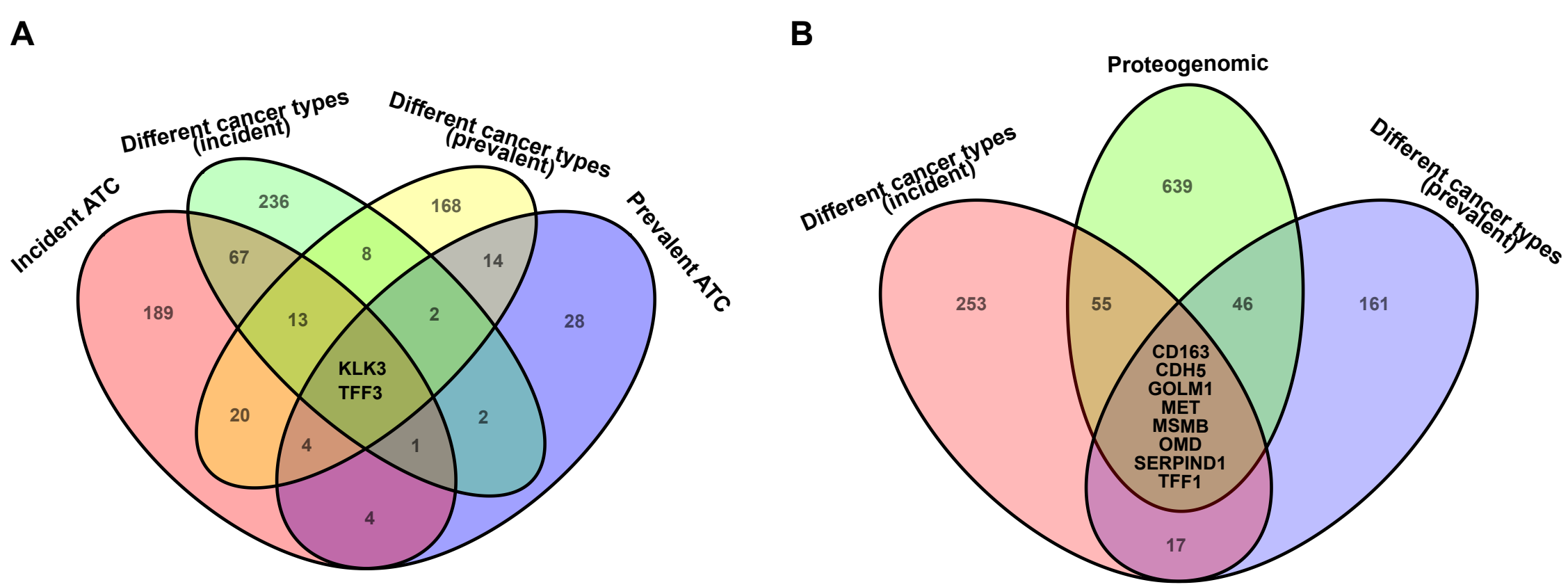

A
Incident ATC
Different cancer types (incident)
Different cancer types (prevalent)
Prevalent ATC
189
236
168
28
67
8
14
13
2
20
2
4
1
4
KLK3
TFF3
B
Proteogenomic
Different cancer types (incident)
Different cancer types (prevalent)
639
253
55
46
161
CD163
CDH5
GOLM1
MET
MSMB
OMD
SERPIND1
TFF1
17

Figure 7

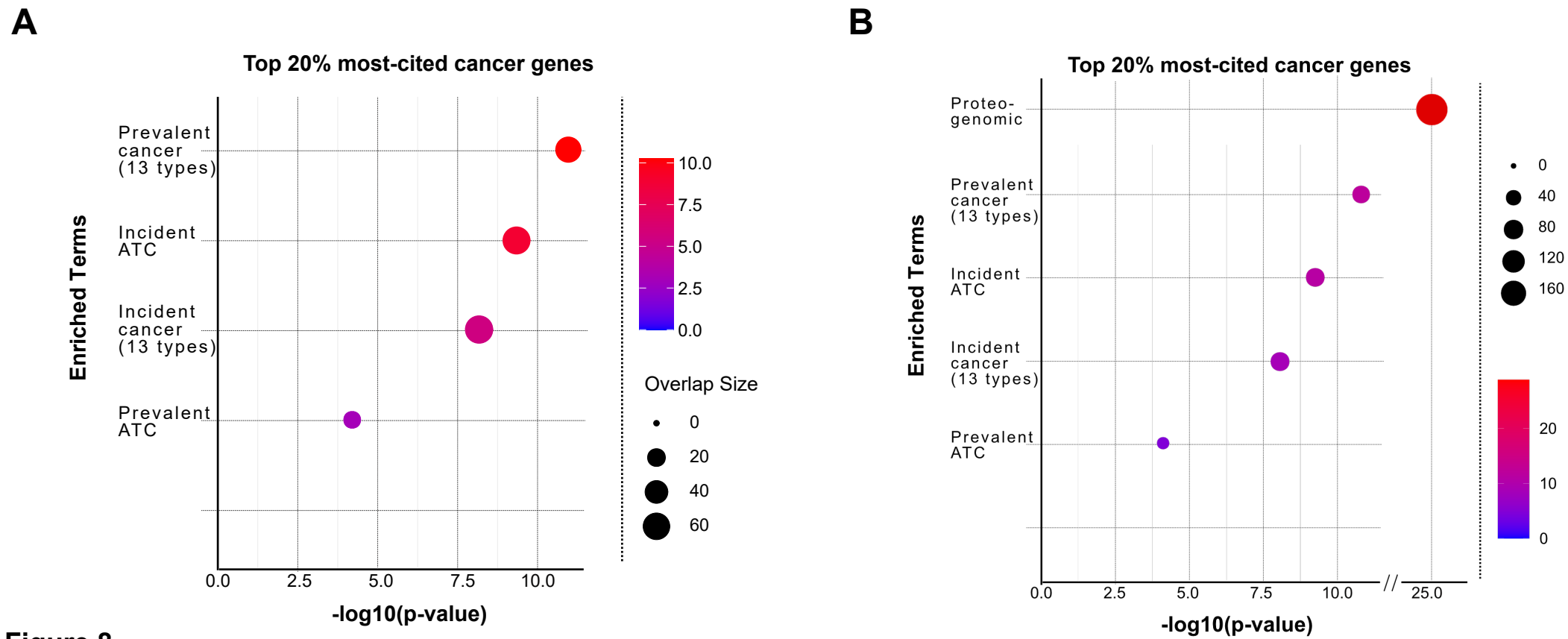
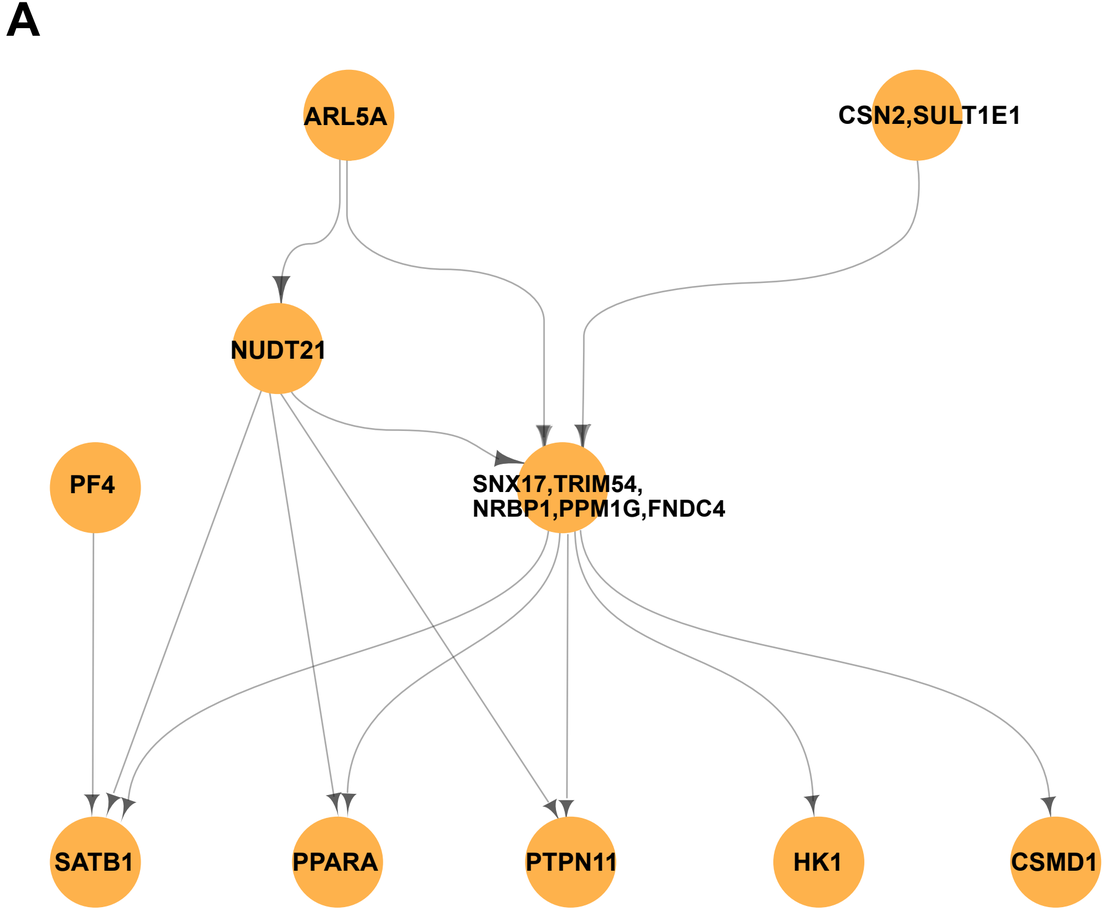

A
Top 20% most-cited cancer genes
Enriched Terms
Prevalent cancer (13 types)
Incident ATC
Incident cancer (13 types)
Prevalent ATC
0.0
2.5
5.0
7.5
10.0
-log10(p-value)
10.0
7.5
5.0
2.5
0.0
Overlap Size
0
20
40
60
B
Top 20% most-cited cancer genes
Enriched Terms
Proteo-genomic
Prevalent cancer (13 types)
Incident ATC
Incident cancer (13 types)
Prevalent ATC
0.0
2.5
5.0
7.5
10.0
25.0
-log10(p-value)
0
40
80
120
160
20
10
0

Figure 8


A
ARL5A
CSN2,SULT1E1
NUDT21
PF4
SNX17,TRIM54,
NRBP1,PPM1G,FNDC4
SATB1
PPARA
PTPN11
HK1
CSMD1

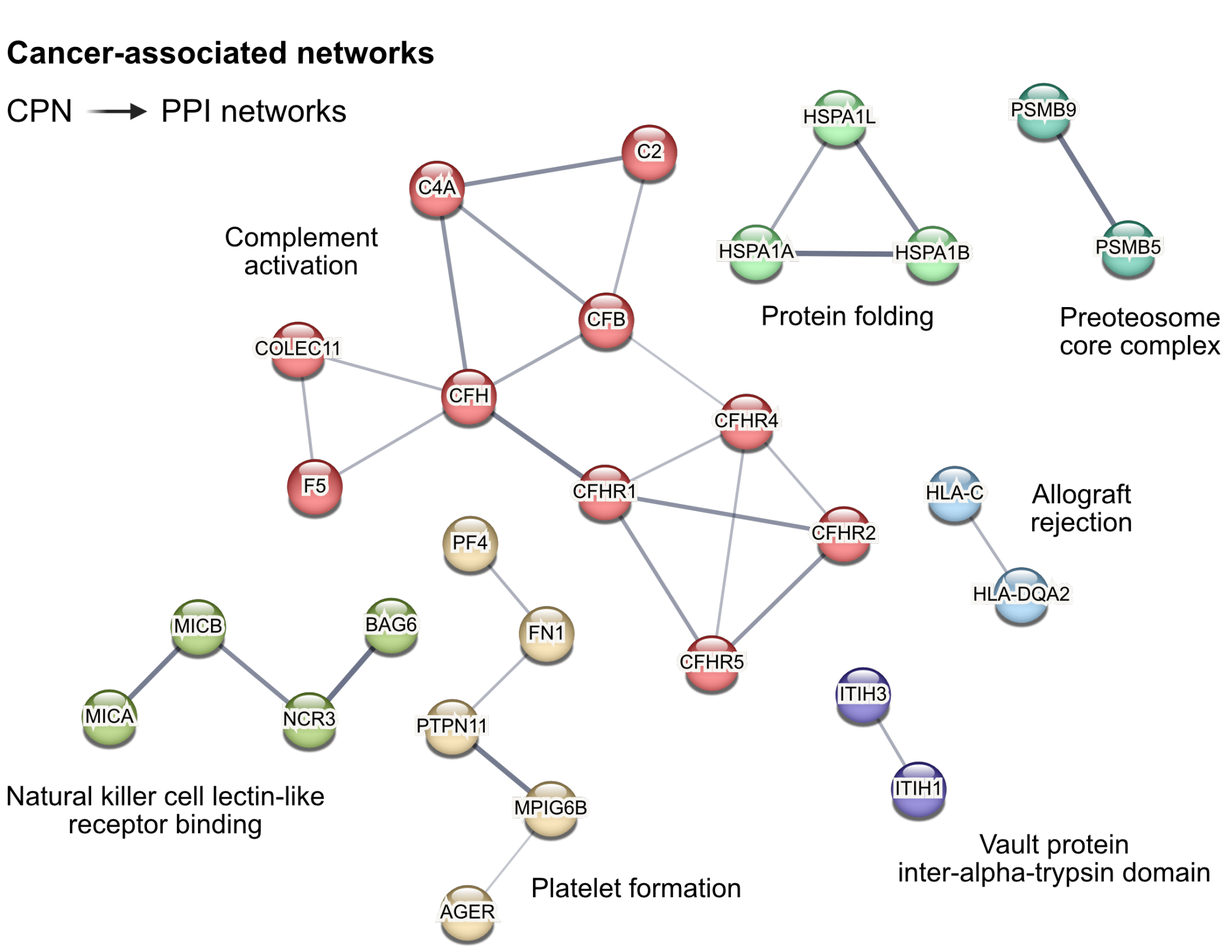

B
Cancer-associated networks
CPN → PPI networks
Complement activation
C4A
C2
CFB
COLEC11
CFH
F5
CFHR4
CFHR1
CFHR2
CFHR5
HSPA1L
HSPA1A
HSPA1B
Protein folding
PSMB9
PSMB5
Preoteosome core complex
HLA-C
HLA-DQA2
Allograft rejection
PF4
FN1
PTPN11
MPIG6B
AGER
Platelet formation
MICB
BAG6
MICA
NCR3
Natural killer cell lectin-like receptor binding
ITIH3
ITIH1
Vault protein inter-alpha-trypsin domain

Figure 9

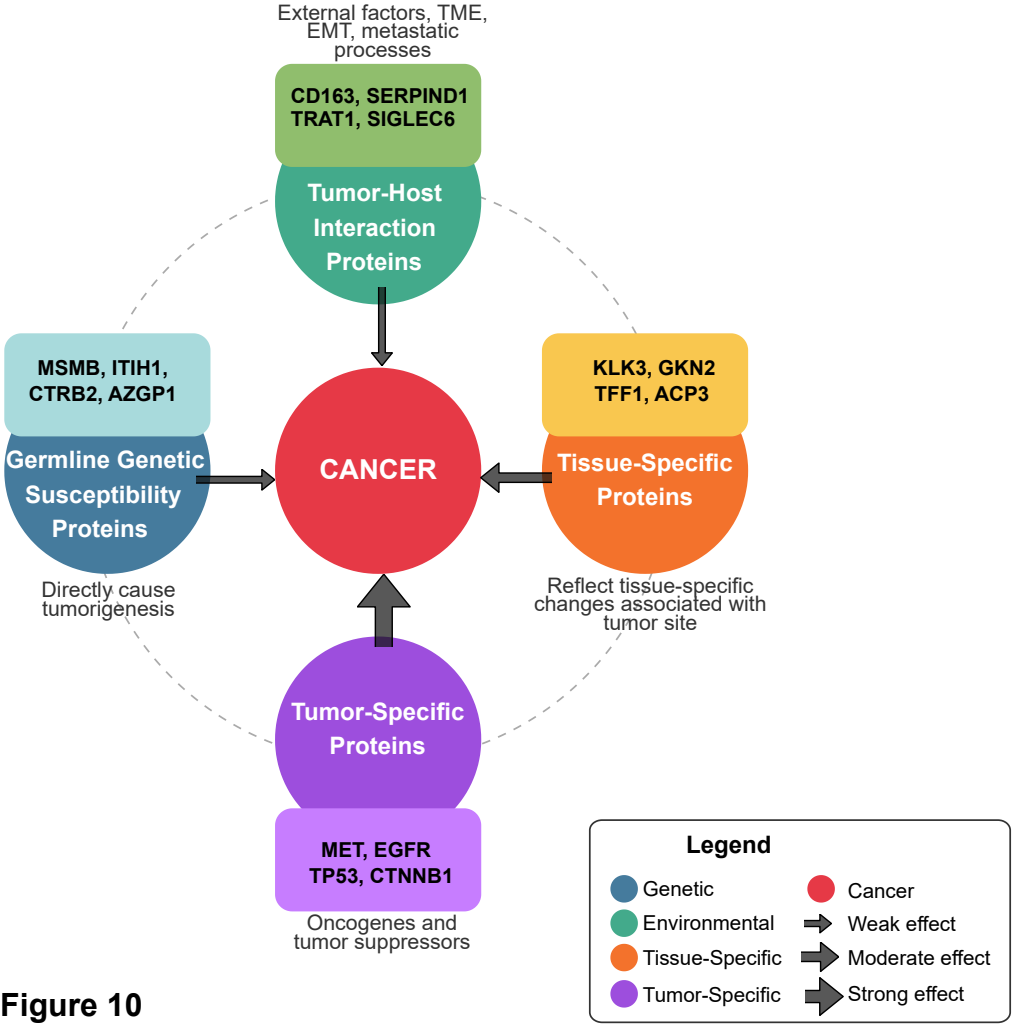

External factors, TME, EMT, metastatic processes
CD163, SERPIND1 TRAT1, SIGLEC6
Tumor-Host Interaction Proteins
MSMB, ITIH1, CTRB2, AZGP1
Germline Genetic Susceptibility Proteins
Directly cause tumorigenesis
CANCER
KLK3, GKN2 TFF1, ACP3
Tissue-Specific Proteins
Reflect tissue-specific changes associated with tumor site
Tumor-Specific Proteins
MET, EGFR TP53, CTNNB1
Oncogenes and tumor suppressors
Legend
Genetic
Environmental
Tissue-Specific
Tumor-Specific
Cancer
Weak effect
Moderate effect
Strong effect

Figure 10