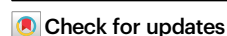


Deciphering the role of complement system genes in pancreatic cancer susceptibility and prognosis

Received: 27 November 2024

Accepted: 23 October 2025

Published online: 28 November 2025



A list of authors and their affiliations appears at the end of the paper

Pancreatic ductal adenocarcinoma (PDAC) genetic susceptibility is partially identified. The complement system (CS) influences carcinogenesis and participates in immunological defense and homeostasis; however, its role in PDAC genetic susceptibility and prognosis is underexplored. The association of SNPs within 111 CS-related genes with PDAC risk is assessed in the PanGenEU study and validated in the UKBiobank. We investigate the association between the CS-related gene variation and PDAC risk, followed by an in-depth functional in silico study using TCGA and ICGC data. We assess whether CS-related genes are associated with prognosis at the germline and somatic levels. We investigate the immune infiltration of PDAC tumors according to their transcriptomic profile. Genetic variation in *FCN1* and *PLAT* is significantly associated with PDAC risk. PDAC patients with elevated expression of *IGHG3*, *IGKC*, *IGHM*, *F2R*, *F2RL2*, *CFI*, *A2M*, or *C4A* display improved survival and higher infiltration of CD8⁺, B cells, and Th1 cells. Individuals with high expression levels of either *FGA*, *SERPINE1*, *FGG*, or *F3* exhibit poorer survival, higher infiltration of Tregs, and lower infiltration of CD8⁺ cells. Results from this study suggest that CS-related genes play a role in PDAC genetic susceptibility and survival through specific immune cell infiltration.

Pancreatic cancer remains one of the most lethal cancers, which is predicted to become the second leading cause of cancer-related deaths by 2030^{1,2}. Pancreatic ductal adenocarcinoma (PDAC) accounts for 95% of pancreatic cancer cases, and patients with PDAC display a dismal 5 year survival rate of 5–10%^{3,4}. Unfortunately, this bleak scenario will not change in the future unless multiple actions are taken to control PDAC, including the identification of PDAC high-risk populations to improve primary and secondary prevention interventions. To this end, a better understanding of the role of genetic and non-genetic factors involved in pancreas carcinogenesis is central.

The immune system plays a critical role in detecting and eliminating cancer cells; however, it can also promote tumor growth and progression through various mechanisms^{5,6}. Among the immune-related pathways, multiple pieces of evidence support the role of the Major Histocompatibility Complex (MHC) region in PDAC risk. While most studies have focused on the MHC class I and class II regions, the

MHC III region has barely been explored. This region is responsible for inflammatory processes and the CS response, a multistep cascade operating as a key component of the innate immune response. Its activation triggers the formation of the cell-killing membrane attack complex (MAC), which disrupts the cell membrane and ultimately leads to the lysis of target cells. This process is orchestrated through three distinct activation pathways: classical, lectin, and alternative pathways^{7–9}. Further to this role, the complement cascade interacts closely with the clotting system. There is strong evidence that several proteases of the coagulation system can activate C3 and C5, indirectly leading to the activation of the CS¹⁰. Notably, the CS can be activated by cancer cells and their microenvironment, producing inflammatory and immunomodulatory molecules that can promote tumor growth and suppress immune surveillance^{5,11,12}. Genetic variation within CS-related genes, such as *C3*, *C5*¹³, or *CD35*¹⁴, has been associated with the prognosis of different cancer types, including PDAC¹⁵. In fact, some CS

✉ e-mail: melopezdm@cniio.es; nmalats@cniio.es

genes have been suggested as biomarkers for PDAC prognosis and treatment response¹⁶. However, the role of the CS in cancer risk and prognosis, particularly in the context of PDAC, is still not well understood. This is particularly relevant given the fact that PDAC risk is also associated with asthma, allergies, autoimmune diseases, and other inflammatory conditions^{17,18}.

In this hypothesis-driven study, we report on the association between variation in CS-related genes and susceptibility to, as well as prognosis of, PDAC. Considering that the impact of a single genetic variant might have a small effect on a complex disease like PDAC, we conducted a comprehensive gene-based association analysis by aggregating all variants within each gene to provide a broader perspective on the overall gene-level effects. To gain deeper insights into the identified susceptibility loci in PDAC, we have also integrated the information on the association of SNPs with other omics data, such as expression and splicing quantitative trait loci (eQTL and sQTL, respectively) in normal pancreas, using colocalization and performing functional in silico analysis. Overall, our findings suggest that several CS-related genes play a significant role in the genetic susceptibility to PDAC. Finally, we demonstrate that the expression of different CS-related genes not only influences PDAC survival but also correlates with immune cell infiltration patterns in PDAC. We also provide a signature of genes whose expression is associated with improved PDAC prognosis. By utilizing large-scale studies and state-of-the-art analytical methodologies, our study contributes to a deeper understanding of the role of CS-related genes in pancreas carcinogenesis and progression, shedding light on the potential implications for disease susceptibility and prognosis.

Results

Characteristics of the study populations

The demographic and clinical characteristics of the PanGenEU and UK Biobank study populations are displayed in Supplementary Table 3. PanGenEU individuals were from Spain (58.8%), Italy (22.6%), Sweden (7.9%), Germany (6.7%), the UK (3.7%), and Ireland (0.2%). The average age of cases and controls was comparable within each study population, although the UK Biobank controls were younger. Moreover, a slightly lower proportion of females was observed in both cases and controls in PanGenEU (n. s.). In terms of risk factors, cases in both PanGenEU and UK Biobank populations had a higher prevalence of smoking and diabetes compared to controls. On the other hand, the prevalence of asthma and allergies was higher among controls than among cases in both studies (Supplementary Table 3). We observed significant differences in the proportion of stage I-II and III-IV PDAC tumors across studies, PanGenEU was the study population with the highest proportion of stage III-IV PDAC tumors and, therefore, more representative of the full spectrum of the disease. The median survival

of the TCGA patients was significantly shorter than survival in the other cohorts (PanGenEU, ICGC-AU, and ICGC-CA).

Single variant association analyses with PDAC risk

Results of the meta-analysis of the CS-related SNPs and PDAC risk are presented in Supplementary Data 1. A total of 653 SNPs located within 41 CS-related genes were found to be significantly associated with PDAC risk at a nominal $P < 0.05$. However, after adjusting for multiple testing (Benjamini–Hochberg), none of the SNPs remained statistically significant (minimum adjusted $P = 0.09$). The five SNPs with the lowest P -value map to *FL3A1* (rs6597196), *A2M* (rs34803501, rs12427063, and rs12426061), and *CFH* (rs10922096).

Gene-based association analysis with PDAC risk and established PDAC risk factors

Utilizing SKAT-O, we identified two genes significantly associated with PDAC risk in the meta-analysis, ficolin-1 (*FCN1*, adjusted $P = 9.9 \times 10^{-3}$) and tissue-type plasminogen activator (*PLAT*, adjusted $P = 1.56 \times 10^{-2}$). Furthermore, four genes showed a borderline significant association: vitronectin (*VTN*, adjusted $P = 7.64 \times 10^{-2}$), the complement regulatory protein *CD46* (adjusted $P = 7.24 \times 10^{-2}$), coagulation factor II receptor-like 2 (*F2RL2*, adjusted $P = 7.24 \times 10^{-2}$), and alpha-2-macroglobulin (*A2M*, adjusted $P = 7.64 \times 10^{-2}$) (Table 1). Notably, *PLAT* and *F2RL2* were also associated with asthma and *VTN* with nasal allergies (Supplementary Table 4).

Functional in silico analyses

The functional in silico analyses provided valuable insights into the CS-related genes associated with PDAC risk. Gene set enrichment analysis conducted using FUMA GWAS revealed that pathways related to humoral immune responses, coagulation, and regulation of immune responses were enriched with the significant set of CS-related genes (Supplementary Fig. 1). Exploring the DisGeNET database, we found that many of the CS-related genes associated with PDAC risk had previously been associated with digestive system diseases and neoplasms, as well as with nervous diseases and mental disorders (Supplementary Fig. 2).

By using GTEx expression data, we discovered that within the *VTN* gene, eight out of the eleven SNPs associated with an increased risk of PDAC (nominal $P < 0.05$) were also eQTLs associated with a higher expression of *VTN* in the normal pancreas: rs2227726, rs2227721, rs2227720, rs2227718, rs111502247, rs2071378, rs2071377, and rs113635549 (max. PDAC OR = 1.17, $P = 4.58 \times 10^{-2}$; max. eQTL OR = 2.34, $P = 3.66 \times 10^{-11}$) (Supplementary Fig. 3a). Additionally, a single SNP within *CD46* was identified as a sQTL, meaning that it is also associated with differential splicing (*CD46*-rs2796278, PDAC OR = 1.08, $P = 5.79 \times 10^{-2}$, sQTL OR = 0.51, $P = 5.73 \times 10^{-16}$) (Supplementary Fig. 3b). Results of colocalization analyses (Supplementary Table 5) suggest that there is a moderate likelihood that the causal variants associated with a higher risk of PDAC (rs2111023, rs12427063, and rs12426061) are also causal of differential expression of the *A2M* gene. The posterior probability for colocalization of PDAC-associated SNPs and eQTLs corresponding to the *A2M* gene was 46.3%. For *VTN*, the posterior probability that the SNPs are causal for both PDAC risk and expression was much lower (13.5%). For *PLAT*, *FCN1*, *CD46*, and *F2RL2* the posterior probabilities for sharing a causal variant were lower than 0.1, as well as for all the genes in the colocalization analysis of PDAC and sQTLs (Supplementary Table 6).

Prognostic associations of CS-related genes in PDAC patients

We first assessed the association between the CS-related SNPs and genes with PDAC survival in both PanGenEU and UK Biobank populations independently. To further assess their association with survival, we conducted a meta-analysis incorporating the findings from each population. This identified 583 SNPs with a nominal P -value

Table 1 | Number of SNPs (nSNPs) within each gene and results from the SKAT-O association test between complement system-related genes and PDAC risk

Genes	nSNPs	P-value	Adj. P-value
<i>FCN1</i>	61	3.30E-04	9.90E-03
<i>PLAT</i>	148	1.04E-03	1.56E-02
<i>CD46</i>	188	8.21E-03	7.24E-02
<i>F2RL2</i>	43	9.66E-03	7.24E-02
<i>VTN</i>	42	1.33E-02	7.64E-02
<i>A2M</i>	214	1.53E-02	7.64E-02

P-values were calculated using a two-sided SKAT-O test, which jointly evaluates the effect of multiple SNPs within each gene on the risk of PDAC. SKAT-O does not provide effect sizes and confidence intervals. P-values were adjusted for multiple comparisons using the Benjamini–Hochberg FDR method (Adj. P-value).

significantly associated with PDAC survival. However, after adjusting for multiple comparisons, these associations did not retain statistical significance (Supplementary Data 2). Additionally, through our SKAT-O model, we also observed that three CS-related genes (*C8B*, *F2*, and *KLKB1*) were significantly associated with PDAC survival at a nominal *P*-value ($P < 0.05$, Supplementary Table 6).

Prognostic CS-based gene signatures for patient stratification

Initially, we tested the association between the expression of CS-related genes and PDAC OS at the individual study level (PanGenEU, TCGA, ICGC-CA, and ICGC-AU). We then performed a meta-analysis combining the results from each study. Out of 17 CS-related genes associated with PDAC OS at nominal $P < 0.05$, 12 remained significant after correcting for multiple testing: Ig heavy chain gamma 3 (*IGHG3*), Ig heavy chain mu (*IGHM*), Ig kappa chain (*IGKC*), coagulation factor II thrombin receptor (*F2R*), *F2RL2*, complement factor I (*CFI*), alpha-2 macroglobulin (*A2M*), complement component 4A (*C4A*), serpin family E member 1 (*SERPINE1*), fibrinogen alpha chain (*FGA*), fibrinogen gamma chain (*FGG*), coagulation factor III, tissue factor (*F3*). Results showed that an elevated expression of *IGHG3*, *IGHM*, *IGKC*, *F2R*, *F2RL2*, *CFI*, *A2M*, and *C4A* was significantly associated with better OS (Fig. 1). In contrast, higher expression of *SERPINE1*, *FGA*, *FGG*, and *F3* was significantly associated with poorer OS (Fig. 2).

We also investigated the association of two gene expression signatures comprising genes whose expression was associated with better (*Signature 1*: *IGHG3*, *IGKC*, *IGHM*, *F2R*, *F2RL2*, *CFI*, and *A2M*) or worse (*Signature 2*: *FGA*, *SERPINE1*, *FGG*, *F3*) OS in the PanGenEU, TCGA, ICGC-CA, and ICGC-AU cohorts, followed by a meta-analysis. Patients with high expression levels of *Signature 1* genes experienced a 32% reduction in mortality risk compared to those with low expression levels ($HR = 0.68$, $P = 4.0 \times 10^{-4}$, Fig. 3a). Conversely, patients with high expression levels of *Signature 2* genes displayed a less favorable OS compared to their low expression counterparts. However, this association was not statistically significant ($HR = 1.16$, $P = 2.01 \times 10^{-1}$, Fig. 3b).

Patients stratified according to CS-related gene expression have distinct immune cell infiltration profiles

We found that patients with elevated expression of any of the CS-related genes associated with better survival in the meta-analysis (*IGHG3*, *IGKC*, *IGHM*, *F2R*, *F2RL2*, *CFI*, and *A2M*), had significantly higher infiltration of TCD8⁺, B cells, and Th1 (Fig. 4). These PDAC tumors also exhibited an increased abundance of state-S1 TCD8⁺, state-S2 TCD4⁺, or state-S1 macrophages (Fig. 5), as well as a higher abundance of C10 ecotype (Fig. 6).

Patients with high expression of *Signature 1* also had increased infiltration by TCD8⁺, Th1, or B cells, and a reduced presence of cells like M2 macrophages (Fig. 4), higher abundance of state-S1 TCD8⁺, state-S1 and state-S2 TCD4⁺, and state-S1 B cells (Fig. 5), along with a higher abundance of ecotype CE10 (Fig. 6). Interestingly, PDAC tumors with high *Signature 1* levels showed elevated clonal expansion of B cells compared to those with low *Signature 1* levels (Supplementary Fig. 4).

Conversely, tumors expressing higher levels of *FGA* or *F3* genes, associated with worse outcomes, had significantly reduced infiltration of TCD8⁺ (Fig. 4). Tumors with high expression of *F3* or *FGG* had a lower abundance of state-S1 NK (Fig. 5) and CE10 ecotype (Fig. 6), and higher abundance of state-S1 mast cells and CE2 ecotype.

Discussion

The development of pancreatic ductal adenocarcinoma is influenced by several genetic factors as described before^{19,20}. The present hypothesis-driven study aimed to investigate the potential role of CS-related genes in PDAC risk and prognosis. Understanding the genetic factors involved in PDAC development is crucial to defining high-risk populations suitable for screening and primary prevention targeted interventions.

Our findings provide valuable insights suggesting that CS is associated with PDAC risk as well as with that of allergy and asthma through a meta-analysis of PanGenEU and UKBiobank study results, both at the SNP and gene levels. Functional in silico analyses of identified susceptibility SNPs/genes have offered additional evidence supporting the involvement of CS-related genes in PDAC and shed light on their potential mechanisms of action, including immune regulation, expression modulation, and tumor-specific expression changes. In particular, eight SNPs in *VTN* identified as risk variants were also associated with a higher *VTN* expression in the normal pancreas. Our study also identified moderate and weak colocalization between PDAC and eQTL signals for *A2M* and *VTN*, respectively, suggesting that the SNPs within these genes might share a functional effect on PDAC risk and gene expression.

At the gene level, we identified six (four with borderline significance) susceptibility CS-related genes, including *FCN1*, which is expressed in leukocytes and involved in elastin-binding^{21,22}; *PLAT* converts plasminogen into plasmin, crucial in cell migration and tissue remodeling/degradation²³; *CD46* protects host cells from complement self-cell damage^{24,25}; *F2RL2* is key in thrombosis²⁶; *VTN* promotes cell adhesion²⁷ and inhibits the MAC²⁸; and *A2M*, which works as a protease inhibitor^{29,30}. Remarkably, our study provides evidence of the association between these CS-related genes and PDAC risk. Interestingly, *PLAT* is among the top overexpressed genes in PDAC compared with normal pancreas^{31,32}. Functional studies have shown its role in tumor growth and invasion in vitro and in vivo^{33,34} through a variety of mechanisms including non-catalytic effects³⁵. We also discovered associations between *PLAT* and *F2RL2* and the risk of asthma, and *VTN* with the risk of allergy, highlighting the interplay between complement-related genetic variations and immune-related conditions also associated with PDAC risk.

Pathways related to humoral immune responses and coagulation are enriched with the identified CS-related genes. Notably, *A2M*, *F2RL2*, *VTN*, and *PLAT* are related to thrombosis, a condition that is prevalent in cancer patients, especially among those with PDAC³⁶, who frequently suffer thromboembolic events^{37–39} due to the release of procoagulant factors like thrombin⁴⁰. *A2M* plays a crucial role in regulating thrombin activity, while thrombin can cleave the *F2RL2* receptor⁴¹, enhancing platelet activation, promoting a procoagulant state, and contributing to thromboembolic diseases⁴². *VTN* binds to fibrin and platelets participating in platelet aggregation⁴³ and inhibiting *PLAT*, responsible for dissolving blood clots^{44,45}. The interplay between these genes and the coagulation cascade components warrants further investigation to elucidate their contribution to the development and progression of PDAC. Furthermore, DisGeNET analysis showed that some of these genes were associated with central nervous system diseases. *PLAT* and *A2M*, previously linked to depression^{46,47}, were also associated with PDAC risk. Interestingly, PDAC has the highest prevalence of depression symptoms preceding its diagnosis^{48–50} compared to other cancer types, suggesting a potential connection between genetic variations in *PLAT* and *A2M*, depression, and PDAC. Furthermore, *A2M* has been linked to neuronal disorders like Alzheimer's disease^{51,52}. PDAC is a neurotropic tumor^{53–55} with a high incidence of perineural invasion. Given this neurotropic nature of PDAC, the association between *A2M* and PDAC becomes even more relevant for its potential involvement in Alzheimer's disease, supporting the observed inverse relationship between neurodegenerative diseases and cancer^{56–60}. However, further research is needed to explore the interplay between these genes, neuronal and mental diseases, and the development of PDAC.

Next, we aimed to identify robust prognostic biomarkers considering the CS-related genetic variation and transcriptomic data. Although no significant variants at the germline level were identified, our results found biomarkers at the somatic level. We observed that overexpression of eight CS-related genes was associated with a

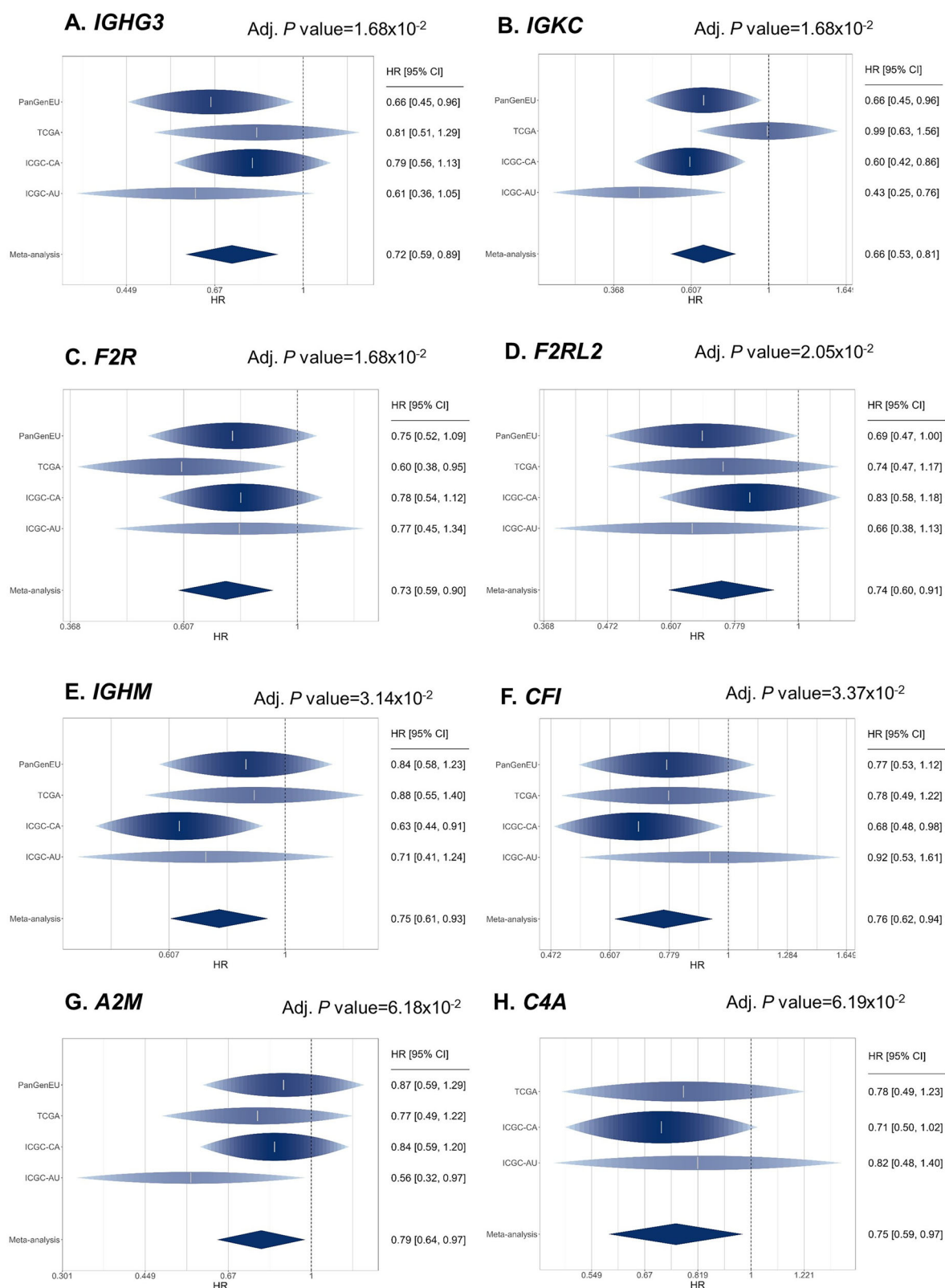


Fig. 1 | Forest plots from the meta-analysis of complement system-related genes associated with improved overall survival when considering PanGenEU, TCGA, ICGC-CA, and ICGC-AU (Fig. 1A–1G) and when including data from TCGA, ICGC-CA, and ICGC-AU (Fig. 1H). The Hazard ratios (HR) and 95% confidence

intervals (CI) were estimated using two-sided Cox proportional-hazards models under a dominance mode of inheritance. Meta-analysis p -values were adjusted for multiple comparisons using the Benjamini–Hochberg false discovery rate method.

favorable prognosis. Among them, *IGHG3*, *IGKC*, and *IGHM* code for proteins essential for immunoglobulin (IG) integrity, indicating a robust immune response potentially through the clonal expansion of B cells and increased IG levels to target tumor cells. We also found that high expression levels of *F2R* and *F2RL2*, crucial for platelet activation and thrombotic responses, were also associated with anti-tumor immune cell infiltration, including TCD8⁺, TCD4⁺, B cells, and DCs. We hypothesize that these inflammation processes, in the context of “cold” tumors like PDAC, might favor the activation and infiltration of the abovementioned TCD8⁺, DCs, and B cells, leading to improved patient survival. Conversely, Wu et al. (2024)⁶¹ identified *F2R* as a potential biomarker of adverse outcomes in gastric adenocarcinoma due to inflammatory processes. Elevated expression of *F2RL2* has been associated with improved prognosis in esophageal squamous cell carcinoma⁶², although with poorer survival in colorectal adenocarcinoma⁶³ and gliomas⁶⁴. Furthermore, we showed that *A2M* expression was associated with enhanced overall survival outcomes and anti-tumor immune cell infiltration (TCD8⁺, Th1, and B cells), consistent with findings in intrahepatic cholangiocarcinoma, where its low expression was associated with unfavorable prognosis⁶⁵. Lastly, we found that greater expression of *CFI* was associated with enhanced patient survival and infiltration of tumor-tolerant immune cells (TCD8⁺, Th1, and NK cells). Other studies have associated *CFI* with shorter breast cancer survival⁶⁶, and tumor progression in cutaneous squamous cell carcinoma and gliomas^{67,68}. Considering that the dual role of these genes might be cancer-specific, further research on the roles of *F2R*, *F2RL2*, *A2M*, *CFI*, and *C4A* in PDAC prognosis is crucial. Additionally, we propose a prognostic gene expression signature

composed of the genes associated with a favorable prognosis: *Signature 1*. Interestingly, patients with higher levels of this signature had tumors with enhanced clonal expansion, higher immune infiltration of tumor-killing immune cells, and state-S1 TCD8⁺, state-S2 TCD4⁺, or state-S1 macrophages or ecotype CE10, previously associated with better overall survival⁶⁹, suggesting a more robust immune response. Based on these findings, we anticipate that these patients may respond favorably to immunotherapy.

We also identified four biomarkers associated with unfavorable prognosis: *FGA*, *FGG*, *SERPINE1*, and *F3*. These genes are involved in blood clot formation through the coagulation cascade. Higher levels of *FGA* and *FGG* are linked to poor prognosis in gastric cancer⁷⁰. Serum levels of *FGG* are also suggested to act as a biomarker for castration-resistant prostate cancer⁷¹. *SERPINE1* expression has been associated with poorer prognosis⁷² and cancer progression in colon⁷³ and gastric⁷⁴ cancer. *F3*, which activates the coagulation cascade via coagulation factor VII interaction, is associated with colorectal cancer⁷⁵. We observed that PDAC tumors displaying high expression of these genes also had a decreased presence of anti-tumor immune cells like TCD8⁺, decreased abundance of state-S1 mast cells or NK cells, and higher abundance of cell communities such as CE1 and CE2⁶⁹ ecotypes, possibly creating an environment conducive to tumor progression, leading to poorer patient outcomes⁶⁹. Interestingly, a signature comprising *SERPINE1* and *F3* among other genes, has been recently associated with worse prognosis in lower-grade glioma⁷⁶. Based on our results and previous evidence, we speculate that targeting *SERPINE1*, *F3*, *FGG*, or *FGA* genes in combination with chemotherapies could enhance treatment effectiveness in PDAC tumors. This is particularly relevant as

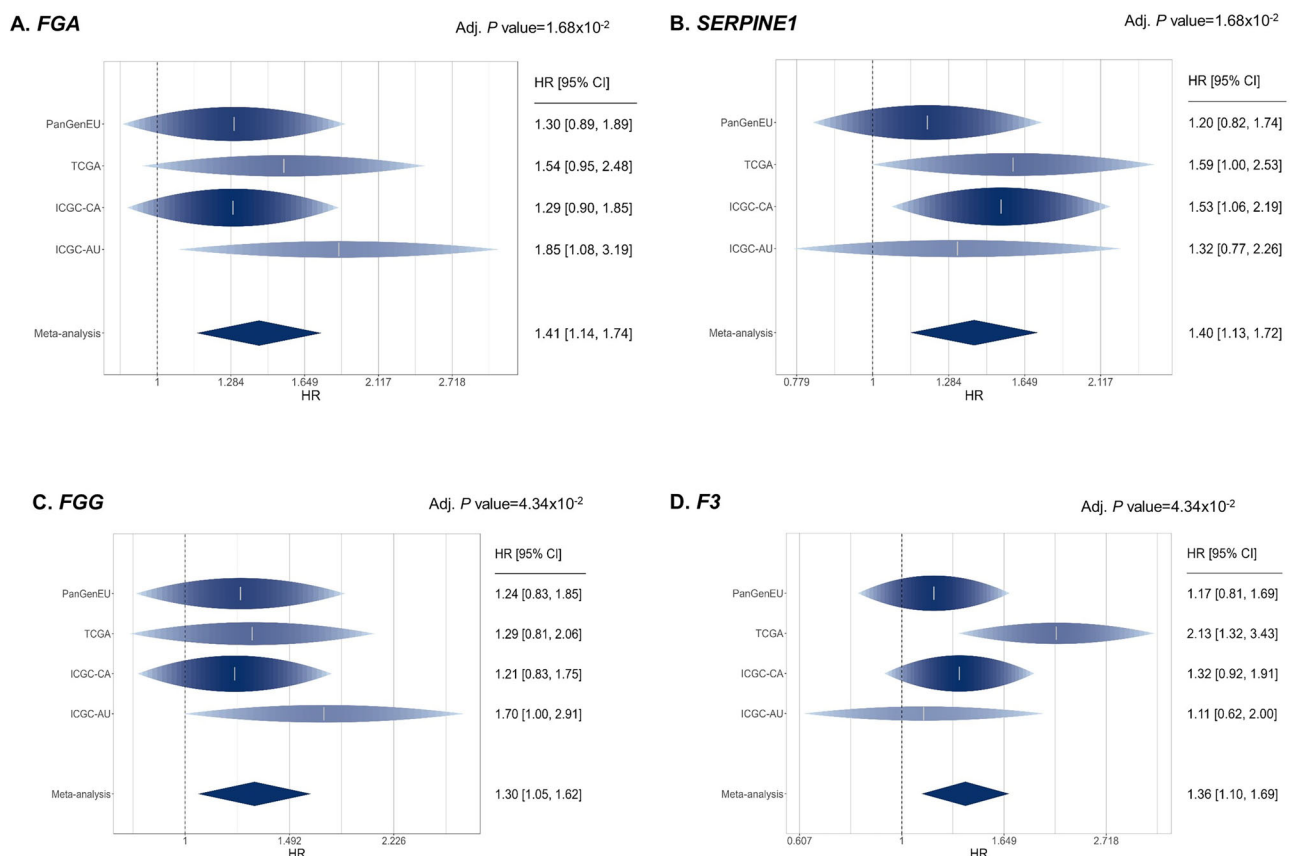


Fig. 2 | Forest plots from the meta-analysis of the complement system-related genes indicate an association with poorer overall survival when considering PanGenEU, TCGA, ICGC-CA, and ICGC-AU (Fig. 2A–2D). Hazard ratios (HR) and 95% confidence intervals (CI) were estimated using two-sided Cox proportional-

hazards regression models under a dominance mode of inheritance. Meta-analysis *p*-values were adjusted for multiple comparisons using the Benjamini–Hochberg false discovery rate method. Source data are provided as a Source Data file.

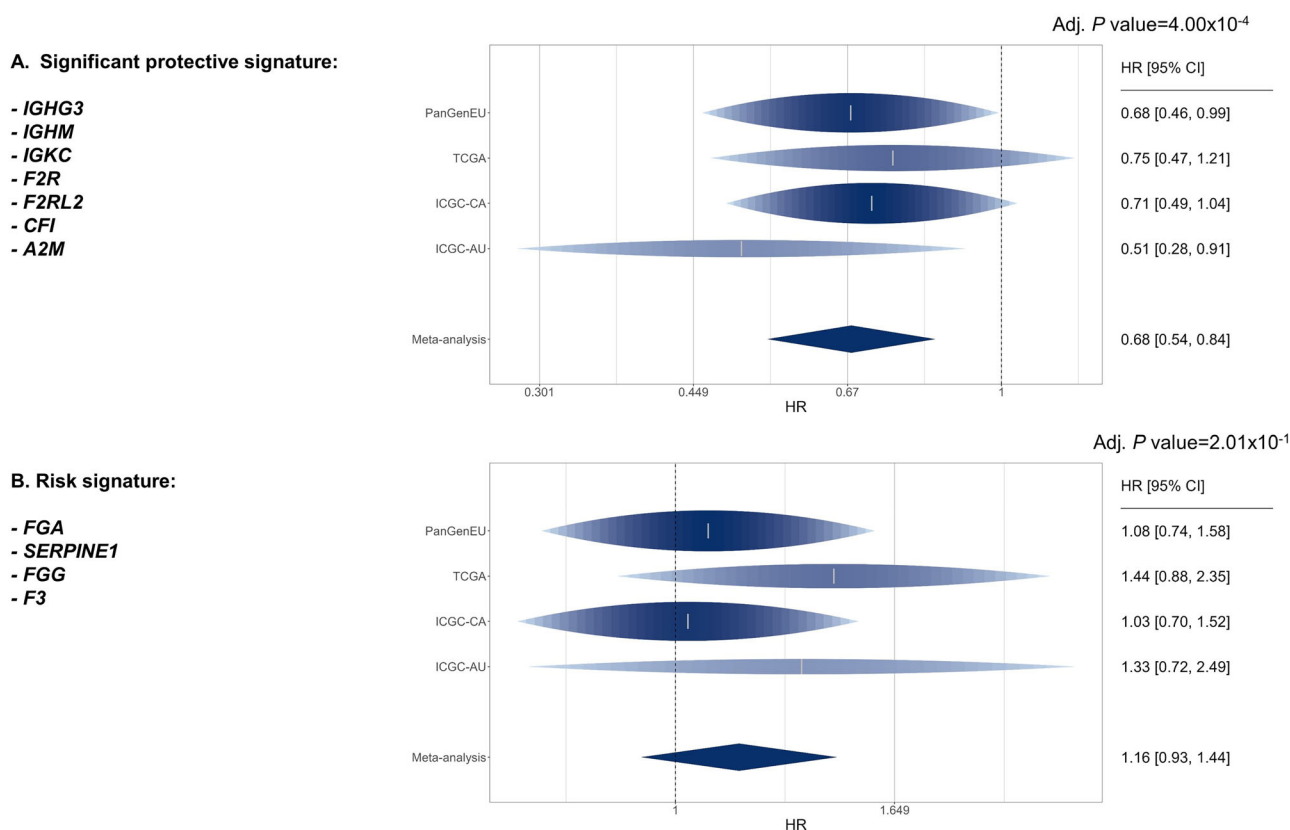


Fig. 3 | Forest plots from the meta-analysis of complement system-related genes illustrate the protective signature that includes genes linked to improved overall survival, as analyzed in the PanGenEU, TCGA, ICGC-CA, and ICGC-AU databases (Fig. 3A). Additionally, the risk signature considers CS-related genes tied to an increased risk of death in the TCGA, ICGC-CA, and ICGC-AU cohorts

(Fig. 3 B). Hazard ratios (HR) and 95% confidence intervals (CI) are estimated using two-sided Cox proportional-hazards regression models under a dominance mode of inheritance. Meta-analysis *p*-values were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate method. Source data are provided as a Source Data file.

some *F3* targeted therapies, like Genmab's tisotumab vedotin, are already approved to treat cervical cancer⁷⁷. Additionally, *SERPINE1* has been shown to reduce the efficacy of chemotherapy agents, promoting drug resistance in triple-negative breast tumors⁷⁸, while *FGA* has been proposed for targeted therapy responsiveness in NSCLC patients⁷⁹.

This study has some limitations. First, our analyses focused on the European population, and therefore, the generalizability to other ancestry groups remains uncertain. Additionally, the study relied on self-reported epidemiological information, which may introduce misclassification of risk factors and comorbidity. However, the use of a meta-analysis approach incorporating data from different populations helped mitigate this potential bias. In this study, we also encountered one of the key limitations affecting association studies in complex diseases, which relates to the small effect sizes of SNP association estimates, resulting in a lack of significant results. The gene-based analysis using the SKAT-O approach alleviates this issue, although it does not elucidate the direction of the effect. We minimized this limitation by applying functional in silico analyses. We applied the Benjamini-Hochberg procedure to control false discoveries. Lastly, we focused on common variants; future studies should explore rare variants through exome or genome sequencing or targeted analysis to gain deeper insights.

Importantly, our study has several strengths. We conducted an extensive analysis of the association between the CS-related genes and PDAC risk utilizing advanced methodologies in two large studies with distinct designs, PanGenEU and UK Biobank, enhancing the statistical power and robustness of our findings. Additionally, we thoroughly examined the association of CS-related genes with various PDAC risk factors using comprehensive epidemiological and clinical data, which

provides a deeper understanding of the complex interplay between CS-related genes and PDAC risk. The use of four independent study populations to explore the role of CS-based genes on PDAC prognosis allowed us to identify a strong signature indicative of a better prognosis, which is also supported by immune infiltration characteristics within the tumor in addition to the statistical evidence.

In conclusion, our work exhaustively assesses the association of six CS-related genes (*CD46*, *A2M*, *VTN*, *F2RL2*, *FCN1*, and *PLAT*) with PDAC risk. The functional relevance of these genetic variations, particularly in gene expression regulation and disease pathways, underscores their role in PDAC. These findings highlight their importance as potential biomarkers for PDAC risk profiling, leading to early diagnosis. Furthermore, we showed that gene expression patterns of CS-related genes may help to discriminate PDAC patients based on their prognosis. We also shed light on the molecular mechanisms underlying pancreatic carcinogenesis, which might provide valuable insights for personalized treatment strategies. Future research should focus on unraveling the intricate mechanisms through which these complement-related genes contribute to PDAC development, as well as exploring their potential as therapeutic targets in PDAC tumors.

Methods

Study populations

We utilized the resources of the PanGenEU case-control study, the UK Biobank, the International Cancer Genome Consortium (ICGC), Pancreatic Cancer Canada (ICGC-CA) and Pancreatic Cancer Australia (ICGC-AU) repositories (<https://dcc.icgc.org/repositories>), and The Cancer Genome Atlas (TCGA) repository (<https://portal.gdc.cancer.gov>). A description of these studies and populations can be found in

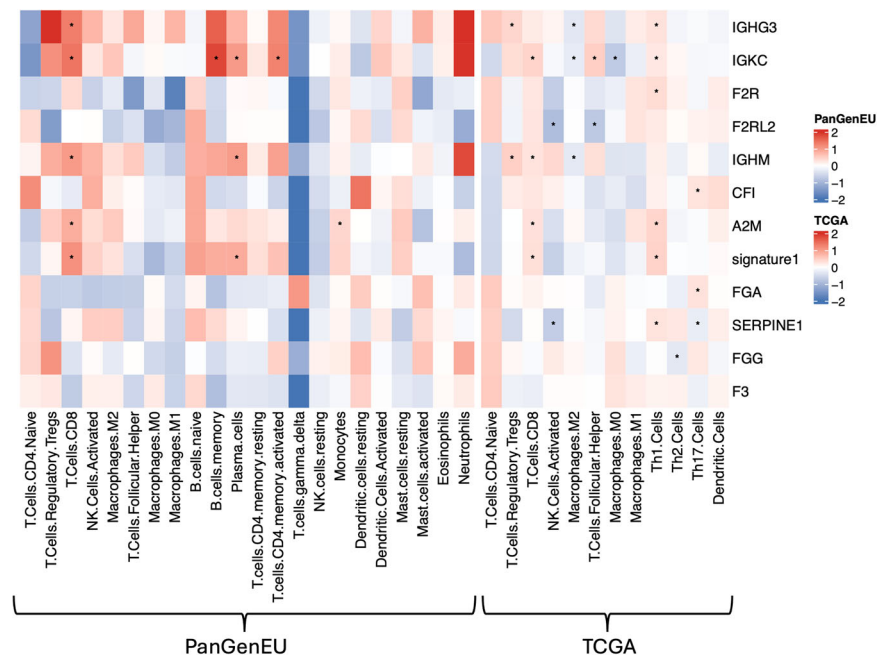


Fig. 4 | Complement-related gene expression is linked to distinct immune-cell infiltration patterns in PDAC. A heatmap displays the $\log_2$ fold-change ($\log_2FC$) in the relative abundance of immune-cell populations when comparing samples with high expression of each complement system-related gene to those with low expression. The left block (PanGenEU, $n = 142$) corresponds to 21 immune cells derived from CIBERSORTx. The right block (TCGA-PAAD, $n = 124$) contains 12 immune cells inferred by Thorsson. Columns are ordered based on hierarchical clustering of the cell types shared between the two cohorts; the brace below indicates the two data sources. Rows represent the 12 complement system-related

genes that were included in both expression matrices (*IGHG3*, *IGKC*, *F2R*, *F2RL2*, *IGHM*, *CFI*, *A2M*, signature 1, *FGA*, *SERPINE1*, *FGG*, *F3*). Color intensity reflects $\log_2FC$ (red = higher, blue = lower abundance in the high-expression group). Gray squares represent cell types for which all values were missing in one comparison group (e.g., Th1 cells in TCGA). Asterisks indicate significant differences that remain after the Benjamini–Hochberg correction ($FDR < 0.05$, two-sided Wilcoxon rank-sum test). One extreme outlier for T cells CD4 Naïve in TCGA was set to NA for that specific comparison only; the sample was retained for all other cell-type tests. Source data are provided as a Source Data file.

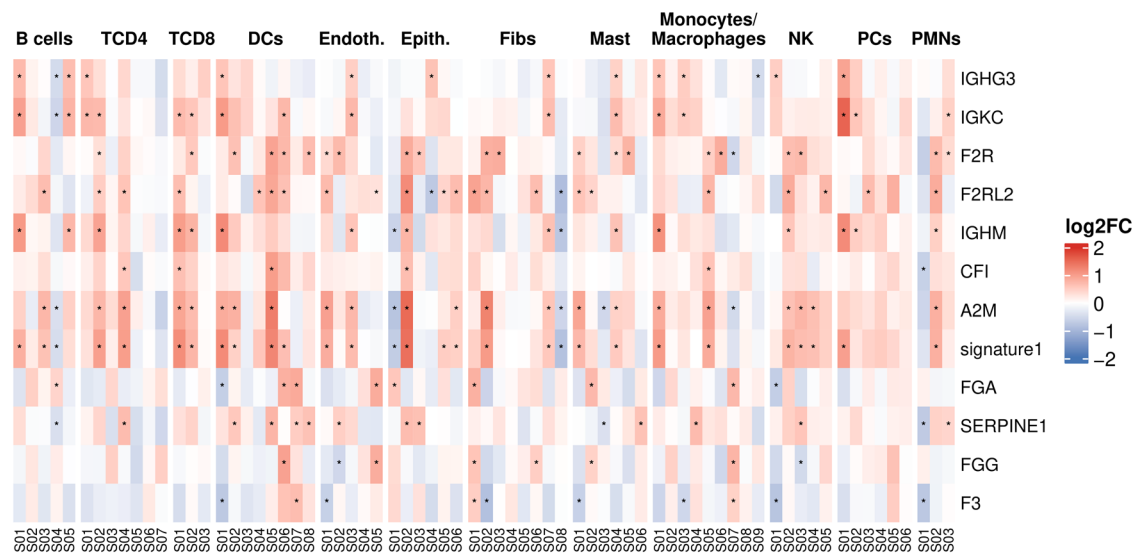


Fig. 5 | The heatmap illustrates the relationships between gene-cell state types, highlighting significant differences in the abundance of 12 immune cell types in PanGenEU. This is based on a comparison of high expression levels of complement system-related genes versus low expression levels. Color intensity reflects

$\log_2FC$ (red = higher, blue = lower abundance in the high-expression group). Asterisks indicate significant differences that remain after the Benjamini–Hochberg correction ($FDR < 0.05$, two-sided Wilcoxon rank-sum test). Source data are provided as a Source Data file.

Supplementary Methods. Sex information was determined by self-report. For the statistical analyses, we prioritized individuals with genetic European ancestry belonging to the PangenEU and UKBiobank study populations, and those whose self-reported race was white in TCGA and ICGC.

Single variant association analysis with PDAC risk
For the single variant association analysis with PDAC risk, we considered both imputed and genotyped SNPs annotated within the 111 complement-related genes preselected by Qian et al. (2019)⁸⁰ (Supplementary Table 2), for both PanGenEU (197,050 SNPs) and UK

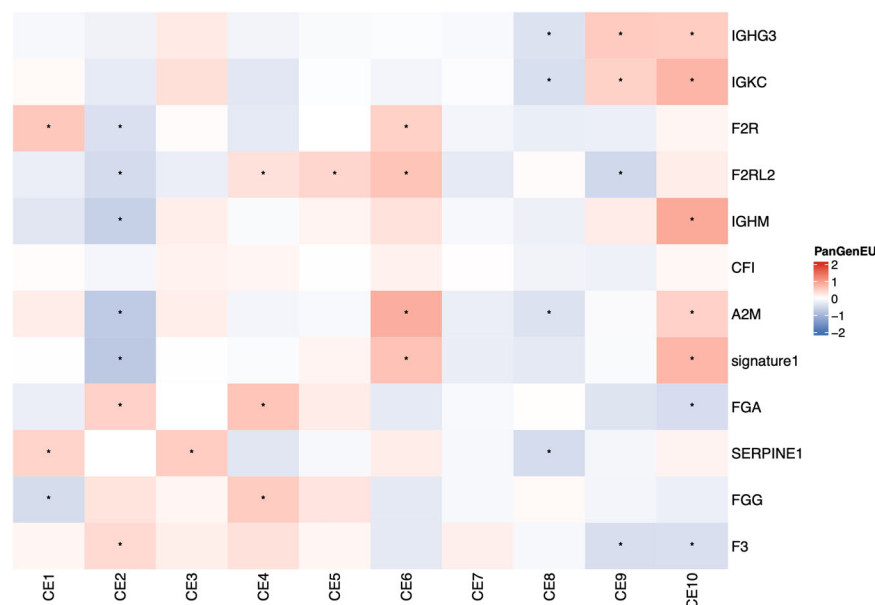


Fig. 6 | Heatmap illustrating gene-ecotype relationships that highlights the significant differences in the abundance of 10 immune ecotypes in PanGenEU by comparing high expression levels of the complement system-related genes to low expression levels. T Color intensity reflects log₂FC (red = higher, blue =

lower abundance in the high-expression group). Asterisks indicate significant differences that remain after the Benjamini–Hochberg correction (FDR < 0.05, two-sided Wilcoxon rank-sum test). Source data are provided as a Source Data file.

Biobank (183,849 SNPs) studies. In the PanGenEU, we prioritized SNPs with a minor allele frequency (MAF) $\geq 5\%$ and high imputation quality (info > 0.3). To estimate the OR and 95% CI for the selected SNPs with PDAC risk in each study population, we conducted logistic regression models adjusted for the first five principal components to account for population stratification, age, sex, and the European region as covariates in the case of the PanGenEU study population. For the UK Biobank population, we adjusted the models for the aforementioned variables except for the European region, in addition to the genotype array type and center. Finally, we combined the individual study estimates in a meta-analysis assessing the summary statistics of the association analysis of 5,932 SNPs obtained in both populations using the function *rma.umi* (metafor R package, v. 4.2-0)⁸¹. To limit the number of false positives resulting from the multiple comparisons performed in the meta-analysis, given the hypothesis-driven approach we followed, we applied the Benjamini-Hochberg procedure, one of the most widely used false discovery rate (FDR) methods.

Gene-based association analysis with PDAC risk

We conducted a gene-based association analysis in each population using the Sequence Kernel Association Test (SKAT-O), implemented in the R package *seqMeta* (v1.6.7) (<https://github.com/DavisBrian/seqMeta>). To do this, we employed the SNPs included in the single variant association analysis, followed by a filtering step based on linkage disequilibrium (LD > 0.7). We adjusted the SKAT-O models for the same covariates as mentioned previously. Furthermore, to detect potential associations and gain more robust insights into the role of the complement-related genes in PDAC risk, we performed a meta-analysis of the summary statistics from the 111 complement-related genes derived from both study populations. We then assessed the association (OR and 95%CI) between the CS-related genes linked with PDAC risk in our analysis and the well-known PC risk factors, including asthma, allergies, tobacco, family history of cancer, and body mass index (BMI) in each study population using logistic regression models adjusted for the same covariates as previous models. Furthermore, we conducted a meta-analysis using a random-effects model to integrate the results from both study populations and applied FDR as previously to correct for multiple comparisons.

To understand the functional implications of the genes significantly associated with PDAC risk, we conducted an *in silico* functional analysis using various bioinformatics tools (see **Supplementary Methods**).

Survival analysis of CS-related genes at germline and somatic levels

We examined the association between the SNPs within 111 CS-related genes and PDAC overall survival (OS) in the PanGenEU and UK Biobank populations (see **Supplementary Methods**).

Similarly, we used data on the expression levels of the 111 CS-related genes in PDAC tumors to examine their association with the overall survival of 134 PDAC patients from TCGA, classifying them as low-expression and high-expression groups based on the median. To validate these findings, we analyzed survival data from (PanGenEU ($n = 122$), ICGC-CA ($n = 160$), and ICGC-AU ($n = 86$)) employing Cox PH models with adjustment for sex, age at diagnosis, and tumor stage in each population. Next, we meta-analyzed the findings from the four study populations, focusing on CS-related genes with a nominal $P < 0.1$ in any of them. We also created gene expression signatures based on the meta-analysis results: *Signature 1*, comprising genes whose expression was associated with improved overall survival after applying multiple test correction; *Signature 2*, comprising genes whose expression was associated with a decreased OS after applying FDR. These signatures were created by converting gene expression levels into a score, with patients being classified into low-expression or high-expression groups based on the median expression levels of the signature. Then, we assessed the association between these signature scores and PDAC survival in each study population and conducted a comprehensive meta-analysis combining findings from the four cohorts.

Finally, we explored whether the observed associations between CS-related genes at the individual and signature levels with survival might be partially explained by specific immune infiltration patterns within PDAC tumors (see **Supplementary Methods**).

Ethics Statement

IRB ethical approval was obtained by all participating centers contributing to PanGenEU (Supplementary Note 1). Written informed

consent was obtained from all study participants. The study was conducted in accordance with the Helsinki Declaration.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The International Cancer Genome Consortium (ICGC), Pancreatic Cancer Canada (ICGC-CA), and Pancreatic Cancer Australia (ICGC-AU) repositories (<https://dcc.icgc.org/repositories>), under the project codes PACA-CA and PACA-AU; and the Cancer Genome Atlas (TCGA) repository (<https://portal.gdc.cancer.gov>), under the following project code TCGA-PAAD. ICGC and TCGA publicly available data supporting this publication can be accessed through the following hyperlink: <https://pancreasexpression.org/analytics/public-data/>. Data supporting these findings from UK Biobank are available through (<https://www.ukbiobank.ac.uk/enable-your-research/register>) and can be found in the UK Biobank Data Showcase, <https://biobank.ctsu.ox.ac.uk/crystal/search.cgi> under the ICD codes displayed in Table S1. Access to this data is available from the authors with UKB permission. The PanGenEU datasets analyzed during the current study are not publicly available due to the sensitive nature of genetic data. Still, they are available from the corresponding author, Núria Malats, (nmalats@cnio.es) upon request, with permission from the PanGenEU consortia. Any request will be reviewed by the “Data Access Committee (DAC)” to ensure patient consent and privacy. The DAC assesses requests within a week and responds within a month. The length of the data use will be established for each request. Databases of eQTL and sQTL from GTEx are available online (<https://www.gtexportal.org/home/>). The source data of the figures are provided in the source data file. Source data are provided with this paper.

Code availability

No custom code or mathematical algorithms were developed for this study. All analyses were performed using publicly available software and packages, as described in the Methods section.

References

- Rahib, L., Wehner, M. R., Matrisian, L. M. & Nead, K. T. Estimated projection of US cancer incidence and death to 2040. *JAMA Netw. Open* **4**, e214708 (2021).
- Rahib, L. et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).
- Arnold, M. et al. Global burden of 5 major types of gastrointestinal cancer. *Gastroenterology* **159**, 335–349.e15 (2020).
- Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA A Cancer J. Clin.* **74**, 12–49 (2024).
- Cho, M. S. et al. Autocrine effects of tumor-derived complement. *Cell Rep.* **6**, 1085–1095 (2014).
- Fan, Z., Qin, J., Wang, D. & Geng, S. Complement C3a promotes proliferation, migration and stemness in cutaneous squamous cell carcinoma. *J. Cell Mol. Med.* **23**, 3097–3107 (2019).
- Walport, M. J. Complement. *N. Engl. J. Med.* **344**, 1058–1066 (2001).
- Chaplin, D. D. 1. Overview of the immune response. *J. Allergy Clin. Immunol.* **111**, S442–S459 (2003).
- Murphy, K. & Weaver, C. *Janeway's Immunobiology* 9th edn, Vol. 904 (Garland Science, 2016).
- Amara, U. et al. Interaction between the coagulation and complement system. In *Current Topics in Complement I* (ed. Lambris, J. D.) 632 68–76 (Springer US, New York, NY, 2008).
- Markiewski, M. M. et al. Modulation of the anti-tumor immune response by complement. *Nat. Immunol.* **9**, 1225–1235 (2008).
- Vadrevu, S. K. et al. Complement C5a receptor facilitates cancer metastasis by altering T-Cell responses in the metastatic niche. *Cancer Res.* **74**, 3454–3465 (2014).
- Lawal, B. et al. Pan-cancer analysis of immune complement signature C3/C5/C3AR1/C5AR1 in association with tumor immune evasion and therapy resistance. *Cancers* **13**, 4124 (2021).
- Luo, L. et al. Genetic polymorphisms in CD35 gene contribute to the susceptibility and prognosis of hepatocellular carcinoma. *Front. Oncol.* **11**, 700711 (2021).
- Shimazaki, R. et al. Complement factor B regulates cellular senescence and is associated with poor prognosis in pancreatic cancer. *Cell Oncol. (Dordr.)* **44**, 937–950 (2021).
- Zhang, X. et al. Enhanced complement expression in the tumor microenvironment following neoadjuvant therapy: implications for immunomodulation and survival in pancreatic ductal adenocarcinoma. *Res. Sq.* **13**, rs.3.rs-4104258 (2024).
- Gomez-Rubio, P. et al. Reduced risk of pancreatic cancer associated with asthma and nasal allergies. *Gut* **66**, 314–322 (2017).
- Gomez-Rubio, P. et al. Pancreatic cancer and autoimmune diseases: an association sustained by computational and epidemiological case-control approaches. *Int. J. Cancer* **144**, 1540–1549 (2019).
- López de Maturana, E. et al. A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. *Genome Med.* **13**, 15 (2021).
- Pantaleo, A. et al. Understanding the genetic landscape of pancreatic ductal adenocarcinoma to support personalized medicine: a systematic review. *Cancers* **16**, 56 (2024).
- Endo, Y., Sato, Y., Matsushita, M. & Fujita, T. Cloning and characterization of the human lectin P35 gene and its related gene. *Genomics* **36**, 515–521 (1996).
- Frankenberger, M., Schwaebler, W. & Ziegler-Heitbrock, L. Expression of M-Ficolin in human monocytes and macrophages. *Mol. Immunol.* **45**, 1424–1430 (2008).
- Chapman, H. A. Plasminogen activators, integrins, and the coordinated regulation of cell adhesion and migration. *Curr. Opin. Cell Biol.* **9**, 714–724 (1997).
- Oglesby, T. J., Allen, C. J., Liszewski, M. K., White, D. J. & Atkinson, J. P. Membrane cofactor protein (CD46) protects cells from complement-mediated attack by an intrinsic mechanism. *J. Exp. Med.* **175**, 1547–1551 (1992).
- Liszewski, M. K., Farries, T. C., Lublin, D. M., Rooney, I. A. & Atkinson, J. P. Control of the complement system. *Adv. Immunol.* **61**, 201–283 (1996).
- Ishihara, H. et al. Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* **386**, 502–506 (1997).
- Madsen, C. D., Ferraris, G. M. S., Andolfo, A., Cunningham, O. & Sidenius, N. uPAR-induced cell adhesion and migration: vitronectin provides the key. *J. Cell Biol.* **177**, 927–939 (2007).
- Milis, L., Morris, C. A., Sheehan, M. C., Charlesworth, J. A. & Pussell, B. A. Vitronectin-mediated inhibition of complement: evidence for different binding sites for C5b-7 and C9. *Clin. Exp. Immunol.* **92**, 114–119 (1993).
- Downing, M. R., Bloom, J. W. & Mann, K. G. Comparison of the inhibition of thrombin by three plasma protease inhibitors. *Biochemistry* **17**, 2649–2653 (1978).
- Sottrup-Jensen, L., Sand, O., Kristensen, L. & Fey, G. H. The α -macroglobulin bait region: sequence diversity and localization of cleavage sites for proteinases in five mammalian α -macroglobulins. *J. Biol. Chem.* **264**, 15781–15789 (1989).
- Paciucci, R. et al. Isolation of tissue-type plasminogen activator, cathepsin H, and non-specific cross-reacting antigen from SK-PC-1 pancreas cancer cells using subtractive hybridization. *FEBS Lett.* **385**, 72–76 (1996).

32. Paciucci, R., Torà, M., Díaz, V. M. & Real, F. X. The plasminogen activator system in pancreas cancer: role of t-PA in the invasive potential in vitro. *Oncogene* **16**, 625–633 (1998).
33. Díaz, V. M., Planaguma, J., Thomson, T. M., Reventós, J. & Paciucci, R. Tissue plasminogen activator is required for the growth, invasion, and angiogenesis of pancreatic tumor cells. *Gastroenterology* **122**, 806–819 (2002).
34. Aguilar, S. et al. Tissue plasminogen activator in murine exocrine pancreas cancer: selective expression in ductal tumors and contribution to cancer progression. *Am. J. Pathol.* **165**, 1129–1139 (2004).
35. Roda, O. et al. Galectin-1 is a novel functional receptor for tissue plasminogen activator in pancreatic cancer. *Gastroenterology* **136**, 1379–1390, e1–5 (2009).
36. Mulder, F. I. et al. Venous thromboembolism in cancer patients: a population-based cohort study. *Blood* **137**, 1959–1969 (2021).
37. Khorana, A. A., Francis, C. W., Culakova, E., Kuderer, N. M. & Lyman, G. H. Frequency, risk factors, and trends for venous thromboembolism among hospitalized cancer patients. *Cancer* **110**, 2339–2346 (2007).
38. Cronin-Fenton, D. P. et al. Hospitalisation for venous thromboembolism in cancer patients and the general population: a population-based cohort study in Denmark, 1997–2006. *Br. J. Cancer* **103**, 947–953 (2010).
39. Shaib, W., Deng, Y., Zilberman, D., Lundberg, B. & Saif, M. W. Assessing risk and mortality of venous thromboembolism in pancreatic cancer patients. *Anticancer Res.* **30**, 4261–4264 (2010).
40. Rudroff, C., Striegler, S., Schilli, M. & Scheele, J. Thrombin enhances adhesion in pancreatic cancer in vitro through the activation of the thrombin receptor PAR 1. *Eur. J. Surg. Oncol. (EJSO)* **27**, 472–476 (2001).
41. Hänzelmann, S. et al. Thrombin stimulates insulin secretion via protease-activated receptor-3. *Islets* **7**, e118195 (2016).
42. Mosley, J. D. et al. Mechanistic phenotypes: an aggregative phenotyping strategy to identify disease mechanisms using GWAS data. *PLoS One* **8**, e81503 (2013).
43. Podor, T. J. et al. Incorporation of vitronectin into fibrin clots: evidence for a binding interaction between vitronectin and γ A/ γ' fibrinogen*. *J. Biol. Chem.* **277**, 7520–7528 (2002).
44. Arroyo De Prada, N. et al. Interaction of plasminogen activator inhibitor type-1 (PAI-1) with vitronectin. *Eur. J. Biochem.* **269**, 184–192 (2002).
45. Reheman, A. et al. Vitronectin stabilizes thrombi and vessel occlusion but plays a dual role in platelet aggregation. *J. Thrombosis Haemost.* **3**, 875–883 (2005).
46. Matthews, K. A. et al. Associations between depressive symptoms and inflammatory/hemostatic markers in women during the menopausal transition. *Psychosom. Med.* **69**, 124–130 (2007).
47. Cattaneo, A. et al. FoxO1, A2M, and TGF- β 1: three novel genes predicting depression in gene X environment interactions are identified using cross-species and cross-tissues transcriptomic and miRNomic analyses. *Mol. Psychiatry* **23**, 2192–2208 (2018).
48. Kenner, B. J. Early detection of pancreatic cancer. *Pancreas* **47**, 363–367 (2018).
49. Jarrin Jara, M. D., Gautam, A. S., Peesapati, V. S. R., Sadik, M. & Khan, S. The role of interleukin-6 and inflammatory cytokines in pancreatic cancer-associated depression. *Cureus* **12**, e9969 (2020).
50. Seoud, T. et al. Depression preceding and following a diagnosis of pancreatic cancer: results from a national, population-based study. *Pancreas* **49**, 1117–1122 (2020).
51. Kovacs, D. M. α 2-Macroglobulin in late-onset Alzheimer's disease. *Exp. Gerontol.* **35**, 473–479 (2000).
52. Varma, V. R. et al. Alpha-2 macroglobulin in Alzheimer's disease: a marker of neuronal injury through the RCAN1 pathway. *Mol. Psychiatry* **22**, 13–23 (2017).
53. Li, J. et al. Hyperglycemic tumor microenvironment induces perineural invasion in pancreatic cancer. *Cancer Biol. Ther.* **16**, 912–921 (2015).
54. Chen, S.-H. et al. Perineural invasion of cancer: a complex crosstalk between cells and molecules in the perineural niche. *Am. J. Cancer Res.* **9**, 1–21 (2019).
55. Tan, X. et al. Nerve fibers in the tumor microenvironment in neurotropic cancer—pancreatic cancer and cholangiocarcinoma. *Oncogene* **40**, 899–908 (2021).
56. Catalá-López, F. et al. Inverse and direct cancer comorbidity in people with central nervous system disorders: a meta-analysis of cancer incidence in 577,013 participants of 50 observational studies. *Psychother. Psychosom.* **83**, 89–105 (2014).
57. Ibáñez, K., Boullosa, C., Tabarés-Seisdedos, R., Baudot, A. & Valencia, A. Molecular evidence for the inverse comorbidity between central nervous system disorders and cancers detected by transcriptomic meta-analyses. *PLoS Genet* **10**, e1004173 (2014).
58. Sánchez-Valle, J. et al. A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's disease, Glioblastoma and Lung cancer. *Sci. Rep.* **7**, 4474 (2017).
59. Greco, A. et al. Molecular inverse comorbidity between Alzheimer's disease and lung cancer: new insights from matrix factorization. *Int. J. Mol. Sci.* **20**, 3114 (2019).
60. Forés-Martos, J. et al. Transcriptomic and genetic associations between Alzheimer's disease, parkinson's disease, and cancer. *Cancers (Basel)* **13**, 2990 (2021).
61. Wu, X., Wang, S., Wang, C., Wu, C. & Zhao, Z. Bioinformatics analysis identifies coagulation factor II receptor as a potential biomarker in stomach adenocarcinoma. *Sci. Rep.* **14**, 2468 (2024).
62. Zhao, M. et al. Weighted correlation network analysis of cancer stem cell-related prognostic biomarkers in esophageal squamous cell carcinoma. *Technol. Cancer Res. Treat.* **21**, 153303382211170 (2022).
63. Zhou, R., Gao, Z. & Ju, Y. Novel six-gene prognostic signature based on colon adenocarcinoma immune-related genes. *BMC Bioinforma.* **23**, 418 (2022).
64. Lvu, W., Fei, X., Chen, C. & Zhang, B. In silico identification of the prognostic biomarkers and therapeutic targets associated with cancer stem cell characteristics of glioma. *Biosci. Rep.* **40**, BSR20201037 (2020).
65. Zhang, G. et al. A2M is a potential core gene in intrahepatic cholangiocarcinoma. *BMC Cancer* **22**, 5 (2022).
66. Okroj, M. et al. Local expression of complement factor I in breast cancer cells correlates with poor survival and recurrence. *Cancer Immunol. Immunother.* **64**, 467–478 (2015).
67. Riihilä, P. et al. Complement factor I promotes progression of cutaneous squamous cell carcinoma. *J. Invest. Dermatol.* **135**, 579–588 (2015).
68. Cai, X. et al. A candidate prognostic biomarker complement factor I promotes malignant progression in Glioma. *Front. Cell. Dev. Biol.* **8**, 615970 (2021).
69. Luca, B. A. et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184**, 5482–5496.e28 (2021).
70. Duan, S. et al. Novel prognostic biomarkers of gastric cancer based on gene expression microarray: COL12A1, GSTA3, FGA and FGG. *Mol. Med. Rep.* **18**, 3727–3736 (2018).
71. Peng, H. H. et al. Elevated serum FGG levels prognosticate and promote the disease progression in prostate cancer. *Front. Genet.* **12**, 651647 (2021).
72. Wang, Y., Wang, J., Gao, J., Ding, M. & Li, H. The expression of SERPINE1 in colon cancer and its regulatory network and prognostic value. *BMC Gastroenterol.* **23**, 33 (2023).
73. Wang, S., Pang, L., Liu, Z. & Meng, X. SERPINE1 associated with remodeling of the tumor microenvironment in colon cancer progression: a novel therapeutic target. *BMC Cancer* **21**, 767 (2021).

74. Chen, S. et al. SERPINE1 overexpression promotes malignant progression and poor prognosis of gastric cancer. *J. Oncol.* **2022**, 2647825 (2022).
75. Gao, Q. et al. CD142 plays an important role in the mobility of colorectal cancer cells. *Biosci. Biotechnol. Biochem.* **84**, 1856–1860 (2020).
76. Wen, X., Bai, S., Fang, Z. & Zhu, W. Integrated pan-cancer and scRNA-seq analyses identify a prognostic coagulation-related gene signature associated with tumor microenvironment in lower-grade glioma. *Discov. Oncol.* **15**, 256 (2024).
77. Ahmadi, S. E. et al. Tissue factor (coagulation factor III): a potential double-edge molecule to be targeted and re-targeted toward cancer. *Biomark. Res.* **11**, 60 (2023).
78. Yan, M. et al. Knockdown of PLAT enhances the anticancer effect of gefitinib in non-small cell lung cancer. *J. Thorac. Dis.* **12**, 712–723 (2020).
79. Shang, Z. et al. FGA isoform as an indicator of targeted therapy for EGFR mutated lung adenocarcinoma. *J. Mol. Med. (Berl.)* **97**, 1657–1668 (2019).
80. Qian, D. et al. Potentially functional genetic variants in the complement-related immunity gene-set are associated with non-small cell lung cancer survival. *Int. J. Cancer* **144**, 1867–1876 (2019).
81. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

Acknowledgements

The authors are thankful to the patients, coordinators, field and administrative workers, and technicians of the European Study into Digestive Illnesses and Genetics (PanGenEU) study. We also thank Guillermo Pita and Anna González-Neira, from CEGEN-CNIO, and Joe Dennis and Laura Fachal, from the University of Cambridge, for genotyping PanGenEU samples, performing variant calling and SNP imputation, and editing data. This research has been conducted using the UK Biobank Resource under Application Number 47884. RR gratefully acknowledge funding from NIH (R35 CA253126.) and SU2C Convergence 3.14. This research has been conducted using the UK Biobank Resource under Application Number 47884. The results here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

The work was partially supported by Fondo de Investigaciones Sanitarias (FIS), Instituto de Salud Carlos III, Spain (#PI061614, #PI11/01542, #PI0902102, #PI12/01635, #PI12/00815, #PI15/01573, #PI18/01347, #PI21/00495); Ministerio de Ciencia, Innovación y Universidades, Madrid, Spain (#RTI2018-101071-B-I00 and #PID2021-128125OB-I00); Red Temática de Investigación Cooperativa en Cáncer, Spain (#RD12/0036/0034, #RD12/0036/0050, #RD12/0036/0073); Fundación Científica de la AECC, Spain; European Cooperation in Science and Technology - COST Action #BM1204: EUPancreas. EU-6FP Integrated Project (#018771-MOLDIAG-PACA), EU-FP7-HEALTH (#259737-CANCERIALIA, #256974-EPC-TM-Net); Associazione Italiana Ricerca sul Cancro (#12182); Cancer Focus Northern Ireland and Department for Employment and Learning; and ALF (#SLL20130022), Sweden; Pancreatic Cancer Collective (PCC): Lustgarten Foundation & Stand-Up to Cancer, USA (SU2C #6179).

Author contributions

Study conception: NM, ELM. Design of the work: AL, RR, ELM, NM. Data acquisition: IF, RL, AC, RAG, MI, XM, JML, CWM, JP, MO, VMB, AT, AF,

LMB, TCJ, EDM, TMG, WG, LS, JB, EC, JK, BK, JM, DO, AS, WY, FXR, NM, and the rest of the PanGenEU Investigators. Data analysis: AL, LA, SS, AM, ELM. Interpretation of data: AL, RR, FXR, ELM, NM. Creation of new software used in the work: No one. Drafting the work or substantively revising it: AL, RR, FXR, ELM, NM. Approval of the submitted version (and any substantially modified version that involves the author's contribution to the study): ALL AUTHORS. Agreement on both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature: ALL AUTHORS.

Competing interests

R.R. is a founder of Genotwin and a member of the SAB of DiaTech and Flahy. None of these activities are related to the work described in this manuscript. The remaining authors declare that they have no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65811-y>.

Correspondence and requests for materials should be addressed to Evangelina López de Maturana or Núria Malats.

Peer review information *Nature Communications* thanks Yasuyuki Mizutani, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2026

Alberto Langtry¹, Raul Rabadan², Lola Alonso^{1,3}, Ioan Filip², Sergio Sabroso-Lasa¹, Ane Moreno-Oya^{1,3}, Rita Lawlor^{4,5}, Alfredo Carrato^{3,6,7,8}, Rafael Alvarez-Gallego^{9,10}, Mar Iglesias^{3,11}, Xavier Molero^{12,13,14}, Matthias J. Löhre¹⁵,

Christoph W. Michalski^{16,17}, José Perea^{18,19}, Michael O'Rorke^{20,21}, Victor M. Barberà²², Adonina Tardón^{23,24,25}, Antoni Farré²⁶, Luís Muñoz-Bellvis²⁷, Tatjana Crnogorac-Jurcevic²⁸, Enrique Domínguez-Muñoz²⁹, Thomas M. Gress³⁰, William Greenhalf³¹, Linda Sharp^{32,33}, Joaquim Balsells³⁴, Eithne Costello³¹, Jörg Kleeff^{16,35}, Bo Kong^{16,17}, Josefina Mora³⁶, Damian O'Driscoll³², Aldo Scarpa^{4,5}, Weimin Ye³⁷, Francisco X. Real^{3,38,39}, Evangelina López de Maturana^{1,3,40}✉, Núria Malats^{1,3,40}✉, On behalf of the PanGenEU Consortium Investigators

¹Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), Madrid, Spain. ²Department of Systems Biology, Columbia University, New York, NY, USA. ³CIBERONC, Madrid, Spain. ⁴ARC-Net Center for Applied Research on Cancer, Department of Diagnostics and Public Health, Verona, Italy. ⁵Department of Engineering for Innovation Medicine, University of Verona, Verona, Italy. ⁶Department of Oncology, Ramón y Cajal University Hospital, Madrid, Spain. ⁷Instituto Ramón y Cajal de investigación sanitaria (IRYCIS), Madrid, and CIBERONC, Madrid, Spain. ⁸Alcala University, Madrid, Spain. ⁹HM CIOCC Madrid (Centro Integral Oncológico Clara Campal). Hospital Universitario HM Sanchinarro. Instituto de Investigación Sanitaria HM Hospitales, Madrid, Spain. ¹⁰Facultad HM de Ciencias de la Salud de la Universidad Camilo José Cela, Madrid, Spain. ¹¹Department of Pathology, Hospital del Mar—Parc de Salut Mar, Barcelona, Spain. ¹²Department of Gastroenterology, Hospital Vall Hebron, Barcelona, Spain. ¹³Hospitals Universitaris Arnau de Vilanova i Santa Maria, IRBLleida, UdL, Lleida, Spain. ¹⁴CIBEREHD, Madrid, Spain. ¹⁵Gastrocentrum, Karolinska Institutet and University Hospital, Stockholm, Sweden. ¹⁶Department of Surgery, Technical University of Munich, Munich, Germany. ¹⁷Department of General, Visceral and Transplantation Surgery, Medical Faculty Heidelberg, University Heidelberg, Heidelberg, Germany. ¹⁸Department of Surgery, Hospital 12 de Octubre, Madrid, Spain. ¹⁹Institute of Biomedical Research of Salamanca, Salamanca, Spain. ²⁰Centre for Public Health, Belfast, Queen's University Belfast, Belfast, UK. ²¹College of Public Health, The University of Iowa, Iowa City, IA, US. ²²Molecular Genetics Laboratory, General University Hospital of Elche, Elche, Spain. ²³Department of Medicine, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), Oviedo, Spain. ²⁴Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain. ²⁵CIBERESP, Madrid, Spain. ²⁶Department of Gastroenterology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. ²⁷Department of Surgery, Hospital Universitario de Salamanca – IBSAL, Universidad de Salamanca, Salamanca; and CIBERONC, Madrid, Spain. ²⁸Barts Cancer Institute, Centre for Cancer Biomarkers and Biotherapeutics, Queen Mary University of London, London, UK. ²⁹Department of Gastroenterology, University Clinical Hospital of Santiago de Compostela, Santiago de Compostela, Spain. ³⁰Department of Gastroenterology, University Hospital of Giessen and Marburg, Marburg, Germany. ³¹Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK. ³²National Cancer Registry Ireland and HRB Clinical Research Facility, University College Cork, Cork, Ireland. ³³Newcastle University, Institute of Health & Society, Newcastle, UK. ³⁴Department of Surgery, Hospital Vall Hebron, Barcelona, Spain. ³⁵Department of Visceral, Vascular and Endocrine Surgery, Martin-Luther-University Wittenber-Halle (Saale), Halle (Saale), Germany. ³⁶Department of Clinical Biochemistry, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. ³⁷Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ³⁸Epithelial Carcinogenesis Group, Spanish National Cancer Research Center (CNIO), Madrid, Spain. ³⁹Pompeu Fabra University, Barcelona, Spain. ⁴⁰These authors contributed equally: Evangelina López de Maturana, Núria Malats.

✉ e-mail: melopezdm@cnio.es; nmalats@cnio.es

the PanGenEU Consortium Investigators

Alberto Langtry¹, Lola Alonso^{1,3}, Sergio Sabroso-Lasa¹, Ane Moreno-Oya^{1,3}, Paulina Gómez-Rubio¹, Esther Molina-Montes¹, Mirari Márquez¹, Roger Milne¹, Ana Alfaro¹, Tania Lobato¹, Lidia Estudillo¹, Francisco X. Real^{3,38,39}, Núria Malats^{1,3}, Aldo Scarpa^{4,5}, Rita Lawlor^{4,5}, Stefania Beghelli^{4,5}, Rafael Alvarez-Gallego^{9,10}, Antonio Cubillo^{9,10}, Manuel Hidalgo^{9,10}, Paloma Peinado^{9,10}, Isabel Salas^{9,10}, Jesús Rodríguez Pascual^{9,10}, Alfredo Carrato^{3,6,7,8}, Julie Earl^{6,7}, Alejandra Caminoa^{6,7}, Carmen Guillén-Ponce^{6,7}, Mercedes Rodríguez-Garrote^{6,7}, Federico Longo-Muñoz^{6,7}, Reyes Ferreiro^{6,7}, Vanessa Pachón^{6,7}, M. Ángeles Vaz^{6,7}, María Muñoz^{6,7}, Mar Iglesias^{3,11}, Lucas Ilzarbe¹¹, Cristina Álvarez-Urturi¹¹, Xavier Bessa¹¹, Felipe Bory¹¹, Lucía Márquez¹¹, Ignasi Poves¹¹, Fernando Burdío¹¹, Luis Grande¹¹, Javier Gimeno¹¹, Xavier Molero^{12,13,14}, Luisa Guarner¹², Joaquín Balcells¹², Mayte Salcedo¹², Xavier Merino¹², Christoph W. Michalski^{16,17}, Irene Esposito¹⁶, Jörg Kleeff^{16,35}, Bo Kong^{16,17}, Carmen Mota¹⁶, Matthias J. Lohr¹⁵, Jiaqui Huang¹⁵, Caroline Verbeke¹⁵, Weimin Ye³⁷, Jingru Yu³⁷, Carmelo Loinaz¹⁸, José Perea^{18,19}, José Luis Rodríguez-Peralto¹⁸, Ana Teijo¹⁸, Pablo Peláez¹⁸, Antoni Farré²⁶, Josefina Mora²⁶, Marta Martín²⁶, Vicenç Artigas²⁶, Carlos Guarner²⁶, Francesc J. Sancho²⁶, Mar Concepción²⁶, Teresa Ramón y Cajal²⁶, Justyna Szafranska²⁶, Michael O'Rorke^{20,21}, Liam Murray²⁰, Marie Cantwell²⁰, Victor M. Barberà²², Javier Gallego²², Adonina Tardón^{23,24,25}, Luis Barneo²³, MM Rodríguez-Suarez²³, Enrique Domínguez-Muñoz²⁹, Yolanda Pazos²⁹, Antonio Lozano²⁹, Maria Luaces²⁹, Francisco Blanco-Antona²⁷, J. M. Sayagués Manzano²⁷, M. L. Gutiérrez Troncoso²⁷, A. Orfao de Matos²⁷, Luís Muñoz-Bellvis²⁷, Thomas M. Gress³⁰, Malte Buchholz³⁰, Albrecht Neesse³⁰, William Greenhalf³¹, Eithne Costello³¹, Tatjana Crnogorac-Jurcevic²⁸, Hemant M. Kocher²⁸, Satyajit Bhattacharya²⁸, Ajit T. Abraham²⁸, Darren Ennis²⁸, Thomas Dowe²⁸, Tomasz Radon²⁸, Damian O'Driscoll³² & Linda Sharp³²