medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Title: *HLA-A*02:01* allele is associated with decreased risk and a longer survival in pancreatic cancer: Results from an exhaustive analysis of the *HLA* variation in PDAC

Authors: Alberto Langtry[1], Raul Rabadan[2], Lola Alonso[1], Casper van Eijck[1,3], Teresa Macarulla[4], Rita T Lawlor[5], Alfredo Carrato[6], Rafael Alvarez-Gallego[7], Mar Iglesias[8], Xavier Molero[9], J Matthias Löhr[10], Christopher W Michalski[11], José Perea[12], Michael O'Rorke[13], Víctor M Barberà[14], Adonina Tardón[15], Antoni Farré[16], Luís Muñoz-Bellvís[17], Tatjana Crnogorac-Jurcevic[18], Enrique Domínguez-Muñoz[19], Thomas Gress[20], William Greenhalf[21], Linda Sharp[22], Sergio Sabroso-Lasa[1], Ioan Filip[2], Gaby Strijk[3], Florian Castet[4], Joaquim Balsells[9], Eithne Costello[21], Jörg Kleeff[23], Bo Kong[24], Josefina Mora[16], Damian O'Driscoll[25], Aldo Scarpa[5], Weimin Ye[26], Francisco X. Real[27], Núria Malats[1*], Evangelina López de Maturana[1*], on behalf of the PanGenEU Investigators[28]

Authors' affiliations:

(1) Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), and CIBERONC, Madrid, Spain.
(2) Department of Systems Biology, Columbia University, New York, NY, USA.
(3) Department of Surgery, Erasmus Medical Center, Rotterdam, The Netherlands.
(4) Vall d'Hebron Institute of Oncology (VHIO), Hospital Universitari Vall d'Hebron, Barcelona, Barcelona, Spain.
(5) ARC-Net Center for Applied Research on Cancer, Department of Diagnostics and Public Health and Department of Engineering for Innovation Medicine , University of Verona, Verona, Italy.
(6) Department of Oncology, Ramón y Cajal University Hospital, IRYCIS, Alcala University, Madrid, and CIBERONC, Spain.
(7) HM CIOCC Madrid (Centro Integral Oncológico Clara Campal). Hospital Universitario HM Sanchinarro. Instituto de Investigación Sanitaria HM Hospitales. Facultad HM de Ciencias de la Salud de la Universidad Camilo José Cela.
(8) Hospital del Mar—Parc de Salut Mar, Barcelona, and CIBERONC, Spain.
(9) Hospitals Universitaris Arnau de Vilanova i Santa Maria, IRBLleida, UdL, 25198, Lleida; and CIBEREHD, Madrid, Spain.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(10) Gastrocentrum, Karolinska Institutet and University Hospital, Stockholm, Sweden.

(11) Department of Surgery, Technical University of Munich, Munich; and Department of General, Visceral and Transplantation Surgery, Medical Faculty Heidelberg, University Heidelberg, Germany.

(12) Department of Surgery, Hospital 12 de Octubre, Madrid; and Institute of Biomedical Research of Salamanca, Salamanca, Spain.

(13) Center for Public Health, Belfast, Queen's University Belfast, UK; and College of Public Health, The University of Iowa, Iowa City, IA, US.

(14) Molecular Genetics Laboratory, General University Hospital of Elche, Spain.

(15) Department of Medicine, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, and CIBERESP, Spain.

(16) Department of Gastroenterology and Clinical Biochemistry, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.

(17) Department of Surgery, Hospital Universitario de Salamanca – IBSAL. Universidad de Salamanca and CIBERONC, Spain.

(18) Barts Cancer Institute, Center for Cancer Biomarkers and Biotherapeutics, Queen Mary University of London, London, UK.

(19) Department of Gastroenterology, University Clinical Hospital of Santiago de Compostela, Spain.

(20) Department of Gastroenterology, University Hospital of Giessen and Marburg, Marburg, Germany.

(21) Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK.

(22) National Cancer Registry Ireland and HRB Clinical Research Facility, University College Cork, Cork, Ireland; and Newcastle University, Institute of Health & Society, Newcastle, UK.

(23) Department of Surgery, Technical University of Munich, Munich; and Department of Visceral, Vascular and Endocrine Surgery, Martin-Luther-University Wittenberg-Halle (Saale), Germany.

(24) Department of Surgery, Technical University of Munich, Munich

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(25) National Cancer Registry Ireland and HRB Clinical Research Facility, University College Cork, Cork, Ireland.

(26) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

(27) Epithelial Carcinogenesis Group, Spanish National Cancer Research Center (CNIO) and CIBERONC, Madrid; and Pompeu Fabra University, Barcelona, Spain.

(28) In Annex 1.

*Co-senior authors

Correspondence to: Núria Malats, and Evangelina López de Maturana Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center (CNIO), C/Melchor Fernández Almagro 3, 28029-Madrid, Spain, Email: nmalats@cnio.es, and melopezdm@cnio.es. Phone: +34 917328000

ABSTRACT

Genetic susceptibility loci are associated with PDAC risk and survival, but the impact of germline HLA region variation remains largely unexplored. This study examined *HLA* I-II alleles within the PanGenEU study and validated our findings using external datasets (UK Biobank, TCGA, PAN-NGS trial, and Caris trial). *HLA-A*02:01and HLA-B*49* alleles were linked to a decreased risk of PDAC, whereas *HLA-B*39*, *HLA-DPB1*04*, and *HLA-A*26:01* were directly associated with increased risk. PDAC patients carrying the *HLA-A*02:01* allele also showed lower mortality rates, with the effect being more pronounced in those with $KRAS^{G12V}$ mutations, pointing to a host*tumor genetic interaction. This research highlights *HLA-A*02:01*, found in 20% of Europeans, as a marker for reduced PDAC risk and mortality, especially in $KRAS^{G12V}$ mutated tumors. Results from this study could enhance personalized medicine for PDAC by identifying patients who may benefit from regular screenings

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

through tailored risk assessments. Importantly, our findings are crucial for stratifying PDAC patients based on their genetic background and tumor mutational profile, which can guide treatment strategies.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Pancreatic cancer has a dismal prognosis with a 5-year survival rate of 5-10% (1–3) and a 10-year survival rate of 1% (4). Worryingly, it is expected to become the second cause of cancer-related deaths in the USA by 2030 if action is not taken (5). Pancreatic ductal adenocarcinoma (PDAC) is the most common form of these neoplasms. So far, great efforts are being made to advance its diagnosis and detect tumors at early stages to improve survival (6–8). Identifying the genetic and non-genetic factors involved in PDAC etiology is crucial for defining high-risk individuals who would benefit from detecting the tumor when it is still curable. However, there is no definition of high-risk individuals for sporadic PDAC, those representing 90% of the burden of the disease.

We previously reported an inverse association between PDAC risk and nasal allergies, asthma, and autoimmune diseases (9,10), conditions characterized by an overactive immune response. Furthermore, our recent multilayered GWAS approach (11) identified a hotspot in the Major Histocompatibility Complex (*MHC*) region, where the alternative allele of the top targeted Human Leukocyte Antigen *(HLA) G* gene was associated with increased PDAC risk (p-value=1.8×10^{-4}). In addition, several GWAS hits at this region have been associated with the risk of other cancer types like breast (12) or lung cancer (13,14), as well as with well-established PDAC risk and protective factors such as chronic immune-related diseases (15,16), diabetes (17,18), smoking (19), asthma (20,21), and nasal allergies (22). The lack of GWAS-reported PDAC genetic susceptibility loci in the *MHC* region might be because SNPs are not fully capturing the variability of the *MHC*. On the other hand, at the tumor level, Filip et al. (2023) (23) reported that Allele-Specific Expression loss of the *HLA* class I genes in PDAC, among other cancers, is associated with a worse prognosis.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The *MHC* is essential for the proper response of the immune system against tumor cells (24). The *MHC* genetic complex is a dense cluster of genes located at the 6p21.1-21.3 chromosome cytoband (28,477,797-33,448,354 bp, GRCh37), and it includes three *HLA* sub-regions. *HLA* class I can be classified into classical (*HLA*-A, -B, and -C) and non-classical genes (*HLA-E*, *-F*, and *-G*). *HLA* class II genes are *HLA-DR*, *-DQ*, and *-DP*. HLA class I and II proteins present intracellular and extracellular peptides to $CD8^+$ and $CD4^+$ T-cells, respectively (25,26). The *MHC* region also shows an extensive linkage disequilibrium (27). As a consequence of its complexity, the role of this region has not yet been studied in-depth in relation to the risk of cancer and, more specifically, PDAC.

Here, we report an exhaustive study on PDAC genetic susceptibility associated with the *MHC* region variation on *HLA* class I and II alleles, haplotypes, and amino acids (**Fig. 1**). We performed association analyses at the *HLA* allele and haplotype levels in PanGenEU, one of the largest PDAC case-control studies with extensive standardized clinical and epidemiological data annotation. We then tested these associations in an independent population, the UK Biobank (UKB), and computed global risk estimates by meta-analyzing the results from the two consortia. We also deliver for the first time an estimate of the heritability for PDAC risk explained by the *HLA* class I and II variation. Furthermore, we used the resources from TCGA, UKB, and PanGenEU to conduct a meta-analysis on the HLA alleles' overall survival (OS) associations. Finally, we assessed the association between the most promising alleles and common *KRAS* mutations using TCGA data and validated results with PAN-NGS and Caris trials. Altogether, our findings underscore the role of the *HLA* region in PDAC genetic susceptibility and prognosis. Our findings may have implications for the design of mutant KRAS-based vaccines in PDAC.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

RESULTS

Characteristics of the study populations

The characteristics of the PanGenEU, UK Biobank, and TCGA study populations are available in **Supplementary Table S1**. PanGenEU participants were Caucasian from Spain (58.8%), Italy (22.6%), Sweden (7.9%), Germany (6.7%), UK (3.7%), and Ireland (0.2%). The average age was similar between cases and controls in each study population; however, UKB controls were, on average, younger than cases. A slightly lower proportion of females across the studies was observed for both cases and controls. In the PanGenEU and UKB populations, cases were more frequently smokers and diabetic, while controls had a higher prevalence of asthma and allergies. Finally, the median survival was shorter for UKB patients than for TCGA and PanGenEU patients. As expected based on study design, TCGA included a higher proportion of patients with early-stage tumors than PanGenEU cases. We also observed significant differences in the stage of the tumors across populations, being the Caris clinical trial the population with the highest proportion of stage III-IV PDAC tumors. The median survival time of the UKB patients was significantly shorter compared to the PanGenEU, TCGA, Erasmus, and Caris patients. Finally, neoadjuvant and adjuvant settings were significantly different between the Erasmus and Caris populations.

Association analyses between HLA alleles and PDAC risk

HLA class I and class II alleles explain 11% of the PDAC risk. Bayesian multi kernel-based regression allowed us to obtain the genetic and residual variance components (σ^2_{HLA-I}= 0.013, σ^2_{HLA-II}= 0.010, and σ^2_e= 0.174) and therefore, the heritability estimates for *HLA* class I and class II alleles were: h^2_{HLA-I}= 0.06 and h^2_{HLA-II}= 0.05.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Supplementary Fig. S1A and **S1B** show the landscape of *HLA* class I and II alleles at two-field level in PanGenEU and UKB populations, respectively. *HLA-B* was the most polymorphic *MHC* gene in both populations: 97/298 (32.6%), two-field *HLA* alleles in the PanGenEU and 108/306 (35.3%) in UKB. Some alleles, like *HLA-DPA1*01:03*, showed large variability across populations. In general, the AF matched expected European frequencies (28). The AF in PanGenEU showed good correlations with the T1DGC reference panel (R=0.92 and R=0.86, for one- (allele group) and two-fields (HLA protein) *HLA* alleles, see, **Supplementary Fig. S2A and S2B**) and moderate correlation with UKB (R=0.57, see **Supplementary Fig. S3**). **Supplementary Table S2** summarizes the number of *HLA* class I and II alleles considered in the association analyses after filtering out those with poor imputation quality and low AF.

In our comprehensive meta-analysis of the PanGenEU and UKB populations, the following four alleles yielded significant associations with PDAC risk: *HLA-A**02:01 (OR=0.86, p-value=1.88×10^{-3}), *HLA-B**49 (OR=0.68, p-value=3.81×10^{-2}), *HLA-B**39 (OR=1.42, p-value=1.99x10-^{2}), and *HLA-DPB1**04 (OR=1.1, p-value=2.86×10^{-2}) (**Fig. 2)**. After multiple testing correction, only the *HLA-A**02:01 remained significant (OR=0.86, corrected p-value=2.25×10^{-2}, **Supplementary Table S3**). We also explored the association between the above-mentioned *HLA* alleles and the established PDAC risk factors. Interestingly, we observed a significant inverse association between *HLA-B**39 and nasal allergies, which is protective against PDAC (OR=0.84, p-value=1.15×10^{-4}, see **Fig. 2C**).

Furthermore, the PanGenEU study identified one allele significantly associated with PDAC: the *HLA* class I *A*26:01* (OR=1.75, corrected p-value=4.17×10^{-2}). Individuals carrying this allele, present in 5% of the PanGenEU study population,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

showed a higher risk of developing PDAC (OR=1.74, **Supplementary Table S4**). Additionally, when both controls and cases were combined to increase statistical power, we found that this allele was significantly and inversely associated with nasal allergies (OR=0.56, p-value=2.87×10^{-2}) and, although non-significantly, with asthma (OR=0.68, p-value=2.2×10^{-1}, **Supplementary Fig. S4A**). However, we did not replicate the association between *HLA-A**26:01 and PDAC risk in the UKB study (**Supplementary Table S5**) nor the meta-analysis, although the direction of the association was the same (OR=1.07, p-value=9.88×10^{-1}).

Cumulative PDAC risk score for the most promising alleles

We estimated cumulative PDAC risk by combining known risk and protective factors and adding PDAC-associated *HLA* alleles. In a meta-analysis with the UK Biobank, individuals with the abovementioned non-genetic high-risk conditions had 2.8 times higher PDAC risk. By including either *HLA-B**39 or *HLA-DPB1**04 alleles, the ORs increased to 5.28 and 3.34, respectively (**Table 1**), though not statistically significant (p-value=1.17×10^{-1}, p-value=1.33×10^{-1}, respectively). Conversely, asthmatic or allergic individuals who did not smoke, had no diabetes, nor had a family history of PDAC, experienced a 62% reduction in the odds (OR=0.38) of developing PDAC. Subjects who had one/two alleles of *HLA-A**02:01 further decreased their ORs to 0.32 and 0.27, respectively (p-value=1.18×10^{-2}). By including one/two copies of the *HLA-B*49* allele in the model, the OR significantly decreased to 0.19 and 0.1 (p-value=2.4×10^{-2}, **Table 1**). Additionally, in the PanGenEU study, diabetics, smokers, and individuals with a family history of cancer but without asthma or allergies had 4.33 times higher risk of PDAC than the population in the lowest risk category. When those individuals at high-risk harbor one or two copies of the *HLA-A**26:01 allele, the

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

ORs increased to 7.73 and 13.81, respectively (p-value=2.28×10^{-2}, see **Supplementary Fig. S4B**).

Association analyses between HLA haplotypes and PDAC risk

Given the critical role of *HLA* haplotypes in immune responses, we assessed their association with PDAC risk. We identified 18 haplotypes significantly associated with PDAC risk, nine of which remained significant after multiple testing correction (**Supplementary Table S6**). Two of the top five protective haplotypes carried the *HLA-A*02* allele, present in 20% of the European population and associated with decreased PDAC risk in our analyses. Haplotypes harboring the *HLA-A*02* allele were all protective for PDAC, some reaching an OR~0.4). Among the haplotypes with the lowest effect size *HLA-B*35* and *HLA-A*02* alleles were common: *HLA-B*35*/*HLA-A*02* (OR=0.41, 95%CI=0.18-0.94), *HLA-DRB1*13*/*HLA-B*35* (OR=0.43, 95%CI=0.38-0.5), and *HLA-DRB1*01*/*HLA-A*02* (OR=0.49, 95%CI=0.27-0.89). On the contrary, the highest risk haplotypes harbored the *HLA-DRB1*04* allele: *HLA-DRB1*04*/*HLA-B*44*/*HLA-C*07* (OR=1.19, 95%CI=1.12-1.26) and *HLA-DRB*04*/*HLA-B*14* (OR=1.28, 95%CI=1.19-1.38) (**Supplementary Table S6**).

The *HLA-A*26*, *HLA-B*39*, and *HLA-B*49* alleles were not part of the haplotypes significantly associated with PDAC risk in the meta-analysis, probably because of their low AF: *HLA-A*26*=0.02, *HLA-B*39*=0.02, and *HLA-B*49*=0.01.

HLA alleles are associated with overall survival of PDAC patients

We further assessed the role of *HLA* alleles in PDAC prognosis through a meta-analysis of the results from the survival analyses in PanGenEU, UK Biobank, and TCGA studies. 156 *HLA* class I and II alleles overlapped between the three studies

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(**Supplementary Fig. S5**). The results obtained under a dominant MoI revealed that patients who carried the *HLA-A*02:01* allele presented better survival than non-carriers (HR=0.87, p-value=$3.3x10^{-2}$, **Fig. 3**), indicating that a single copy of this allele is enough to improve survival. On the other hand, homozygous individuals for *HLA-DPB1*04* displayed a borderline significantly worse survival under a codominance model (HR=1.19, p-value=$6.3x10^{-2}$, **Supplementary Fig. S6**). Additionally, carriers of *HLA-A*30:01* (HR=1.75, p-value=$1.2x10^{-2}$), and *HLA-DPB1*09:01* (HR=3.36, p-value=$1.5x10^{-3}$) were also associated with worse overall survival compared to non-carriers (**Supplementary Fig. S7**).

*HLA-A*02:01 interacts with KRAS^{G12V} mutations for PDAC survival*

More than 90% of PDAC harbor *KRAS* mutations. We found that the PDAC protective allele *HLA-A**02:01 was associated with tumors harboring *KRAS*^{G12V} vs. other mutations in PDAC, NSCLC, and colorectal cancers (OR=1.55, p-value=$4.65x10^{-2}$), and with *KRAS*^{G12D} mutations in gastric tumors (OR=7.27, p-value=$6.42x10^{-3}$) (**Supplementary Table S7**).

Stratified analysis revealed that *HLA-A*02:01* carriers with *KRAS*^{G12V} tumors had significantly better survival (HR=0.19, p-value=$1.3x10^{-2}$, **Fig. 4A**) than *HLA-A**02:01 carriers with other *KRAS* mutations; importantly, survival did not differ for patients with *KRAS*^{G12V} mutated tumors who did not carry the *HLA-A*02:01* allele (HR=0.87, p-value=$8.1x10^{-1}$, **Fig. 4B**). Conversely, carriers of *HLA-A*02:01* and *KRAS*^{G12D} had significantly worse survival (HR=5.51, p-value=$1.5x10^{-3}$, **Fig. 4C**) than non-carriers (HR=1.13, p-value=$6.2x10^{-1}$, **Fig. 4D**). Further stratified analyses focusing on individuals carrying either the *KRAS*^{G12V} or *KRAS*^{G12D} mutations, along with the presence or absence of the *HLA-A*02:01* allele, showed that the longer

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

survival of *KRAS*^{G12V} carriers appear to be driven by the interaction with the *HLA-A*02:01* allele (HR=0.13, p-value=4.4x10^{-2}, **Fig. 5A**). However, for those with *KRAS*^{G12D} mutated tumors, the decreased survival was mainly due to the mutation itself, regardless of the *HLA* allele (HR=1.21, p-value=2.2x10^{-1}, **Fig. 5B**). Interestingly, we discovered that *KRAS*^{G12D} mutated tumors, displayed a significant higher expression level of IFN-γ in comparison to tumors carrying any of the other *KRAS* mutations (**Supplementary Fig. S8**).

The interaction model in the TCGA study showed that carriers of the *HLA-A*02:01* allele who had a *KRAS*^{G12V} mutated tumor had a significantly lower risk of death (HR=0.14, p-value=2.4x10^{-2}) compared to patients with any of the other *KRAS* mutations, adjusted for gender, age at diagnosis, stage, and Moffit tumor subtype. We then performed a meta-analysis of the interaction between the *A*02:01* allele and *KRAS*^{G12V} mutations, including the PAN-NGS and Caris populations (**Supplementary Fig. S9**). This interaction held significant in early-stage tumors (HR=0.25, p-value=1.7x10^{-2}, **Fig. 6A**) but not in advanced-stage (HR=0.93, p-value=8.7x10^{-1}, **Fig. 6B**).

DISCUSSION

The present study represents the most exhaustive analysis of the role of the *MHC* region at the *HLA* allele and haplotype levels associated with PDAC risk and prognosis. We show that different *MHC* genetic profiles recognize distinct PDAC risk populations. We identified five novel susceptibility alleles in the *HLA* class I and II regions associated with PDAC risk: *HLA-A**26:01, *HLA-A**02:01, *HLA-B**49, *HLA-B**39, and *HLA-DPB1**04. Among these, *HLA-A*02:01* remained significant after correcting for multiple testing, showing a notably reduced risk of PDAC

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(metaOR=0.86, corrected p-value=2.25×10^{-2}. This allele is prevalent in 20% of the European population, and was present in all the protective haplotypes. Interestingly, HLA-A*02:01 has been found to be associated also with a decreased risk of melanoma (29). Importantly, we found that germline *HLA-A*02:01* carriers with PDAC from the PanGenEU, UKB, TCGA, and Caris had a survival benefit (metaHR=0.87, p-value=3.3×10^{-2}; **Fig. 3B**). These findings go along with the observations that *HLA-A*02:01* carriers have better survival in NSCLC (30), and might have improved outcomes to immunotherapy (31), pointing to *HLA-A*02:01* as a good prognostic marker for PDAC. Notably, the *HLA-A*02:01* allele is among the eligibility criteria in different cancer immunotherapy clinical trials for esophageal carcinoma (32), acute myeloid leukemia (33), NSCLC (34), and glioblastoma (currently in phase III) (35).

Since *HLA-A*02:01* has shown a high predicted binding affinity for the *KRAS* codon 12 mutation in silico-in vitro studies (36), we also assessed the association of *HLA-A*02:01* with the most common hotspot *KRAS* mutations (at codons 12, 13, and 61) in *KRAS*-mutated cancers (PDAC, NSCLC, stomach, and colorectal) using TCGA data. After adjusting for cancer type, we discovered that *HLA-A*02:01* is preferentially associated with tumors harboring $KRAS^{G12V}$ mutations (OR=1.55, p-value=4.65×10^{-2}). Remarkably, we observed that patients with $KRAS^{G12V}$ mutated tumors who carried the *HLA-A*02:01*, comprising approximately 5% of the PDAC patients, experienced significantly better survival compared to those without the allele. This novel interaction was confirmed in an independent series of resectable (or early-stage) PDAC cases (PAN-NGS). Supporting this host-tumor interaction, the NEPdb (https://nep.whu.edu.cn/) database, which includes data on over 17K validated human immunogenic neoantigens, shows that only $KRAS^{G12V}$ and $KRAS^{G12D}$ mutations have

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

been validated to bind the *HLA-A*02:01* protein and elicit an immune response. Conversely, regardless of the HLA allele, patients with *KRAS*^{G12D}-mutant tumors had significantly worse survival than those without this mutation. However, the PAN-NGS and Caris independent datasets could not validate this finding.

Therefore, this study is a pioneer in showing that the presence of *HLA-A*02:01* allele not only is associated with protection against PDAC risk but it is also associated with better PDAC survival, mainly in patients with early-stage *KRAS*^{G12V} mutant tumours. Our findings could lead to an *HLA-A*02:01-KRAS*^{G12V} classification system for nearly 2% of PDAC patients with better survival. Also, we anticipate that these patients may benefit from immunotherapy based on the results observed in xenograft models, with the *HLA-A*02:01* showing a better response to immunotherapy due to its ability to bind *KRAS*^{G12V} neoantigens (31). In contrast, approximately 40% of PDAC patients with *KRAS*^{G12D} mutated tumors demonstrated the poorest survival outcomes in TCGA, independent of their *HLA* allele. This significantly narrows the therapeutic options for these patients, potentially restricting them to systemic chemotherapy alone.

Given that *KRAS* mutations are neoantigens capable of binding the *HLA-A*02:01* protein and initiating an immune response, we hypothesize that *HLA-A*02:01-KRAS*^{G12V} individuals might have an enhanced immune response characterized by increased infiltration of anti-tumor immune cells, particularly CD8^{+} T cells, compared to those with the allele and another *KRAS* mutation. Conversely, patients with *KRAS*^{G12D} mutated tumors might either produce less immunogenic neoantigens or exhibit a higher infiltration of cells able to promote tumor tolerance. This results in a reduced ability of cytotoxic T cells to recognize and target the tumor cells. Interestingly, we discovered that *KRAS*^{G12D} PDAC tumors exhibited higher expression levels of IFN-γ in comparison to other *KRAS* mutated tumors. The

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

differences in survival that we found in *KRAS*^{G12D} PDAC tumors might be due to elevated IFN-γ levels upregulate PD-L1 expression, which binds to PD-1 receptors on T cells, inhibiting their activity and the recruitment of regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs) to the tumor microenvironment (37). In addition, high IFN PDAC tumors have been associated with worse overall survival, due to a higher activation of stroma contributing to tumor growth (38). Altogether might explain the differences in survival that we found in *KRAS*^{G12D} PDAC tumors. Consistently, a recent study published by Till et al., 2024, demonstrated that the clearance of ct*KRAS*^{G12D} is associated with improved overall outcomes in metastatic PDAC. This might be due to the immune system no longer recognizing the G12D neoantigen, leading to reduced IFN-γ levels (39).

Beyond A*02:01, we identified three additional alleles associated with PDAC risk in our meta-analysis. Among them, *HLA-DPB1*04* was associated with an increased risk of PDAC, with haplotypes containing this allele also showing an increased risk. Limited information is available on the role of this allele in cancer (40). Previous studies have shown that the risk allele *HLA-B*39* is associated with Grave's disease (41) while the protective allele *HLA-B*49* has been inversely associated with COVID-19 (42). Nevertheless, our study is the first to report an association between these alleles and cancer risk. Finally, *HLA-A*26:01* was associated with an increased risk of PDAC and inversely associated with nasal allergies and asthma in the PanGenEU population. These results point to a possible explanation of previous findings showing an inverse association between atopic comorbidities and PDAC risk (10). While the meta-analysis combining PanGenEU and UKB data threw a risk estimate (OR=1.07), it lacked statistical significance. This could be accounted for by the low AF of *HLA-A*26:01* in UKB population (0.02) vs the AF in PanGenEU

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(0.06), the latter AF being similar in most European countries with UK differing in the AF by >50%. *HLA-A*26:01* has been reported to be associated with an increased risk of non-Hodgkin lymphoma (OR=8.5, p-value=2.6×10^{-2}) (43). Importantly, including these alleles in a non-genetic-based cumulative risk score significantly altered the risk of developing PDAC, either increasing or decreasing it substantially.

Our study has some limitations. First, the analyses employed self-reported epidemiological information from subjects included in the studies that could lead to misclassification of the risk factors and comorbidities, such as asthma and rhinitis. However, our meta-analysis using the estimates of both PanGenEU and UKB populations allowed us to minimize this potential bias. In addition, we restricted the analyses to the Caucasian population; therefore, the impact of *HLA* alleles on PDAC risk and prognosis in other ethnic groups should be further assessed. Additionally, the *MHC* region is one of the most polymorphic areas of the human genome, making it challenging for association analyses to detect rare alleles that may play a crucial role in PDAC risk.

Despite the above-mentioned limitations, our work has several strengths. We performed an in-depth analysis of the relationship between the *HLA* region and PDAC risk by employing state-of-the-art methodology using important resources from two large and independent study populations with complementary designs: case-control and cohort. We corrected for population stratification by adjusting for the PCAs. Additionally, the association with PDAC prognosis was first conducted with PanGenEU study and further validated with data from three additional independent datasets: UKB, TCGA, PAN-NGS, and Caris.

Our findings provide a strong basis for *HLA-A*02:01*, a single biomarker to be considered in improving PDAC prevention and risk stratification, allowing for

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

identifying patients who may require more aggressive monitoring or treatment due to their lower genetic risk of mortality. These latter results warrant improving clinical trials based on *HLA* allele and *KRAS* mutation profiling. Future studies should also include PDAC subtypes in the association analysis and extend the scope to other *KRAS*-related cancer types. Since tumors can lose the expression of the *HLA* alleles, we suggest assessing whether this loss is happening in specific alleles with a protective role. Finnaly, given the crucial role of *HLA* alleles in immune responses, we suggest evaluating the immunotherapy response in PDAC patients according to their *HLA* allelic profile.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

MATERIAL AND METHODS

Ethics Statement

IRB ethical approval and written informed consent were obtained by all participating centers contributing to PanGenEU, and study participants, respectively. The study was conducted in accordance with the Helsinki Declaration.

Study populations

In this study, we employed the resources of the PanGenEU case-control study, the UKB cohort, and datasets from The Cancer Genome Atlas (TCGA), PAN-NGS trial from Erasmus Medical Center (EMC), Rotterdam, and Caris trial from Vall d'Hebron University Hospital (VHIO), Barcelona.

PanGenEU study details have been reported elsewhere (9–11,44). It is a multicentric case-control study on PDAC conducted in 28 centers across Spain, Italy, Sweden, Germany, the UK, and Ireland from 2009-2014 and 2016-2018. Newly diagnosed PDAC patients over 18 years of age were recruited. Eligible controls were hospital inpatients with primary diagnoses not associated with any of the well-known PDAC risk factors. Controls for Ireland and Sweden were population-based. Epidemiological information was collected through direct in-person interviews by trained health monitors using structured questionnaires, including information about lifestyle, environmental exposures, and medical history (9). The individuals were genotyped with the Infinium OncoArray-500K (45) and missing genotypes were imputed using IMPUTE2 v2 (46) and 1000Genomes (Phase 3, v1) as reference (47). Quality control filters included the missing call rate, unexpected heterozygosity, discordance between reported and genotyped gender, unexpected relatedness, and estimated European ancestry < 80%. Once those samples were filtered out, genotyped

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

data from 1,317 cases and 700 controls were employed for *HLA* allele imputation and association analyses.

UKB is a prospective cohort study of approximately 500,000 individuals from the United Kingdom (48). The UKB Axiom array was used for genotyping 450,000 subjects, and the remaining 50,000 individuals were genotyped with the UK BiLEVE Axiom array. Based on their primary International Classification of Diseases (ICD9 and ICD10) codes, we selected 935 PDAC patients (**Supplementary Table S8**). We discarded the self-reported cases (N=45) to ensure PDAC diagnosis reliability and those individuals with no Caucasian genetic ethnicity (N=129), resulting in a final set of 761 PDAC cases. After selecting Caucasian individuals without a history of cancer and with the same ICD9 and ICD10 codes as PanGenEU controls (10), the final dataset of controls included 95,050 subjects.

TCGA, established in 2006 by the National Cancer Institute and the National Human Genome Research Institute, is one of the largest databases of cancer genomics data (49). TCGA has genetic information for 33 cancer types, genotyped with the Affymetrix 6.0 SNP array. We collected information on outcome, *HLA* alleles typed with the Optitype tool (50), and *KRAS* mutations from PDAC, colorectal, stomach adenocarcinoma, and Non-Small Cell Lung Cancer (NSCLC) patients. We kept for the analysis Caucasian individuals and those patients with complete data, 101 PDAC, 185 colorectal, 196 stomach adenocarcinoma, and 620 NSCLC. Additionally, we employed data from Thorsson et al, 2018 to assess the immune landscape of the 101 TCGA PDAC tumors.

The PAN-NGS trial (CCMO ID NL75415.078.20) includes PDAC patients from twelve centers throughout the Netherlands who were prospectively recruited from April 2021 to June 2023. Local review committees approved the trial. Eligible

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

patients had cytologically or histologically confirmed PDAC, were aged 18-60 years at diagnosis, had an Eastern Cooperative Oncology Group performance score of 0-2, and had an estimated life expectancy of at least twelve weeks. Patients were excluded when they did not want to know if alterations might be associated with genetic predisposition of cancer or in case of locally advanced or local recurrence of PDAC without histological tissue readily available. Clinical data were collected at participating centers and supplemented as needed. Clinical information was obtained at the site of the participating centers, and missing information was requested if were also treated in other hospitals. A final dataset of 108 PDAC patients was employed for the subsequent analysis.

Caris clinical trial is a retrospective single-center dataset involving patients with PDAC treated at Vall d'Hebron University Hospital between 2021 and 2023. Eligible participants were over 18 years old, had histologically confirmed PDAC, and underwent molecular analysis with the Caris Life Sciences sequencing panel. Exclusion criteria were non-adenocarcinoma pancreatic tumors, poor-quality samples unsuitable for sequencing, inadequate follow-up, or lack of written informed consent. Clinical data were collected at diagnosis, and a multidisciplinary team assessed the clinical stage. The Ethics Committee of the Vall d'Hebron University Hospital (PR(AG)651/2023) approved the study. Complete data was obtained for 145 PDAC patients and included in the analysis.

HLA (or MHC) alleles assessment

HLA class I and II alleles were first imputed using genotyped data from the PanGenEU study. We took advantage of the imputed *HLA* alleles from UKB, TCGA, Erasmus MC, and Caris subjects.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

PanGenEU study: The imputation was done by *SNP2HLA* software (version v 1.0.3) (51) considering 5,790 SNPs within the *MHC* region with an info parameter ≥0.9. This info parameter ranges from 0 to 1 and assesses the quality/reliability for a particular variant imputed through IMPUTE2. We focused on SNPs overlapping with those from the Type 1 Diabetes Genetic Consortium (T1DGC) V1.0.3, the largest European reference panel (52). This dataset includes 5,225 individuals with genotyping data for 7,238 SNPs, 424 *HLA* alleles (126 1-field and 298 2-field alleles), and 1,276 HLA amino acids. It allowed an accurate imputation of classical *HLA* alleles and amino acids of class I (*HLA-A*, *-B*, and *-C*) and class II (*HLA-DPB1*, *-DPA1*, *-DQB1*, *-DQA1*, and *-DRB1*) genes. We excluded *HLA* alleles that were not in Hardy-Weinberg equilibrium (HWE) in controls (p-value<10^{-6}), calculated using the *HardyWeinberg* R package V1.6.3 (53), those with a bad imputation quality (info<0.3), as well as those with an allele frequency (AF) <5%, ending up with 40 alleles of 1-field, and 39 alleles of 2-fields.

UKB information on 362 *HLA* alleles across eleven loci (*HLA-A*, *-B*, *-C*, *-DPB1*, *-DPA1*, *-DQB1*, *-DQA1*, *-DRB1*, *-DRB2*, *-DRB3*, *-DRB4*, and *-DRB5*) was available and obtained through the HLA*IMP:02 program (54).

TCGA: *HLA* class I alleles from the colorectal, stomach, and NSCLC tumors were inferred from normal WES using polysolver (55). Then, xHLA (56) was employed to infer the *HLA* I and *HLA* II alleles (*HLA-A*, *-B*, *-C*, *-DPB1*, *-DQB1*, and *-DRB1*) from normal WES of the PDAC patients.

Erasmus MC: *HLA* genotype results were determined based on information from mRNA isolated from a formalin-fixed paraffin-embedded tumor sample using the Agilent SureSelectXT Low Input Library prep chemistry, optimized for FFPE

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

tissue, in conjunction with the SureSelect Human All Exon V7 bait panel (48.2 Mb) and the Illumina NovaSeq.

Caris clinical trial: HLA genotypes were obtained from Caris Life Sciences reports for all patients. Genomic DNA was isolated from microdissected, formalin-fixed paraffin-embedded tumor samples and analyzed using the Illumina NovaSeq 6000 sequencers. A hybrid pull-down panel of baits designed to enrich for more than 700 clinically relevant genes at high coverage and high read-depth was used, along with another panel designed to enrich for additional >20,000 genes at a lower depth. To assist with gene amplification/deletion measurements and other analyses, a 500Mb SNP backbone panel (Agilent Technologies) was included. Analytical validation for *HLA* genotyping showed >99% concordance with a validated comparator method.

HLA haplotype assessment

We inferred the haplotypes independently in each population using the *haplo.stats* R package (57) after phasing with BEAGLE (58) and considering the *HLA* gene combinations prevalent in the Net Allele Database (28). A final set of 3,129 haplotypes were inferred in the PanGenEU individuals using 262 1-field *HLA* alleles across all eight loci with an info parameter reflecting imputation quality >0.3. In the UKB study, a final set of 148 and 57 1-field *HLA* class I and *HLA* class II haplotypes, respectively, were identified using 141 1-field *HLA* alleles. No *HLA* haplotype assessment was done for the TCGA population.

Statistical analyses

We assessed the association between the *HLA* region and PDAC risk for the individual *HLA* alleles and haplotypes within each study population. Subsequently, we combined

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the summary statistics using a meta-analysis approach. Then, we studied the association of the *HLA* alleles with well-established PDAC risk factors. Additionally, we investigated the heritability explained by *HLA* class I and class II alleles (**Fig. 1**). We did not consider further the analysis at the amino-acid level because significant results at the individual level were not found (**Supplementary Table S9**). Finally, prognostic analyses integrated *HLA* alleles and *KRAS* mutational profiles.

HLA allele-based association analysis with PDAC risk. Association analyses were done using the *PLINK* V1.9 software (59). In the PanGenEU setting, we tested the association between the *HLA* alleles and PDAC risk at the individual level by estimating the Odds Ratios (OR) and 95% Confidence Intervals (CI) through logistic regression models including age, gender, the European region, and the first five principal components (PCAs) to correct for population stratification. We considered the three Modes of Inheritance (MoI) for 424 PanGenEU *HLA* class I and II alleles with AF >0.05. In UKB, we included the I- and II-fields *HLA* alleles with an AF >0.1% (60,61). We performed logistic regression models adjusting for the same covariates as in the PanGenEU models, plus the type of genotyping array and center of inclusion.

A posterior meta-analysis using a random-effects model was performed with the *rma.uni* function of the *metafor* R package (62), combining the risk estimates of *HLA* alleles for PDAC with a nominal p-value<0.1 from the association analysis in PanGenEU and UKB studies, followed by multiple testing correction.

HLA allele-based association analysis with established PDAC risk factors. We used logistic regression models to assess the association between *HLA* alleles significantly

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

associated with PDAC risk in the previous meta-analyses with established PDAC risk factors such as Family History of Pancreatic Cancer (FHPC), T2DM, obesity, and tobacco and alcohol consumption, as well as with PDAC protective factors (asthma and nasal allergies). We restricted these analyses to the controls from each study. Then, we performed a meta-analysis by applying a random-effects model. A cumulative PDAC risk score for the most promising alleles was computed by using the estimates obtained in the meta-analysis including PanGenEU and UKB.

Haplotype association analysis. We used the *haplo.stats* R package (version 1.7.9) to perform the statistical association analysis of the PanGenEU and UKB *HLA* haplotypes with PDAC risk. We employed a logistic regression model adjusted for the same covariates as in the previous analyses. We considered the three MoIs in each study population. Multiple testing correction was applied for each gene combination separately after discarding the low prevalence *HLA* haplotypes (i.e., frequencies either <5% or <0.1% in PanGenEU and UKB, respectively).

Estimation of the heritability explained by HLA class I and class II alleles

We applied a Bayesian multi-kernel-based regression (63) to estimate the genetic variance explained by *HLA* class I and *HLA* class II alleles using the resources of the PanGenEU study:

$$y = X\beta + Zh_I + Zh_II + \varepsilon$$

where the response variable y is a vector of case/control outcomes, $\boldsymbol{X}$ is an incidence or design matrix relating the adjusting covariates to each individual, $\boldsymbol{\beta}$ is the corresponding regression coefficient, $\boldsymbol{Z}$ is a design matrix allocating records (($\boldsymbol{y}$) to the vector of *HLA* class I/II associated risk scores (h_I and h_{II}), and $\boldsymbol{\varepsilon}$ is a vector of random

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

normal deviates with null mean and variance σ_ε^2 that includes model misspecification and environmental effects not considered in this model. The two genetic and residual variances were assumed to follow the distributions $h_I \sim N(0, HI\sigma_{HLA-I}^2)$, $N(0, H_{II}\sigma_{HLA-II}^2)$ and $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$, where H_I and H_{II} correspond to the *HLA* class I and *HLA* class II alleles-derived relationship matrices between individuals, σ_{HLA-I}^2 and σ_{HLA-II}^2 are the genetic variances explained by *HLA* class I and II alleles, **I** is an identity matrix, and σ_ε^2 corresponds to the residual variance. The prior distribution for the genetic and residual variances were inverted chi-square distributions.

H_I and H_{II} similarity matrices were constructed using a Gaussian kernel with the *G.matrix* function of *snpReady* R package (64). Estimates of the unknown parameters were obtained from their posterior distributions using a Gibbs sampling implemented in the *BGLR* R package. We ran a McMC chain of 500,000 iterations, and the first 100,000 were discarded as burn-in.

HLA allele-based survival analysis with PDAC risk. We assessed patient survival based on the presence or absence of the *HLA* alleles, with a strong prior on *HLA-A*02:01*. To this end, we used the survival data of 425 PDAC cases from PanGenEU, 506 from UKB, 101 from TCGA (98 when we account for the stage of the tumor), and 145 from the Caris clinical trial. First, we assessed the overlap of the *HLA* class I and II alleles between the three study populations. We computed Hazard Ratios (HR) and 95%CI using Cox proportional-hazards regression models for each population separately under (co-)dominance MoIs. Adjustment covariates in PanGenEU models were sex, age at diagnosis, center, tumor stage, and the first PCAs; for UK Biobank, we included sex, age at diagnosis, genotyping array type, recruitment center and the first five PCAs; for the TCGA and Caris models, we considered sex, age at diagnosis,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and tumor stage. Finally, we meta-analyzed the results from the four populations using the function *rma.uni* from the *metafor* R package.

*Association analysis between HLA-A*02:01 and KRAS common mutations.* Based on previous reports pointing to a joint effect of *HLA-A*02:01* and *KRAS* mutations (31,36), we further explored the interaction between these biomarkers using TCGA data. We first assessed the association between *HLA-A*02:01* and well-established *KRAS* missense mutations at codons 12 (G12A, G12C, G12D, G12F, G12R, G12S, G12V), 13 (G13C, G13D), and 61 (Q61H, Q61K, Q61L, Q61R) using logistic regression models in tumors with high proportion of *KRAS* mutations (PDAC, colorectal, stomach, and NSCLC) (**Supplementary Table S10**). As the reference category, we used the non-G12V mutations for PDAC and colorectal cancer and the non-G12C for NSCLC. We initially assessed the association within each specific cancer type individually. Subsequently, we pooled the data from all cancer types into a single model, adjusting for cancer type. Finally, we used a Cox proportional-hazards regression model to evaluate whether the interaction between *HLA-A*02:01* and the specific $KRAS^{G12V}$ and $KRAS^{G12D}$ mutations was associated with overall survival in PDAC. The same covariates mentioned in the previous paragraph and the Moffit subtype of the tumor were employed to adjust these models. We used a Wilcoxon test to evaluate whether the associations found in overall survival could be explained by infiltrated immune components. Differences were considered significant at p-value<0.05.

To validate the interaction between the HLA-A*02:01 and the *KRAS* mutations, we used data from two independent populations, the Erasmus MC study and the Caris clinical trial. We analyzed the interaction between the *HLA-A*02:01*

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

allele and the *KRAS*^{G12V} mutation in each population, adjusting for variables such as sex, age at diagnosis, tumor stage, and treatment for Caris, and the same variables plus center of inclusion for Erasmus MC. We then conducted a meta-analysis using a random-effects model. Finally, we examined the effect of the interaction within each population by stratifying patients based on tumor stage (stages I-II and stages III-IV) and conducted separate meta-analyses for these subgroups.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DECLARATIONS

Availability of data and materials: The datasets generated and/or analyzed during the current study are not publicly available due to the sensitive nature of genetic data but are available from the corresponding author upon reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: The work was partially supported by Pancreatic Cancer Collective (PCC): Lustgarten Foundation & Stand-Up to Cancer (SU2C #6179); 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); Red Temática de Investigación Cooperativa en Cáncer, Spain (#RD12/0036/0034, #RD12/0036/0050, #RD12/0036/0073); EU-6FP Integrated Project (#018771-MOLDIAG-PACA), EU-FP7-HEALTH (#259737-CANCERALIA, #256974-EPC-TM-Net); Associazione Italiana Ricerca sul Cancro (IG 26343); Fondazione Cariverona: Oncology Biobank Project "Antonio Schiavi" (prot. 203885/2017); Italian Ministry of Health through Fondazione Italiana Malattie Pancreas (FIMP_CUP J37G22000230001); Cancer Focus Northern Ireland and Department for Employment and Learning; and ALF (#SLL20130022), Sweden, Cancer Research UK (C7690/A26881) and Pancreatic Cancer UK.

Authors' contributions: Study conception: NM, ELM. Design of the work: AL, RR, ELM, NM. Data acquisition: RTL, AC, CvE, TM, MH, MI, XM, JML, CM, JP, MO, VMB, AT, AF, LMB, TCJ, EFM, TG, WG, LS, IF, GS, LLC, JB, EC, LI, JK, BK, JM, DO, AS, WY, JY, FXR, NM, and rest of the PanGenEU Investigators. Data analysis: AL, LA, SS, ELM. Interpretation of data: AL, RR, FXR, ELM, NM. Creation of new software used in the work: No one. Drafting the work or substantively

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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.

Acknowledgements The authors thank the patients, coordinators, field and administrative workers, and technicians of the European Study into Digestive Illnesses and Genetics (PanGenEU) and TCGA studies. This research has been conducted using the UK Biobank Resource under Application Number 47884. This research was performed using resources generated by Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supplied by NIDDK Central Repository (NIDDK-CR). This manuscript was not prepared under the auspices of the T1DGC study and does not necessarily reflect the opinions or views of the T1DGC study, study sponsors, NIDDK-CR, or NIH. The results here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Connor AA, Gallinger S. Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data. Nat Rev Cancer. Nature Publishing Group; 2022;22:131–42.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians. 2020;70:7–30.

3. Raimondi S, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: an overview. Nat Rev Gastroenterol Hepatol. Nature Publishing Group; 2009;6:699–708.

4. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: a population-based study. Lancet. 2015;385:1206–18.

5. Smith BD, Smith GL, Hurria A, Hortobagyi GN, Buchholz TA. Future of Cancer Incidence in the United States: Burdens Upon an Aging, Changing Nation. JCO. Wolters Kluwer; 2009;27:2758–65.

6. Singhi AD, Koay EJ, Chari ST, Maitra A. Early Detection of Pancreatic Cancer: Opportunities and Challenges. Gastroenterology. 2019;156:2024–40.

7. Ho WJ, Jaffee EM, Zheng L. The tumour microenvironment in pancreatic cancer — clinical challenges and opportunities. Nat Rev Clin Oncol. Nature Publishing Group; 2020;17:527–40.

8. Gobbi PG, Bergonzi M, Comelli M, Villano L, Pozzoli D, Vanoli A, et al. The prognostic role of time to diagnosis and presenting symptoms in patients with pancreatic cancer. Cancer Epidemiology. 2013;37:186–90.

9. Gomez-Rubio P, Zock J-P, Rava M, Marquez M, Sharp L, Hidalgo M, et al. Reduced risk of pancreatic cancer associated with asthma and nasal allergies. Gut. BMJ Publishing Group; 2017;66:314–22.

10. Gomez-Rubio P, Piñero J, Molina-Montes E, Gutiérrez-Sacristán A, Marquez M, Rava M, et al. Pancreatic cancer and autoimmune diseases: An association sustained by computational and epidemiological case–control approaches. International Journal of Cancer. 2019;144:1540–9.

11. López de Maturana E, Rodríguez JA, Alonso L, Lao O, Molina-Montes E, Martín-Antoniano IA, et al. A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. Genome Med. 2021;13:15.

12. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. Nature Publishing Group; 2017;551:92–4.

13. Lan Q, Hsiung CA, Matsuo K, Hong Y-C, Seow A, Wang Z, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat Genet. Nature Publishing Group; 2012;44:1330–5.

14. Shiraishi K, Okada Y, Takahashi A, Kamatani Y, Momozawa Y, Ashikawa K, et al. Association of variations in HLA class II and other loci with susceptibility to EGFR-mutated lung adenocarcinoma. Nat Commun. Nature Publishing Group; 2016;7:12451.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

15. Aterido A, Cañete JD, Tornero J, Ferrándiz C, Pinto JA, Gratacós J, et al. Genetic variation at the glycosaminoglycan metabolism pathway contributes to the risk of psoriatic arthritis but not psoriasis. Annals of the Rheumatic Diseases. BMJ Publishing Group Ltd; 2019;78:355–64.

16. Merkel PA, Xie G, Monach PA, Ji X, Ciavatta DJ, Byun J, et al. Identification of Functional and Expression Polymorphisms Associated With Risk for Antineutrophil Cytoplasmic Autoantibody–Associated Vasculitis. Arthritis & Rheumatology. 2017;69:1054–66.

17. Mendoza-Ramírez P, López-Olaiz MA, Morales-Fernández AL, Flores-Echiveste MI, de Jesus Casillas-Navarro A, Pérez-Rodríguez MA, et al. Class I MHC Polymorphisms Associated with Type 2 Diabetes in the Mexican Population. Genes (Basel). 2022;13:772.

18. Jacobi T, Massier L, Klöting N, Horn K, Schuch A, Ahnert P, et al. HLA Class II Allele Analyses Implicate Common Genetic Components in Type 1 and Non-Insulin-Treated Type 2 Diabetes. J Clin Endocrinol Metab. 2020;105:dgaa027.

19. Purohit A, Purohit BM, Mani A, Bhambal A. Genetic Implications of HLA-DR and HLA-DQ Genotype on Tobacco Smoking and Oral Submucous Fibrosis. Oral Health Prev Dent. 2020;18:455–60.

20. Movahedi M, Moin M, Gharagozlou M, Aghamohammadi A, Dianat S, Moradi B, et al. Association of HLA class II alleles with childhood asthma and Total IgE levels. Iran J Allergy Asthma Immunol. 2008;7:215–20.

21. Ribeyre C, Carlini F, René C, Jordier F, Picard C, Chiaroni J, et al. HLA-G Haplotypes Are Differentially Associated with Asthmatic Features. Frontiers in Immunology [Internet]. 2018 [cited 2022 Sep 29];9. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2018.00278

22. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. Nat Genet. Nature Publishing Group; 2018;50:1072–80.

23. Filip I, Wang A, Kravets O, Orenbuch R, Zhao J, Perea-Chamblee TE, et al. Pervasiveness of HLA allele-specific expression loss across tumor types. Genome Med. 2023;15:8.

24. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-I Genotype Restricts the Oncogenic Mutational Landscape. Cell. 2017;171:1272-1283.e15.

25. Kloetzel PM. Generation of major histocompatibility complex class I antigens: functional interplay between proteasomes and TPPII. Nat Immunol. Nature Publishing Group; 2004;5:661–9.

26. Unanue ER. From antigen processing to peptide-MHC binding. Nat Immunol. Nature Publishing Group; 2006;7:1277–9.

27. Stenzel A, Lu T, Koch WA, Hampe J, Guenther SM, De La Vega FM, et al. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. Hum Genet. 2004;114:377–85.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

28. Gonzalez-Galarza FF, McCabe A, Melo dos Santos EJ, Takeshita L, Ghattaoraya G, Jones AR, et al. Allele Frequency Net Database. In: Boegel S, editor. HLA Typing: Methods and Protocols [Internet]. New York, NY: Springer; 2018 [cited 2022 Sep 12]. page 49–62. Available from: https://doi.org/10.1007/978-1-4939-8546-3_4

29. Talwar JV, Laub D, Pagadala MS, Castro A, Lewis M, Luebeck GE, et al. Autoimmune alleles at the major histocompatibility locus modify melanoma susceptibility. The American Journal of Human Genetics. Elsevier; 2023;110:1138–61.

30. Dyckhoff G, Herold-Mende C, Scherer S, Plinkert PK, Warta R. Human Leucocyte Antigens as Prognostic Markers in Head and Neck Squamous Cell Carcinoma. Cancers. Multidisciplinary Digital Publishing Institute; 2022;14:3828.

31. Rive CM, Yung E, Hughes CS, Brown SD, Sharma G, Dreolini L, et al. Recombinant T cell receptors specific for HLA-A*02:01-restricted neoepitopes containing KRAS codon 12 hotspot mutations [Internet]. bioRxiv; 2020 [cited 2022 Sep 13]. page 2020.06.15.149021. Available from: https://www.biorxiv.org/content/10.1101/2020.06.15.149021v1

32. Shenzhen BinDeBio Ltd. EGFRvIII/DR5/NY-ESO-1/Mesothelin CAR-T/TCR-T Cells Immunotherapy for Solid Malignancies [Internet]. clinicaltrials.gov; 2021 Feb. Report No.: NCT03941626. Available from: https://clinicaltrials.gov/ct2/show/NCT03941626

33. Yamamoto Y. Investigator-initiated Clinical Trial (Phase II) of Cancer Vaccine “DSP-7888” for Acute Myeloid Leukemia Patients. [Internet]. clinicaltrials.gov; 2021 Feb. Report No.: NCT04747002. Available from: https://clinicaltrials.gov/ct2/show/NCT04747002

34. PDC*line Pharma SAS. An Open-label, Dose-escalation, Phase I/II Study to Assess the Safety, the Tolerability, the Immunogenicity and the Preliminary Clinical Activity of the Therapeutic Cancer Vaccine, PDC*lung01, Associated or Not With Anti-PD-1 Treatment in Patients With Non-small-cell Lung Cancer (NSCLC) [Internet]. clinicaltrials.gov; 2022 Jul. Report No.: NCT03970746. Available from: https://clinicaltrials.gov/ct2/show/NCT03970746

35. Greenwich LifeSciences, Inc. A Randomized, Multicenter, Placebo-controlled, Phase 3 Study to Evaluate the Efficacy and Safety of HER2/Neu Peptide GLSI-100 (GP2 + GM-CSF) in HER2/Neu Positive Subjects With Residual Disease or High-Risk PCR After Both Neoadjuvant and Postoperative Adjuvant Trastuzumab-based Therapy (FLAMINGO-01) [Internet]. clinicaltrials.gov; 2022 Nov. Report No.: NCT05232916. Available from: https://clinicaltrials.gov/ct2/show/NCT05232916

36. Mishto M, Mansurkhodzhaev A, Ying G, Bitra A, Cordfunke RA, Henze S, et al. An in silico—in vitro Pipeline Identifying an HLA-A*02:01+ KRAS G12V+ Spliced Epitope Candidate for a Broad Tumor-Immune Response in Cancer Patients. Frontiers in Immunology [Internet]. 2019 [cited 2022 Sep 13];10. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02572

37. Fenton SE, Saleiro D, Platanias LC. Type I and II Interferons in the Anti-Tumor Immune Response. Cancers (Basel). 2021;13:1037.

38. Espinet E, Gu Z, Imbusch CD, Giese NA, Büscher M, Safavi M, et al. Aggressive PDACs Show Hypomethylation of Repetitive Elements and the Execution of an Intrinsic IFN Program Linked to a Ductal Cell of Origin. Cancer Discovery. 2021;11:638–59.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

39. Till JE, McDaniel L, Chang C, Long Q, Pfeiffer SM, Lyman JP, et al. Circulating KRAS G12D but not G12V is associated with survival in metastatic pancreatic ductal adenocarcinoma. Nat Commun. Nature Publishing Group; 2024;15:5763.

40. Rivera Pirela S, Echeverría M, Márquez G, Carrillo Z, Parra Y, Salcedo P, et al. HLA-DPA1*/DPB1* en asociación con leucemias linfoides agudas y leucemias mieloides crónicas en mestizos venezolanos. Inmunología. 2014;33:81–6.

41. Zawadzka-Starczewska K, Tymoniuk B, Stasiak B, Lewiński A, Stasiak M. Actual Associations between HLA Haplotype and Graves' Disease Development. J Clin Med. 2022;11:2492.

42. Deng H, Yan X, Yuan L. Human genetic basis of coronavirus disease 2019. Signal Transduct Target Ther. 2021;6:344.

43. Sayad A, Akbari MT, Mehdizadeh M, Taheri M, Hajifathali A. HLA-A*26 and susceptility of iranian patients with non-Hodgkin lymphoma. Iran J Immunol. 2014;11:269–74.

44. Molina-Montes E, Gomez-Rubio P, Márquez M, Rava M, Löhr M, Michalski CW, et al. Risk of pancreatic cancer associated with family history of cancer and other medical conditions by accounting for smoking among relatives. International Journal of Epidemiology. 2018;47:473–83.

45. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev. 2017;26:126–35.

46. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLOS Genetics. Public Library of Science; 2009;5:e1000529.

47. Durbin RM, Altshuler D, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. A map of human genome variation from population-scale sequencing. Nature. Nature Publishing Group; 2010;467:1061–73.

48. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.

49. Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The Cancer Genome Atlas Pan-Cancer Analysis Project. Nat Genet. 2013;45:1113–20.

50. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics. 2014;30:3310–6.

51. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. PLOS ONE. Public Library of Science; 2013;8:e64683.

52. Rich SS, Concannon P, Erlich H, Julier C, Morahan G, Nerup J, et al. The Type 1 Diabetes Genetics Consortium. Annals of the New York Academy of Sciences. 2006;1079:1–8.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

53. Graffelman J. Exploring Diallelic Genetic Markers: The HardyWeinberg Package. Journal of Statistical Software. 2015;64:1–23.

54. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. PLoS Comput Biol. 2013;9:e1002877.

55. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. Nature Publishing Group; 2015;33:1152–8.

56. Xie C, Yeo ZX, Wong M, Piper J, Long T, Kirkness EF, et al. Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences; 2017;114:8059–64.

57. JP S, DJ S. haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous [Internet]. 2022 [cited 2022 Sep 12]. Available from: https://CRAN.R-project.org/package=haplo.stats

58. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. The American Journal of Human Genetics. 2021;108:1880–90.

59. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

60. Venkataraman GR, Olivieri JE, DeBoever C, Tanigawa Y, Justesen JM, Dilthey A, et al. Pervasive additive and non-additive effects within the HLA region contribute to disease risk in the UK Biobank [Internet]. bioRxiv; 2020 [cited 2022 Sep 12]. page 2020.05.28.119669. Available from: https://www.biorxiv.org/content/10.1101/2020.05.28.119669v2

61. Krebs K, Bovijn J, Zheng N, Lepamets M, Censin JC, Jürgenson T, et al. Genome-wide Study Identifies Association between HLA-B∗55:01 and Self-Reported Penicillin Allergy. Am J Hum Genet. 2020;107:612–21.

62. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software. 2010;36:1–48.

63. Wahba G. Spline models for observational data. Philadelphia, Pa: Society for Industrial and Applied Mathematics; 1990.

64. Granato I, Fritsche-Neto R. snpReady: Preparing Genotypic Datasets in Order to Run Genomic Analysis [Internet]. 2018 [cited 2022 Sep 12]. Available from: https://CRAN.R-project.org/package=snpReady

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Cumulative PDAC risk estimates (OR: odds ratio and CI: confidence interval) in the extreme scenarios according to the presence/absence of known risk factors (Type II diabetes, smoking, and family history of pancreatic cancer, FHPC) or protective factors (asthma and nasal allergies) in addition to the significant *HLA* alleles identified in the meta-analysis.

	OR	L. CI	U. CI	p-value
High-risk scenario [a]	2.80	1.24	6.35	
High-risk scenario + *HLA-B**39				1.17×10^{-1}
1 allele	3.85	1.20	12.32	
2 alleles	5.28	1.65	16.92	
High-risk scenario + *HLA-DPB1**04				1.33×10^{-1}
1 allele	3.06	1.22	7.67	
2 alleles	3.34	1.33	8.37	
Low-risk scenario [b]	0.38	0.13	1.09	
Low-risk scenario + *HLA-A**02:01				1.18×10^{-2}
1 allele	0.32	0.1	1.03	
2 alleles	0.27	0.08	0.88	
Low-risk scenario + *HLA-B**49				2.4×10^{-2}
1 allele	0.19	0.03	1.07	
2 alleles	0.1	0.02	0.55	

Scenario a: smokers, diabetics or individuals with family history of cancer who did not have asthma nor allergies in the meta-analysis of the PangenEU and UK Biobank population.
Scenario b: asthmatic or allergic individuals who did not smoke, did not have diabetes, nor family history of cancer.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

MAIN FIGURE LEGENDS

Figure 1. Study Flowchart: Overview of the complementary approaches adopted in this study to explore the role of *MHC* region in the PDAC susceptibility and prognosis.

Figure 2. Odds ratio (OR) and 95% confidence interval (CI) of the association analyses between PDAC susceptibility alleles, as well as the allele frequency (AF): *HLA-A*02:01* (Figure 2A), *HLA-B*49* (Figure 2B), *HLA-B*39* (Figure 2C), and *HLA-DPB1*04* (Figure 2D) obtained in the meta-analysis and known risk/protective factors for PDAC.

Figure 3. Hazard ratio (HR) and 95% confidence interval (CI) of the Cox proportional-hazards regression models under a dominance MoI obtained in the meta-analysis for the allele *HLA-A*02:01* when considering PanGenEU, UK Biobank, and TCGA (Figure 3A), and when including data from the Caris clinical trial in the meta-analysis (Figure 3B).

Figure 4. Kaplan-Meier curves of patients according to their KRAS mutational status in *HLA-A*02:01* carriers with PDAC tumors harboring $KRAS^{G12V}$ patients whose tumors have other KRAS mutations who are carriers of *HLA-A*02:01* (Figures 4A and Figure 4C) and non-carriers of that allele (Figures 4B and Figure 4D).

Figure 5. Kaplan-Meier curves of patients according to the presence or absence of the *HLA-A*02:01* allele in $KRAS^{G12V}$ (Figures 5A) and $KRAS^{G12V}$ (Figures 5B) mutated tumors.

Figure 6. Hazard ratio (HR) and 95% confidence interval (CI) of the Cox proportional-hazards regression models under a dominance MoI obtained in the meta-

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

analysis for the allele *HLA-A*02:01* for (A) stages I&II PDAC tumors from Erasmus and TCGA study populations and (B) stages III&IV tumors from Erasmus MC and Caris study populations.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

925 **Additional Files**

Annex 1. PanGenEU Investigators.

Table S1. Baseline characteristics of the study populations.

Table S2. Number of alleles of *HLA* Class I and Class II genes considered in the association analysis in each study population after filtering out the alleles with poor imputation quality.

Table S3. *HLA* allele, 95% confidence intervals bounds (ci.lb and ci.ub), odds ratios (OR), p-value, and adjusted p-value and of the association meta-analysis between the *HLA* alleles and pancreatic ductal adenocarcinoma (PDAC) risk including the PanGenEU and UK Biobank study populations.

Table S4. *HLA* allele, frequency (Freq), Imputation quality (INFO), odds ratios (OR), p-value, and adjusted p-value and of the association between the *HLA* alleles and pancreatic ductal adenocarcinoma (PDAC) risk in the PanGenEU study population.

Table S5. *HLA* allele, frequency (Freq), odds ratios (OR), p-value, and adjusted p-value and of the association meta-analysis between the *HLA* alleles and pancreatic ductal adenocarcinoma (PDAC) risk in the UK Biobank study population.

Table S6. Frequency in both populations (freq), Odd ratio (OR), confidence interval (CI), and p-value (p) of the inferred haplotypes from *MHC* I and II *HLA* genes significantly associated with PDAC risk by genome-wide association studies considering an additive mode of inheritance.

Table S7. Odd ratio (OR) and p-value (p) of the association between *HLA-A**02:01 and *KRAS* mutations across three different tumor types from TCGA data: PDAC, colorectal, and NSCLC (we excluded stomach because there was only one individual with G12V mutations).

Table S8. ICD9 and ICD10 codes for PDAC in UK Biobank

Table S9. Amino acid (AA), base pair (BP), Frequency (Freq), odds ratios (OR), p-value, and adjusted p-value and of the association between the HLA amino acids and pancreatic ductal adenocarcinoma (PDAC) risk in the PanGenEU study populations.

Table S10. *KRAS* mutational status of PDAC, colorectal, stomach, and NSCLC cancer types from the TGCA study.

Figure S1. Whole landscape of the alleles of *HLA* class I and II genes at two-fields in PanGenEU (Figure S1A) and in UKBiobank (Figure S1B) study populations. The x-axis shows each of the imputed *HLA* alleles across the 8 studied loci. The y-axis, represents the frequency in the PanGenEU

medRxiv preprint doi: https://doi.org/10.1101/2024.08.29.24312704; this version posted August 29, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

and UK Biobank populations for each *HLA* allele.

Figure S2. Scatter plots of the HLA allele frequencies in PanGenEU study population versus in T1DGC, which was used as a reference panel to impute the 1-field (Figure S2A) and 2- fields (Figure S2B) *HLA* alleles in PanGenEU.

Figure S3. Scatter Plot of the allele frequencies in PanGenEU and UK Biobank populations, showing a moderate agreement.

Figure S4. Odds ratio (OR) and 95% confidence interval (CI) of the association analyses between *HLA-A*26:01* and PDAC risk and PDAC risk/protective factors in PanGenEU population. FHPC: Family history of PC; BMI: body mass index.

Figure S5. Venn diagram illustrating the overlap of *HLA* Class I and II gene alleles at two fields in the PanGenEU, UKBiobank, and TCGA study populations.

Figure S6. Hazard ratio (HR) and 95% confidence interval (CI) of the Cox proportional-hazards regression models under a codominance MoI obtained in the meta-analysis for the allele *HLA-DPB1*04* when considering PanGenEU, UK Biobank, and TCGA.

Figure S7. Hazard ratio (HR) and 95% confidence interval (CI) of the Cox proportional-hazards regression models obtained under the additive MoI in the meta-analysis of three populations for the 2-field alleles *HLA-A*30:01* and *HLA-DPB1*09:01*.

Figure S8. Boxplot showing the significant differences in IFN-γ levels between PDAC tumors with $KRAS^{G12D}$ mutations vs any of the other KRAS mutated tumors obtained through the Wilcoxon test.

Figure S9. Hazard ratio (HR) and 95% confidence interval (CI) of the Cox proportional-hazards regression models obtained under the additive MoI in the meta-analysis of three populations for the interaction between the *HLA-A*02:01* allele with $KRAS^{G12V}$ mutations.

926

Figure 1

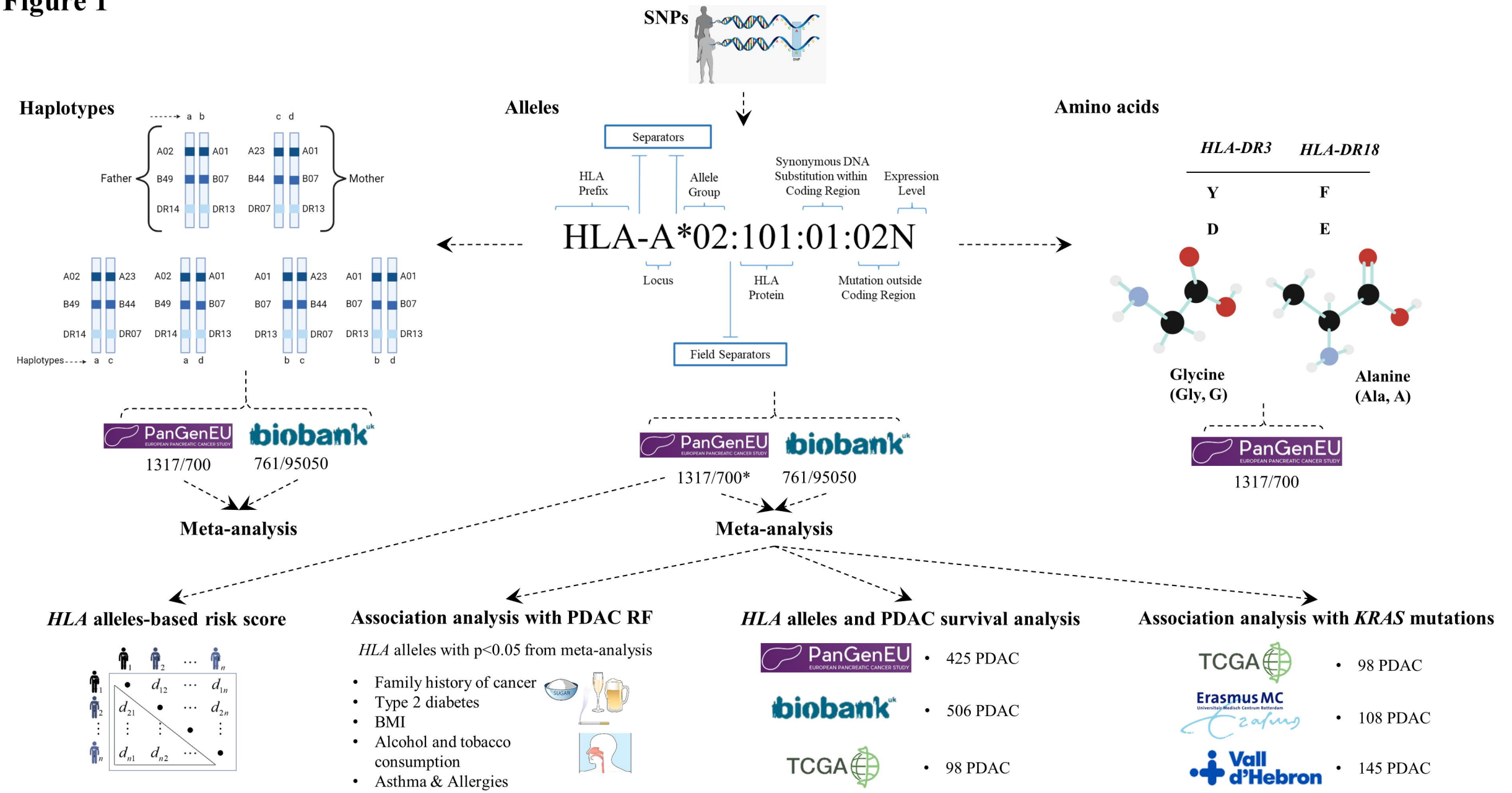

SNPs
Haplotypes
Alleles
Amino acids
Separators
HLA Prefix
Allele Group
Synonymous DNA Substitution within Coding Region
Expression Level
HLA-A*02:101:01:02N
Locus
HLA Protein
Mutation outside Coding Region
Field Separators
Father
Mother
a b
c d
A02 A01 A23 A01
B49 B07 B44 B07
DR14 DR13 DR07 DR13
Haplotypes
HLA-DR3
HLA-DR18
Y F
D E
Glycine (Gly, G)
Alanine (Ala, A)
PanGenEU
biobank
1317/700
761/95050
1317/700*
Meta-analysis
HLA alleles-based risk score
Association analysis with PDAC RF
HLA alleles with p<0.05 from meta-analysis
Family history of cancer
Type 2 diabetes
BMI
Alcohol and tobacco consumption
Asthma & Allergies
HLA alleles and PDAC survival analysis
425 PDAC
506 PDAC
98 PDAC
Association analysis with KRAS mutations
TCGA
Erasmus MC
Vall d'Hebron
98 PDAC
108 PDAC
145 PDAC
*Cases/controls

Figure 2

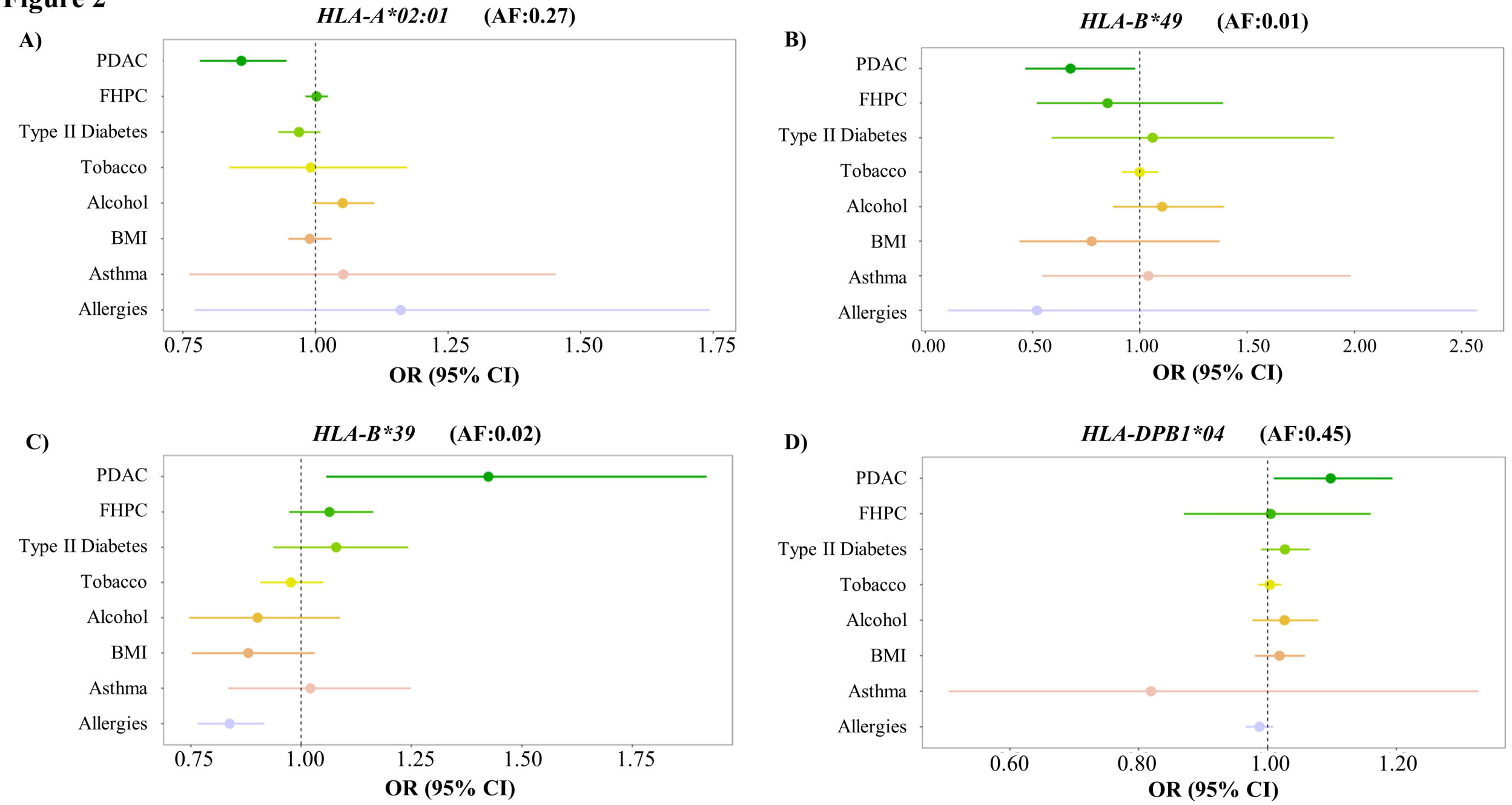

A)
HLA-A*02:01 (AF:0.27)
PDAC
FHPC
Type II Diabetes
Tobacco
Alcohol
BMI
Asthma
Allergies
0.75
1.00
1.25
1.50
1.75
OR (95% CI)
B)
HLA-B*49 (AF:0.01)
PDAC
FHPC
Type II Diabetes
Tobacco
Alcohol
BMI
Asthma
Allergies
0.00
0.50
1.00
1.50
2.00
2.50
OR (95% CI)
C)
HLA-B*39 (AF:0.02)
PDAC
FHPC
Type II Diabetes
Tobacco
Alcohol
BMI
Asthma
Allergies
0.75
1.00
1.25
1.50
1.75
OR (95% CI)
D)
HLA-DPB1*04 (AF:0.45)
PDAC
FHPC
Type II Diabetes
Tobacco
Alcohol
BMI
Asthma
Allergies
0.60
0.80
1.00
1.20
OR (95% CI)

Figure 3

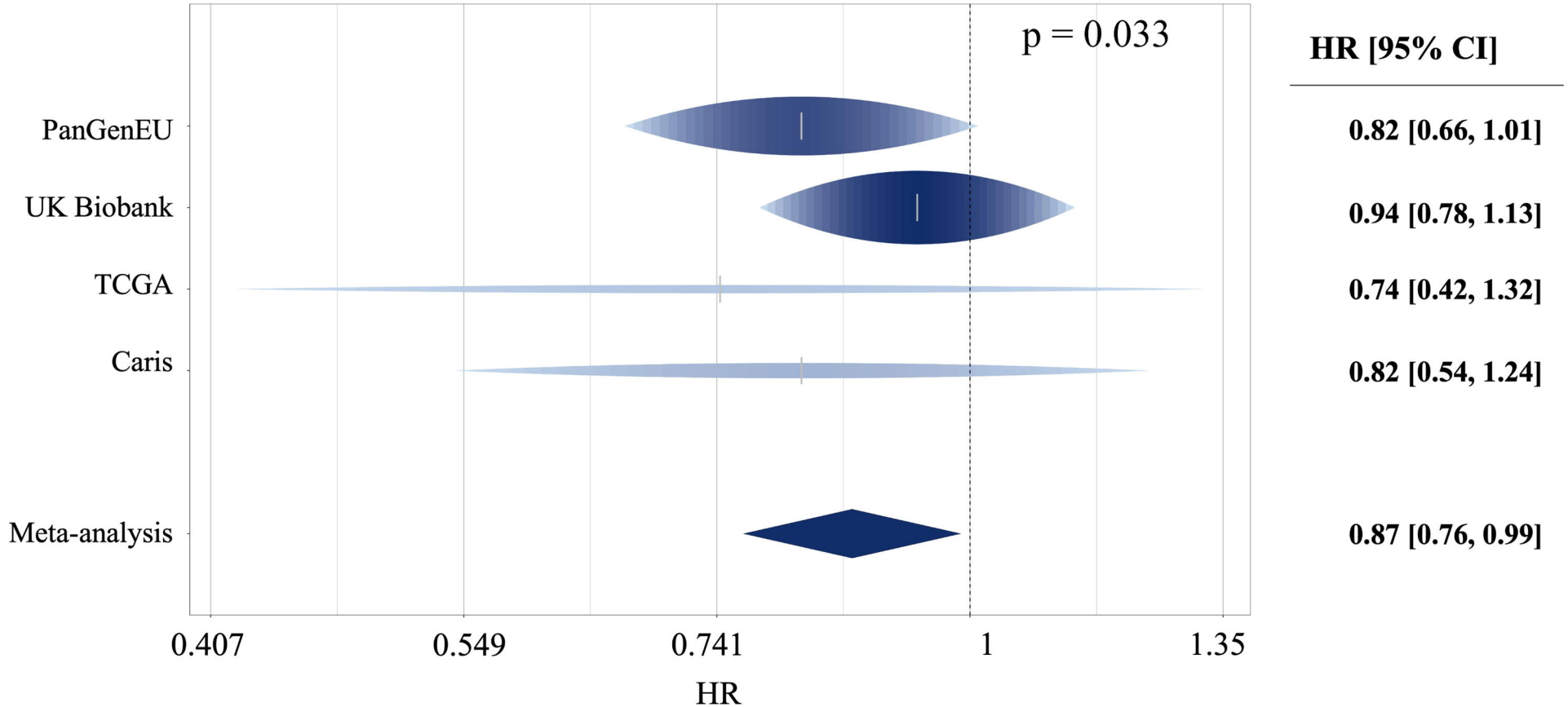

p = 0.033
HR [95% CI]
PanGenEU
0.82 [0.66, 1.01]
UK Biobank
0.94 [0.78, 1.13]
TCGA
0.74 [0.42, 1.32]
Caris
0.82 [0.54, 1.24]
Meta-analysis
0.87 [0.76, 0.99]
0.407
0.549
0.741
1
1.35
HR

Figure 4

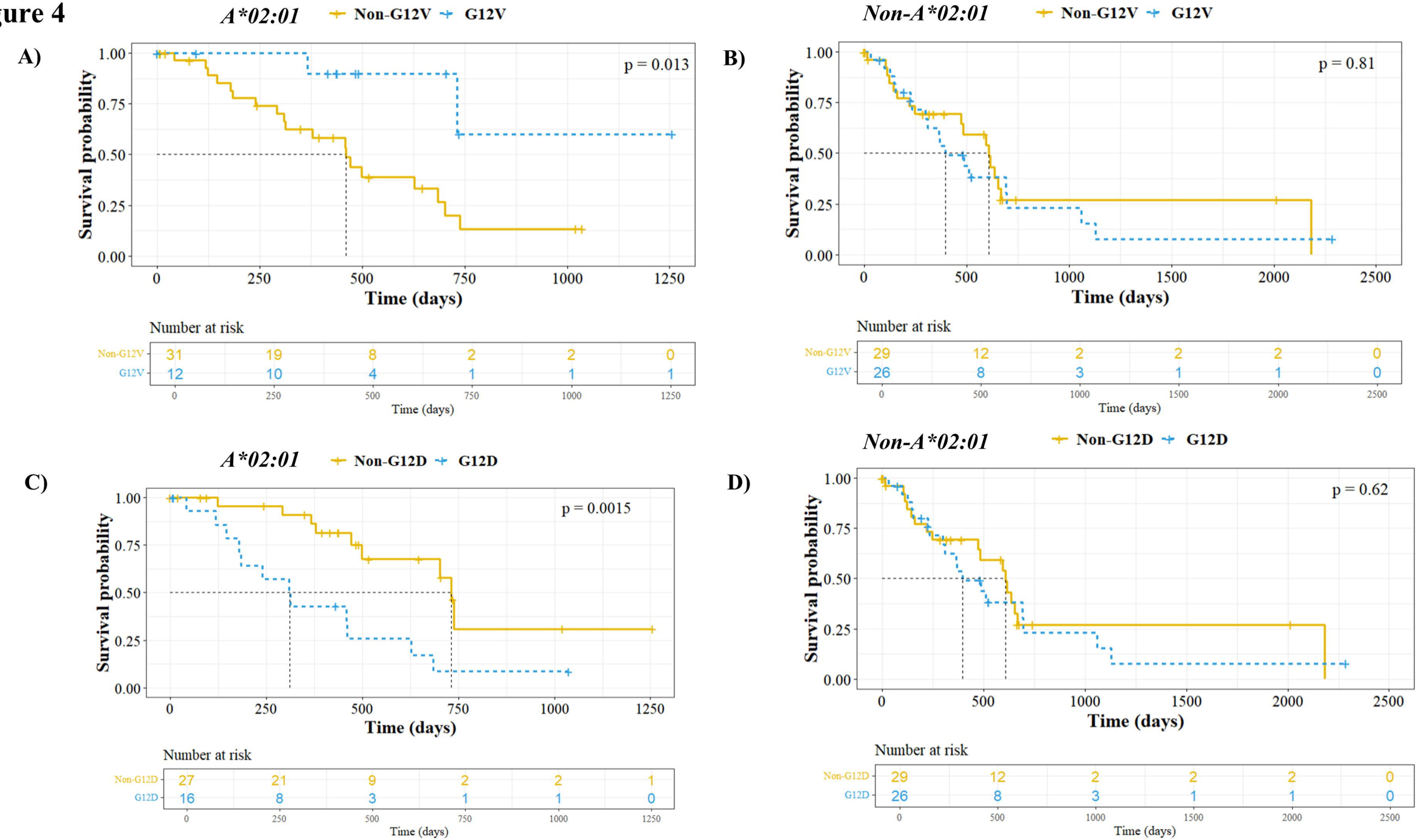
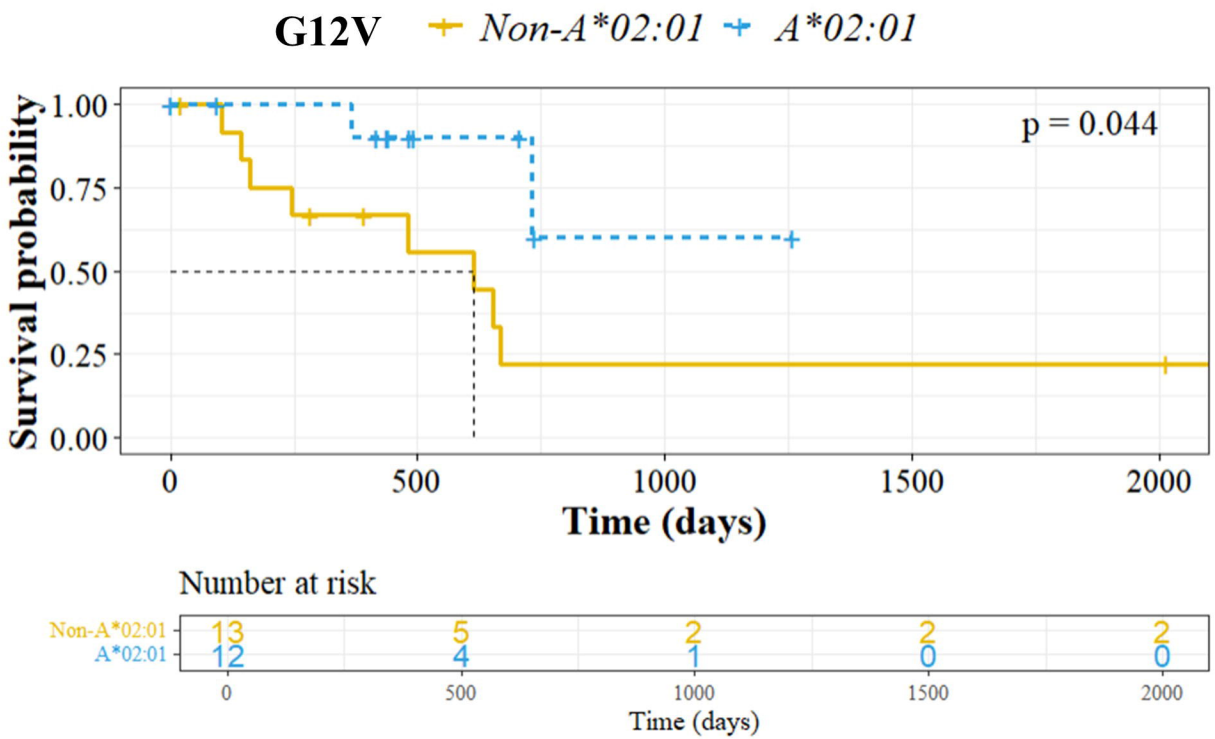

A)
A*02:01
Non-G12V
G12V
p = 0.013
Survival probability
Time (days)
Number at risk
Non-G12V 31 19 8 2 2 0
G12V 12 10 4 1 1 1
B)
Non-A*02:01
Non-G12V
G12V
p = 0.81
Survival probability
Time (days)
Number at risk
Non-G12V 29 12 2 2 2 0
G12V 26 8 3 1 1 0
C)
A*02:01
Non-G12D
G12D
p = 0.0015
Survival probability
Time (days)
Number at risk
Non-G12D 27 21 9 2 2 1
G12D 16 8 3 1 1 0
D)
Non-A*02:01
Non-G12D
G12D
p = 0.62
Survival probability
Time (days)
Number at risk
Non-G12D 29 12 2 2 2 0
G12D 26 8 3 1 1 0

Figure 5


G12V
Non-A*02:01
A*02:01
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.044
0
500
1000
1500
2000
Time (days)
Number at risk
Non-A*02:01
A*02:01
13
12
5
4
2
1
2
0
2
0
0
500
1000
1500
2000
Time (days)

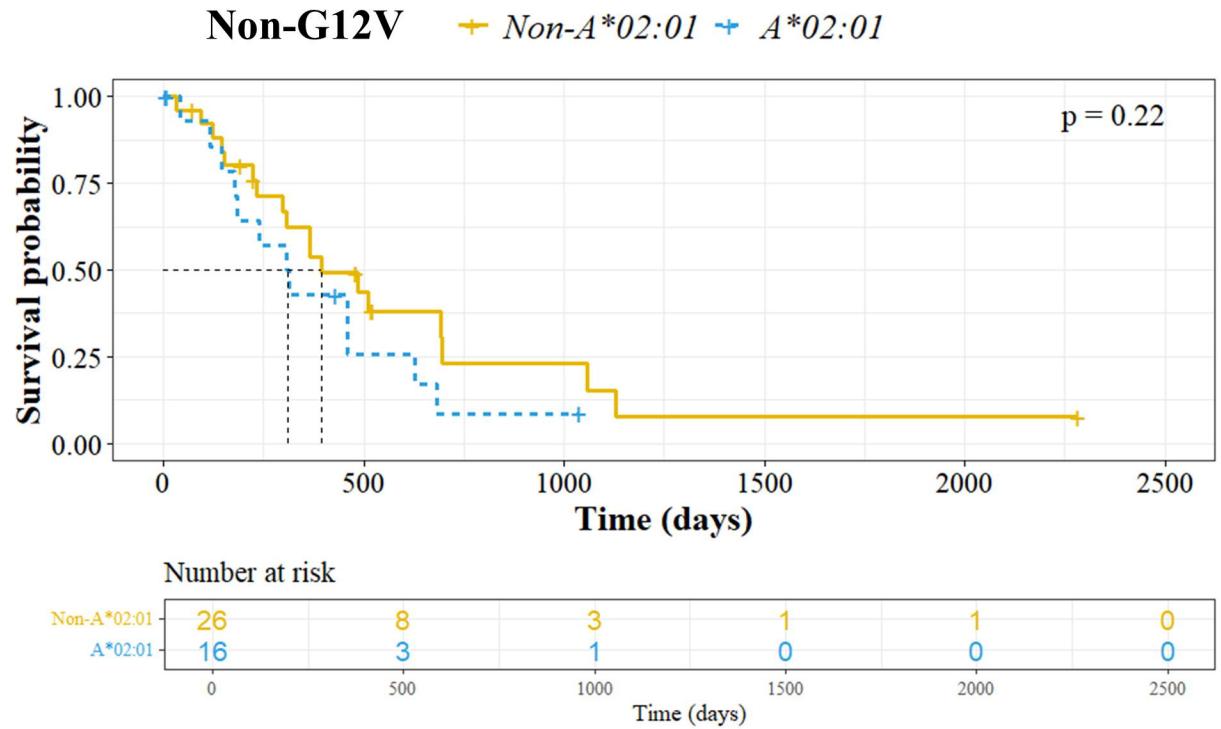
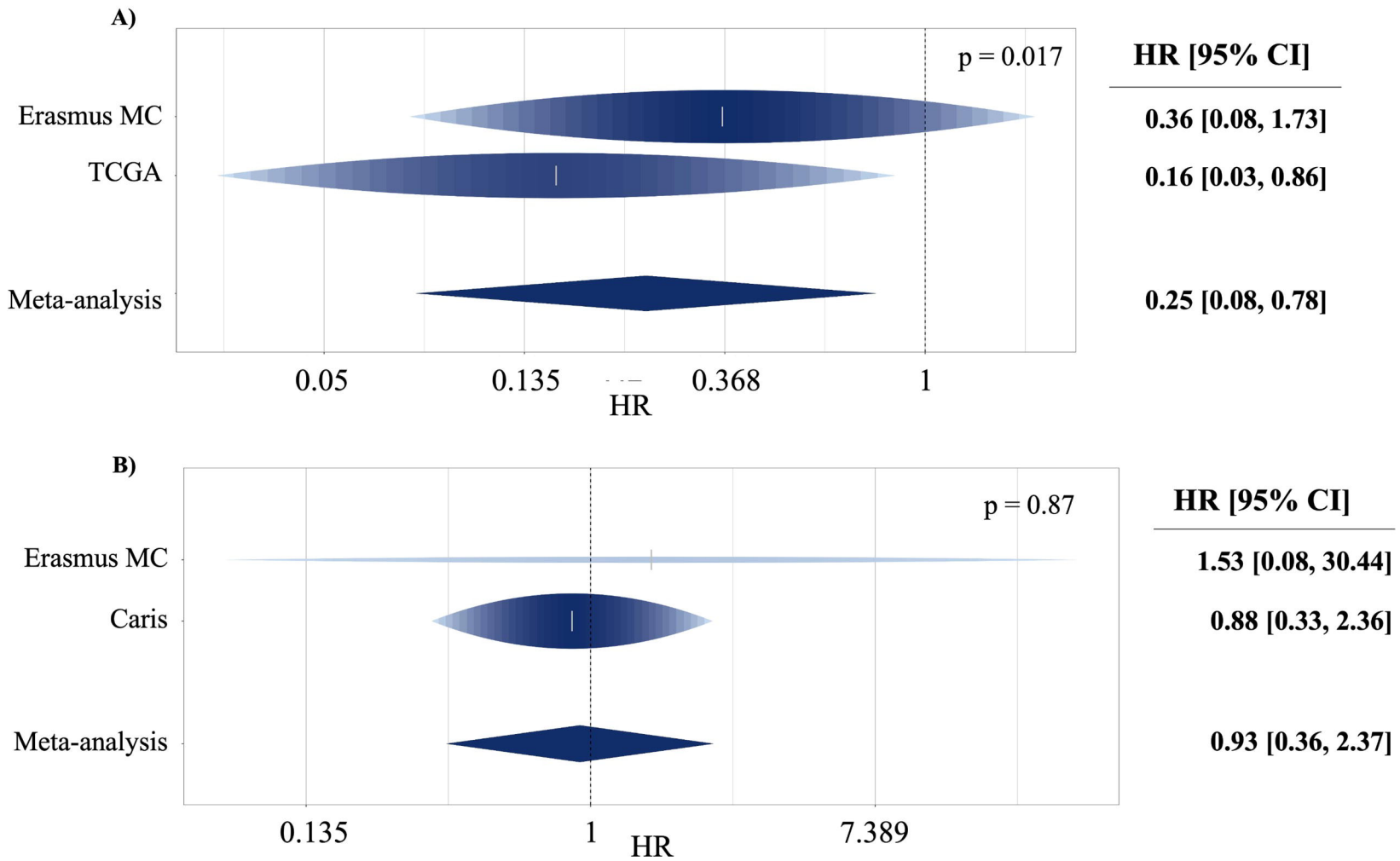

Non-G12V
Non-A*02:01
A*02:01
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.22
0
500
1000
1500
2000
2500
Time (days)
Number at risk
Non-A*02:01
A*02:01
26
16
8
3
3
1
1
0
1
0
0
0
0
500
1000
1500
2000
2500
Time (days)

Figure 6


A)
p = 0.017
HR [95% CI]
Erasmus MC
0.36 [0.08, 1.73]
TCGA
0.16 [0.03, 0.86]
Meta-analysis
0.25 [0.08, 0.78]
0.05
0.135
0.368
1
HR
B)
p = 0.87
HR [95% CI]
Erasmus MC
1.53 [0.08, 30.44]
Caris
0.88 [0.33, 2.36]
Meta-analysis
0.93 [0.36, 2.37]
0.135
1
7.389
HR