

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

3
4 **Authors:** Alberto Langtry¹, Raul Rabadan², Lola Alonso¹, Casper van Eijck^{1,3},
5 Teresa Macarulla⁴, Rita T Lawlor⁵, Alfredo Carrato⁶, Rafael Alvarez-Gallego⁷, Mar
6 Iglesias⁸, Xavier Molero⁹, J Matthias Löhr¹⁰, Christopher W Michalski¹¹, José Perea¹²,
7 Michael O’Rorke¹³, Víctor M Barberà¹⁴, Adonina Tardón¹⁵, Antoni Farré¹⁶, Luís
8 Muñoz-Bellvís¹⁷, Tatjana Crnogorac-Jurcevic¹⁸, Enrique Domínguez-Muñoz¹⁹,
9 Thomas Gress²⁰, William Greenhalf²¹, Linda Sharp²², Sergio Sabroso-Lasa¹, Ioan
10 Filip², Gaby Strijk³, Florian Castet⁴, Joaquim Balsells⁹, Eithne Costello²¹, Jörg
11 Kleeff²³, Bo Kong²⁴, Josefina Mora¹⁶, Damian O’Driscoll²⁵, Aldo Scarpa⁵, Weimin
12 Ye²⁶, Francisco X. Real²⁷, Núria Malats^{1*}, Evangelina López de Maturana^{1*}, on behalf
13 of the PanGenEU Investigators²⁸

14
15 **Authors’ affiliations:**

16 (1) Genetic and Molecular Epidemiology Group, Spanish National Cancer
17 Research Center (CNIO), and CIBERONC, Madrid, Spain.

18 (2) Department of Systems Biology, Columbia University, New York, NY, USA.

19 (3) Department of Surgery, Erasmus Medical Center, Rotterdam, The Netherlands.

20 (4) Vall d’Hebron Institute of Oncology (VHIO), Hospital Universitari Vall
21 d’Hebron, Barcelona, Barcelona, Spain.

22 (5) ARC-Net Center for Applied Research on Cancer, Department of Diagnostics
23 and Public Health and Department of Engineering for Innovation Medicine,
24 University of Verona, Verona, Italy.

25 (6) Department of Oncology, Ramón y Cajal University Hospital, IRYCIS, Alcala
26 University, Madrid, and CIBERONC, Spain.

27 (7) HM CIOCC Madrid (Centro Integral Oncológico Clara Campal). Hospital
28 Universitario HM Sanchinarro. Instituto de Investigación Sanitaria HM
29 Hospitales. Facultad HM de Ciencias de la Salud de la Universidad Camilo
30 José Cela.

31 (8) Hospital del Mar—Parc de Salut Mar, Barcelona, and CIBERONC, Spain.

32 (9) Hospitals Universitaris Arnau de Vilanova i Santa Maria, IRBLleida, UdL,
33 25198, Lleida; and CIBEREHD, Madrid, Spain.

- 34 (10) Gastrocentrum, Karolinska Institutet and University Hospital,
35 Stockholm, Sweden.
- 36 (11) Department of Surgery, Technical University of Munich, Munich; and
37 Department of General, Visceral and Transplantation Surgery, Medical Faculty
38 Heidelberg, University Heidelberg, Germany.
- 39 (12) Department of Surgery, Hospital 12 de Octubre, Madrid; and Institute
40 of Biomedical Research of Salamanca, Salamanca, Spain.
- 41 (13) Center for Public Health, Belfast, Queen's University Belfast, UK; and
42 College of Public Health, The University of Iowa, Iowa City, IA, US.
- 43 (14) Molecular Genetics Laboratory, General University Hospital of Elche,
44 Spain.
- 45 (15) Department of Medicine, Instituto Universitario de Oncología del
46 Principado de Asturias (IUOPA), Instituto de Investigación Sanitaria del
47 Principado de Asturias (ISPA), Oviedo, and CIBERESP, Spain.
- 48 (16) Department of Gastroenterology and Clinical Biochemistry, Hospital de
49 la Santa Creu i Sant Pau, Barcelona, Spain.
- 50 (17) Department of Surgery, Hospital Universitario de Salamanca – IBSAL.
51 Universidad de Salamanca and CIBERONC, Spain.
- 52 (18) Barts Cancer Institute, Center for Cancer Biomarkers and
53 Biotherapeutics, Queen Mary University of London, London, UK.
- 54 (19) Department of Gastroenterology, University Clinical Hospital of
55 Santiago de Compostela, Spain.
- 56 (20) Department of Gastroenterology, University Hospital of Giessen and
57 Marburg, Marburg, Germany.
- 58 (21) Department of Molecular and Clinical Cancer Medicine, University of
59 Liverpool, Liverpool, UK.
- 60 (22) National Cancer Registry Ireland and HRB Clinical Research Facility,
61 University College Cork, Cork, Ireland; and Newcastle University, Institute of
62 Health & Society, Newcastle, UK.
- 63 (23) Department of Surgery, Technical University of Munich, Munich; and
64 Department of Visceral, Vascular and Endocrine Surgery, Martin-Luther-
65 University Wittenberg-Halle (Saale), Germany.
- 66 (24) Department of Surgery, Technical University of Munich, Munich

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

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

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

74 (28) In Annex 1.

75

76 *Co-senior authors

77

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

82

83 **ABSTRACT**

84 Genetic susceptibility loci are associated with PDAC risk and survival, but the impact
85 of germline HLA region variation remains largely unexplored. This study examined
86 *HLA* I-II alleles within the PanGenEU study and validated our findings using external
87 datasets (UK Biobank, TCGA, PAN-NGS trial, and Caris trial). *HLA-A*02:01* and
88 *HLA-B*49* alleles were linked to a decreased risk of PDAC, whereas *HLA-B*39*,
89 *HLA-DPB1*04*, and *HLA-A*26:01* were directly associated with increased risk.
90 PDAC patients carrying the *HLA-A*02:01* allele also showed lower mortality rates,
91 with the effect being more pronounced in those with *KRAS*^{G12V} mutations, pointing to
92 a host*tumor genetic interaction. This research highlights *HLA-A*02:01*, found in
93 20% of Europeans, as a marker for reduced PDAC risk and mortality, especially in
94 *KRAS*^{G12V} mutated tumors. Results from this study could enhance personalized
95 medicine for PDAC by identifying patients who may benefit from regular screenings

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

99

100 INTRODUCTION

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

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

125 The *MHC* is essential for the proper response of the immune system against
126 tumor cells (24). The *MHC* genetic complex is a dense cluster of genes located at the
127 6p21.1-21.3 chromosome cytoband (28,477,797-33,448,354 bp, GRCh37), and it
128 includes three *HLA* sub-regions. *HLA* class I can be classified into classical (*HLA*-A, -
129 B, and -C) and non-classical genes (*HLA*-E, -F, and -G). *HLA* class II genes are *HLA*-
130 *DR*, -*DQ*, and -*DP*. *HLA* class I and II proteins present intracellular and extracellular
131 peptides to CD8⁺ and CD4⁺ T-cells, respectively (25,26). The *MHC* region also shows
132 an extensive linkage disequilibrium (27). As a consequence of its complexity, the role
133 of this region has not yet been studied in-depth in relation to the risk of cancer and,
134 more specifically, PDAC.

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

150 RESULTS

151 *Characteristics of the study populations*

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

169

170 *Association analyses between HLA alleles and PDAC risk*

171 HLA class I and class II alleles explain 11% of the PDAC risk. Bayesian multi kernel-
172 based regression allowed us to obtain the genetic and residual variance components
173 ($\sigma_{HLA-I}^2 = 0.013$, $\sigma_{HLA-II}^2 = 0.010$, and $\sigma_e^2 = 0.174$) and therefore, the heritability
174 estimates for HLA class I and class II alleles were: $h_{HLA-I}^2 = 0.06$ and $h_{HLA-II}^2 = 0.05$.

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

187 In our comprehensive meta-analysis of the PanGenEU and UKB populations,
188 the following four alleles yielded significant associations with PDAC risk: *HLA-*
189 *A*02:01* ($OR=0.86$, $p\text{-value}=1.88\times 10^{-3}$), *HLA-B*49* ($OR=0.68$, $p\text{-value}=3.81\times 10^{-2}$),
190 *HLA-B*39* ($OR=1.42$, $p\text{-value}=1.99\times 10^{-2}$), and *HLA-DPBI*04* ($OR=1.1$, $p\text{-}$
191 $value=2.86\times 10^{-2}$) (**Fig. 2**). After multiple testing correction, only the *HLA-A*02:01*
192 remained significant ($OR=0.86$, corrected $p\text{-value}=2.25\times 10^{-2}$, **Supplementary Table**
193 **S3**). We also explored the association between the above-mentioned *HLA* alleles and
194 the established PDAC risk factors. Interestingly, we observed a significant inverse
195 association between *HLA-B*39* and nasal allergies, which is protective against PDAC
196 ($OR=0.84$, $p\text{-value}=1.15\times 10^{-4}$, see **Fig. 2C**).

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

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

208

209 *Cumulative PDAC risk score for the most promising alleles*

210 We estimated cumulative PDAC risk by combining known risk and protective factors
211 and adding PDAC-associated *HLA* alleles. In a meta-analysis with the UK Biobank,
212 individuals with the abovementioned non-genetic high-risk conditions had 2.8 times
213 higher PDAC risk. By including either *HLA-B*39* or *HLA-DPBI*04* alleles, the ORs
214 increased to 5.28 and 3.34, respectively (**Table 1**), though not statistically significant
215 (p-value= 1.17×10^{-1} , p-value= 1.33×10^{-1} , respectively). Conversely, asthmatic or
216 allergic individuals who did not smoke, had no diabetes, nor had a family history of
217 PDAC, experienced a 62% reduction in the odds (OR=0.38) of developing PDAC.
218 Subjects who had one/two alleles of *HLA-A*02:01* further decreased their ORs to 0.32
219 and 0.27, respectively (p-value= 1.18×10^{-2}). By including one/two copies of the *HLA-*
220 *B*49* allele in the model, the OR significantly decreased to 0.19 and 0.1 (p-
221 value= 2.4×10^{-2} , **Table 1**). Additionally, in the PanGenEU study, diabetics, smokers,
222 and individuals with a family history of cancer but without asthma or allergies had
223 4.33 times higher risk of PDAC than the population in the lowest risk category. When
224 those individuals at high-risk harbor one or two copies of the *HLA-A*26:01* allele, the

225 ORs increased to 7.73 and 13.81, respectively (p-value=2.28x10⁻², see
226 **Supplementary Fig. S4B**).

227

228 *Association analyses between HLA haplotypes and PDAC risk*

229 Given the critical role of *HLA* haplotypes in immune responses, we assessed their
230 association with PDAC risk. We identified 18 haplotypes significantly associated with
231 PDAC risk, nine of which remained significant after multiple testing correction
232 (**Supplementary Table S6**). Two of the top five protective haplotypes carried the
233 *HLA-A*02* allele, present in 20% of the European population and associated with
234 decreased PDAC risk in our analyses. Haplotypes harboring the *HLA-A*02* allele were
235 all protective for PDAC, some reaching an OR~0.4). Among the haplotypes with the
236 lowest effect size *HLA-B*35* and *HLA-A*02* alleles were common: *HLA-B*35/HLA-*
237 *A*02* (OR=0.41, 95%CI=0.18-0.94), *HLA-DRB1*13/HLA-B*35* (OR=0.43,
238 95%CI=0.38-0.5), and *HLA-DRB1*01/HLA-A*02* (OR=0.49, 95%CI=0.27-0.89). On
239 the contrary, the highest risk haplotypes harbored the *HLA-DRB1*04* allele: *HLA-*
240 *DRB1*04/HLA-B*44/HLA-C*07* (OR=1.19, 95%CI=1.12-1.26) and *HLA-*
241 *DRB*04/HLA-B*14* (OR=1.28, 95%CI=1.19-1.38) (**Supplementary Table S6**).

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

245

246 *HLA alleles are associated with overall survival of PDAC patients*

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

250 (**Supplementary Fig. S5**). The results obtained under a dominant MoI revealed that
251 patients who carried the *HLA-A*02:01* allele presented better survival than non-
252 carriers (HR=0.87, p-value= 3.3×10^{-2} , **Fig. 3**), indicating that a single copy of this
253 allele is enough to improve survival. On the other hand, homozygous individuals for
254 *HLA-DPBI*04* displayed a borderline significantly worse survival under a
255 codominance model (HR=1.19, p-value= 6.3×10^{-2} , **Supplementary Fig. S6**).
256 Additionally, carriers of *HLA-A*30:01* (HR=1.75, p-value= 1.2×10^{-2}), and *HLA-*
257 *DPBI*09:01* (HR=3.36, p-value= 1.5×10^{-3}) were also associated with worse overall
258 survival compared to non-carriers (**Supplementary Fig. S7**).

259

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

261 More than 90% of PDAC harbor *KRAS* mutations. We found that the PDAC
262 protective allele *HLA-A*02:01* was associated with tumors harboring *KRAS*^{G12V} vs.
263 other mutations in PDAC, NSCLC, and colorectal cancers (OR=1.55, p-
264 value= 4.65×10^{-2}), and with *KRAS*^{G12D} mutations in gastric tumors (OR=7.27, p-
265 value= 6.42×10^{-3}) (**Supplementary Table S7**).

266 Stratified analysis revealed that *HLA-A*02:01* carriers with *KRAS*^{G12V} tumors
267 had significantly better survival (HR=0.19, p-value= 1.3×10^{-2} , **Fig. 4A**) than *HLA-*
268 *A*02:01* carriers with other *KRAS* mutations; importantly, survival did not differ for
269 patients with *KRAS*^{G12V} mutated tumors who did not carry the *HLA-A*02:01* allele
270 (HR=0.87, p-value= 8.1×10^{-1} , **Fig. 4B**). Conversely, carriers of *HLA-A*02:01* and
271 *KRAS*^{G12D} had significantly worse survival (HR=5.51, p-value= 1.5×10^{-3} , **Fig. 4C**) than
272 non-carriers (HR=1.13, p-value= 6.2×10^{-1} , **Fig. 4D**). Further stratified analyses
273 focusing on individuals carrying either the *KRAS*^{G12V} or *KRAS*^{G12D} mutations, along
274 with the presence or absence of the *HLA-A*02:01* allele, showed that the longer

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

282 The interaction model in the TCGA study showed that carriers of the *HLA*-
283 *A*02:01* allele who had a *KRAS*^{G12V} mutated tumor had a significantly lower risk of
284 death (HR=0.14, p-value=2.4x10⁻²) compared to patients with any of the other *KRAS*
285 mutations, adjusted for gender, age at diagnosis, stage, and Moffit tumor subtype. We
286 then performed a meta-analysis of the interaction between the *A*02:01* allele and
287 *KRAS*^{G12V} mutations, including the PAN-NGS and Caris populations (**Supplementary**
288 **Fig. S9**). This interaction held significant in early-stage tumors (HR=0.25, p-
289 value=1.7x10⁻², **Fig. 6A**) but not in advanced-stage (HR=0.93, p-value=8.7x10⁻¹, **Fig.**
290 **6B**).

291

292 **DISCUSSION**

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

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

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

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

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

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

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

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

375 (0.06), the latter AF being similar in most European countries with UK differing in the
376 AF by >50%. *HLA-A*26:01* has been reported to be associated with an increased risk
377 of non-Hodgkin lymphoma (OR=8.5, p-value= 2.6×10^{-2}) (43). Importantly, including
378 these alleles in a non-genetic-based cumulative risk score significantly altered the risk
379 of developing PDAC, either increasing or decreasing it substantially.

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

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

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

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

409

410 MATERIAL AND METHODS

411 *Ethics Statement*

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

415

416 *Study populations*

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

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

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

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

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

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

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

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

480

481 *HLA (or MHC) alleles assessment*

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

485 *PanGenEU study*: The imputation was done by *SNP2HLA* software (version v
486 1.0.3) (51) considering 5,790 SNPs within the *MHC* region with an *info* parameter
487 ≥ 0.9 . This *info* parameter ranges from 0 to 1 and assesses the quality/reliability for a
488 particular variant imputed through IMPUTE2. We focused on SNPs overlapping with
489 those from the Type 1 Diabetes Genetic Consortium (T1DGC) V1.0.3, the largest
490 European reference panel (52). This dataset includes 5,225 individuals with
491 genotyping data for 7,238 SNPs, 424 *HLA* alleles (126 1-field and 298 2-field alleles),
492 and 1,276 *HLA* amino acids. It allowed an accurate imputation of classical *HLA*
493 alleles and amino acids of class I (*HLA-A*, *-B*, and *-C*) and class II (*HLA-DPBI*, *-*
494 *DPAI*, *-DQBI*, *-DQAI*, and *-DRBI*) genes. We excluded *HLA* alleles that were not in
495 Hardy-Weinberg equilibrium (HWE) in controls (p -value $<10^{-6}$), calculated using the
496 *HardyWeinberg* R package V1.6.3 (53), those with a bad imputation quality
497 (*info* <0.3), as well as those with an allele frequency (AF) $<5\%$, ending up with 40
498 alleles of 1-field, and 39 alleles of 2-fields.

499 *UKB* information on 362 *HLA* alleles across eleven loci (*HLA-A*, *-B*, *-C*, *-*
500 *DPBI*, *-DPAI*, *-DQBI*, *-DQAI*, *-DRBI*, *-DRB2*, *-DRB3*, *-DRB4*, and *-DRB5*) was
501 available and obtained through the *HLA*IMP:02* program (54).

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

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

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

511 *Caris clinical trial:* HLA genotypes were obtained from Caris Life Sciences
512 reports for all patients. Genomic DNA was isolated from microdissected, formalin-
513 fixed paraffin-embedded tumor samples and analyzed using the Illumina NovaSeq
514 6000 sequencers. A hybrid pull-down panel of baits designed to enrich for more than
515 700 clinically relevant genes at high coverage and high read-depth was used, along
516 with another panel designed to enrich for additional >20,000 genes at a lower depth.
517 To assist with gene amplification/deletion measurements and other analyses, a 500Mb
518 SNP backbone panel (Agilent Technologies) was included. Analytical validation for
519 *HLA* genotyping showed >99% concordance with a validated comparator method.

520

521 *HLA haplotype assessment*

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

530

531 *Statistical analyses*

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

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

540

541 *HLA allele-based association analysis with PDAC risk.* Association analyses were
542 done using the *PLINK* V1.9 software (59). In the PanGenEU setting, we tested the
543 association between the *HLA* alleles and PDAC risk at the individual level by
544 estimating the Odds Ratios (OR) and 95% Confidence Intervals (CI) through logistic
545 regression models including age, gender, the European region, and the first five
546 principal components (PCAs) to correct for population stratification. We considered
547 the three Modes of Inheritance (MoI) for 424 PanGenEU *HLA* class I and II alleles
548 with AF >0.05. In UKB, we included the I- and II-fields *HLA* alleles with an AF
549 >0.1% (60,61). We performed logistic regression models adjusting for the same
550 covariates as in the PanGenEU models, plus the type of genotyping array and center of
551 inclusion.

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

556

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

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

566

567 *Haplotype association analysis.* We used the *haplo.stats* R package (version 1.7.9) to
568 perform the statistical association analysis of the PanGenEU and UKB *HLA*
569 haplotypes with PDAC risk. We employed a logistic regression model adjusted for the
570 same covariates as in the previous analyses. We considered the three MoIs in each
571 study population. Multiple testing correction was applied for each gene combination
572 separately after discarding the low prevalence *HLA* haplotypes (i.e., frequencies either
573 <5% or <0.1% in PanGenEU and UKB, respectively).

574

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

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

$$y = X\beta + Zh_I + Zh_{II} + \varepsilon$$

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

583 normal deviates with null mean and variance σ_ε^2 that includes model misspecification
584 and environmental effects not considered in this model. The two genetic and residual
585 variances were assumed to follow the distributions $h_I \sim N(0, H_I \sigma_{HLA-I}^2)$,
586 $N(0, H_{II} \sigma_{HLA-II}^2)$ and $\varepsilon \sim N(0, I \sigma_\varepsilon^2)$, where H_I and H_{II} correspond to the *HLA* class I and
587 *HLA* class II alleles-derived relationship matrices between individuals, σ_{HLA-I}^2 and
588 σ_{HLA-II}^2 are the genetic variances explained by *HLA* class I and II alleles, \mathbf{I} is an
589 identity matrix, and σ_ε^2 corresponds to the residual variance. The prior distribution for
590 the genetic and residual variances were inverted chi-square distributions.

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

596

597 *HLA allele-based survival analysis with PDAC risk.* We assessed patient survival
598 based on the presence or absence of the *HLA* alleles, with a strong prior on *HLA*-
599 *A*02:01*. To this end, we used the survival data of 425 PDAC cases from PanGenEU,
600 506 from UKB, 101 from TCGA (98 when we account for the stage of the tumor), and
601 145 from the Caris clinical trial. First, we assessed the overlap of the *HLA* class I and
602 II alleles between the three study populations. We computed Hazard Ratios (HR) and
603 95%CI using Cox proportional-hazards regression models for each population
604 separately under (co-)dominance MoIs. Adjustment covariates in PanGenEU models
605 were sex, age at diagnosis, center, tumor stage, and the first PCAs; for UK Biobank,
606 we included sex, age at diagnosis, genotyping array type, recruitment center and the
607 first five PCAs; for the TCGA and Caris models, we considered sex, age at diagnosis,

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

610

611 *Association analysis between HLA-A*02:01 and KRAS common mutations.* Based on
612 previous reports pointing to a joint effect of *HLA-A*02:01* and *KRAS* mutations
613 (31,36), we further explored the interaction between these biomarkers using TCGA
614 data. We first assessed the association between *HLA-A*02:01* and well-established
615 *KRAS* missense mutations at codons 12 (G12A, G12C, G12D, G12F, G12R, G12S,
616 G12V), 13 (G13C, G13D), and 61 (Q61H, Q61K, Q61L, Q61R) using logistic
617 regression models in tumors with high proportion of *KRAS* mutations (PDAC,
618 colorectal, stomach, and NSCLC) (**Supplementary Table S10**). As the reference
619 category, we used the non-G12V mutations for PDAC and colorectal cancer and the
620 non-G12C for NSCLC. We initially assessed the association within each specific
621 cancer type individually. Subsequently, we pooled the data from all cancer types into a
622 single model, adjusting for cancer type. Finally, we used a Cox proportional-hazards
623 regression model to evaluate whether the interaction between *HLA-A*02:01* and the
624 specific *KRAS*^{G12V} and *KRAS*^{G12D} mutations was associated with overall survival in
625 PDAC. The same covariates mentioned in the previous paragraph and the Moffit
626 subtype of the tumor were employed to adjust these models. We used a Wilcoxon test
627 to evaluate whether the associations found in overall survival could be explained by
628 infiltrated immune components. Differences were considered significant at p-
629 value<0.05.

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

633 allele and the *KRAS*^{G12V} mutation in each population, adjusting for variables such as
634 sex, age at diagnosis, tumor stage, and treatment for Caris, and the same variables plus
635 center of inclusion for Erasmus MC. We then conducted a meta-analysis using a
636 random-effects model. Finally, we examined the effect of the interaction within each
637 population by stratifying patients based on tumor stage (stages I-II and stages III-IV)
638 and conducted separate meta-analyses for these subgroups.
639

640 **DECLARATIONS**

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

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

645 **Funding:** The work was partially supported by Pancreatic Cancer Collective (PCC):
646 Lustgarten Foundation & Stand-Up to Cancer (SU2C #6179); Fondo de
647 Investigaciones Sanitarias (FIS), Instituto de Salud Carlos III, Spain (#PI061614,
648 #PI11/01542, #PI0902102, #PI12/01635, #PI12/00815, #PI15/01573, #PI18/01347,
649 #PI21/00495); Red Temática de Investigación Cooperativa en Cáncer, Spain
650 (#RD12/0036/0034, #RD12/0036/0050, #RD12/0036/0073); EU-6FP Integrated
651 Project (#018771-MOLDIAG-PACA), EU-FP7-HEALTH (#259737-CANCERALIA,
652 #256974-EPC-TM-Net); Associazione Italiana Ricerca sul Cancro (IG 26343);
653 Fondazione Cariverona: Oncology Biobank Project “Antonio Schiavi” (prot.
654 203885/2017); Italian Ministry of Health through Fondazione Italiana Malattie
655 Pancreas (FIMP_CUP J37G22000230001); Cancer Focus Northern Ireland and
656 Department for Employment and Learning; and ALF (#SLL20130022), Sweden,
657 Cancer Research UK (C7690/A26881) and Pancreatic Cancer UK.

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

664 revising it: AL, RR, FXR, ELM, NM. Approval of the submitted version (and any
665 substantially modified version that involves the author's contribution to the study):
666 ALL AUTHORS. Agreement on both to be personally accountable for the author's
667 own contributions and to ensure that questions related to the accuracy or integrity of
668 any part of the work, even ones in which the author was not personally involved, are
669 appropriately investigated, resolved and the resolution documented in the literature:
670 ALL AUTHORS.

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

685

686 **REFERENCES**

- 687 1. Connor AA, Gallinger S. Pancreatic cancer evolution and heterogeneity: integrating
688 omics and clinical data. *Nat Rev Cancer*. Nature Publishing Group; 2022;22:131–42.
- 689 2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: A Cancer Journal for*
690 *Clinicians*. 2020;70:7–30.
- 691 3. Raimondi S, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: an
692 overview. *Nat Rev Gastroenterol Hepatol*. Nature Publishing Group; 2009;6:699–708.
- 693 4. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all
694 cancers combined and survival adjusted for age and sex for each cancer in England and
695 Wales, 1971-2011: a population-based study. *Lancet*. 2015;385:1206–18.
- 696 5. Smith BD, Smith GL, Hurria A, Hortobagyi GN, Buchholz TA. Future of Cancer
697 Incidence in the United States: Burdens Upon an Aging, Changing Nation. *JCO*. Wolters
698 Kluwer; 2009;27:2758–65.
- 699 6. Singhi AD, Koay EJ, Chari ST, Maitra A. Early Detection of Pancreatic Cancer:
700 Opportunities and Challenges. *Gastroenterology*. 2019;156:2024–40.
- 701 7. Ho WJ, Jaffee EM, Zheng L. The tumour microenvironment in pancreatic cancer —
702 clinical challenges and opportunities. *Nat Rev Clin Oncol*. Nature Publishing Group;
703 2020;17:527–40.
- 704 8. Gobbi PG, Bergonzi M, Comelli M, Villano L, Pozzoli D, Vanoli A, et al. The
705 prognostic role of time to diagnosis and presenting symptoms in patients with pancreatic
706 cancer. *Cancer Epidemiology*. 2013;37:186–90.
- 707 9. Gomez-Rubio P, Zock J-P, Rava M, Marquez M, Sharp L, Hidalgo M, et al. Reduced
708 risk of pancreatic cancer associated with asthma and nasal allergies. *Gut*. BMJ
709 Publishing Group; 2017;66:314–22.
- 710 10. Gomez-Rubio P, Piñero J, Molina-Montes E, Gutiérrez-Sacristán A, Marquez M, Rava
711 M, et al. Pancreatic cancer and autoimmune diseases: An association sustained by
712 computational and epidemiological case–control approaches. *International Journal of*
713 *Cancer*. 2019;144:1540–9.
- 714 11. López de Maturana E, Rodríguez JA, Alonso L, Lao O, Molina-Montes E, Martín-
715 Antoniano IA, et al. A multilayered post-GWAS assessment on genetic susceptibility to
716 pancreatic cancer. *Genome Med*. 2021;13:15.
- 717 12. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association
718 analysis identifies 65 new breast cancer risk loci. *Nature*. Nature Publishing Group;
719 2017;551:92–4.
- 720 13. Lan Q, Hsiung CA, Matsuo K, Hong Y-C, Seow A, Wang Z, et al. Genome-wide
721 association analysis identifies new lung cancer susceptibility loci in never-smoking
722 women in Asia. *Nat Genet*. Nature Publishing Group; 2012;44:1330–5.
- 723 14. Shiraishi K, Okada Y, Takahashi A, Kamatani Y, Momozawa Y, Ashikawa K, et al.
724 Association of variations in HLA class II and other loci with susceptibility to EGFR-
725 mutated lung adenocarcinoma. *Nat Commun*. Nature Publishing Group; 2016;7:12451.

- 726 15. Aterido A, Cañete JD, Tornero J, Ferrándiz C, Pinto JA, Gratacós J, et al. Genetic
727 variation at the glycosaminoglycan metabolism pathway contributes to the risk of
728 psoriatic arthritis but not psoriasis. *Annals of the Rheumatic Diseases*. BMJ Publishing
729 Group Ltd; 2019;78:355–64.
- 730 16. Merkel PA, Xie G, Monach PA, Ji X, Ciavatta DJ, Byun J, et al. Identification of
731 Functional and Expression Polymorphisms Associated With Risk for Antineutrophil
732 Cytoplasmic Autoantibody–Associated Vasculitis. *Arthritis & Rheumatology*.
733 2017;69:1054–66.
- 734 17. Mendoza-Ramírez P, López-Olaiz MA, Morales-Fernández AL, Flores-Echiveste MI, de
735 Jesus Casillas-Navarro A, Pérez-Rodríguez MA, et al. Class I MHC Polymorphisms
736 Associated with Type 2 Diabetes in the Mexican Population. *Genes (Basel)*.
737 2022;13:772.
- 738 18. Jacobi T, Massier L, Klöting N, Horn K, Schuch A, Ahnert P, et al. HLA Class II Allele
739 Analyses Implicate Common Genetic Components in Type 1 and Non-Insulin-Treated
740 Type 2 Diabetes. *J Clin Endocrinol Metab*. 2020;105:dga027.
- 741 19. Purohit A, Purohit BM, Mani A, Bhambal A. Genetic Implications of HLA-DR and
742 HLA-DQ Genotype on Tobacco Smoking and Oral Submucous Fibrosis. *Oral Health*
743 *Prev Dent*. 2020;18:455–60.
- 744 20. Movahedi M, Moin M, Gharagozlou M, Aghamohammadi A, Dianat S, Moradi B, et al.
745 Association of HLA class II alleles with childhood asthma and Total IgE levels. *Iran J*
746 *Allergy Asthma Immunol*. 2008;7:215–20.
- 747 21. Ribeyre C, Carlini F, René C, Jordier F, Picard C, Chiaroni J, et al. HLA-G Haplotypes
748 Are Differentially Associated with Asthmatic Features. *Frontiers in Immunology*
749 [Internet]. 2018 [cited 2022 Sep 29];9. Available from:
750 <https://www.frontiersin.org/articles/10.3389/fimmu.2018.00278>
- 751 22. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide
752 association and HLA fine-mapping studies identify risk loci and genetic pathways
753 underlying allergic rhinitis. *Nat Genet*. Nature Publishing Group; 2018;50:1072–80.
- 754 23. Filip I, Wang A, Kravets O, Orenbuch R, Zhao J, Perea-Chamblee TE, et al.
755 Pervasiveness of HLA allele-specific expression loss across tumor types. *Genome Med*.
756 2023;15:8.
- 757 24. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, et al. MHC-
758 I Genotype Restricts the Oncogenic Mutational Landscape. *Cell*. 2017;171:1272-
759 1283.e15.
- 760 25. Kloetzel PM. Generation of major histocompatibility complex class I antigens:
761 functional interplay between proteasomes and TPPII. *Nat Immunol*. Nature Publishing
762 Group; 2004;5:661–9.
- 763 26. Unanue ER. From antigen processing to peptide-MHC binding. *Nat Immunol*. Nature
764 Publishing Group; 2006;7:1277–9.
- 765 27. Stenzel A, Lu T, Koch WA, Hampe J, Guenther SM, De La Vega FM, et al. Patterns of
766 linkage disequilibrium in the MHC region on human chromosome 6p. *Hum Genet*.
767 2004;114:377–85.

- 768 28. Gonzalez-Galarza FF, McCabe A, Melo dos Santos EJ, Takeshita L, Ghattaoraya G,
769 Jones AR, et al. Allele Frequency Net Database. In: Boegel S, editor. HLA Typing:
770 Methods and Protocols [Internet]. New York, NY: Springer; 2018 [cited 2022 Sep 12].
771 page 49–62. Available from: https://doi.org/10.1007/978-1-4939-8546-3_4
- 772 29. Talwar JV, Laub D, Pagadala MS, Castro A, Lewis M, Luebeck GE, et al. Autoimmune
773 alleles at the major histocompatibility locus modify melanoma susceptibility. The
774 American Journal of Human Genetics. Elsevier; 2023;110:1138–61.
- 775 30. Dyckhoff G, Herold-Mende C, Scherer S, Plinkert PK, Warta R. Human Leucocyte
776 Antigens as Prognostic Markers in Head and Neck Squamous Cell Carcinoma. Cancers.
777 Multidisciplinary Digital Publishing Institute; 2022;14:3828.
- 778 31. Rive CM, Yung E, Hughes CS, Brown SD, Sharma G, Dreolini L, et al. Recombinant T
779 cell receptors specific for HLA-A*02:01-restricted neoepitopes containing KRAS codon
780 12 hotspot mutations [Internet]. bioRxiv; 2020 [cited 2022 Sep 13]. page
781 2020.06.15.149021. Available from:
782 <https://www.biorxiv.org/content/10.1101/2020.06.15.149021v1>
- 783 32. Shenzhen BinDeBio Ltd. EGFRvIII/DR5/NY-ESO-1/Mesothelin CAR-T/TCR-T Cells
784 Immunotherapy for Solid Malignancies [Internet]. clinicaltrials.gov; 2021 Feb. Report
785 No.: NCT03941626. Available from: <https://clinicaltrials.gov/ct2/show/NCT03941626>
- 786 33. Yamamoto Y. Investigator-initiated Clinical Trial (Phase II) of Cancer Vaccine “DSP-
787 7888” for Acute Myeloid Leukemia Patients. [Internet]. clinicaltrials.gov; 2021 Feb.
788 Report No.: NCT04747002. Available from:
789 <https://clinicaltrials.gov/ct2/show/NCT04747002>
- 790 34. PDC*line Pharma SAS. An Open-label, Dose-escalation, Phase I/II Study to Assess the
791 Safety, the Tolerability, the Immunogenicity and the Preliminary Clinical Activity of the
792 Therapeutic Cancer Vaccine, PDC*lung01, Associated or Not With Anti-PD-1
793 Treatment in Patients With Non-small-cell Lung Cancer (NSCLC) [Internet].
794 clinicaltrials.gov; 2022 Jul. Report No.: NCT03970746. Available from:
795 <https://clinicaltrials.gov/ct2/show/NCT03970746>
- 796 35. Greenwich LifeSciences, Inc. A Randomized, Multicenter, Placebo-controlled, Phase 3
797 Study to Evaluate the Efficacy and Safety of HER2/Neu Peptide GLSI-100 (GP2 + GM-
798 CSF) in HER2/Neu Positive Subjects With Residual Disease or High-Risk PCR After
799 Both Neoadjuvant and Postoperative Adjuvant Trastuzumab-based Therapy
800 (FLAMINGO-01) [Internet]. clinicaltrials.gov; 2022 Nov. Report No.: NCT05232916.
801 Available from: <https://clinicaltrials.gov/ct2/show/NCT05232916>
- 802 36. Mishto M, Mansurkhodzhaev A, Ying G, Bitra A, Cordfunke RA, Henze S, et al. An in
803 silico—in vitro Pipeline Identifying an HLA-A*02:01+ KRAS G12V+ Spliced Epitope
804 Candidate for a Broad Tumor-Immune Response in Cancer Patients. Frontiers in
805 Immunology [Internet]. 2019 [cited 2022 Sep 13];10. Available from:
806 <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02572>
- 807 37. Fenton SE, Saleiro D, Platanius LC. Type I and II Interferons in the Anti-Tumor
808 Immune Response. Cancers (Basel). 2021;13:1037.
- 809 38. Espinet E, Gu Z, Imbusch CD, Giese NA, Büscher M, Safavi M, et al. Aggressive
810 PDACs Show Hypomethylation of Repetitive Elements and the Execution of an Intrinsic
811 IFN Program Linked to a Ductal Cell of Origin. Cancer Discovery. 2021;11:638–59.

- 812 39. Till JE, McDaniel L, Chang C, Long Q, Pfeiffer SM, Lyman JP, et al. Circulating KRAS
813 G12D but not G12V is associated with survival in metastatic pancreatic ductal
814 adenocarcinoma. *Nat Commun. Nature Publishing Group*; 2024;15:5763.
- 815 40. Rivera Pirela S, Echeverría M, Márquez G, Carrillo Z, Parra Y, Salcedo P, et al. HLA-
816 DPA1*/DPB1* en asociación con leucemias linfoides agudas y leucemias mieloides
817 crónicas en mestizos venezolanos. *Inmunología*. 2014;33:81–6.
- 818 41. Zawadzka-Starczewska K, Tymoniuk B, Stasiak B, Lewiński A, Stasiak M. Actual
819 Associations between HLA Haplotype and Graves' Disease Development. *J Clin Med*.
820 2022;11:2492.
- 821 42. Deng H, Yan X, Yuan L. Human genetic basis of coronavirus disease 2019. *Signal*
822 *Transduct Target Ther*. 2021;6:344.
- 823 43. Sayad A, Akbari MT, Mehdizadeh M, Taheri M, Hajifathali A. HLA-A*26 and
824 susceptibility of iranian patients with non-Hodgkin lymphoma. *Iran J Immunol*.
825 2014;11:269–74.
- 826 44. Molina-Montes E, Gomez-Rubio P, Márquez M, Rava M, Löhr M, Michalski CW, et al.
827 Risk of pancreatic cancer associated with family history of cancer and other medical
828 conditions by accounting for smoking among relatives. *International Journal of*
829 *Epidemiology*. 2018;47:473–83.
- 830 45. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The
831 OncoArray Consortium: A Network for Understanding the Genetic Architecture of
832 Common Cancers. *Cancer Epidemiol Biomarkers Prev*. 2017;26:126–35.
- 833 46. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation
834 Method for the Next Generation of Genome-Wide Association Studies. *PLOS Genetics*.
835 *Public Library of Science*; 2009;5:e1000529.
- 836 47. Durbin RM, Altshuler D, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. A
837 map of human genome variation from population-scale sequencing. *Nature. Nature*
838 *Publishing Group*; 2010;467:1061–73.
- 839 48. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an
840 open access resource for identifying the causes of a wide range of complex diseases of
841 middle and old age. *PLoS Med*. 2015;12:e1001779.
- 842 49. Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The
843 Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat Genet*. 2013;45:1113–20.
- 844 50. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType:
845 precision HLA typing from next-generation sequencing data. *Bioinformatics*.
846 2014;30:3310–6.
- 847 51. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, et al. Imputing
848 Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLOS ONE. Public Library*
849 *of Science*; 2013;8:e64683.
- 850 52. Rich SS, Concannon P, Erlich H, Julier C, Morahan G, Nerup J, et al. The Type 1
851 Diabetes Genetics Consortium. *Annals of the New York Academy of Sciences*.
852 2006;1079:1–8.

- 853 53. Graffelman J. Exploring Diallelic Genetic Markers: The HardyWeinberg Package.
854 Journal of Statistical Software. 2015;64:1–23.
- 855 54. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population
856 classical HLA type imputation. PLoS Comput Biol. 2013;9:e1002877.
- 857 55. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al.
858 Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes.
859 Nat Biotechnol. Nature Publishing Group; 2015;33:1152–8.
- 860 56. Xie C, Yeo ZX, Wong M, Piper J, Long T, Kirkness EF, et al. Fast and accurate HLA
861 typing from short-read next-generation sequence data with xHLA. Proceedings of the
862 National Academy of Sciences. Proceedings of the National Academy of Sciences;
863 2017;114:8059–64.
- 864 57. JP S, DJ S. haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates
865 when Linkage Phase is Ambiguous [Internet]. 2022 [cited 2022 Sep 12]. Available from:
866 <https://CRAN.R-project.org/package=haplo.stats>
- 867 58. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale
868 sequence data. The American Journal of Human Genetics. 2021;108:1880–90.
- 869 59. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK:
870 a tool set for whole-genome association and population-based linkage analyses. Am J
871 Hum Genet. 2007;81:559–75.
- 872 60. Venkataraman GR, Olivieri JE, DeBoever C, Tanigawa Y, Justesen JM, Dilthey A, et al.
873 Pervasive additive and non-additive effects within the HLA region contribute to disease
874 risk in the UK Biobank [Internet]. bioRxiv; 2020 [cited 2022 Sep 12]. page
875 2020.05.28.119669. Available from:
876 <https://www.biorxiv.org/content/10.1101/2020.05.28.119669v2>
- 877 61. Krebs K, Bovijn J, Zheng N, Lepamets M, Censin JC, Jürgenson T, et al. Genome-wide
878 Study Identifies Association between HLA-B*55:01 and Self-Reported Penicillin
879 Allergy. Am J Hum Genet. 2020;107:612–21.
- 880 62. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of
881 Statistical Software. 2010;36:1–48.
- 882 63. Wahba G. Spline models for observational data. Philadelphia, Pa: Society for Industrial
883 and Applied Mathematics; 1990.
- 884 64. Granato I, Fritsche-Neto R. snpReady: Preparing Genotypic Datasets in Order to Run
885 Genomic Analysis [Internet]. 2018 [cited 2022 Sep 12]. Available from:
886 <https://CRAN.R-project.org/package=snpReady>
- 887

888 **Table 1.** Cumulative PDAC risk estimates (OR: odds ratio and CI: confidence
 889 interval) in the extreme scenarios according to the presence/absence of known risk
 890 factors (Type II diabetes, smoking, and family history of pancreatic cancer, FHPC) or
 891 protective factors (asthma and nasal allergies) in addition to the significant *HLA*
 892 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 + <i>HLA-B*39</i>				1.17 x 10 ⁻¹
1 allele	3.85	1.20	12.32	
2 alleles	5.28	1.65	16.92	
High-risk scenario + <i>HLA-DPBI*04</i>				1.33 x 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 + <i>HLA-A*02:01</i>				1.18 x 10 ⁻²
1 allele	0.32	0.1	1.03	
2 alleles	0.27	0.08	0.88	
Low-risk scenario + <i>HLA-B*49</i>				2.4 x 10 ⁻²
1 allele	0.19	0.03	1.07	
2 alleles	0.1	0.02	0.55	

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

895 Scenario b: asthmatic or allergic individuals who did not smoke, did not have diabetes, nor family history of cancer.

896
 897

898 **MAIN FIGURE LEGENDS**

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

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

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

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

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

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

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

924

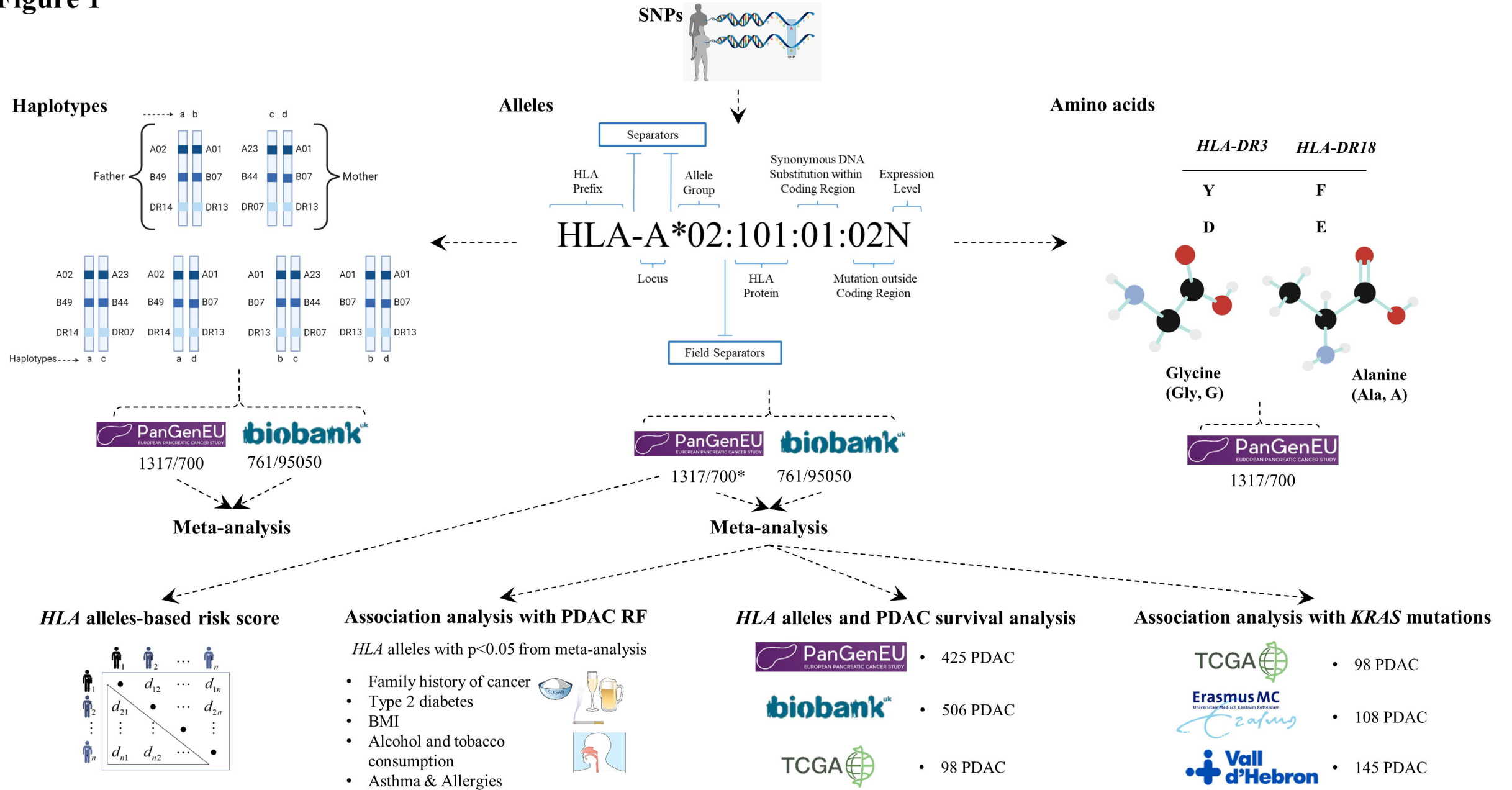
925 Additional Files

- 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 TCGA 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

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-DPBI*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-DPBI*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.

Figure 1



*Cases/controls

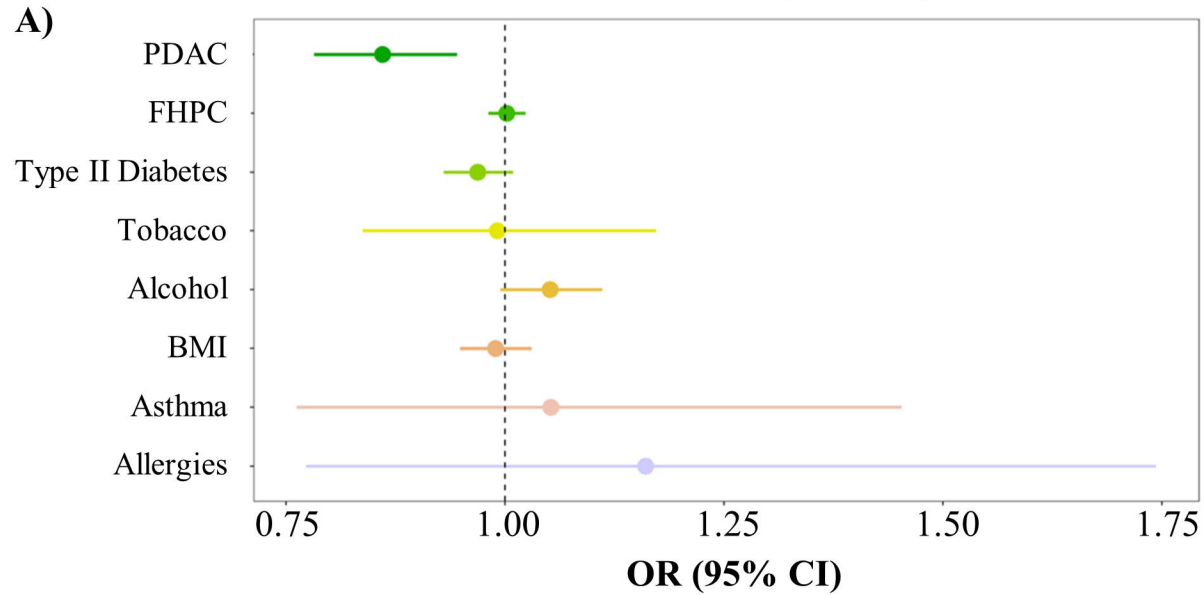
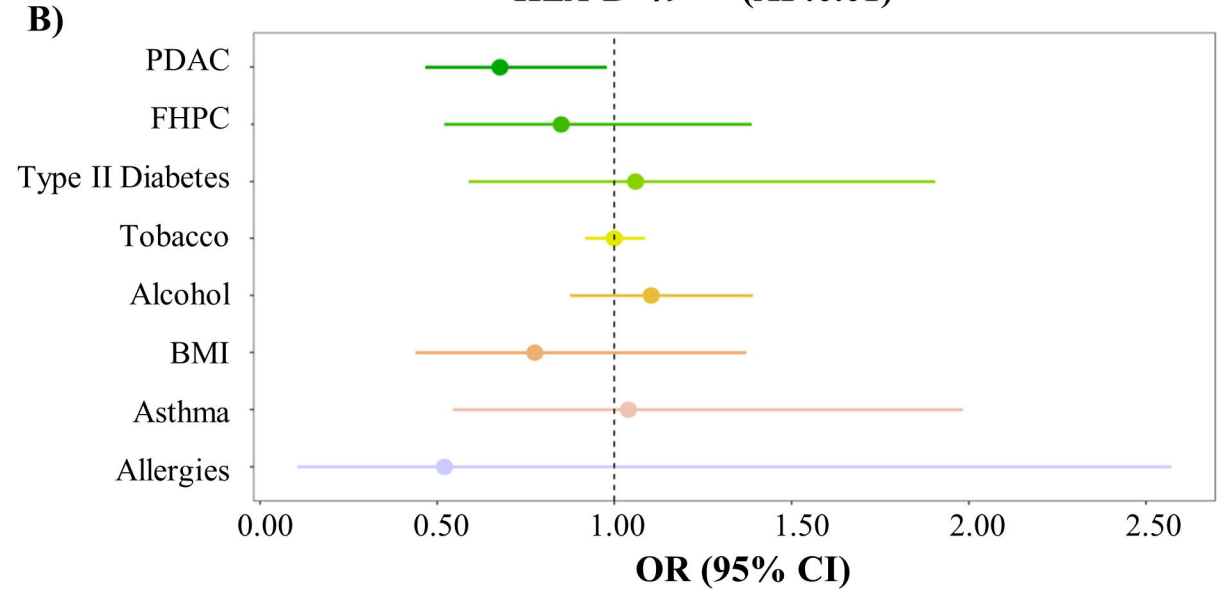
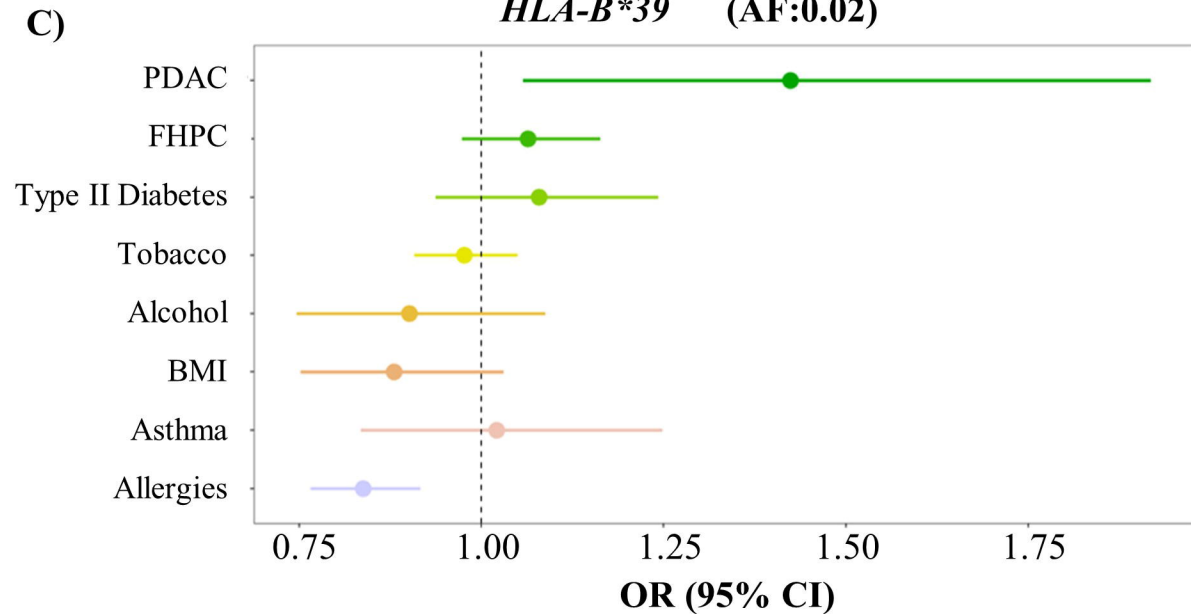
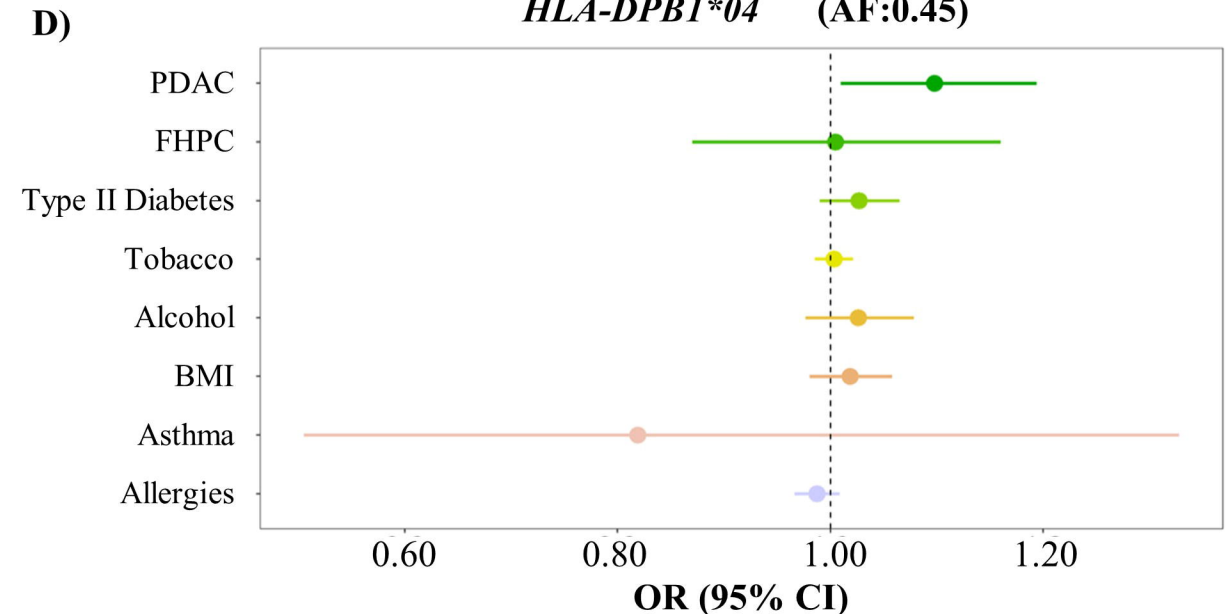
Figure 2***HLA-A*02:01* (AF:0.27)*****HLA-B*49* (AF:0.01)*****HLA-B*39* (AF:0.02)*****HLA-DPB1*04* (AF:0.45)**

Figure 3

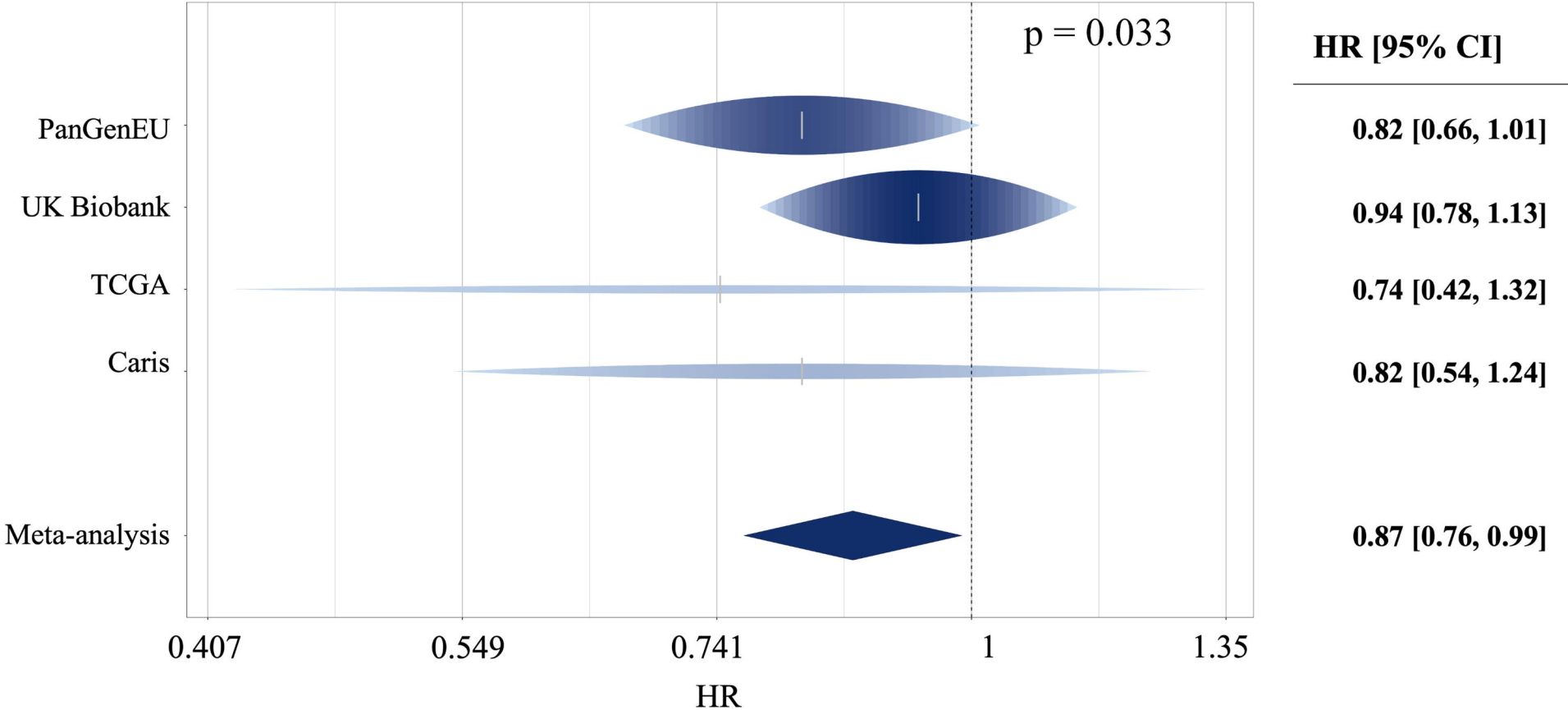
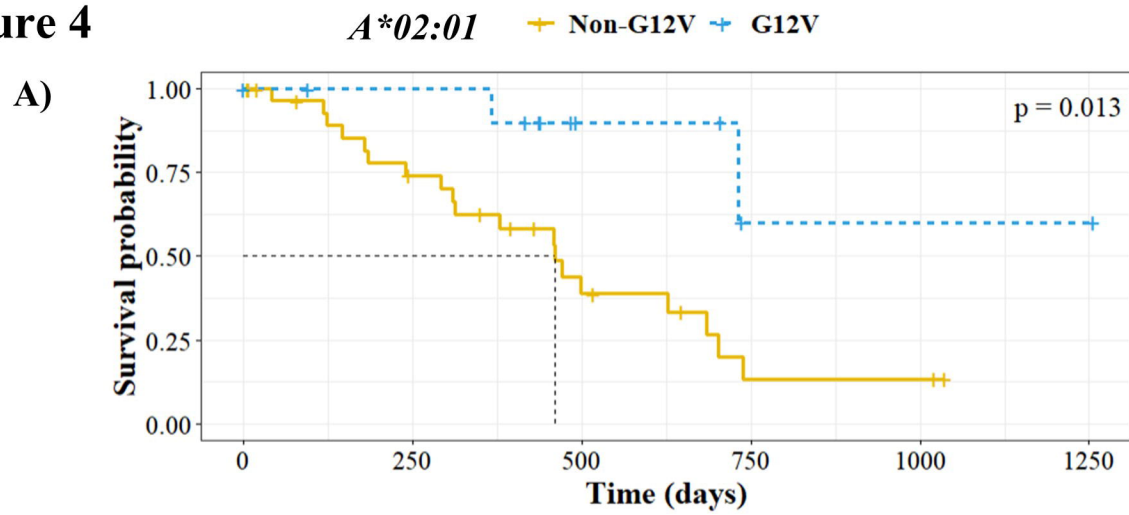
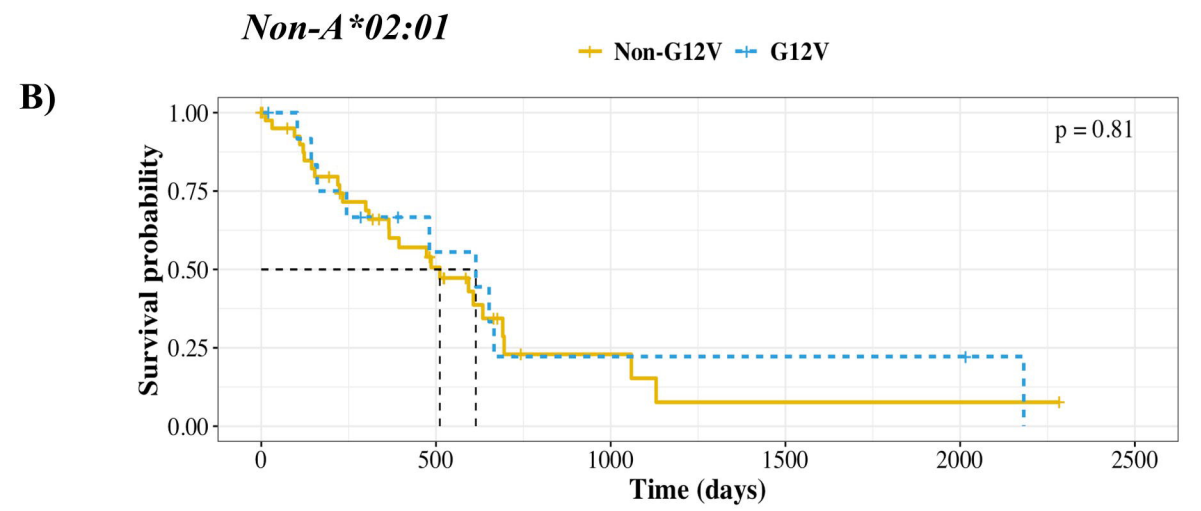


Figure 4



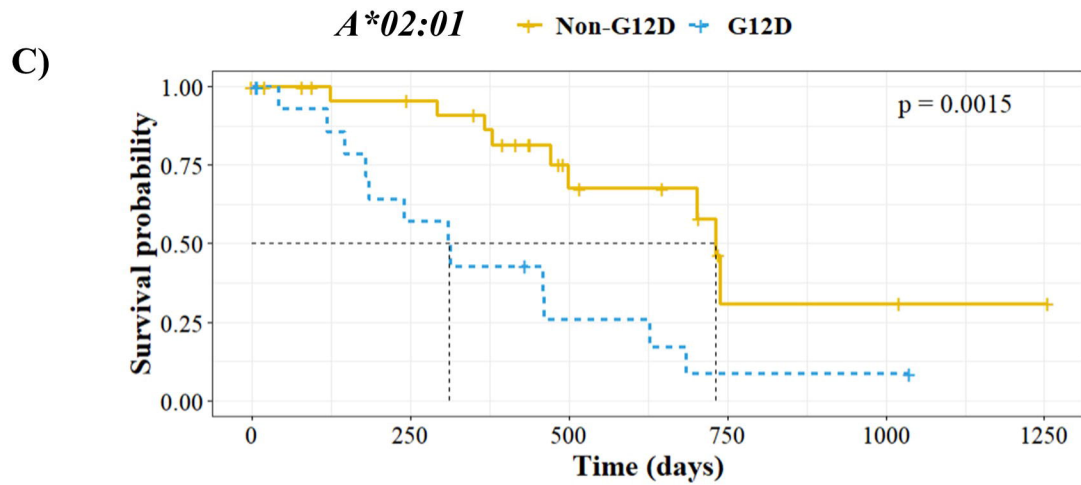
Number at risk

Non-G12V	31	19	8	2	2	0
G12V	12	10	4	1	1	1
	0	250	500	750	1000	1250



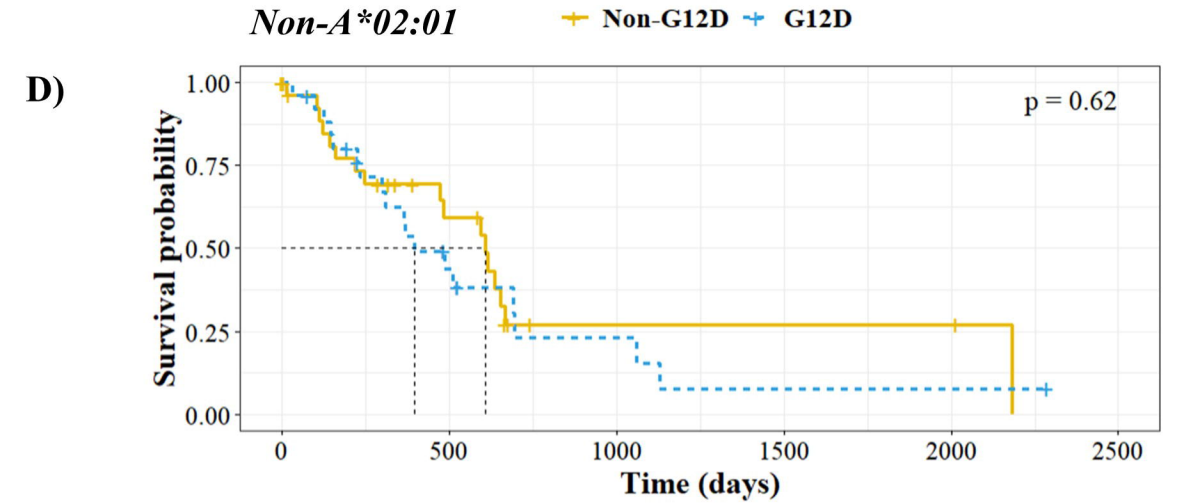
Number at risk

Non-G12V	42	15	3	1	1	0
G12V	13	5	2	2	2	0
	0	500	1000	1500	2000	2500



Number at risk

Non-G12D	27	21	9	2	2	1
G12D	16	8	3	1	1	0
	0	250	500	750	1000	1250

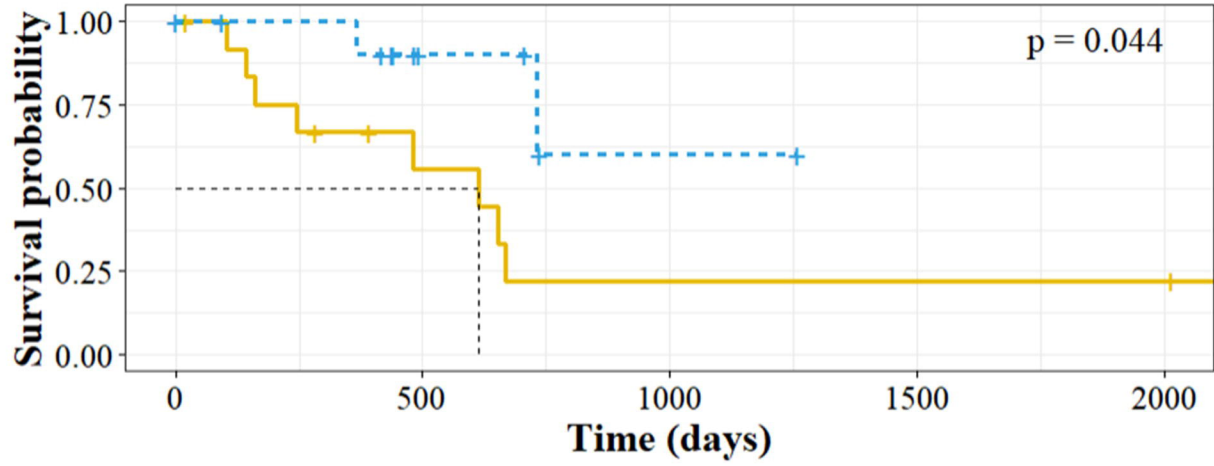


Number at risk

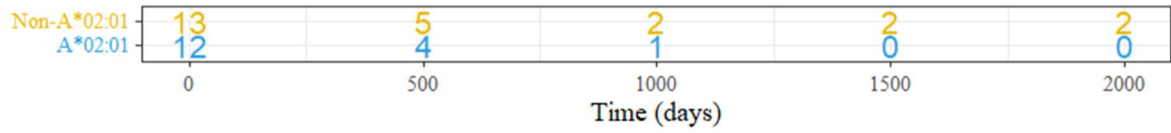
Non-G12D	29	12	2	2	2	0
G12D	26	8	3	1	1	0
	0	500	1000	1500	2000	2500

Figure 5

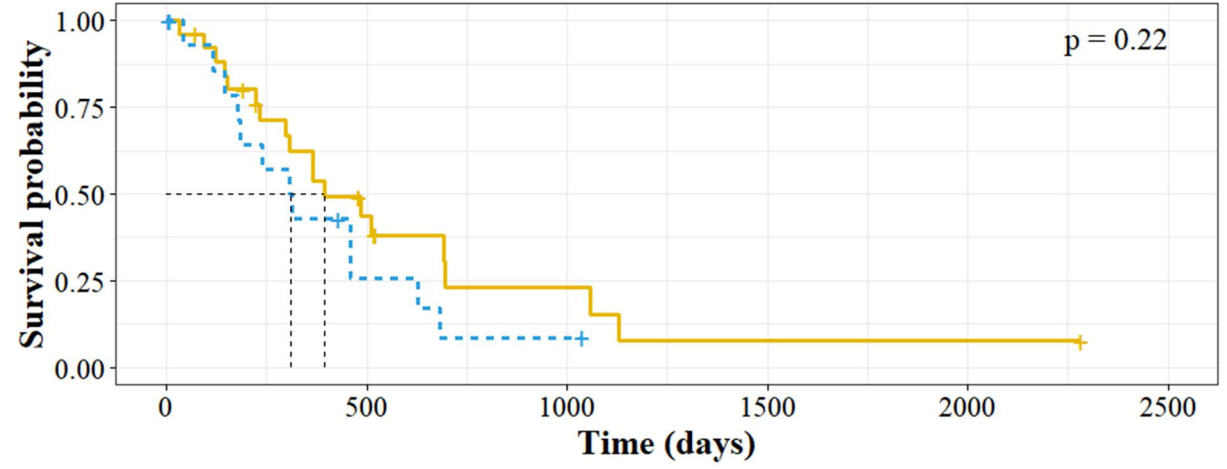
G12V + *Non-A*02:01* + *A*02:01*



Number at risk



Non-G12V + *Non-A*02:01* + *A*02:01*



Number at risk

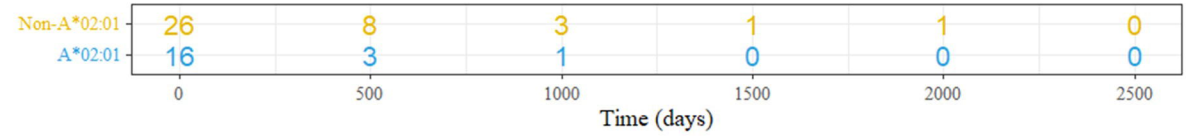


Figure 6

