Original Article
Association of HLA-A, B, and C alleles and cancer susceptibility in 179 solid malignancies

Miguel Carabaño[1*], Tao Liu[2*], Abraham Liu[2], Jim ChunHao Lee[1], Liang Cheng[1], Li-Juan Wang[1], Chiung-Kuei Huang[3], Shaolei Lu[1]

[1]Department of Pathology and Laboratory Medicine, Alpert Medical School of Brown University, Providence, RI 02903, USA; [2]Department of Biostatistics, Brown University, Providence, RI 02912, USA; [3]Department of Pathology and Laboratory Medicine, Tulane University School of Medicine, New Orleans, LA 70112, USA. *Equal contributors.

Abstract: Background: The major histocompatibility complex (MHC) genes are known to be capable of influencing the susceptibility of many cancers. All mammalian cells, including cancer cells, express MHC class I molecules consisting of human leukocyte antigens (HLA) A, B, and C. The tumor susceptibility of HLA-A, B, and C alleles has not been studied extensively in solid tumors. Methods: HLA-A, B, and C genotypes of 179 solid tumors were collected from Caris Comprehensive Tumor Profiling reports, including 45 GU, 44 GI, 28 pancreaticobiliary, 21 thoracic, 15 breast, 13 Gyn, among others. The tumors were mainly from Caucasians (82%). The HLA allele frequencies in the tumors were compared to those of respective ethnic populations in the US National Marrow Donor Program (NMDP) database. Fisher's exact tests were performed, adjusted *P* values were calculated using Benjamini-Hochberg's method for false discovery rate (FDR), and Prevalence ratios (PRs) were calculated to quantify associations. Results: Twenty-one alleles were not listed in the NMDP. Among them, A*11:303 alone was present in 11 carcinomas, and B*08:222 was seen in 4 tumors. Among the alleles listed in the NMDP, C*08:02, B*14:02, A*03:02, and B*44:06 were significantly associated with tumors in Caucasian Americans (PR: 2.50-170), while B*44:02 appeared protective (PR: 0.36). Alleles with less significant associations were listed. Conclusions: From the HLA-A, B, and C data of the 179 tumors, we identified several susceptible alleles and one protective allele. Of interest, 21 alleles were not listed in the NMDP. The limited cases prevented our analysis from identifying cancer-susceptible alleles in other races.

Keywords: HLA class I, cancer, biomarker

Introduction

Major Histocompatibility Complex (MHC) proteins present antigenic peptides on the cell surface. MHC class I proteins are present in all somatic cells, including cancer and mesenchymal cells, while MHC class II proteins are limited to the hematopoietic cells. MHC class I mediated antigen presentation is critical for CD8+ T-cell responses. Class I genes are responsible for encoding the MHC class I proteins and are subdivided into HLA-A, B, and C. HLA genes are highly polymorphic and expressed in a co-dominant manner. This polymorphic characteristic provides the immune system with an advantage in diverse selectivity. It enables HLA genes to play essential roles in regulating the

immune response to transplantation, infectious diseases, autoimmune diseases, and tumors [1].

Cancer is currently one of the leading causes of death globally [2]. It is proven that MHC proteins play essential roles in tumor immunity. Class I MHC antigens are pivotal for cytotoxic T lymphocyte recognition of cancer and mesenchymal tumor cells and subsequently in immunotherapy response [3, 4]. It has been shown that structural and functional alterations in HLA can result in mechanisms that cause tumor cells to escape immune response, such as loss of expression of tumor antigens, lack of co-stimulatory molecules, production of immuno-suppressive cytokines, and an immune nega-

tive selection process for mutated cells are some of the possible mechanisms that cause tumor cells to escape immune response [5-7]. One would expect tumor cells to gain better survival by eliminating the HLA expression altogether. However, natural killer (NK) cells recognize cells with low HLA Class I proteins and subsequently lyze them. As the ligands of the killer Ig-like receptors (KIRs) on the NK cells, HLA Class I proteins need to be at a certain level binding to KIRs to deactivate the NK cells, which is known as the "missing self" hypothesis [8, 9]. Many tumor cells are known to evade NK cell-induced tumor surveillance by influencing NK cell activities, and modulating NK cell functions have been explored to enhance the efficacy of immunotherapy [10].

The association between HLA polymorphism and cancer susceptibility has been an intriguing question for decades. It was initially believed that HLA-dependent immunity mainly influences the susceptibility of virally associated malignancies [11]. More studies later showed that the HLA Class I genotype [12], HLA Class II allele [13], and HLA-G [14, 15] were also associated with oncogenic mutations or susceptibility of certain cancers, while recent studies suggest certain HLA Class I alleles might be predictor of response to immunotherapy [16, 17].

Given the central role of HLA molecules in anti-tumor immunity, we hypothesize that polymorphisms at the HLA loci could differentially affect the tumor genesis and progression. This study aims to determine HLA Class I allele representation among patients presenting with different types of malignant tumors in comparison with the general population. The susceptibility of HLA-A, B, and C alleles has not been extensively studied in solid tumors.

Methods

Patient cohort

One hundred seventy-nine cases of malignant solid tumors were collected from Lifespan Pathology archives (Rhode Island Hospital and The Miriam Hospital) from 2012-2022. The series consisted of all the cases that were submitted to Caris for Comprehensive Cancer Profiling (carislifesciences.com) for HLA profiling, including 45 genitourinary (GU), 44 gastro-

intestinal (GI), 28 Pancreatobiliary, 21 thoracic, 15 breast, 13 gynecologic (GYN) cancers as well as 5 sarcomas, 4 central nerve system (CNS) malignancies, and 4 head and neck (H&N) cancers. The tissue sources included 95 primary and 84 metastatic tumors. Most metastatic sites were liver (31%), lymph nodes (22%), bone (10%), and lung (8%). Caris report provided 4-digit HLA-A, B, and C genotypes for each patient. The racial composition of the patients included 11 (6.1%) Black or African Americans, 147 (82%) Caucasians, 16 (8.9%) Hispanics or Latinos, and 5 (2.8%) Asians and Pacific Islanders. The study was approved by the Institutional Review Board of Lifespan Health System.

Control

Four-digit allele frequencies of each control race group were obtained from the US National Marrow Donor Program (NMDP) through Allele Frequency Net Database (allelefrequencies. net). The data from NMDP were based on the HLA typing of 416,581 Black or African Americans, 1,242,890 Caucasians, 146,714 Hispanics, 99,672 Chinese, and 27,978 Southeast Asians, which represent the composition of the general population of the United States [18].

Statistical analysis

HLA allele frequency estimation: HLA-A, B, and C allele frequencies were estimated by direct counting for the entire patient cohort and stratified by their race. The percentage of each allele was calculated as its frequency divided by 2N, where N was the corresponding number of patients in each stratum.

Comparison between the control and patient cohorts: The patient cohort and the NMDP were then matched by ethnicity. Standard contingency tables that compared HLA alleles between the patient cohort and the NMDP controls were constructed, and prevalence ratios (PRs) were calculated. Statistical significance was performed by a two-tailed Fisher's exact test. False discovery rate (FDR)-adjusted P values were calculated by accounting for the total number of comparisons (Benjamini-Hochberg (B-H) method). Only alleles carried by two or more tumors were reported in the studies. An adjusted P value of <0.05 was considered statisti-

Table 1. Cancer types and the patients' racial distribution

Cancer type	Overall N=179	Caucasian N=147 (82%)	Hispanic or Latino N=16 (8.9%)	Black or African American N=11 (6.1%)	Asian American/ Pacific Islander N=5 (2.8%)
Cancer type based on organ system, n (%)					
Breast cancer	15 (8.4)	14 (9.5)	1 (6.2)	0 (0)	0 (0)
CNS malignancy	4 (2.2)	4 (2.7)	0 (0)	0 (0)	0 (0)
GI cancer	44 (25)	36 (24)	3 (19)	3 (27)	2 (40)
GU cancer	45 (25)	38 (26)	4 (25)	2 (18)	1 (20)
Gyn cancer	13 (7.3)	11 (7.5)	1 (6.2)	1 (9.1)	0 (0)
H&N cancer	4 (2.2)	3 (2.0)	1 (6.2)	0 (0)	0 (0)
Pancreatobiliary cancer	28 (16)	21 (14)	3 (19)	2 (18)	2 (40)
Sarcoma	5 (2.8)	4 (2.7)	0 (0)	1 (9.1)	0 (0)
Thoracic cancer	21 (12)	16 (11)	3 (19)	2 (18)	0 (0)
Cancer site, n (%)					
Metastatic	84 (47)	71 (48)	5 (31)	5 (45)	3 (60)
Primary	95 (53)	76 (52)	11 (69)	6 (55)	2 (40)

cally significant. All statistical analyses were conducted using R (version 4.2.1).

Results

HLA type I alleles of different tumor types with different racial background

The cancer patient cohort included Caucasian Americans (82%), Hispanic or Latino Americans (8.9%), Black/African Americans (6.1%), and Asian Americans and Pacific Islanders (2.8%). The most common cancer types based on organ system were GU (25%), GI (25%), pancreaticobiliary (16%), and thoracic cancers (12%), accounting for about three-fourths of the entire patient cohort (**Table 1**). The frequencies appeared similar in different race groups, although not all the tumors were represented in less populated groups. About half of the tumors were metastases (**Table 1**).

One hundred forty-six (146) HLA type I alleles were found in the tumors, including 44 HLA-A, 72 HLA-B, and 30 HLA-C alleles (Supplementary Table 1). The 15 most common alleles were included in **Figure 1A**, among which the most prevalent allele was A*02:01 (n=78; 21.8%), followed by C*07:01 (n=65; 18.1%) and A*03:01 (n=57; 15.9%). All other alleles collectively accounted for <15%. When stratified by race, tumor type, and cancer site, not all the alleles had a balanced distribution (**Figure 1B-D**). For example, A*02:01 and C*04:01,

and C*06:02 occurred more frequently in Hispanic or Latino Americans; A*03:01 and B*35:01 were more frequently present in Black/African Americans; A*02:01, A*03:01, and C*06:02 appeared more common in thoracic tumors. The prevalence of all alleles in the cohort were listed in Supplementary Table 1.

Tumor HLA type I alleles not included in the NMDP database

Not all the 146 alleles in this cohort have been included in the current NMDP database (accessed in September 2022). **Table 2** listed 21 HLA type I alleles that were not curated in the NMDP database, accounting for 14.4% of all the 146 alleles detected. The most striking finding in our study was allele A*11:303. It is not curated in the NMDP; however, it was found in 11 individual cancers, including four GI tract cancers, three GU cancers, two cholangiocarcinomas, and others (**Table 2**). The patients carrying the allele consisted of 8 Caucasians, 2 Hispanic or Latinos, and 1 Asian. One prostate cancer had homozygous A*11:303, likely due to loss of heterogeneity in the tumor. The next frequent allele was B*08*222, which was seen in four different tumors (one breast cancer, one renal cell carcinoma, one high-grade serous carcinoma, and one pancreatic cancer) from two races (**Table 2**). Alleles that occurred in two different tumors included A*02:215 (two renal cell carcinomas), A*02:771 (one breast cancer and one squamous cell carcinoma of
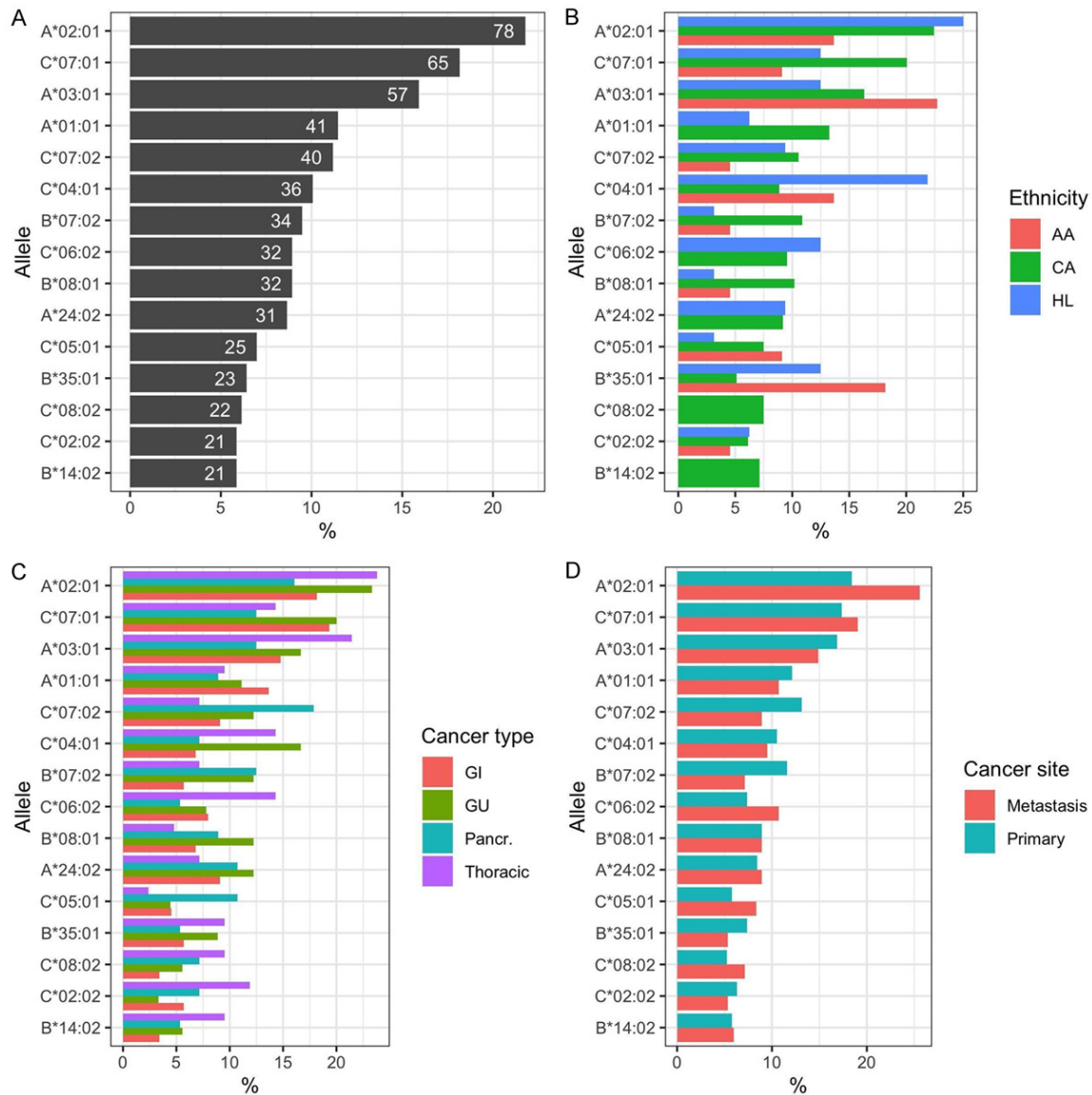
Figure 1. The fifteen most prevalent alleles in tumor tissues. A. The percentages of the most prevalent alleles (The numbers on the bars indicate the counts of corresponding alleles). B. Relative prevalence (%) of the alleles as stratified by ethnicity. AA: Black/African American; CA: Caucasian American; HL: Hispanic or Latino. C. Relative prevalence (%) as stratified by the four common cancer types. D. Relative prevalence (%) as stratified by tumor sites.

the esophagus), A*24:01 (one colon cancer and one endometrial cancer), B*44:323 (one urothelial carcinoma and one endocervical cancer), B*51:239 (one breast cancer and one high-grade serous carcinoma) (**Table 2**).

Tumor HLA type I alleles included in the NMDP database

For the alleles included in the NMDP, we compared the relative prevalence of the alleles between the cohort and NMDP, stratified by

HLA type and race. The differences were quantified by prevalence ratios and tested using Fisher's exact text. **Figure 2A** exhibited a part of the comparison results, where only those alleles with P<0.05 and with counts ≥2 cancers were included. There were 17 alleles (from 13 Caucasians, 3 Black/African Americans, and 1 Hispanic or Latino American) that exhibited significant (unadjusted *P* value <0.05) prevalence differences between our cohort and NMDP controls. After adjusting for FDR, the prevalence of five alleles among Caucasians re-

Table 2. HLA type I alleles that are not included in the NMDP database

HLA-A		HLA-B		HLA-C	
Caucasian					
A*11:303	Renal cell carcinoma ×2 Prostate cancer Breast cancer Esophageal adenocarcinoma Cholangiocarcinoma Endocervical adenocarcinoma SMARCA4-deficient malignant neoplasm	B*08:222	Renal cell carcinoma Pancreatic cancer Breast cancer	C*02:10	Colon cancer
A*01:215	Renal cell carcinoma ×2	B*44:323	Urothelial carcinoma Endocervical adenocarcinoma	C*03:32	Breast cancer
A*02:771	Breast cancer Esophageal squamous cell carcinoma	B*51:239	Breast cancer High-grade serous carcinoma	C*04:356	Colon cancer
A*24:01	Endometrial adenocarcinoma Colon cancer	B*35:208	High-grade serous carcinoma	C*06:227	Renal cell carcinoma
A*02:724	Breast cancer	B*44:225	Nasopharyngeal carcinoma	C*07:18	SMARCA4-deficient malignant neoplasm
A*02:802	Glioblastoma	B*44:269	Malignant PEComa		
A*02:705	Urothelial carcinoma	B*46:24	Breast cancer	C*12:143	Colon cancer
A*02:89	Renal cell carcinoma	B*51:260	Colon cancer		
Hispanic or Latino					
A*11:303	Colon cancer Cholangiocarcinoma				
Black or African American					
A*01:252	Stomach cancer	B*08:222	High-grade serous carcinoma	C*08:32	Urothelial carcinoma
Asian American and Pacific Islander					
A*11:303	Colon cancer				

mained significantly different, which were highlighted in the red box in **Figure 2A**: B*44:02 (PR=0.36, adjusted P=0.03), C*08:02 (PR= 1.94, adjusted P=0.047), B*14:02 (PR=2.5, adjusted P=0.003), A*03:02 (PR=7.09, adjusted P=0.014), and B*44:06 (PR=170, adjusted P=0.003). The five alleles were all found in tumors from Caucasians.

Figure 2B provided a scatterplot of all the alleles, where -log10 (unadjusted P values) of each comparison were plotted against log10 (PR) for each allele. Alleles with small P values tended to appear in the upper half of the figure, and alleles with significant differences in PR between cancer and NMDP controls appeared in the right or left parts of the diagram, away from zero. For reference, we added a horizontal line in **Figure 2B** indicating -log10 (unadjusted P=0.05) and two arbitrary vertical lines indicating log10 (PR=1/4) and log10 (PR=4). In **Figure 2B**, the alleles highlighted in red were those with highly significant differences (both unadjusted P<0.05 and BH-adjusted P<0.05). Moreover, five alleles were highlighted in yellow with PRs of >4 and unadjusted P<0.05 but BH-adjusted P>0.05. These five alleles were

B*42:02 and B*18:01 for Black/African Americans, C*03:02 and A*02:02 for Caucasians, and A*02:02 for Hispanic or Latino Americans. These five alleles may be worth further investigation. The only significant allele associated with decreased prevalence in the cohort was B*44:02 (PR=0.36).

We further compared the prevalence of alleles between the tumors and NMDP controls, stratified by race and cancer organ system. Besides the five alleles that showed significant difference collectively for all tumors, we further identified 15 alleles with significant differences from the NMDP controls. These included four alleles in breast cancers (including C*08:02), one in CNS malignancies, two in GI cancers, two in H&N cancers, five in pancreaticobiliary cancers, and one in sarcomas of Caucasians, one allele in pancreaticobiliary cancers of Black/African Americans, and one allele in breast cancers and one in GI cancers of Hispanic or Latino Americans (**Table 3**). Of note, all 15 alleles were associated with increased prevalence in the current cohort. A more inclusive list of all the alleles with unadjusted P values of <0.05 was presented in <u>Supplementary</u>
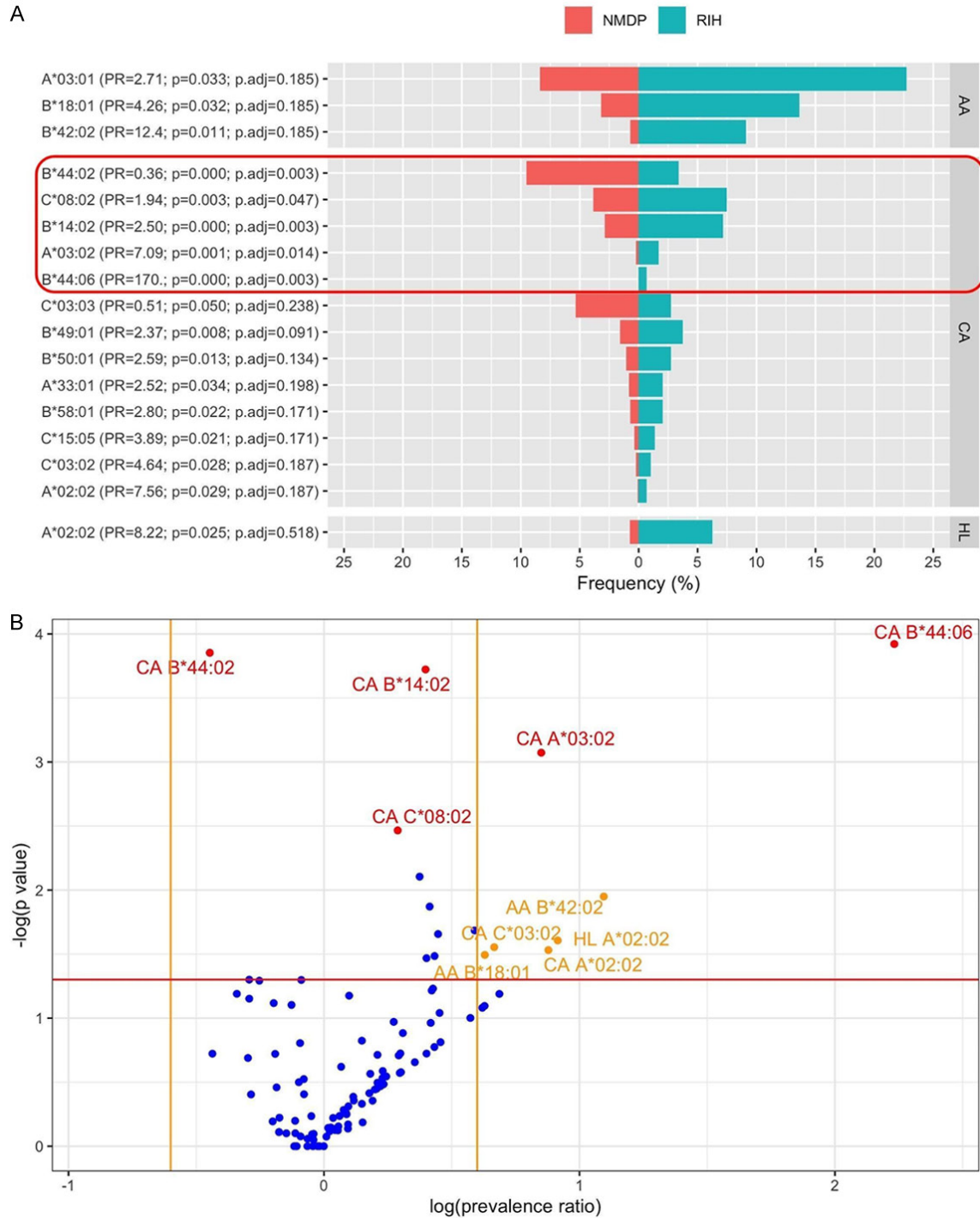
Figure 2. Comparison of allele prevalence between the cancer patient cohort and NMDP controls. A. Pyramid plot of allele prevalence. AA: Black or African American; CA: Caucasian American; HL: Hispanic or Latino; p: P-value; p.adj: "B-H" adjusted P-value. The red box indicates alleles with significant differences (after adjustment). B. -log (unadjusted P-value) versus log (Prevalence Ratio). The red horizontal line indicates unadjusted P=0.05; so all alleles above the red line have unadjusted P<0.05. The left yellow vertical line indicates PR=1/4; the right line indicates PR=4; so the alleles outside the two vertical lines exhibit a clinically major difference (PR<1/4 or PR>4). We further highlight those alleles with both unadjusted P<0.05 and adjusted P<0.05 in red. Alleles with PR>4 and unadjusted P<0.05 but adjusted P>0.05 are highlighted in yellow.

Table 3. Alleles in Caucasian with adjusted *P* value less than 0.05 and occurred in 2 or more tumors

Allele	Count in Tumor Samples	Prevalence in Patient Series (%)	Prevalence in NMDP (%)	Prevalence Ratio	*P* value	Adjusted *P* value
Caucasian - All tumors						
C*08:02	22	7.48	3.85	1.94	0.003	0.047
B*14:02	21	7.14	2.86	2.50	0.000	0.003
B*44:02	10	3.40	9.52	0.36	0.000	0.003
A*03:02	5	1.70	0.24	7.09	0.001	0.014
B*44:06	2	0.68	0.00	170	0.000	0.003
Caucasian - Breast cancer						
C*08:02	5	17.8	3.85	4.64	0.004	0.022
B*14:02	4	14.3	2.86	5.00	0.008	0.034
B*44:06	2	7.14	<.01	>100	<.001	<.001
B*47:01	2	7.14	0.27	26.5	0.003	0.022
Caucasian - CNS malignancy						
B*49:01	2	25.0	1.58	15.8	0.007	0.033
Caucasian - GI cancer						
C*15:05	3	4.17	0.35	11.9	0.002	0.044
A*02:02	2	2.78	0.09	30.9	0.002	0.044
Caucasian - H&N cancer						
A*02:01	5	83.3	27.6	3.02	0.007	0.015
B*50:01	2	33.3	1.05	31.8	0.002	0.006
Caucasian - Pancreatobiliary cancer						
A*68:02	3	7.14	0.84	8.50	0.005	0.031
B*58:01	3	7.14	0.73	9.78	0.004	0.031
C*03:02	3	7.14	0.22	32.5	0.000	0.003
A*03:02	2	4.76	0.24	19.8	0.005	0.031
A*69:01	2	4.76	0.14	34.0	0.002	0.024
Caucasian - Sarcoma						
A*03:02	2	25.0	0.24	104	0.000	0.001
Black or African American - Pancreatobiliary cancer						
C*16:01	2	50.0	9.69	5.16	0.049	0.049
Hispanic - Breast adenocarcinoma						
A*02:01	2	100	20.9	4.77	0.044	0.044
Hispanic - GI cancer						
C*12:03	2	33.3	4.15	8.03	0.023	0.046

Table 2. Comparisons with stratification by race and individual cancer type were shown in Supplementary Table 3. Only alleles carried by two or more tumors were included in the tables.

Discussion

Next-generation sequencing has provided powerful tools and resulted in massive data for the HLA community. HLA typing has been through multiple generations of methodology, and the next generation sequencing has provided great opportunities and posed challenges at the same time to HLA typing [19, 20]. In the current study, we extracted HLA-A, B, and C data from the Caris Comprehensive Tumor Profiling report. Paraffin-embedded tumor tissue was used by Caris to perform the HLA typing. The details of Caris's methodology are not known to us. Traditionally, peripheral blood or oral swabs are the classic sources of DNA for HLA typing. The current study is the first to our knowledge that used tumor HLA for a tumor susceptibility study. The concern is that cancer cells could acquire HLA mutations to facilitate immune evasion [21], which may void the analysis. In addi-

tion, a comprehensive study has evaluated the somatic mutation in HLA genes [22]. Based on TCGA and other large datasets, the study discovered 73 mutations in 64 alleles of 2545 persons. If we extrapolated their findings to our cohort, only five mutations could have occurred in our 179 patients. Although we cannot identify these somatic mutations, their impacts on our data and results appeared acceptable.

The general way to obtain a balanced control group is to recruit patients from the geographic vicinity. In our study, the NMDP data was used. NMDP consisted of HLA typing data from a large number of individuals of each race. Therefore, its allele frequencies should be well-representative of the general population of the United States. Although the detailed racial composition of our served regional population is not available, prevalence ratios were calculated for each major race to avoid selection bias. NMDP data has been used in similar studies to reveal the relationship between alleles and the prevalence of certain tumors [23].

The fact that our cohort had 21 alleles that have not been listed in the NMDP database is striking. Among them, A*11:303 is the most frequent. Very little is known about the A*11:303. It was initially submitted to Immuno Polymorphism Database (IPD) from a lab in the United States based on a buccal swab sample (https://www.ebi.ac.uk/ipd/imgt/hla/cells/cell/?cellid=47998). A*11:303 was unlikely to be caused by HLA somatic mutations, given its presence in multiple tumors in patients of various races. Its high prevalence in our cohort surely warrants further studies to better characterize these tumors, to examine the patients' germline HLA genotype, and to determine the efficacy of antigen presentation. The next most frequent allele, B*8*222 was carried by four different tumors in patients of two races. It is not listed in IPD, and there is no published information about the allele to our knowledge. All the remaining alleles in **Table 2** are not listed in IPD except A*02:215, which was identified from a group of stem cell donors [24]. The study found that out of 2127 new alleles, 156 (7.3%) new alleles were found in multiple people, suggesting that these alleles were not necessarily rare [25]. The exact frequencies of the unlisted alleles in our cohort may not be as low as expected and are worth further investigation.

Due to the small sample sizes, few alleles were found to be associated with non-Caucasian people. In Caucasians, five alleles showed tumor association. C*08:02, B*14:02, A*03:02, and B*44:06 appeared susceptible with prevalence ratios ranging from 1.94 to 170, while B*44:02 was protective when all the cancers were considered (**Figure 2** and **Table 3**). For individual cancers, patients with B*44:06 were susceptible to breast cancer, and C*03:02 patients were susceptible to pancreaticobiliary cancers.

C*08:02 allele was enriched in our cohort and associated with overall tumor susceptibility and breast cancer when analyzed individually (**Table 3**). C*08:02 T cells have been found capable of targeting Kirsten rat sarcoma virus (KRAS) with G12D mutation [25, 26]. However, the KRAS G12D mutated colon cancer in C*08:02 patients did not have any unique immune activity [27]. The underlying mechanisms of the association between C*08:02 and certain tumor are still not clear. No studies have been reported regarding any association between C*08:02 and breast cancer.

B*14:02 was seen in 21 tumors in the current series with an odd ratio of 2.61. It has been found to be associated with Gag-specific CD8+ T-cell response against HIV-1 [28]. It was also found to be able to bind Phox2b, a neuroblastoma-specific marker, and to be used in peptide-centric chimeric antigen receptors (CARs) for specific tumor targeting [29]. B*14:02 has also been shown to be associated with the development of cervical dysplasia [30, 31]. In the current study, B*14:02 was associated with individual cancers such as breast cancer (P= 0.0079, adjusted P=0.034) and thoracic cancer (P=0.0127, adjusted P=0.152) (**Table 3** and Supplementary Table 2).

In the current cohort, A*03:02 was seen in pancreatic cancers, metastatic GIST, and endometrial cancer. The implication of A*03:02 in cancer has been reported. It could present a mutated PIK3CA (H1047L) peptide for neoantigen targeting [32]; however, the real-life effect of the targeting has not been examined. HLA-A*03 has been shown to be a predictive biomarker of poor response to immune checkpoint inhibitors (ICIs) [16]. The exact role of A*03:02 needs further evaluation.

B*44:02 was demonstrated to be one of the protective alleles for multiple sclerosis [33]. It was also reported that B*44:02 positively influenced the overall survival of melanoma patients treated with ICIs [17]. B*44:02 was also found as a protective factor for medically refractory pediatric acute lymphoblastic leukemia [34]. B*44:02, along with other HLA alleles, was associated with a lower incidence of relapse of acute myeloid leukemia after autologous hematopoietic cell transplantation [35]. Although closely related to B*44:02, the susceptibility of B*44:06 to breast cancer is a new finding; no study has been reported on the effect of B*44:06 on cancer or leukemia.

C*03:02 was found to be associated with long-term non-progression of HIV disease in African Pediatric populations [36]. To our knowledge, no tumor association had been found about this allele.

Table 3 included some candidate alleles with tumor association, such as A*02:01 in H&N cancer, C*15:05 in GI cancer, among others. These associations were based on relatively low tumor samples, and they need to be confirmed by further studies. Different HLA molecules differ in their efficacies to present the antigens. It was found in advanced cancer patients homozygous MHC type I molecules reduced the survival and immune checkpoint inhibitor antitumor response. The same study revealed that linked MHC type I heterozygosity broadened the tumor-specific antigen presentations [37]. Furthermore, Marty et al. indicated that individual MHC type I genotypes could predict cancer susceptibility, as certain oncogenic mutations linked to this genotype may limit possible tumor mutations [12].

Our studies have several limitations. First, due to the highly polymorphic nature of HLA genes, our relative sample number (179) might not properly represent all the different disease/allele combinations; but those alleles are not equally distributed in different populations, as the few most common alleles are present in the vast majority of individuals, and this is reflected in our results. We used Benjamini-Hochberg method in our statistical analysis to limit the false discovery rate. Second, as mentioned above, the HLA genotype in cancer may be the same as the germline HLA in the patient. Although the predicted mutations are limited, they could cause over-or under-estimate in some correlation studies. However, such bias is likely minimal. HLA-typing of the tumor could be helpful or even necessary when HLA-based targeted therapy is available in the future. Third, not all the tumors in our health system were tested by Caris cancer profiling. The tumors in the cohort were profiled mainly to meet the requirements of ongoing clinical trials or for the purpose of targeted therapy. Therefore, it may introduce some selection biases, such as possible preferences towards certain patient groups or tumor stages. Fourth, homozygous alleles resulting from loss of heterozygosity (LOH) in the tumor cells could lead to the over-representation of specific alleles. For example, one prostate cancer had two copies of allele A*11:303. Given the low frequency of A*11:303 in the general population, it was very likely an example of LOH. It is one of the limitations of using tumor tissue to genotype HLA alleles. Fifth, to increase the robustness of the analysis, we excluded alleles that only carried in one tumor from our cohort, which could miss some significant alleles. Finally, due to limited amounts of cancers and limited general HLA information from minority groups, particularly Asians and Pacific Islanders, the most important findings were limited to the Caucasian population.

In summary, we collected HLA-A, B, and C data from 179 malignant solid tumors and identified several susceptible alleles and one protective allele. Of interest, 21 alleles were not reported in the NMDP, one of the largest allele databases of the American population. The identified alleles might not be limited only to Caucasian Americans if we could have more cases from other races.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Shaolei Lu, Department of Pathology and Laboratory Medicine, Alpert Medical School of Brown University, 593 Eddy Street, Providence, RI 02806, USA. Tel: 401-444-7120; ORCID: 0000-0003-2870-5793; E-mail: slu@lifespan.org

References

[1] Shiina T, Hosomichi K, Inoko H and Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 2009; 54: 15-39.

[2] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[3] Thor Straten P and Garrido F. Targetless T cells in cancer immunotherapy. J Immunother Cancer 2016; 4: 23.

[4] Garrido F. HLA class-I expression and cancer immunotherapy. Adv Exp Med Biol 2019; 1151: 79-90.

[5] Aptsiauri N, Cabrera T, Garcia-Lora A, Lopez-Nevot MA, Ruiz-Cabello F and Garrido F. MHC class I antigens and immune surveillance in transformed cells. Int Rev Cytol 2007; 256: 139-89.

[6] Rodríguez JA. HLA-mediated tumor escape mechanisms that may impair immunotherapy clinical outcomes via T-cell activation. Oncol Lett 2017; 14: 4415-4427.

[7] Martínez-Jiménez F, Priestley P, Shale C, Baber J, Rozemuller E and Cuppen E. Genetic immune escape landscape in primary and metastatic cancer. Nat Genet 2023; 55: 820-831.

[8] Höglund P and Brodin P. Current perspectives of natural killer cell education by MHC class I molecules. Nat Rev Immunol 2010; 10: 724-34.

[9] Cruz-Muñoz ME, Valenzuela-Vázquez L, Sánchez-Herrera J and Santa-Olalla Tapia J. From the "missing self" hypothesis to adaptive NK cells: insights of NK cell-mediated effector functions in immune surveillance. J Leukoc Biol 2019; 105: 955-971.

[10] Hazini A, Fisher K and Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. J Immunother Cancer 2021; 9: e002899.

[11] Little AM and Stern PL. Does HLA type predispose some individuals to cancer? Mol Med Today 1999; 5: 337-42.

[12] Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB, de Prisco N, Ideker T, Hildebrand WH, Font-Burgada J and Carter H. MHC-I genotype restricts the oncogenic mutational landscape. Cell 2017; 171: 1272-1283, e15.

[13] Chaudhuri S, Cariappa A, Tang M, Bell D, Haber DA, Isselbacher KJ, Finkelstein D, Forcione D and Pillai S. Genetic susceptibility to breast cancer: HLA DQB*03032 and HLA DRB1*11 may represent protective alleles. Proc Natl Acad Sci U S A 2000; 97: 11451-11454.

[14] Vaquero-Yuste C, Juarez I, Molina-Alejandre M, Molanes-López EM, López-Nares A, Suárez-Trujillo F, Gutiérrez-Calvo A, López-García A, Lasa I, Gómez R, Fernández-Cruz E, Rodrígez-Sainz C, Arnaiz-Villena A and Martín-Villa JM. HLA-G 3'UTR polymorphisms are linked to susceptibility and survival in Spanish gastric adenocarcinoma patients. Front Immunol 2021; 12: 698438.

[15] Dhouioui S, Laaribi AB, Boujelbene N, Jelassi R, Ben Salah H, Bellali H, Ouzari HI, Mezlini A, Zemni I, Chelbi H and Zidi I. Association of HLA-G 3'UTR polymorphisms and haplotypes with colorectal cancer susceptibility and prognosis. Hum Immunol 2022; 83: 39-46.

[16] Naranbhai V, Viard M, Dean M, Groha S, Braun DA, Labaki C, Shukla SA, Yuki Y, Shah P, Chin K, Wind-Rotolo M, Mu XJ, Robbins PB, Gusev A, Choueiri TK, Gulley JL and Carrington M. HLA-A*03 and response to immune checkpoint blockade in cancer: an epidemiological biomarker study. Lancet Oncol 2022; 23: 172-184.

[17] Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, Kuo F, Kendall SM, Requena D, Riaz N, Greenbaum B, Carroll J, Garon E, Hyman DM, Zehir A, Solit D, Berger M, Zhou R, Rizvi NA and Chan TA. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. Science 2018; 359: 582-587.

[18] Gragert L, Madbouly A, Freeman J and Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Hum Immunol 2013; 74: 1313-1320.

[19] Bravo-Egana V, Sanders H and Chitnis N. New challenges, new opportunities: next generation sequencing and its place in the advancement of HLA typing. Hum Immunol 2021; 82: 478-487.

[20] Creary LE, Sacchi N, Mazzocco M, Morris GP, Montero-Martin G, Chong W, Brown CJ, Dinou A, Stavropoulos-Giokas C, Gorodezky C, Narayan S, Periathiruvadi S, Thomas R, De Santis D, Pepperall J, ElGhazali GE, Al Yafei Z, Askar M, Tyagi S, Kanga U, Marino SR, Planelles D, Chang CJ and Fernández-Viña MA. High-resolution HLA allele and haplotype frequencies in several unrelated populations determined by next generation sequencing: 17th International HLA and Immunogenetics Workshop joint report. Hum Immunol 2021; 82: 505-522.

[21] Montesion M, Murugesan K, Jin DX, Sharaf R, Sanchez N, Guria A, Minker M, Li G, Fisher V, Sokol ES, Pavlick DC, Moore JA, Braly A, Singal G, Fabrizio D, Comment LA, Rizvi NA, Alexander BM, Frampton GM, Hegde PS and Albacker LA. Somatic HLA class I loss is a widespread mechanism of immune evasion which refines the use of tumor mutational burden as a bio-

marker of checkpoint inhibitor response. Cancer Discov 2021; 11: 282-292.

[22] Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, Stevens J, Lane WJ, Dellagatta JL, Steelman S, Sougnez C, Cibulskis K, Kiezun A, Hacohen N, Brusic V, Wu CJ and Getz G. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol 2015; 33: 1152-1158.

[23] Beksac M, Gragert L, Fingerson S, Maiers M, Zhang MJ, Albrecht M, Zhong X, Cozen W, Dispenzieri A, Lonial S and Hari P. HLA polymorphism and risk of multiple myeloma. Leukemia 2016; 30: 2260-2264.

[24] Hernández-Frederick CJ, Giani AS, Cereb N, Sauter J, Silva-González R, Pingel J, Schmidt AH, Ehninger G and Yang SY. Identification of 2127 new HLA class I alleles in potential stem cell donors from Germany, the United States and Poland. Tissue Antigens 2014; 83: 184-189.

[25] Tran E, Robbins PF, Lu YC, Prickett TD, Gartner JJ, Jia L, Pasetto A, Zheng Z, Ray S, Groh EM, Kriley IR and Rosenberg SA. T-cell transfer therapy targeting mutant KRAS in cancer. N Engl J Med 2016; 375: 2255-2262.

[26] Choi J, Goulding SP, Conn BP, McGann CD, Dietze JL, Kohler J, Lenkala D, Boudot A, Rothenberg DA, Turcott PJ, Srouji JR, Foley KC, Rooney MS, van Buuren MM, Gaynor RB, Abelin JG, Addona TA and Juneja VR. Systematic discovery and validation of T cell targets directed against oncogenic KRAS mutations. Cell Rep Methods 2021; 1: 100084.

[27] Rech AJ and Vonderheide RH. T-cell transfer therapy targeting mutant KRAS. N Engl J Med 2017; 376: e11.

[28] Leitman EM, Willberg CB, Tsai MH, Chen H, Buus S, Chen F, Riddell L, Haas D, Fellay J, Goedert JJ, Piechocka-Trocha A, Walker BD, Martin J, Deeks S, Wolinsky SM, Martinson J, Martin M, Qi Y, Sáez-Cirión A, Yang OO, Matthews PC, Carrington M and Goulder PJR. HLA-B*14:02-restricted Env-specific CD8+ T-cell activity has highly potent antiviral efficacy associated with immune control of HIV infection. J Virol 2017; 91: e00544-17.

[29] Yarmarkovich M, Marshall QF, Warrington JM, Premaratne R, Farrel A, Groff D, Li W, di Marco M, Runbeck E, Truong H, Toor JS, Tripathi S, Nguyen S, Shen H, Noel T, Church NL, Weiner A, Kendsersky N, Martinez D, Weisberg R, Christie M, Eisenlohr L, Bosse KR, Dimitrov DS, Stevanovic S, Sgourakis NG, Kiefel BR and Maris JM. Cross-HLA targeting of intracellular oncoproteins with peptide-centric CARs. Nature 2021; 599: 477-484.

[30] Wang SS, Hildesheim A, Gao X, Schiffman M, Herrero R, Bratti MC, Sherman ME, Barnes WA, Greenberg MD, McGowan L, Mortel R, Schwartz PE, Zaino RJ, Glass AG, Burk RD, Karacki P and Carrington M. Comprehensive analysis of human leukocyte antigen class I alleles and cervical neoplasia in 3 epidemiologic studies. J Infect Dis 2002; 186: 598-605.

[31] Bahls L, Yamakawa R, Zanão K, Alfieri D, Flauzino T, Delongui F, de Abreu A, Souza R, Gimenes F, Reiche E, Borelli S and Consolaro M. Human leukocyte antigen class I and class II polymorphisms and serum cytokine profiles in cervical cancer. Int J Mol Sci 2017; 18: 1478.

[32] Chandran SS, Ma J, Klatt MG, Dündar F, Bandlamudi C, Razavi P, Wen HY, Weigelt B, Zumbo P, Fu SN, Banks LB, Yi F, Vercher E, Etxeberria I, Bestman WD, Da Cruz Paula A, Aricescu IS, Drilon A, Betel D, Scheinberg DA, Baker BM and Klebanoff CA. Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated PIK3CA. Nat Med 2022; 28: 946-957.

[33] Healy BC, Liguori M, Tran D, Chitnis T, Glanz B, Wolfish C, Gauthier S, Buckle G, Houtchens M, Stazzone L, Khoury S, Hartzmann R, Fernandez-Vina M, Hafler DA, Weiner HL, Guttmann CR and De Jager PL. HLA B*44: protective effects in MS susceptibility and MRI outcome measures. Neurology 2010; 75: 634-640.

[34] Klitz W, Gragert L and Trachtenberg E. Spectrum of HLA associations: the case of medically refractory pediatric acute lymphoblastic leukemia. Immunogenetics 2012; 64: 409-419.

[35] Boudreau JE, Giglio F, Gooley TA, Stevenson PA, Le Luduec JB, Shaffer BC, Rajalingam R, Hou L, Hurley CK, Noreen H, Reed EF, Yu N, Vierra-Green C, Haagenson M, Malkki M, Petersdorf EW, Spellman S and Hsu KC. KIR3DL1/HLA-B subtypes govern acute myelogenous leukemia relapse after hematopoietic cell transplantation. J Clin Oncol 2017; 35: 2268-2278.

[36] Kyobe S, Mwesigwa S, Kisitu GP, Farirai J, Katagirya E, Mirembe AN, Ketumile L, Wayengera M, Katabazi FA, Kigozi E, Wampande EM, Retshabile G, Mlotshwa BC, Williams L, Morapedi K, Kasvosve I, Kyosiimire-Lugemwa J, Nsangi B, Tsimako-Johnstone M, Brown CW, Joloba M, Anabwani G, Bhekumusa L, Mpoloka SW, Mardon G, Matshaba M, Kekitiinwa A and Hanchard NA. Exome sequencing reveals a putative role for HLA-C*03:02 in control of HIV-1 in African pediatric populations. Front Genet 2021; 12: 720213.

[37] Stokidis S, Fortis SP, Kogionou P, Anagnostou T, Perez SA and Baxevanis CN. HLA class I allele expression and clinical outcome in De Novo metastatic prostate cancer. Cancers (Basel) 2020; 12: 1623.

Supplementary Table 1. Alleles prevalence by ethnicity

Alleles	Overall N=179	Caucasian N=147 (82%)	Hispanic or Latino N=16 (8.9%)	Black or African American N=11 (6.1%)	Asian American/ Pacific Islander N=5 (2.8%)
HLA-A Allele, n (%)					
A*01:01	41 (11)	0 (0)	39 (13)	2 (6.2)	0 (0)
A*01:02	1 (0.3)	1 (4.5)	0 (0)	0 (0)	0 (0)
A*01:03	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*01:215	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
A*01:252	1 (0.3)	1 (4.5)	0 (0)	0 (0)	0 (0)
A*02:01	78 (22)	3 (14)	66 (22)	8 (25)	1 (10)
A*02:02	4 (1.1)	0 (0)	2 (0.7)	2 (6.2)	0 (0)
A*02:03	2 (0.6)	0 (0)	1 (0.3)	0 (0)	1 (10)
A*02:05	4 (1.1)	0 (0)	3 (1.0)	1 (3.1)	0 (0)
A*02:34	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*02:705	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*02:724	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*02:771	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
A*02:802	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*02:89	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*03:01	57 (16)	5 (23)	48 (16)	4 (12)	0 (0)
A*03:02	5 (1.4)	0 (0)	5 (1.7)	0 (0)	0 (0)
A*11:01	11 (3.1)	0 (0)	10 (3.4)	0 (0)	1 (10)
A*11:303	12 (3.4)	0 (0)	9 (3.1)	2 (6.2)	1 (10)
A*23:01	7 (2.0)	3 (14)	3 (1.0)	1 (3.1)	0 (0)
A*24:01	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
A*24:02	31 (8.7)	0 (0)	27 (9.2)	3 (9.4)	1 (10)
A*24:07	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
A*24:10	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
A*24:25	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
A*24:32	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*25:01	6 (1.7)	0 (0)	5 (1.7)	0 (0)	1 (10)
A*26:01	11 (3.1)	0 (0)	10 (3.4)	0 (0)	1 (10)
A*26:08	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*29:02	12 (3.4)	2 (9.1)	8 (2.7)	2 (6.2)	0 (0)
A*30:01	8 (2.2)	3 (14)	5 (1.7)	0 (0)	0 (0)
A*30:02	5 (1.4)	1 (4.5)	3 (1.0)	1 (3.1)	0 (0)
A*31:01	4 (1.1)	0 (0)	4 (1.4)	0 (0)	0 (0)
A*32:01	10 (2.8)	0 (0)	9 (3.1)	1 (3.1)	0 (0)
A*33:01	6 (1.7)	0 (0)	6 (2.0)	0 (0)	0 (0)
A*33:03	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*34:01	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
A*34:02	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
A*66:01	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
A*68:01	9 (2.5)	0 (0)	9 (3.1)	0 (0)	0 (0)
A*68:02	5 (1.4)	1 (4.5)	4 (1.4)	0 (0)	0 (0)
A*68:03	2 (0.6)	0 (0)	0 (0)	2 (6.2)	0 (0)
A*69:01	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
A*74:01	3 (0.8)	2 (9.1)	0 (0)	1 (3.1)	0 (0)

HLA type I molecules of solid malignancies

HLA-B Allele, n (%)

Allele					
B*07:02	34 (9.5)	1 (4.5)	32 (11)	1 (3.1)	0 (0)
B*07:05	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*08:01	32 (8.9)	1 (4.5)	30 (10)	1 (3.1)	0 (0)
B*08:12	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*08:222	4 (1.1)	1 (4.5)	3 (1.0)	0 (0)	0 (0)
B*13:01	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
B*13:02	8 (2.2)	0 (0)	8 (2.7)	0 (0)	0 (0)
B*14:01	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*14:02	21 (5.9)	0 (0)	21 (7.1)	0 (0)	0 (0)
B*15:01	17 (4.7)	0 (0)	16 (5.4)	1 (3.1)	0 (0)
B*15:03	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*15:10	1 (0.3)	1 (4.5)	0 (0)	0 (0)	0 (0)
B*15:16	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*15:17	3 (0.8)	0 (0)	3 (1.0)	0 (0)	0 (0)
B*15:24	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*15:35	2 (0.6)	0 (0)	0 (0)	0 (0)	2 (20)
B*18:01	18 (5.0)	3 (14)	13 (4.4)	0 (0)	2 (20)
B*18:02	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
B*27:02	3 (0.8)	0 (0)	3 (1.0)	0 (0)	0 (0)
B*27:05	5 (1.4)	0 (0)	5 (1.7)	0 (0)	0 (0)
B*27:06	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
B*27:08	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*35:01	23 (6.4)	4 (18)	15 (5.1)	4 (12)	0 (0)
B*35:02	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*35:03	5 (1.4)	0 (0)	3 (1.0)	2 (6.2)	0 (0)
B*35:08	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*35:12	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
B*35:208	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*35:22	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*35:29	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*37:01	4 (1.1)	0 (0)	4 (1.4)	0 (0)	0 (0)
B*38:01	8 (2.2)	0 (0)	8 (2.7)	0 (0)	0 (0)
B*39:01	5 (1.4)	0 (0)	5 (1.7)	0 (0)	0 (0)
B*39:05	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
B*39:06	2 (0.6)	0 (0)	1 (0.3)	1 (3.1)	0 (0)
B*39:09	1 (0.3)	0 (0)	0 (0)	0 (0)	1 (10)
B*39:10	1 (0.3)	1 (4.5)	0 (0)	0 (0)	0 (0)
B*40:01	12 (3.4)	0 (0)	10 (3.4)	1 (3.1)	1 (10)
B*40:02	10 (2.8)	0 (0)	6 (2.0)	3 (9.4)	1 (10)
B*41:01	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*42:01	3 (0.8)	2 (9.1)	0 (0)	1 (3.1)	0 (0)
B*42:02	2 (0.6)	2 (9.1)	0 (0)	0 (0)	0 (0)
B*44:02	10 (2.8)	0 (0)	10 (3.4)	0 (0)	0 (0)
B*44:03	12 (3.4)	2 (9.1)	7 (2.4)	3 (9.4)	0 (0)
B*44:05	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*44:06	3 (0.8)	0 (0)	2 (0.7)	1 (3.1)	0 (0)
B*44:22	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*44:225	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)

2

HLA type I molecules of solid malignancies

B*44:269	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*44:323	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*45:01	2 (0.6)	1 (4.5)	0 (0)	1 (3.1)	0 (0)
B*46:24	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*47:01	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*49:01	13 (3.6)	1 (4.5)	11 (3.7)	1 (3.1)	0 (0)
B*50:01	10 (2.8)	0 (0)	8 (2.7)	2 (6.2)	0 (0)
B*50:02	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
B*51:01	19 (5.3)	0 (0)	16 (5.4)	3 (9.4)	0 (0)
B*51:05	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*51:08	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
B*51:239	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*51:260	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*51:42	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*52:01	3 (0.8)	1 (4.5)	2 (0.7)	0 (0)	0 (0)
B*53:01	3 (0.8)	1 (4.5)	2 (0.7)	0 (0)	0 (0)
B*55:01	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*56:01	2 (0.6)	0 (0)	2 (0.7)	0 (0)	0 (0)
B*57:01	7 (2.0)	0 (0)	7 (2.4)	0 (0)	0 (0)
B*57:03	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
B*58:01	6 (1.7)	0 (0)	6 (2.0)	0 (0)	0 (0)
B*58:02	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
B*73:01	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
B*83:01	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
HLA-C Allele, n (%)					
C*01:02	10 (2.8)	0 (0)	10 (3.4)	0 (0)	0 (0)
C*02:02	21 (5.9)	1 (4.5)	18 (6.1)	2 (6.2)	0 (0)
C*02:10	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*03:02	4 (1.1)	1 (4.5)	3 (1.0)	0 (0)	0 (0)
C*03:03	9 (2.5)	0 (0)	8 (2.7)	1 (3.1)	0 (0)
C*03:04	19 (5.3)	2 (9.1)	14 (4.8)	2 (6.2)	1 (10)
C*03:32	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*04:01	36 (10)	3 (14)	26 (8.8)	7 (22)	0 (0)
C*04:356	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*05:01	25 (7.0)	2 (9.1)	22 (7.5)	1 (3.1)	0 (0)
C*06:02	32 (8.9)	0 (0)	28 (9.5)	4 (12)	0 (0)
C*06:227	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*07:01	65 (18)	2 (9.1)	59 (20)	4 (12)	0 (0)
C*07:02	40 (11)	1 (4.5)	31 (11)	3 (9.4)	5 (50)
C*07:04	3 (0.8)	0 (0)	1 (0.3)	0 (0)	2 (20)
C*07:18	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*08:01	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
C*08:02	22 (6.1)	0 (0)	22 (7.5)	0 (0)	0 (0)
C*08:32	1 (0.3)	1 (4.5)	0 (0)	0 (0)	0 (0)
C*12:02	3 (0.8)	1 (4.5)	2 (0.7)	0 (0)	0 (0)
C*12:03	17 (4.7)	1 (4.5)	13 (4.4)	2 (6.2)	1 (10)
C*12:143	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*14:02	8 (2.2)	0 (0)	7 (2.4)	1 (3.1)	0 (0)
C*15:02	9 (2.5)	0 (0)	8 (2.7)	0 (0)	1 (10)

C*15:05	4 (1.1)	0 (0)	4 (1.4)	0 (0)	0 (0)
C*15:11	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*16:01	14 (3.9)	3 (14)	9 (3.1)	2 (6.2)	0 (0)
C*16:02	1 (0.3)	0 (0)	0 (0)	1 (3.1)	0 (0)
C*16:04	1 (0.3)	0 (0)	1 (0.3)	0 (0)	0 (0)
C*17:01	6 (1.7)	4 (18)	1 (0.3)	1 (3.1)	0 (0)

The percentage was calculated as frequency/(2*N).

Supplementary Table 2. Alleles with unadjusted *P* values of <0.05 in cancer organ systems

Allele	Count in Tumor Samples	Prevalence in Patient Series (%)	Prevalence in NMDP (%)	Prevalence Ratio	*P* value	Adjusted *P* value
Black or African American - All tumors						
A*03:01	5	22.73	8.39	2.71	0.033	0.185
B*18:01	3	13.64	3.20	4.26	0.032	0.185
B*42:02	2	9.09	0.73	12.45	0.011	0.185
Black or African American - Pancreatobiliary cancer						
C*16:01	2	50.00	9.69	5.16	0.049	0.049
Black or African American - Thoracic cancer						
B*35:01	2	50.00	6.89	7.26	0.026	0.078
Caucasian - All tumors						
C*08:02	22	7.48	3.85	1.94	0.003	0.047
B*14:02	21	7.14	2.86	2.50	0.000	0.003
B*49:01	11	3.74	1.58	2.37	0.008	0.091
B*44:02	10	3.40	9.52	0.36	0.000	0.003
B*50:01	8	2.72	1.05	2.59	0.013	0.134
C*03:03	8	2.72	5.34	0.51	0.050	0.238
A*33:01	6	2.04	0.81	2.52	0.034	0.198
B*58:01	6	2.04	0.73	2.80	0.022	0.171
A*03:02	5	1.70	0.24	7.09	0.001	0.014
C*15:05	4	1.36	0.35	3.89	0.021	0.171
C*03:02	3	1.02	0.22	4.64	0.028	0.187
A*02:02	2	0.68	0.09	7.56	0.029	0.187
B*44:06	2	0.68	<0.01	170	0.000	0.003
Caucasian - Breast adenocarcinoma						
C*08:02	5	17.86	3.85	4.64	0.004	0.022
B*14:02	4	14.29	2.86	5.00	0.008	0.034
A*33:01	2	7.14	0.81	8.82	0.022	0.073
B*44:06	2	7.14	<0.001	>100	0.000	0.000
B*47:01	2	7.14	0.27	26.45	0.003	0.022
Caucasian - CNS malignancy						
B*49:01	2	25.00	1.58	15.82	0.007	0.033
Caucasian - GI cancer						
B*49:01	4	5.56	1.58	3.52	0.027	0.281
C*15:05	3	4.17	0.35	11.91	0.002	0.044
A*02:02	2	2.78	0.09	30.87	0.002	0.044
B*07:05	2	2.78	0.30	9.26	0.020	0.274
Caucasian - GU cancer						
B*39:01	4	5.26	1.12	4.70	0.011	0.362
B*35:08	2	2.63	0.43	6.12	0.043	0.665

Allele	Count in Tumor Samples	Prevalence in Patient Series (%)	Prevalence in NMDP (%)	Prevalence Ratio	P value	Adjusted P value
Caucasian - Gyn cancer						
C*14:02	2	9.09	1.27	7.16	0.031	0.361
Caucasian - H&N cancer						
A*02:01	5	83.33	27.55	3.02	0.007	0.015
B*50:01	2	33.33	1.05	31.75	0.002	0.006
Caucasian - Pancreatobiliary cancer						
A*68:02	3	7.14	0.84	8.50	0.005	0.031
B*40:02	3	7.14	1.26	5.67	0.016	0.077
B*58:01	3	7.14	0.73	9.78	0.004	0.031
C*03:02	3	7.14	0.22	32.47	0.000	0.003
A*03:02	2	4.76	0.24	19.84	0.005	0.031
A*69:01	2	4.76	0.14	34.01	0.002	0.024
Caucasian - Sarcoma						
A*03:02	2	25.00	0.24	104.16	0.000	0.001
Caucasian - Thoracic cancer						
B*14:02	4	12.50	2.86	4.37	0.013	0.152
C*02:02	4	12.50	4.35	2.87	0.049	0.188
C*08:02	4	12.50	3.85	3.25	0.033	0.188
C*14:02	3	9.38	1.27	7.38	0.008	0.152
A*33:01	2	6.25	0.81	7.72	0.028	0.188
B*50:01	2	6.25	1.05	5.95	0.044	0.188
Hispanic - All tumors						
A*02:02	2	6.25	0.76	8.22	0.025	0.518
Hispanic - Breast adenocarcinoma						
A*02:01	2	100.00	20.95	4.77	0.044	0.044
Hispanic - GI cancer						
C*12:03	2	33.33	4.15	8.03	0.023	0.046
Hispanic - Thoracic cancer						
B*51:01	2	33.33	6.05	5.51	0.047	0.089
C*06:02	2	33.33	6.14	5.43	0.048	0.089

Supplementary Table 3. Alleles (race x cancer) Not needed

Allele	Count in Tumor Samples	Prevalence in Patient Series (%)	Prevalence in NMDP (%)	Prevalence Ratio	P value	Adjusted P value
Black or African American - Colorectal adenocarcinoma						
A*03:01	2	100.00	8.39	11.92	0.007	0.007
Black or African American - Gastric adenocarcinoma						
C*17:01	2	50.00	6.81	7.34	0.025	0.025
Black or African American - Lung adenocarcinoma						
B*35:01	2	100.00	6.89	14.51	0.005	0.005
Black or African American - Pancreatic ductal carcinoma						
C*16:01	2	50.00	9.69	5.16	0.049	0.049
Caucasian - Breast adenocarcinoma						
C*08:02	5	17.86	3.85	4.64	0.004	0.022
B*14:02	4	14.29	2.86	5.00	0.008	0.034
A*33:01	2	7.14	0.81	8.82	0.022	0.073
B*44:06	2	7.14	<0.01	>100	0.000	0.000
B*47:01	2	7.14	0.27	26.45	0.003	0.022

Caucasian - Cholangiocarcinoma

B*49:01	2	16.67	1.58	10.55	0.015	0.133

Caucasian - Colon neuroendocrine carcinoma

A*01:01	2	100.00	16.46	6.08	0.027	0.027

Caucasian - Colorectal adenocarcinoma

A*02:02	2	4.55	0.09	50.51	0.001	0.022

Caucasian - Colorectal neuroendocrine carcinoma

C*03:03	2	100.00	5.34	18.73	0.003	0.003

Caucasian - Gastric adenocarcinoma

B*49:01	2	25.00	1.58	15.82	0.007	0.013

Caucasian - Glioblastoma

B*49:01	2	25.00	1.58	15.82	0.007	0.033

Caucasian - High-grade serous carcinoma

A*25:01	2	25.00	2.10	11.90	0.011	0.034

Caucasian - Lung adenocarcinoma

B*51:01	3	25.00	4.73	5.29	0.017	0.059
C*14:02	3	25.00	1.27	19.69	0.000	0.003

Caucasian - Lung non-small cell carcinoma

C*06:02	2	50.00	9.32	5.36	0.046	0.092

Caucasian - Malignant PEComa

C*05:01	2	100.00	9.39	10.65	0.009	0.009

Caucasian - Metastatic GIST

A*03:02	2	100.00	0.24	>100	0.000	0.000
C*07:01	2	100.00	16.00	6.25	0.026	0.026

Caucasian - Oropharyngeal squamous cell carcinoma

B*50:01	2	100.00	1.05	95.24	0.000	0.000

Caucasian - Panceatic acinar cell carcinoma

A*03:01	2	100.00	13.99	7.15	0.020	0.020

Caucasian - Pancreas neuroendocrine carcinoma

B*07:02	2	100.00	13.06	7.66	0.017	0.017

Caucasian - Pancreatic ductal carcinoma

C*02:02	4	16.67	4.35	3.83	0.019	0.054
B*40:02	3	12.50	1.26	9.92	0.003	0.013
B*58:01	3	12.50	0.73	17.12	0.001	0.005
C*03:02	3	12.50	0.22	56.82	0.000	0.000
A*03:02	2	8.33	0.24	34.72	0.002	0.008
A*68:02	2	8.33	0.84	9.92	0.017	0.054
A*69:01	2	8.33	0.14	59.53	0.001	0.005

Caucasian - Prostatic adenocarcinoma

B*39:01	2	14.29	1.12	12.75	0.010	0.104

Caucasian - Prostatic sarcomatoid carcinoma

B*14:02	2	100.00	2.86	34.97	0.001	0.001

Caucasian - Renal cell carcinoma

B*18:01	3	18.75	4.43	4.23	0.032	0.200
C*12:03	3	18.75	4.86	3.86	0.040	0.200

Caucasian - Small bowel adenocarcinoma

A*03:01	2	100.00		7.15	0.020	0.020

Caucasian - Squamous cell lung carcinoma

B*40:02	2	50.00	1.26	39.68	0.001	0.002
C*02:02	2	50.00	4.35	11.49	0.011	0.011

HLA type I molecules of solid malignancies

Caucasian - Thymoma

C*07:01	2	100.00	16.00	6.25	0.026	0.026

Caucasian - Urachal adenocarcinoma

B*38:01	2	100.00	2.09	47.85	0.000	0.000

Hispanic - Breast adenocarcinoma

A*02:01	2	100.00	20.95	4.77	0.044	0.044

Hispanic - Gastric squamous cell carcinoma

A*24:02	2	100.00	13.16	7.60	0.017	0.017

Hispanic - Urothelial carcinoma

B*40:02	2	50.00	4.78	10.46	0.013	0.013