bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# HLA Gene Expression Mediates Tumor Immunogenicity and Escape

Xutong Gong[1,2] and Rachel Karchin[1,2,3,*]

1 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

2 Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD 21218, USA

3 Department of Oncology, Johns Hopkins Medicine, Baltimore, MD 21287, USA

* Corresponding author

Rachel Karchin, Ph.D.

217A Hackerman Hall

3400 N. Charles St.

Baltimore, MD USA 21218

ph: +1 410 516 5578

fax: +1 410 516 5294

karchin@jhu.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Summary**

Human Leukocyte Antigen (HLA) expression contributes to the activation of anti-tumor immunity through interactions with T cell receptors. However, pan-cancer HLA expression in tumors has not been systematically studied. In a retrospective analysis using the Cancer Genome Atlas, we quantified HLA class I and class II expression across 33 tumor types, which strongly correlated with infiltration of various immune cell types, expression of pro-inflammatory genes, and immune checkpoint markers. Patients with high HLA allelic diversity and gene expression had better survival. Immune microenvironments could be predicted using a neural network model trained on HLA expression data with varied survival outcomes. Furthermore, we identified a subset of tumors which upregulated HLA class I but not class II genes and exploited HLA-mediated escape strategies. Our results suggest the potential of using HLA expression data to predict immunogenicity. Taken together, we emphasize the crucial role of HLA upregulation in shaping prolonged anti-tumor immunity.

**Keywords**

HLA expression, immune microenvironment, immunoediting, TCGA, HLA methylation, HLA LOH, HLA allelic diversity

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Highlights

- Systematic analysis of HLA gene expression across 33 cancer types.
- A machine learning model trained on HLA expression data predicts immunogenicity.
- HLA upregulation correlates with cell-mediated immunity and improved survival.
- A subset of highly immunogenic tumors exhibit HLA-mediated tumor escape.

## Introduction

The tumor microenvironment (TME) is a hostile environment for infiltrating immune cells and poses significant challenges to their proper function. For instance, hypoxia and nutritional depletion hinder lymphocyte viability and dampen the immune response against tumors (Ghosh-Choudhary et al., 2020). Tumors may exploit immunosuppressive strategies to evade immune rejection, including the expression of immune checkpoints like programmed death-ligand 1 (PD-L1) and the recruitment of immune suppressor cells (Binnewies et al., 2018; Thommen and Schumacher, 2018). Shaped by the complex interactions between cellular components in the TME, the immune microenvironment varies significantly among patients and across tumor types. Previously, Thorsson et al. have clustered immuno-oncologic gene signatures of more than 10,000 tumors from The Cancer Genome atlas (TCGA) into six immune subtypes, C1-C6 (Thorsson et al., 2018). Among these immune subtypes, C2 (IFN-γ dominant) and C3 (inflammatory) are characterized with Type I immune response consisting of Tbet+, IFNγ-producing group 1 innate lymphoid cells, CD8 cytotoxic T cells (Tc1), and CD4 helper 1 T cells (Th1), and are associated with improved patient survival (Annunziato et al., 2015). In contrast, C4 (lymphocyte depleted) and C6 (TGF-β dominant), with low immune infiltrates and a macrophage-dominated TME, have the worst prognosis.

In recent years, significant efforts have been made to study the dynamic components in the TME to shape immunogenically hot tumors, which are associated with better response to therapy and improved prognosis (Duan et al., 2020). In the antigen presentation pathway, the interactions between peptide/major histocompatibility complex (MHC) and T cell receptors (TCR) are critical in triggering adaptive immunity (Garrido et al., 2016). The MHC is highly polymorphic, enabling the presentation of a wide variety of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

antigens on the cell surface (Radwan et al., 2020). In particular, the classical MHC I molecules (HLA-A, HLA-B, HLA-C) present endogenous antigens to CD8 T cells, and MHC II molecules (HLA-DR, HLA-DQ, HLA-DM) present exogenous antigens to CD4 helper T cells (Roche and Furuta, 2015; Wieczorek et al., 2017). Whereas classical MHC molecules directly present antigens and trigger an adaptive immune response, the less polymorphic non-classical MHC I molecules (HLA-E, HLA-G) can act as inhibitory ligands to NK cells and contribute to immune tolerance (Halenius et al., 2015). The non-classical MHC II molecules (HLA-DM, HLA-DO) act as chaperones and regulate antigen processing and loading (Mellins and Stern, 2014). In the TME, both the expression of antigen-presenting MHC molecules and the generation of tumor-specific antigens (neoantigens) are important for effective immunity. In 2020, the FDA has approved the PD-1 inhibitor, pembrolizumab, in treating tumors with high tumor mutational burden (TMB), which have been shown to correlate with hot immune signatures and effective response to immune checkpoint inhibitors (ICIs) (Fan et al., 2020; Marabelle et al., 2020; McNamara et al., 2020). A large pool of neoantigens derived from tumor mutations trigger increased T cell response, shaping the formation of hot tumors (Richman et al., 2019; Yarchoan et al., 2017). Deep learning algorithms have been built to predict high-affinity MHC-neoantigen binding, which contributes to the development of neoantigen-based cancer vaccines (Roudko et al., 2020). On the other hand, immunohistochemistry (IHC) results have shown that upregulation of Human Leukocyte Antigen (HLA) class I expression in early-stage tumors may lead to CD8 T cell-mediated anti-tumor immunity (Akazawa et al., 2019). Inducement of HLA I expression in tumors by pro-inflammatory cytokines or by the inhibition of DNA methyltransferase can also result in strong cytotoxic CD8 T cell response (Luo et al., 2018; Seliger et al., 2008).

Although these studies have demonstrated the potential role of HLA genes and neoantigen presentation in triggering immune-mediated tumor rejection, pan-cancer HLA class I and class II gene expression in various immune microenvironments has not been systematically studied. Here, we aimed to assess HLA class I and II gene expression in 33 TCGA tumor types across 6 immune subtypes. We found that HLA expression was positively correlated with immune characteristics, and this correlation was even stronger than TMB, a metric widely used to evaluate immunogenicity (Strickler et al., 2021). HLA

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

allelic diversity in tumors with high HLA expression was associated with improved survival. Moreover, while upregulated HLA I expression was observed in most tumors with Type I immunity (immune subtypes C2 and C3), upregulated HLA II expression was uniquely seen in the inflammatory subtype (C3), and was correlated with helper T cell-mediated immunity and improved survival. In contrast to C3, C2 tumors with upregulation of HLA I but not HLA II genes showed heavy DNA methylation near HLA I genes and HLA LOH under strong immunity, suggesting that they had undergone immunoediting and exploited escape mechanisms. We also took advantage of the variation in HLA expression levels among different immune microenvironments, and trained machine learning models on HLA class I and class II gene expression data to predict the cytolytic activity and immune subtypes with high accuracy. Our results demonstrate the crucial role of HLA class I and II gene upregulation in triggering effective immunity and suggest that HLA gene expression may be a more accurate metric than TMB to evaluate tumor immunogenicity.

## Results

### HLA genes Are Differentially Expressed in Most TCGA Tumor Types, with the Expression of Class I Genes Higher than Class II.

We first assessed HLA class I and class II gene expression in all 33 TCGA tumor types. While most HLA class I genes are expressed ubiquitously, HLA class II expression is restricted to professional antigen-presenting cells (APCs) (Dendrou et al., 2018). Overall, most HLA class I genes demonstrated higher expression than class II genes (Figure 1A). Consistent with the fact that HLA-G and HLA-DO molecules have more tissue-restricted distribution (Carosella et al., 2003; Poluektov et al., 2013), we observed the lowest expression in *HLA-G* and HLA-DO (*HLA-DOA*, *HLA-DOB*) among class I and class II genes, respectively.

We also evaluated HLA differential expression among 24 TCGA tumor types with RNA-seq data of matched normal tissues. Overall HLA class I and class II gene expression were represented as the geometric mean of transcript levels of *HLA-A*, *HLA-B*, *HLA-C*, and HLA-DP (*HLA-DPA1*, *HLA-DPB1*), HLA-DQ (*HLA-DQA1*, *HLA-DQB1*), HLA-DR (*HLA-DRA*, *HLA-DRB1*), respectively. We observed that 10 tumor types had

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

significantly upregulated HLA class I expression and 2 had downregulated class I expression, while 6 tumor types had upregulated class II expression and 6 had downregulated class II expression (Figure 1B-C). Among them, KIRC showed the most upregulation of HLA class I genes and class II genes. LUSC and PAAD had the most downregulation of class I and class II genes, respectively. This variability in HLA gene differential expression is likely due to tissue of origin and reflects a difference in tumor-specific mechanisms to regulate HLA gene expression. Also, we observed no apparent correlation between HLA expression and TMB (Spearman correlation, HLA-I: R=0.18, HLA-II: R=0.045) or neoantigen loads (HLA-I: R=0.17, HLA-II: R=0.079) (Figure S1C-F). This suggests that HLA gene expression could be explored as a metric independent of TMB or neoantigen loads to evaluate the immune microenvironment.

**HLA Class I and Class II Gene Expression Correlate with Tumor Immune Characteristics and Predict Cytolytic Activity.**

We then evaluated whether HLA expression was associated with the immune characteristics in tumors, including the relative abundance of infiltrating immune cells, proinflammatory gene signatures, and immune checkpoint signatures. Across 32 tumor types, both HLA class I and class II expression demonstrated a strong correlation with the relative abundance of CD8 cytotoxic T cells (Tc1), activated CD4 memory T cells (Tmem), helper 1 T cells (Th1), helper 17 T cells (Th17), regulatory T cells (Tregs), activated natural killer (NK) cells, macrophage 1 (M1), and a negative correlation with helper 2 T cells (Th2) (Figure 2A). They were also positively correlated with expression of pro-inflammatory genes (*IDO1*, *CXCL9*, *CXCL10*, *STAT1*, *IFNG*), as well as immune checkpoints (*PDCD1*, *CD274*, *CTLA4*, *LAG3*, *TIGIT*, *HAVCR2*). Intriguingly, HLA gene expression displayed a stronger correlation with the immune characteristics than TMB.

We next explored the relationship between HLA expression levels and cytolytic score (CYT), a metric previously proposed by Rooney et al. to represent immune cytolytic activity (Rooney et al., 2015). While TMB showed weak correlation with CYT (Spearman correlation, R=0.2, p<2.2e-16), both HLA class I and class II expression were strongly associated with CYT (class I: R=0.62, class II: R=0.74, p<2.2e-16). In particular, among all classical and non-classical HLA I genes, *HLA-E* demonstrated the highest correlation

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with CYT (R=0.69), while *HLA-G* showed the lowest (R=0.43) (Figure 2B). Among classical and non-classical HLA II genes, HLA-DQ and HLA-DR showed the highest correlation with CYT (R=0.73), while HLA-DM had the lowest (R=0.64). An unsupervised principal-component analysis (PCA) of all HLA I and HLA II expression data across 32 tumor types showed that tumors with high cytolytic activity (top quantile) clustered together (Figure 2C). We also trained a random forest classifier using as features classical and non-classical HLA I and HLA II gene expression, and were able to classify samples showing strong cytolytic activity with high accuracy (AUC=0.93) (Figure S2A). Consistent with the strength of their correlation with CYT, *HLA-DQA1*, *HLA-E*, and *HLA-DRA1* demonstrated the highest relative feature importance as reported by the classifier, while *HLA-G* had the lowest (Figure S2B). Taken together, HLA class I and II expression were strongly associated with tumor immunogenicity in our study. With a random forest classifier, we demonstrated the potential of using HLA gene expression data to predict cytolytic activity.

**HLA Allelic Diversity in Tumors with High HLA Expression Correlates with Improved Survival.**

Since HLA class I molecules expressed on tumors bind antigens in their peptide-binding grooves to trigger T cell response, a higher diversity of a patient's HLA alleles could potentially allow a larger pool of neoantigens to be presented and hence increase the likelihood of the tumor being rejected by immunity (Pierini and Lenz, 2018). We therefore asked if HLA allelic diversity is correlated with tumor immunogenicity. To this end, we collected computational HLA typing for 8859 patients across 30 tumor types and estimated the biophysiochemical diversity of each patient's HLA class I alleles with the HLA-I evolutionary divergence (HED) score (Chowell et al., 2019). A patient homozygous at all three HLA I loci has a HED score of 0. The greater the functional diversity of HLA I alleles, the higher the patient's HED score (Figure S3A). The median HED of all patients was 6.66, with a low being 0 and a high being 11.54 (Figure 3A). However, we did not observe any correlations between HLA I allelic diversity and infiltration levels of different immune cell types, pro-inflammatory genes, or immune checkpoints in the TME (Figure S3B).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We reasoned that the high allelic diversity might not contribute to the presentation of a large neoantigen pool if HLA expression is low. Hence, we proposed that high allelic divergence of HLA I alleles in tumors with high HLA I expression may more efficiently present neoantigens and lead to tumor rejection. To test this hypothesis, we divided all patients into four quadrants by their HED score and HLA I expression (Figure 3B). Patients within Quadrant I with high HLA I expression and high HED scores demonstrated significantly better survival (log-rank test, HR=0.75, p=0.0015) (Figure 3C). Moreover, among the four quadrants, either high HLA I (Quadrant IV) or high HED (Quadrant II) alone did not show a significant correlation with survival (Figurer 3D). We concluded that high HLA I allelic diversity in tumors with high HLA I expression is associated with better survival.

**HLA Class I and Class II Genes Show Differential Expression Patterns across Various Immune Microenvironments with Different Survival Outcomes.**

Next, we asked if HLA genes were differentially regulated within various immune microenvironments in tumors. To this end, we evaluated HLA class I and II differential gene expression in tumors compared to their matched normal samples across six immune subtypes (C1-C6) previously clustered from the immune signatures in the TME (Thorsson et al., 2018). HLA class I expression was significantly upregulated in C2 (IFN-γ dominant) and C3 (inflammatory) (Wilcoxon rank-sum, p<2e-16), and was downregulated in C1 (wound-healing) (p=0.0091) and C5 (immunologically quiet subtype) (p=0.024) (Figure 4A, 4C). In comparison, HLA class II expression was uniquely upregulated in C3 (p=4.0e-13), while significantly downregulated in C1 (p=8.8e-12), C4 (lymphocyte depleted) (p=0.0027), and C5 (p=0.030) (Figure 4B, 4C).

We considered if HLA gene expression data could be used to predict the immune subtype of a given tumor sample. An artificial neural network model was trained on the transcriptomic data of HLA class I and class II genes from 30 tumor types to predict the six immune subtypes. We did not include HLA expression data from normal samples, because there was only one case available in C5. The model was capable to predict the immune subtype of tumors with high accuracy (micro-average AUC=0.94) (Figure 4D). Among the six immune subtypes, tumors in C3 and C5 were classified with the highest

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

accuracy (C3: AUC=0.94, C5: AUC=0.99), while tumors in C6 were classified with the lowest accuracy (C6: AUC=0.88). Overall, the results demonstrated that HLA class I and class II gene expression were highly variable and had the potential to predict different types of immune microenvironments in tumors.

Since HLA class I molecules interact with CD8 T cells, while class II molecules interact with CD4 helper T cells, we next evaluated if HLA class I and class II gene upregulation were correlated with helper vs. cytotoxic T cell-mediated adaptive immunity. Patients were divided into 4 groups according to their HLA fold change (tumor/normal ratio): high HLA-I and HLA-II fold change, high HLA-I and low HLA-II fold change, low HLA-I and high HLA-II fold change, and low HLA-I and HLA-II fold change. Groups with high HLA-II fold change demonstrated significantly elevated Th1 cell fractions than the groups with low HLA-II fold change (Figure 4E). In comparison, high HLA I fold change only weakly correlated with increased Th1 cell fractions. On the other hand, high HLA-I and HLA-II fold change both correlated with increased cytolytic activity of Tc1 cells, yet it was more significantly elevated in groups with high HLA-II fold change (Figure S4A). CD8 T cell fractions showed similar trends among the 4 groups (Figure S4B). While both the upregulation of HLA class I and class II genes were associated with cytotoxic T cell-mediated immunity, our results suggest a strong correlation between HLA class II upregulation and helper T cell-mediated immunity, which might lead to more effective T cell priming and stronger cytolytic activity in the TME.

We then asked whether HLA expression fold change could predict patient survival. Compared to cytolytic activity (Cox proportional hazards, HR=1.40, p=0.044), HLA class II fold change demonstrated a significant correlation with better survival (HR=0.60, p=0.004) (Figure 4F). Across the four groups defined earlier based on HLA fold change, two groups with high HLA-II fold change were both correlated with improved survival compared to the low HLA-I low HLA-II reference group (Cox proportional hazards, high HLA-I HLA-II: HR=0.68, p=0.018; low HLA-I high HLA-II: HR=0.74, p=0.22) (Figure S4D). Also, among individual tumor types, high HLA-II fold change in LUAD and BRCA was associated with significantly better survival (Log-rank test, LUAD: HR=0.27, p=0.0061; BRCA: HR=0.42, p=0.036) (Figure S4C). High HLA-II expression fold change in other tumor types (THCA, KIRP, UCEC, COAD, ESCA, KIRRC, LIHC, HNSC, BLCA) also

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

correlated with better survival, though the correlation was not statistically significant. Of note, in contrast to HLA fold change, HLA expression in tumor samples alone showed no correlation with survival (Cox proportional hazards, HLA I: HR=1.01, p=0.774; HLA II: HR=0.92, p=0.064). The results suggest that HLA fold change may be more important than HLA expression in determining the survival outcome. We concluded that upregulation of HLA class II genes was correlated with effective immune response and rejection of tumors, potentially contributing to a better prognosis.

**DNA Methylation near HLA Class I Genes Favored under Strong Cytolytic Activity Can Be Alleviated by Class II Gene Upregulation.**

DNA methylation near HLA genes has previously been shown to downregulate gene expression and dampen the immune response as a potential HLA-mediated tumor escape strategy (Luo et al., 2018). We first evaluated the effect of differentially methylated HLA class I genes on immunity. To this end, we calculated the mean β-scores across all CpG sites near HLA I genes and identified aberrant methylation states in tumors compared to the matched normal samples. Overall, there existed a significant inverse correlation between HLA I expression and methylation levels (Spearman correlation, R=-0.61, p<2.2e-16). Tumors with high cytolytic activity were more hypomethylated near HLA I genes, whereas those with low cytolytic activity were hypermethylated (Figure 5A). Moreover, we observed varied HLA I methylation levels across the immune subtypes (Kruskal-Wallis, p<2.2e-16) (Figure 5B). Consistent with low HLA I gene expression observed in C1 (wound healing) and C4 (lymphocyte depleted), HLA I methylation was elevated in C1 and C4. Surprisingly, tumors in C2 (IFN-γ dominant) demonstrated significantly higher HLA I methylation compared to tumors in C3 (inflammatory) (Wilcoxon rank-sum, p=9.5e-6), though their HLA I expression were comparable (Figure S5A). Also, HLA I methylation and expression fold change were the most negatively correlated in C3 (Spearman correlation, R=-0.70, p<2.2e-16), and the least correlated in C2 (R=-0.26, p=0.025) (Figure S5B). This suggests that while tumors in both C2 and C3 upregulated HLA I expression, C2 tumors were at the same time heavily methylated near HLA I genes.

We specifically focused on tumors in C2 and C3, which were both defined by Type I immune response (Thorsson et al., 2018). The immune subtype of these tumors could

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

be predicted using a logistic regression classifier based on the mean β-scores across all CpG sites near HLA I genes (AUC=0.93) (Figure S5C). This result supports the hypothesis that HLA genes are indeed differentially methylated among tumors in C2 and C3, and methylation is predictive of these immune subtypes. In particular, C2 tumors with high cytolytic activity showed significantly heavier HLA I methylation than C3 tumors (Wilcoxon rank-sum, p=2.5e-16), whereas HLA I methylation was comparable between those with low cytolytic activity (Figure 5C). Since HLA class II genes were significantly upregulated in C3 (Figure S5A), we asked if higher HLA II fold change would alleviate HLA I methylation in tumors. To this end, we divided the samples of all immune subtypes into two groups based on HLA II fold change and assessed aberrant methylation states near HLA I genes across the groups. While most tumors with low HLA-II fold change exhibited hypermethylated HLA I genes, more tumors with high HLA-II fold change showed hypomethylated HLA I genes (Two-sided Fisher's Exact Test: p=9.456e-10) (Figure 5D). Hence, we suggest that a subset of tumors with Type I immune response have increased HLA I methylation to evade immune recognition. Our results support that HLA class II gene upregulation may be a mechanism to reduce tumor escape and trigger effective immunity in the TME.

**HLA Loss of Heterozygosity (LOH) Is Associated with Worse Survival but Can Be Counter Balanced by High HLA Expression.**

Lastly, we evaluated how HLA LOH, another potential HLA-mediated tumor escape strategy, correlated with HLA expression and survival among TCGA LUAD and BRCA patients. To this end, we inferred HLA class I alleles' estimated raw copy number and LOH with Loss of Heterozygosity in Human Leukocyte Antigen (LOHHLA) in patients fully heterozygous at all 3 HLA I loci (McGranahan et al., 2017). Overall, HLA LOH correlated with worse survival (Log-rank test, HR=1.36, p=0.025) (Figure S6A). Among BRCA patients, HLA LOH did not show a significant correlation with survival (HR=0.66, p=0.11), whereas in LUAD patients HLA LOH was correlated with worse survival (HR=1.42, p=0.020) (Figure S6B-C). The results suggest that the negative impact of HLA LOH in tumor eradication and patients' survival likely varies across tissues and cancer types.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Next, we proposed that tumors with HLA LOH under strong immune selection were favored to evade recognition. Supporting our hypothesis, the proportion of tumors with high cytolytic activity harboring HLA LOH was significantly higher than that with low cytolytic activity (Two-sided Fisher's exact Test, p=9.52e-8) (Figure 6A). Across the five immune subtypes, tumors in C2 (IFN-γ dominant) and C6 (TGF-β dominant) demonstrated the highest percentage of HLA LOH in BRCA and LUAD (BRCA C2: 24%, C6: 24%; LUAD C2: 43%, C6: 42%), while tumors in C3 (inflammatory) had the lowest (BRCA C3: 13%, LUAD C3: 23%;) (Figure 6B). Consistent with our observations of HLA I methylation, though the tumors in both C2 and C3 were characterized with Type I immune response, a larger fraction of C2 tumors demonstrated HLA LOH (Two-sided Fisher's exact Test, BRCA: p=9.52e-3, LUAD: p=8.96e-4), suggesting potential immunoediting and tumor escape.

Surprisingly, we did not find any correlation between the mean HLA I copy number and overall expression (Spearman correlation, R=0.098, p=0.0017), or between *HLA-A*, *HLA-B*, or *HLA-C* gene copy number and expression individually (Figure 6C, S6D-F). Moreover, there was no significant difference in HLA I expression between tumors with HLA LOH and those without LOH (Figure S6G), in contrast to what was previously proposed by Garrido *et al.* (Garrido et al., 2016). We also did not find association between HLA I expression fold change and LOH (Figure S6H). This suggests that HLA I expression at the RNA level is uncorrelated with HLA I loss at the DNA level. HLA LOH most likely does not downregulate HLA expression as a means of facilitating tumor escape.

Based on the notion that HLA LOH may lead to a reduced pool of neoantigens being presented on the cell surface, we explored if high HLA gene expression may compensate for the allelic loss. To this end, we divided all fully heterozygous patients into groups with high HLA I expression and low HLA I expression. The group with low HLA I expression showed a significant correlation between HLA LOH and worse survival (Log-rank test, HR=1.52, p=0.024). Interestingly, HLA LOH was not correlated with worse survival in the group with high HLA I expression (HR=1.19, p=0.40) (Figure 6D). We concluded that the negative impact of HLA LOH on tumor eradication might be alleviated by high HLA expression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Discussion

Based on the notion that HLA class I and II molecules which present antigens are fundamental for triggering anti-tumor immunity, we aimed to address how HLA expression shapes various immune microenvironments in a pan-cancer analysis of tumor types from TCGA. We quantified overall HLA expression using a simple expression metric of HLA class I and class II genes. Our analysis suggests that elevated HLA expression is likely to drive the formation of immunogenic tumors. Our results also demonstrate the predictive power of HLA gene expression data on the immune microenvironment and survival.

We first considered if HLA expression correlates with tumor immunogenicity. HLA class I and class II gene expression demonstrated a strong positive correlation with various infiltrating immune cells, pro-inflammatory genes, and cytolytic activity. They were also positively associated with immune checkpoint markers, implying active self-tolerance mechanisms in the TME against T cell cytotoxicity. Of note, though TMB is frequently used to predict immunogenicity and survival, TMB showed much weaker correlation with the immune characteristics compared to HLA I and HLA II expression. We reasoned that although immune cells are more likely to recognize tumors with high mutational burden, neoantigens generated in tumors with high TMB but low HLA expression could not be presented to trigger immune activities. Hence, high HLA expression is necessary for effective antigen presentation, and it likely plays a greater role in inducing immune activity than tumor mutations. In recent years, neoantigen-based personalized cancer vaccines and adoptive cell therapy (ACT) are among the advanced immunotherapies to trigger neoantigen-specific T cell activity (Blass and Ott, 2021). Current *in silico* algorithms predict neoantigens based solely on a patient's HLA types and tumor mutations (Lauss et al., 2017; Yamamoto et al., 2019). Here, we suggest that HLA expression is a crucial component that should be incorporated in neoantigen prediction algorithms. We further demonstrated that the expression levels of classical and non-classical HLA genes could be used to train a probabilistic classifier to predict tumors with high cytolytic activity. Interestingly, although the non-classical HLA-E molecule can act as an inhibitor against NK cells (Joosten et al., 2016; Liu et al., 2020), it showed the strongest correlation with cytolytic activity among class I genes and was also the most important feature in the Random Forest classification. HLA-E is known to play a role in both innate and adaptive

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

immunity. Whereas it inhibits NK-mediated lysis by interacting with the CD94/NKG2A complex on NK cells, it can present a specific pool of peptides to activate HLA-E restricted CD8 T cells independent of the classical HLA-restricted T cells (Joosten et al., 2016). Future efforts can be made to explore the specific roles of HLA-E molecules in anti-tumor adaptive immunity. Our results also suggest that inducing HLA expression in tumors systematically may trigger prolonged anti-tumor immunity and enhance the efficacy of immunotherapies.

In addition to HLA gene expression, we also considered HLA allelic diversity. Tumors with higher HLA class I allelic diversity may present a wider pool of biochemically different neoantigens, encouraging immune infiltration and activation (Chowell et al., 2019). Yet, we did not observe any significant correlations between HLA allelic diversity and various immune characteristics or cytolytic activity. Interestingly though, we found that high HLA allelic diversity in tumors expressing high levels of HLA genes correlated with improved survival. This suggests that high HLA allelic diversity may contribute more to tumor elimination through the presentation of a diverse neoantigen pool when HLA expression is high. Yet, the role of HLA allelic diversity alone in triggering immune activities is limited. One possible explanation is that high HLA gene expression despite low allelic diversity may still trigger immunity by presenting sufficient amounts of neoantigens even if the presented neoantigens are of low diversity. In contrast, high HLA allelic diversity, though having the potential to present a larger group of neoantigens, cannot compensate for low HLA expression in cells with reduced neoantigen presentation.

We further evaluated HLA differential expression patterns in tumors compared to their matched normal tissues across six immune subtypes. A neural network model trained on classical and non-classical tumor HLA class I and class II expression was used to predict the immune subtypes of tumors with high accuracy. Tumors with Type I immunity (C2 and C3) showed higher HLA class I expression fold change than non-immunogenic tumors. Interestingly, tumors in C3 had uniquely upregulated HLA II expression, which was not observed in other immune subtypes. While HLA I molecules expressed on tumors present endogenous neoantigens to CD8 cytotoxic T cells, most HLA II molecules expressed on professional APCs present exogenous neoepitopes to

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CD4 helper T cells (Axelrod et al., 2019). CD4 T cell priming is essential in activating CD8 cytotoxic T cells and forming prolonged memory (Tay et al., 2021; Thibodeau et al., 2012). As expected, high HLA II fold change was correlated with helper T cell-mediated immunity, as well as elevated cytolytic activities and CD8 Tc1 infiltration, suggesting effective T cell priming and activation of cytotoxic T cell-mediated immunity. Moreover, high HLA II fold change was correlated with improved survival. Taken together, the results support the crucial role of HLA class II upregulation in T cell priming and generation of strong cytotoxicity for tumor elimination. The expression data of HLA class I and class II genes may potentially be used to classify immune microenvironments and survival. Future work could include training classifiers using HLA expression data from ICI-treated cohorts to predict response to therapy.

Strong immune selective pressure can lead to immunoediting and favor tumors that exploit escape strategies to evade immune recognition. Here, we considered DNA methylation near HLA I genes and HLA LOH as potential HLA I-mediated immune evasion strategies. Tumors with low cytolytic activity showed more hypermethylated HLA genes, hence confirming that tumors having heavy HLA methylation could dampen the immune response. Interestingly, among tumors with dominant Type I immunity (C2 and C3), C2 tumors under strong cytolytic activity were more heavily methylated near HLA genes than those in C3. We reasoned that HLA II upregulation may effectively reduce tumor escape by means of HLA I gene methylation. High HLA-II fold change was associated with more hypomethylated HLA I genes, whereas low fold change was associated with hypermethylated genes. Hence, we suggest that, in addition to its role in triggering effective immunity, HLA II upregulation reduces immunoediting through HLA gene methylation. In contrast, tumors with high HLA I fold change alone (represented by C2) may experience strong T cell cytotoxicity in the absence of helper T cell activation, leading to immunoediting and eventually tumor escape.

Finally, we assessed the role of HLA LOH in anti-tumor immunity. Overall, HLA LOH was correlated with worse survival. A significantly higher fraction of tumors with high cytolytic activity exploited HLA LOH, which supports the hypothesis that tumors with HLA LOH are favored under strong immune selection to evade elimination. Consistent with our previous finding, we again observed high fractions of tumors in C2 harboring HLA LOH,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

suggesting that C2 tumors might have been immunoedited. While previous studies propose that HLA I allelic loss results in tumors evading immune recognition potentially through loss of allele-specific expression (Garrido et al., 2016; McGranahan et al., 2017), we did not observe any correlation between HLA LOH and HLA downregulation. This suggests that HLA LOH most likely results in tumor evasion not through downregulating the overall HLA expression, but by other mechanisms such as reducing the presentation of the most immunogenic neoantigens on the cell surface. Furthermore, we observed that the negative impact of HLA LOH on survival could be alleviated by increased expression of HLA genes. Since the overall HLA expression was not correlated with HLA LOH, this finding suggests a new way of rescuing the anti-tumor immune recognition in patients with HLA LOH. Although HLA LOH, unlike HLA gene methylation, is an irreversible hard lesion at the gene level, therapies that enhance the expression of HLA genes at the RNA and protein levels could be exploited to compensate for the reduced pool of neoantigens being presented. Possible future directions include inducing HLA gene expression with pro-inflammatory cytokines like IFN-γ in tumors with HLA LOH to evaluate if elevated gene expression rescues immune recognition and tumor elimination *in vivo*.

Our study has several limitations. First, the HLA expression metric we used was transcript-based and did not take into account post-translational modification, thus might not accurately capture the final quantity of HLA molecules expressed on the cell surface. We also did not consider specific neoantigen pools in tumors that could bind to the HLA molecules and be presented to T cells. Nonetheless, our analysis of TCGA samples has emphasized the important role of HLA expression in facilitating the formation of prolonged immune response in tumors with improved survival. We also demonstrate the power of HLA expression from RNA-seq data in predicting various immune microenvironments. We anticipate that in the future the effect of HLA expression on the immune signatures could be investigated in patients receiving ICIs, which might enable us to more accurately predict individuals' response to therapy.

**Acknowledgments**

The results published here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Author Contributions

XG conceived the study and performed the experiments, RK supervised the study, XG and RK wrote the paper.

## Declaration of Interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Figure Legends

**Figure 1. HLA Expression Is Differentially Expressed across Tumor Types.** (A) Expression of classical and non-classical HLA class I and class II genes in tumor samples across 33 TCGA tumor types. Expression of gene X is represented as log-transformed Fragments per Kilobase of Transcripts per Million (FPKM). (B) HLA class I expression represented as the geometric mean of *HLA-A*, *-B*, *-C* in FPKM, and (C) HLA class II expression represented as the geometric mean of HLA-DP *(HLA-DPA1*, *HLA-DPB1)*, -DQ *(HLA-DQA1*, *HLA-DQB1)*, -DR *(HLA-DRA*, *HLA-DRB1)* in FPKM, in tumor-normal pairs. Boxes in the box plot represent interquartile ranges and horizontal lines represent $5^{th}$-$95^{th}$ percentile ranges. p values between the pairwise tumor-normal samples are calculated by Wilcoxon rank-sum test and are unadjusted. Asterisks denote significant FDR-adjusted (Benjamini & Hochberg method) p values (*p<0.05; **p<0.01; ***p<0.001).

**Figure 2. HLA Class I and Class II Gene Expression Are Associated with Immune Activity.** (A) Spearman correlations between HLA class I, class II expression, TMB, and infiltrating immune cells (orange), pro-inflammatory genes (blue), and immune checkpoints (green) across 32 TCGA tumor types. Each row within a horizontal block represents a distinct tumor type. CD8 Tc1: CD8 cytotoxic T cells; CD4 Tmem: activated CD4 memory T cells; Th1: T helper 1 cells; Th2: T helper 2 cells; Th17: T helper 17 cells; Tregs: regulatory T cells; NK: natural killer cells; M1: Macrophage 1; M2: Macrophage 2. (B) Spearman correlations between the expression of classical HLA class I *HLA-A*, *-B*, *-C*, non-classical *HLA-E*, *-G*, classical HLA class II HLA-DP (*HLA-DPA1*, *HLA-DPB1*), -DQ (*HLA-DQA1*, *HLA-DQB1*), -DR (*HLA-DRA*, *HLA-DRB1*), non-classical HLA-DM (*HLA-DMA*, *HLA-DMB*), -DO (*HLA-DOA*, *HLA-DOB*), and cytolytic activity (CYT). p values are calculated from Spearman's rank correlation and are unadjusted. (C) Principal Component Analysis (PCA) on 32 tumor types by their expression of classical and non-classical HLA class I and class II genes. Principal Component 1 (Dim1) and Principal Component 2 (Dim2) are shown. High cytolytic activity (CYT high) is defined as the top 75% of all tumors' mean CYT.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 3. HLA-I Evolutionary Divergence (HED) Score in Tumors with High HLA Expression Is Associated with Improved Survival.** (A) HED score distribution among 8801 tumor samples from 30 tumor types. (B) HED score vs. HLA class I expression. High cytolytic activity (CYT high) is defined as the top 75% of all tumor samples. The vertical line represents the median HED score, and the horizontal line represents the $90^{th}$ percentile HLA class I expression. Four quadrants are defined accordingly. (C) Progression-free survival (PFS) by high HLA I expression and HED (Quadrant I) and low HLA I expression or HED (Quadrant II, III, IV). p-value, HR, and 95% confidence interval are calculated by the Log-rank test. (D) The hazard ratio (HR) of the effects of HED and HLA I expression on PFS, with low HED low HLAI as reference. Squares represent the HR and horizontal bars represent 95% confidence intervals. HR and p values are calculated from Cox Proportional-hazards Model. HED: HLA-I evolutionary divergence score; HLAI: HLA class I expression.

**Figure 4. HLA Expression Predicts the Immune Subtypes and Is Associated with Improved Survival.** (A) HLA class I expression, and (B) HLA class II expression between tumors and matched normal samples across six immune subtypes, C1-C6. Boxes in the box plot represent interquartile ranges. P values between pairwise tumor-normal samples are calculated by Wilcoxon rank-sum test and are unadjusted. Asterisks denote significant FDR-adjusted (Benjamini & Hochberg method) p values (*$p<0.05$; **$p<0.01$; ***$p<0.001$). C1: wound-healing; C2: IFN-γ dominant; C3: inflammatory; C4: lymphocyte-depleted; C5: immunologically quiet; C6: TGF-β dominant. (C) Expression fold change of classical and non-classical HLA class I and class II genes between tumors and matched normal samples across six immune subtypes. The heatmap is scaled by columns. (D) Receiver Operating Characteristic (ROC) curve of a neural network model. The model is trained on log-transformed expression levels for classical and non-classical HLA class I and class II genes and predicts the immune subtypes of tumors. Micro-average area under the curve (AUC) as well as AUC for individual immune subtype predictions are presented. (E) Helper 1 T cells (Th1) signatures across 4 groups based on their HLA class I and class II

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expression fold change. High HLA I and high HLA II are defined as the top 50% of all tumor samples, respectively. Pairwise p values are calculated by Wilcoxon rank-sum test and are unadjusted. p-value across all 4 groups is calculated by the Kruskal-Wallis test. (F) The hazard ratio (HR) with CYT, HLA I fold change, HLA II fold change as explanatory variables, and PFS as the dependent variable. Squares denote the HR and horizontal bars represent 95% confidence intervals. HR and p-value are calculated from Cox Proportional-hazards Model. CYT: cytolytic activity; HLA I: HLA I expression fold change; HLA II: HLA II expression fold change.

**Figure 5. HLA Class I Methylation Is Favored in A Subset of Hot Tumors and Dampens Immune Activity.** (A) A kernel density plot of HLA class I methylation levels and expression across groups with high (top 75%) or low cytolytic activity. Methylation level was represented as the tumor:normal ratio of mean β-scores across all CpG sites near HLA I genes. Correlation coefficient and p values are calculated from Spearman's rank correlation. CYT: cytolytic activity; Hypomethylation: HLA I methylation level < 0.75; Hypermethylation: HLA I methylation level > 1.25. (B) HLA class I methylation levels across five immune subtypes (excluding C5 with no available data). Boxes in the violin plot represent interquartile ranges. p values between pairwise immune subtypes are calculated by Wilcoxon rank-sum test and are unadjusted. p-value across six immune subtypes is calculated by the Kruskal-Wallis test. (C) Left panel: the Spearman correlations between cytolytic activity (CYT) and HLA class I methylation in C2 and C3 tumors. p values within each immune subtype are calculated from Spearman's rank correlation. Right panel: HLA class I methylation in tumors with high cytolytic activity (top 50%) across C2 and C3. p-value is calculated by Wilcoxon rank-sum test and is unadjusted. (D) Association between HLA II fold change and aberrant methylation states near HLA I genes. Two groups are defined based on HLA II fold change in all samples. Numbers on pie charts indicate the number of tumor samples showing hypermethylated or hypomethylated HLA I genes. p-value is calculated from two-sided Fisher's Exact Test.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Figure 6. HLA LOH Is Exploited by Hot Tumors with Worse Patients' Survival.** (A) Association between cytolytic activity and HLA LOH. Two groups are defined based on cytolytic activity in Tumors (high CYT: top 50%). Numbers on pie charts indicate the number of tumors harboring HLA LOH. p-value is calculated from two-sided Fisher's Exact Test. (B) Tumor samples harboring HLA LOH across five immune subtypes (excluding C5 with no available data) in TCGA BRCA and LUAD, respectively. Percentage on top of each bar represents the percent of tumor samples with HLA LOH within the corresponding immune subtype. (C) Spearman correlation between the mean copy number of HLA class I genes (*HLA-A*, *HLA-B*, *HLA-C*) and HLA class I expression. The correlation coefficient and p-value are calculated from Spearman's rank correlation. The allelic loss represents the number of lost HLA I alleles (raw copy number<0.5 and p<0.05 from LOHHLA output). (D) The hazard ratio (HR) of the effect of HLA class I expression on PFS. Circles represent the HR and horizontal bars represent 95% confidence intervals. HR and p values are calculated from the log-rank test. High HLAI: high HLA class I expression (top 50%); low HLAI: low HLA class I expression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Methods and Materials

### Study Cohort and Data Acquisition

The samples analyzed in this study include primary tumors of 33 TCGA tumor types. Gene-level RNA-Seq expression data, both tumor and matched normal samples, whole-exome sequencing Bam files of LUAD tumor/normal pairs, and Methylation Array files from Illumina Infinium Human Methylation 450k (HM450) were obtained from GDC Data Portal (https://portal.gdc.cancer.gov/).

### Expression of HLA Genes, Pro-inflammatory Genes, and immune checkpoints in Tumor and Normal Samples

FPKM reads at *HLA-A* (ENSG00000206503), *HLA-B* (ENSG00000234745), *HLA-C* (ENSG00000204525), *HLA-E* (ENSG00000204592), *HLA-G* (ENSG00000204632), *HLA-DRA* (ENSG00000204287), *HLA-DRB1* (ENSG00000196126), *HLA-DQA1* (ENSG00000196735), *HLA-DQB1* (ENSG00000179344), *HLA-DPA1* (ENSG00000231389), *HLA-DPB1* (ENSG00000223865), *HLA-DMA* (ENSG00000204257), *HLA-DMB* (ENSG00000242574), *HLA-DOA* (ENSG00000204252), and *HLA-DOB* (ENSG00000241106) were extracted, and the normalized log 2 value of FPKM reads at each locus was calculated to represent the expression of each gene. Expression of HLA-DP, HLA-DQ, HLA-DR, HLA-DM, HLA-DO were represented as

$$Expression = \frac{log_2(\alpha + 1) + log_2(\beta + 1)}{2}$$

where $\alpha$ and $\beta$ represents the transcript levels of $\alpha$ and $\beta$ subunits in FPKM, respectively. Overall tumor and normal HLA class I gene expression were measured as the geometric mean of *HLA-A*, *HLA-B*, and *HLA-C* expression in FPKM. Tumor and normal HLA class II gene expression was measured as the geometric mean of *HLA-DRA*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, and *HLA-DPB1* expression in FPKM.

Expression of pro-inflammatory genes was represented as the log 2 value of *IDO1* (ENSG00000131203), *CXCL9* (ENSG00000138755), *CXCL10* (ENSG00000169245), *STAT1* (ENSG00000115415), *IFNG* (ENSG00000111537) in FPKM.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Expression of immune checkpoints was represented as the log 2 value of *PDCD1* (ENSG00000188389), *CD274* (ENSG00000120217), *CTLA4* (ENSG00000163599), *LAG3* (ENSG00000089692), *TIGIT* (ENSG00000181847), *HAVCR2* (ENSG00000135077) in FPKM.

### HLA Gene Expression Fold Change in Tumors

HLA class I gene expression fold change was represented as $\frac{tumor\ HLA\ I\ expression}{normal\ HLA\ I\ expression}$, where HLA I expression was represented as the geometric mean of *HLA-A*, *HLA-B*, and *HLA-C* expression in FPKM. HLA class II gene expression fold change was represented as $\frac{tumor\ HLA\ II\ expression}{normal\ HLA\ II\ expression}$, where HLA II expression was represented as the geometric mean of *HLA-DRA*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, and *HLA-DPB1* expression in FPKM. TCGA ACC, DLBC, LAML, LGG, MESO, OV, TGCT, UCS, and UVM were excluded since RNA-seq data of their matched normal tissues were not available.

### HLA Typing

The 4-digit HLA type for each sample was acquired from Thorssen et al. (Thorsson et al., 2018), who used OptiType software (Szolek et al., 2014) taking tumor RNA-seq fastq files as input. The HLA type of a subset of TCGA LUAD samples with no available HLA typing were inferred using HLA-HD (Kawaguchi et al., 2017) taking whole-exome sequencing BAM files as input.

### HLA Gene Methylation in Tumor and Normal Samples

DNA methylation of HLA genes was measured as $\frac{tumor\ HLA\ methylation\ level}{normal\ HLA\ methylation\ level}$, where HLA methylation level was represented as the mean β-scores across all CpGs near HLA I gene loci from the HM450 array. For each sample, HLA genes were hypermethylated if $\frac{tumor\ HLA\ methylation\ level}{normal\ HLA\ methylation\ level} > 1.25$, or hypomethylated if $\frac{tumor\ HLA\ methylation\ level}{normal\ HLA\ methylation\ level} < 0.75$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Immune Cellular Fraction Estimates, Immune Subtypes, tumor mutational burden, and Cytolytic Activity.**

The immune subtype immune cellular fraction estimates, tumor mutational burden, and neoantigen loads were obtained from Thorssen et al. (Thorsson et al., 2018). In brief, six immune subtypes, C1-C6, were clustered from five immune expression signatures (macrophages/monocytes, overall lymphocyte infiltration, TGF-β response, IFN-γ response, and wound healing), based on six TCGA molecular platforms (mRNA, microRNA, and exome sequencing; DNA methylation-, copy number-, and reverse-phase protein arrays). Immune cellular fraction estimates of 22 immune cell types were inferred with CIBERSORT using TCGA RNA-Seq data. The resulting proportions were multiplied by leukocyte fractions (LF) to yield overall fraction estimates. For tumor mutational burden, the mutation data generated by the PanCancer Atlas consortium (https://gdc.cancer.gov/about-data/publications/pancanatlas) was used. Nonsilent mutation rate per Mb was calculated. For neoantigen prediction, neoepitopes were identified from single nucleotide variants (SNVs) and insertion-deletion mutations (Indels) using NetMHCpan v3.0 (Nielsen and Andreatta, 2016) with default settings. HLA calls and mutant peptides were used as inputs, and mutant peptides were identified as potential neoantigens if the predicted binding affinity ($IC_{50}$) to autologous MHC < 500 nM and gene expression > 1.6 transcripts-per-million (TPM).

Cytolytic activity in each tumor sample was represented as cytolytic score (CYT) as previously proposed (Rooney et al., 2015). The expression levels of Granzyme A (*GZMA*: ENSG00000145649) and Perforin-1 (*PRF1*: ENSG00000180644) were extracted as FPKM reads from RNA-seq data. The following formula was applied to represent CYT:

$$CYT = \frac{log_2(GZMA + 1) + log_2(PRF1 + 1)}{2}$$

**Functional Divergence of HLA Alleles and Patient HED**

HED was calculated as described in (Pierini and Lenz, 2018). In brief, protein sequence of exons 2 and 3, namely the peptide-binding domain, of each HLA allele was extracted.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Protein sequences and exon annotations were obtained from the ImMunoGeneTics/HLA (IMGT/HLA) database (https://www.ebi.ac.uk/ipd/imgt/hla/). Pairwise alignment between the protein sequences of two alleles in each HLA-I locus was performed using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004). The Grantham distance was used to measure functional divergence between two alleles taking into account the physicochemical properties of amino acids. It was calculated as the sum of amino acid differences (including the biochemical composition, polarity, and volume of each amino acid) in the pairwise alignment along the protein sequence of peptide-binding domains following the formula by R. Grantham (Grantham, 1974):

$$\text{Grantham Distance} = \sum D_{ij} = \sum \left[\alpha\left(c_i - c_j\right)^2 + \beta\left(p_i - p_j\right)^2 + \gamma\left(v_i - v_j\right)^2\right]^{1/2}$$

where $D$ is the Grantham distance between the aligned sequences, and $i$ and $j$ are the amino acids at a homologous position. $c$, $p$, and $v$ represent biochemical composition, polarity, and volume of the amino acids, respectively. $\alpha$, $\beta$, and $\gamma$ are constants as originally proposed. The Grantham distance was normalized by the length of the sequence alignment. Patient HED was calculated as the mean of Grantham distances at *HLA-A*, *HLA-B*, and *HLA-C*.

**Principal Component Analysis**

The log 2 values of classical HLA class I genes (*HLA-A*, *HLA-B*, *HLA-C*), non-classical HLA class I genes (*HLA-E*, *HLA-G*), classical HLA class II genes (*HLA-DRA*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-DPB1*), and non-classical HLA class II genes (*HLA-DMA*, *HLA-DMB*, *HLA-DOA*, *HLA-DOB*) expression in FPKM were calculated, and each feature was averaged within individual tumor type. The features were centered and projected onto principal components (PCs).

**Classification**

To classify tumors with high cytolytic activity, defined as the top quantile of cytolytic score, one-hot encoding was applied on 33 TCGA tumor type, and log 2 values of classical and non-classical HLA class I (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-G*) and class II gene (*HLA-DRA*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DMA*, *HLA-*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

*DMB*, *HLA-DOA*, *HLA-DOB*) expression in FPKM in tumors were used to train a random forest classifier with n_estimators=200 and criterion='gini'.

To predict the immune subtypes of all tumors, one-hot encoding was applied on 30 TCGA tumor types; their log 2 values of classical HLA class I and class II gene expression in FPKM in tumors were used to train a three-layer neural network model with hyperbolic tangent activation function, Adam optimizer, Sparse-Categorical-Cross-Entropy loss function, and accuracy metric. The input layer contains 128 neurons; the hidden layer contains 96 neurons; the output layer contains 6 neurons with SoftMax activation function.

To predict the immune subtypes of tumors with Type-I immunity (C2 and C3), one-hot encoding was performed on 20 TCGA tumor types. Classical HLA class I and class II gene methylation β values were used to train a logistic regression classifier with solver="liblinear".

For all classifications, area under the curve (AUC) was calculated to evaluate the performance of the classifiers. For multiclass classification, a micro-average AUC was computed to account for class imbalance and evaluate the overall performance.

**HLA Loss of Heterozygosity (LOH) Analysis**

LOHHLA (Loss of Heterozygosity in Human Leukocyte Antigen) was used to assess HLA LOH in BRCA and LUAD patients. The program requires a tumor and germline BAM, patient-specific HLA typing, HLA fasta file, and purity and ploidy estimates. TCGA-LUAD purity and ploidy estimates were acquired from GDC (https://portal.gdc.cancer.gov/). The estimated allele 1 and allele 2 copy number for each patient (HLA_type1copyNum_withBAFBin, HLA_type2copyNum_withBAFBin, respectively), as well as a p-value (PVal_unique), were extracted from the LOHHLA output file.

A patient is identified as fully heterozygous if the individual has distinct alleles at all loci of *HLA-A*, *HLA-B*, *HLA-C*. The classification of HLA LOH in a tumor was the same as originally proposed by (McGranahan et al., 2017). In specific, a tumor sample was identified as harboring HLA LOH if an allele's copy number was less than 0.5 with p value less than 0.05. The number of allelic loss for each patient was calculated as the total number of alleles being classified as LOH.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Survival Analysis**

The values of progression-free interval (PFI) and status were used as obtained from (Liu et al., 2018).

To examine the survival effect of a continuous variable x, samples were split equally into high-x and low-x patients and analyzed using log-rank test and Kaplan-Meier estimator. In HED score analysis, the variable x represents HED score and HLA class I gene expression. In HLA class II expression fold change analysis across individual tumor types, the variable x represents HLA class II fold change in tumor samples with defined immune subtypes. In HLA LOH analysis, all BRCA and LUAD samples were split into groups with high HLA I expression or low HLA I expression, and the variable x represents if the tumor sample harbors HLA LOH within each group.

To compare the performance of cytolytic activity, HLA class I fold change and HLA class II fold change on survival prediction, as well as to compare the four groups defined based on HLA class I and class II expression fold change, Cox proportional-hazards modeling was used.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Supplemental Information

### Supplementary Figure Legends

**Figure S1. HLA Gene Expression Is Uncorrelated to TMB and Neoantigen Loads.** (A) Correlation between HLA class I expression and TMB with Spearman's rank correlation coefficient and p-value. (B) Same as (A), but between HLA class II expression and TMB. (C) Same as (A), but between HLA class I expression and neoantigen loads. (D) Same as (A), but between HLA class II expression and neoantigen loads.

**Figure S2. HLA Class I and Class II Expression Data Predict Cytolytic Activity with High Accuracy.** (A) Receiver operating Characteristic (ROC) curve of a random forest classifier. The random forest classifier is trained on log-transformed expression levels for classical and non-classical HLA class I and class II genes and predicts tumors with high cytolytic activity (high CYT). (B) The relative importance of classical and non-classical HLA gene expression in the random forest classifier. All features used to train the classifier are listed on the y-axis.

**Figure S3. HED Score Is Uncorrelated to Immune Characteristics.** (A) HED score distribution across groups of patients with different number of heterozygous HLA I loci. (B) Spearman's rank correlation between HED score and immune characteristics. Unadjusted two-tailed p values are reported below the diagonal and adjusted p values (Benjamini & Hochberg method) for multiple testing are reported above the diagonal. Correlations not significant ($p > 0.01$) are marked with an X in each corresponding cell.

**Figure 4S. HLA Class II Upregulation Correlates with Strong Immunity and Improved Survival.** (A) Cytolytic activity (CYT) across 4 groups based on their HLA class I and class II expression fold change. High HLA I and high HLA II are defined as the top 50% of all tumor samples, respectively. Pairwise p values are calculated by Wilcoxon rank-sum test and are unadjusted. p-value across all 4 groups is calculated by the Kruskal-Wallis test. (B) Same as (A), but with infiltration levels of CD8 cytotoxic T

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cells. (C) The hazard ratio (HR) of the effect of high HLA class II expression fold change (top 50%) on PFS across 13 TCGA tumor types. Circles represent the HR and horizontal bars represent 95% confidence intervals. HR and p values are calculated from the log-rank test. (D) The hazard ratio (HR) of the effects of different HLA I and HLA II fold change on PFS, with low HLAI low HLAII as reference. Squares represent the HR and horizontal bars represent 95% confidence intervals. HR and p values are calculated from Cox Proportional-hazards Model. HLAI: HLA class I expression fold change; HLAII: HLA class II expression fold change.

**Figure S5. The Immune Subtypes of Tumors with Type I immune Response Can Be Predicted with DNA Methylation Levels near HLA Class I Genes.** (A) HLA class I and class II expression fold change are compared between C2 and C3 tumors. p values are calculated by Wilcoxon rank-sum test and are unadjusted. (B) Spearman correlations between HLA class I (*HLA-A*, *HLA-B*, *HLA-C*) expression fold change and methylation. p values are calculated from Spearman's rank correlation and are unadjusted. (C) Receiver operating Characteristic (ROC) curve of a logistic regression classifier. The logistic regression classifier is trained on the raw β-scores of HLA class I genes from DNA methylation arrays in tumor-normal pairs and predicts the immune subtype (C2 and C3).

**Figure S6. HLA LOH Contributes to Worse Survival but Does Not Downregulate HLA Gene Expression.** (A) Progression-free survival (PFS) by HLA LOH among all heterozygous TCGA BRCA and LUAD patients. (B) same as (A), but among all TCGA BRCA patients. (C) same as (A), but among all TCGA LUAD patients. (D) Spearman correlation between the mean copy number of *HLA-A*, (E) *HLA-B*, (F) *HLA-C*, and expression. The correlation coefficient and p-value are calculated from Spearman's rank correlation. The allelic loss represents the number of lost HLA I alleles (raw copy number<0.5 and p<0.05 from LOHHLA output). (G) HLA class I expression, and (H) expression fold change, between tumors with HLA LOH and those without across BRCA and LUAD. p-value is calculated by Wilcoxon rank-sum test and is unadjusted.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## References

Akazawa, Y., Nobuoka, D., Takahashi, M., Yoshikawa, T., Shimomura, M., Mizuno, S., Fujiwara, T., Nakamoto, Y., and Nakatsura, T. (2019). Higher human lymphocyte antigen class I expression in early-stage cancer cells leads to high sensitivity for cytotoxic T lymphocytes. Cancer Sci *110*, 1842-1852.

Annunziato, F., Romagnani, C., and Romagnani, S. (2015). The 3 major types of innate and adaptive cell-mediated effector immunity. J Allergy Clin Immunol *135*, 626-635.

Axelrod, M.L., Cook, R.S., Johnson, D.B., and Balko, J.M. (2019). Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. Clin Cancer Res *25*, 2392-2402.

Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C.*, et al.* (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med *24*, 541-550.

Blass, E., and Ott, P.A. (2021). Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. Nat Rev Clin Oncol *18*, 215-229.

Carosella, E.D., Moreau, P., Le Maoult, J., Le Discorde, M., Dausset, J., and Rouas-Freiss, N. (2003). HLA-G Molecules: from Maternal–Fetal Tolerance to Tissue Acceptance. In, pp. 199-252.

Chowell, D., Krishna, C., Pierini, F., Makarov, V., Rizvi, N.A., Kuo, F., Morris, L.G.T., Riaz, N., Lenz, T.L., and Chan, T.A. (2019). Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. Nat Med *25*, 1715-1720.

Dendrou, C.A., Petersen, J., Rossjohn, J., and Fugger, L. (2018). HLA variation and disease. Nat Rev Immunol *18*, 325-339.

Duan, Q., Zhang, H., Zheng, J., and Zhang, L. (2020). Turning Cold into Hot: Firing up the Tumor Microenvironment. Trends Cancer *6*, 605-618.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fan, S., Gao, X., Qin, Q., Li, H., Yuan, Z., and Zhao, S. (2020). Association between tumor mutation burden and immune infiltration in ovarian cancer. Int Immunopharmacol *89*, 107126.

Garrido, F., Aptsiauri, N., Doorduijn, E.M., Garcia Lora, A.M., and van Hall, T. (2016). The urgent need to recover MHC class I in cancers for effective immunotherapy. Curr Opin Immunol *39*, 44-51.

Ghosh-Choudhary, S., Liu, J., and Finkel, T. (2020). Metabolic Regulation of Cell Fate and Function. Trends Cell Biol *30*, 201-212.

Grantham, R. (1974). Amino Acid Difference Formula to Help Explain Protein Evolution. Science *185*, 3.

Halenius, A., Gerke, C., and Hengel, H. (2015). Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targets-but how many arrows in the quiver? Cell Mol Immunol *12*, 139-153.

Joosten, S.A., Sullivan, L.C., and Ottenhoff, T.H. (2016). Characteristics of HLA-E Restricted T-Cell Responses and Their Role in Infectious Diseases. J Immunol Res *2016*, 2695396.

Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R., and Matsuda, F. (2017). HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. Hum Mutat *38*, 788-797.

Lauss, M., Donia, M., Harbst, K., Andersen, R., Mitra, S., Rosengren, F., Salim, M., Vallon-Christersson, J., Torngren, T., Kvist, A.*, et al.* (2017). Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. Nat Commun *8*, 1738.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V.*, et al.* (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*, 400-416 e411.

Liu, L., Wang, L., Zhao, L., He, C., and Wang, G. (2020). The Role of HLA-G in Tumor Escape: Manipulating the Phenotype and Function of Immune Cells. Front Oncol *10*, 597468.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Luo, N., Nixon, M.J., Gonzalez-Ericsson, P.I., Sanchez, V., Opalenik, S.R., Li, H., Zahnow, C.A., Nickels, M.L., Liu, F., Tantawy, M.N.*, et al.* (2018). DNA methyltransferase inhibition upregulates MHC-I to potentiate cytotoxic T lymphocyte responses in breast cancer. Nat Commun *9*, 248.

Marabelle, A., Fakih, M., Lopez, J., Shah, M., Shapira-Frommer, R., Nakagawa, K., Chung, H.C., Kindler, H.L., Lopez-Martin, J.A., Miller, W.H.*, et al.* (2020). Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. The Lancet Oncology *21*, 1353-1365.

McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J.*, et al.* (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell *171*, 1259-1271 e1211.

McNamara, M.G., Jacobs, T., Lamarca, A., Hubner, R.A., Valle, J.W., and Amir, E. (2020). Impact of high tumor mutational burden in solid tumors and challenges for biomarker application. Cancer Treat Rev *89*, 102084.

Mellins, E.D., and Stern, L.J. (2014). HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. Curr Opin Immunol *26*, 115-122.

Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med *8*, 33.

Pierini, F., and Lenz, T.L. (2018). Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. Mol Biol Evol *35*, 2145-2158.

Poluektov, Y.O., Kim, A., and Sadegh-Nasseri, S. (2013). HLA-DO and Its Role in MHC Class II Antigen Presentation. Front Immunol *4*, 260.

Radwan, J., Babik, W., Kaufman, J., Lenz, T.L., and Winternitz, J. (2020). Advances in the Evolutionary Understanding of MHC Polymorphism. Trends Genet *36*, 298-311.

Richman, L.P., Vonderheide, R.H., and Rech, A.J. (2019). Neoantigen Dissimilarity to the Self-Proteome Predicts Immunogenicity and Response to Immune Checkpoint Blockade. Cell Syst *9*, 375-382 e374.

Roche, P.A., and Furuta, K. (2015). The ins and outs of MHC class II-mediated antigen processing and presentation. Nat Rev Immunol *15*, 203-216.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell *160*, 48-61.

Roudko, V., Greenbaum, B., and Bhardwaj, N. (2020). Computational Prediction and Validation of Tumor-Associated Neoantigens. Front Immunol *11*, 27.

Seliger, B., Ruiz-Cabello, F., and Garrido, F. (2008). Chapter 7 IFN Inducibility of Major Histocompatibility Antigens in Tumors. In, pp. 249-276.

Strickler, J.H., Hanks, B.A., and Khasraw, M. (2021). Tumor Mutational Burden as a Predictor of Immunotherapy Response: Is More Always Better? Clin Cancer Res *27*, 1236-1241.

Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310-3316.

Tay, R.E., Richardson, E.K., and Toh, H.C. (2021). Revisiting the role of CD4(+) T cells in cancer immunotherapy-new insights into old paradigms. Cancer Gene Ther *28*, 5-17.

Thibodeau, J., Bourgeois-Daigneault, M.C., and Lapointe, R. (2012). Targeting the MHC Class II antigen presentation pathway in cancer immunotherapy. Oncoimmunology *1*, 908-916.

Thommen, D.S., and Schumacher, T.N. (2018). T Cell Dysfunction in Cancer. Cancer Cell *33*, 547-562.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A.*, et al.* (2018). The Immune Landscape of Cancer. Immunity *48*, 812-830 e814.

Wieczorek, M., Abualrous, E.T., Sticht, J., Alvaro-Benito, M., Stolzenberg, S., Noe, F., and Freund, C. (2017). Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. Front Immunol *8*, 292.

Yamamoto, T.N., Kishton, R.J., and Restifo, N.P. (2019). Developing neoantigen-targeted T cell-based treatments for solid tumors. Nat Med *25*, 1488-1499.

Yarchoan, M., Johnson, B.A., 3rd, Lutz, E.R., Laheru, D.A., and Jaffee, E.M. (2017). Targeting neoantigens to augment antitumour immunity. Nat Rev Cancer *17*, 209-222.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

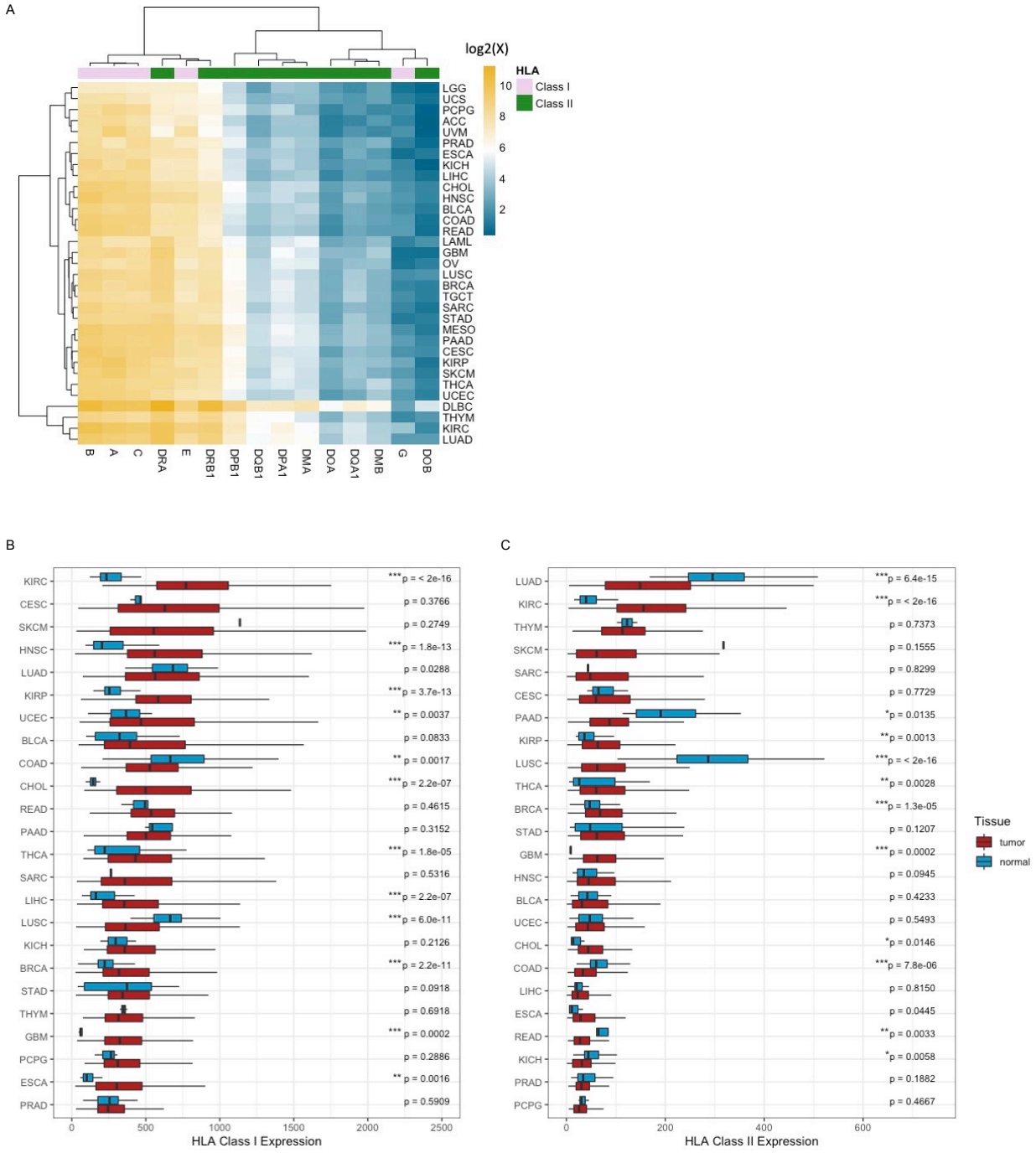


**Figure 1. HLA Expression Is Differentially Expressed across Tumor Types.** (A) Expression of classical and non-classical HLA class I and class II genes in tumor samples across 33 TCGA tumor types. Expression of gene X is represented as log-transformed Fragments per Kilobase of Transcripts per Million (FPKM). (B) HLA class I expression represented as the geometric mean of *HLA-A*, *-B*, *-C* in FPKM, and (C) HLA class II expression represented as the geometric mean of HLA-DP *(HLA-DPA1*, *HLA-DPB1)*, -DQ *(HLA-DQA1*, *HLA-DQB1)*, -DR *(HLA-DRA*, *HLA-DRB1)* in FPKM, in tumor-normal pairs. Boxes in the box plot represent interquartile ranges and horizontal lines represent 5th-95th percentile ranges. p values between the pairwise tumor-normal samples are calculated by Wilcoxon rank-sum test and are unadjusted. Asterisks denote significant FDR-adjusted (Benjamini & Hochberg method) p values (*p<0.05; **p<0.01; ***p<0.001).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

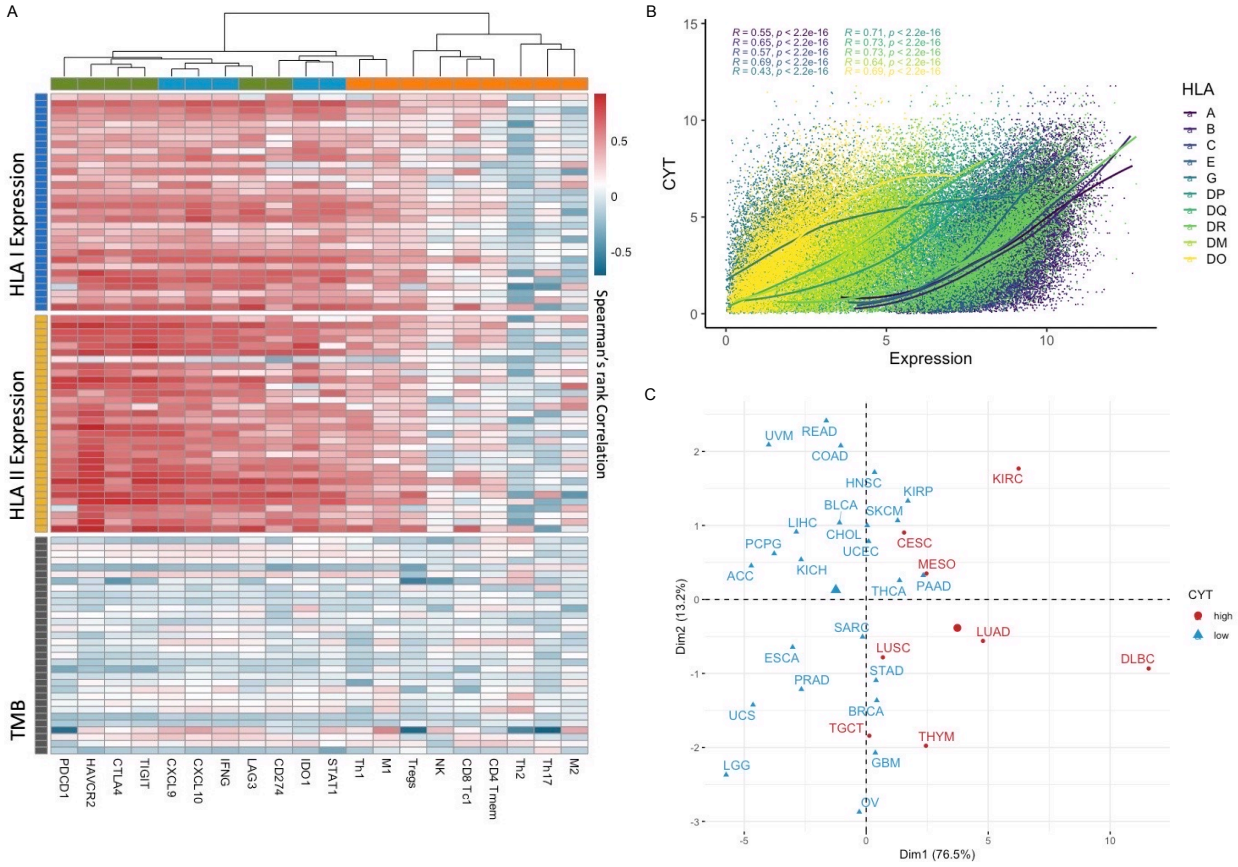


**Figure 2. HLA Class I and Class II Gene Expression Are Associated with Immune Activity.** (A) Spearman correlations between HLA class I, class II expression, TMB, and infiltrating immune cells (orange), pro-inflammatory genes (blue), and immune checkpoints (green) across 32 TCGA tumor types. Each row within a horizontal block represents a distinct tumor type. CD8 Tc1: CD8 cytotoxic T cells; CD4 Tmem: activated CD4 memory T cells; Th1: T helper 1 cells; Th2: T helper 2 cells; Th17: T helper 17 cells; Tregs: regulatory T cells; NK: natural killer cells; M1: Macrophage 1; M2: Macrophage 2. (B) Spearman correlations between the expression of classical HLA class I *HLA-A*, *-B*, *-C*, non-classical *HLA-E*, *-G*, classical HLA class II HLA-DP (*HLA-DPA1*, *HLA-DPB1*), -DQ (*HLA-DQA1*, *HLA-DQB1*), -DR (*HLA-DRA*, *HLA-DRB1*), non-classical HLA-DM (*HLA-DMA*, *HLA-DMB*), -DO (*HLA-DOA*, *HLA-DOB*), and cytolytic activity (CYT). p values are calculated from Spearman's rank correlation and are unadjusted. (C) Principal Component Analysis (PCA) on 32 tumor types by their expression of classical and non-classical HLA class I and class II genes. Principal Component 1 (Dim1) and Principal Component 2 (Dim2) are shown. High cytolytic activity (CYT high) is defined as the top 75% of all tumors' mean CYT.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

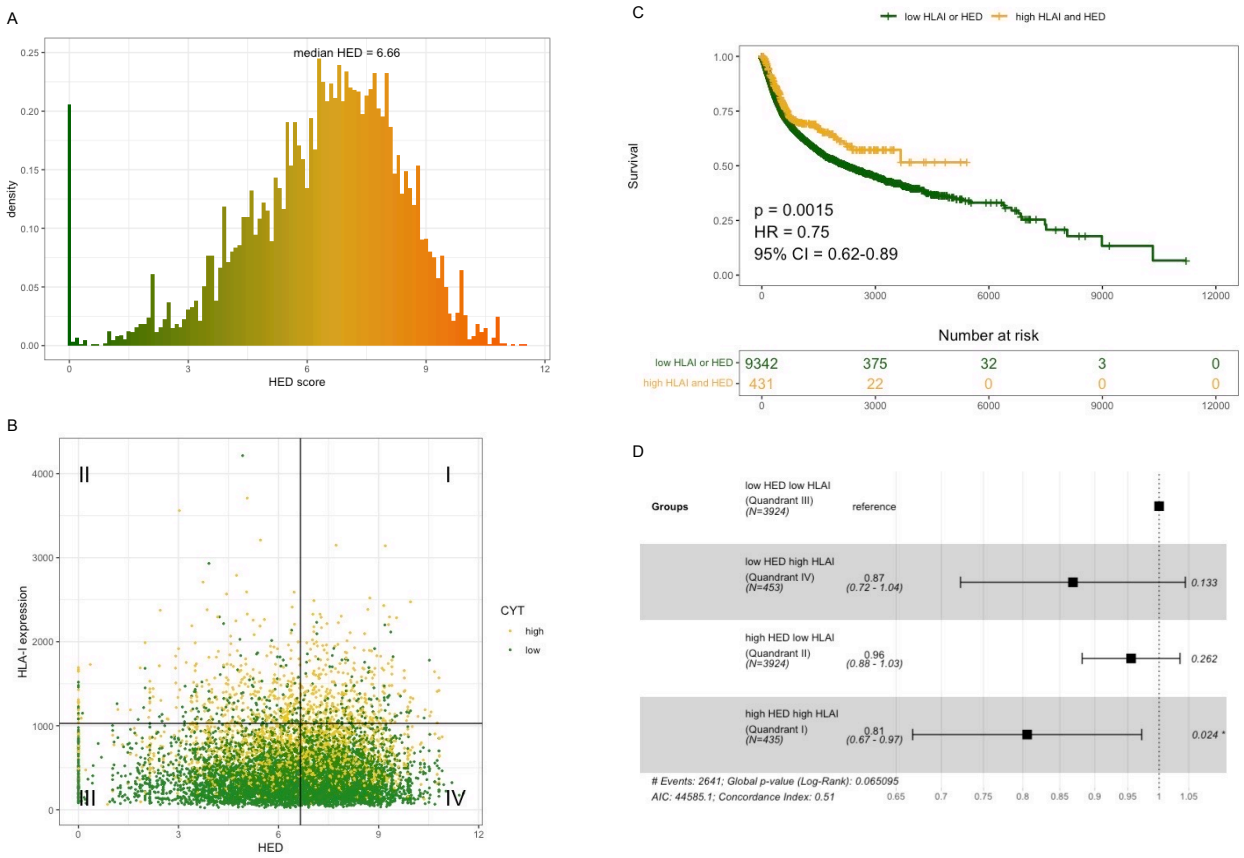


**Figure 3. HLA-I Evolutionary Divergence (HED) Score in Tumors with High HLA Expression Is Associated with Improved Survival.** (A) HED score distribution among 8801 tumor samples from 30 tumor types. (B) HED score vs. HLA class I expression. High cytolytic activity (CYT high) is defined as the top 75% of all tumor samples. The vertical line represents the median HED score, and the horizontal line represents the 90th percentile HLA class I expression. Four quadrants are defined accordingly. (C) Progression-free survival (PFS) by high HLA I expression and HED (Quadrant I) and low HLA I expression or HED (Quadrant II, III, IV). p-value, HR, and 95% confidence interval are calculated by the Log-rank test. (D) The hazard ratio (HR) of the effects of HED and HLA I expression on PFS, with low HED low HLAI as reference. Squares represent the HR and horizontal bars represent 95% confidence intervals. HR and p values are calculated from Cox Proportional-hazards Model. HED: HLA-I evolutionary divergence score; HLAI: HLA class I expression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

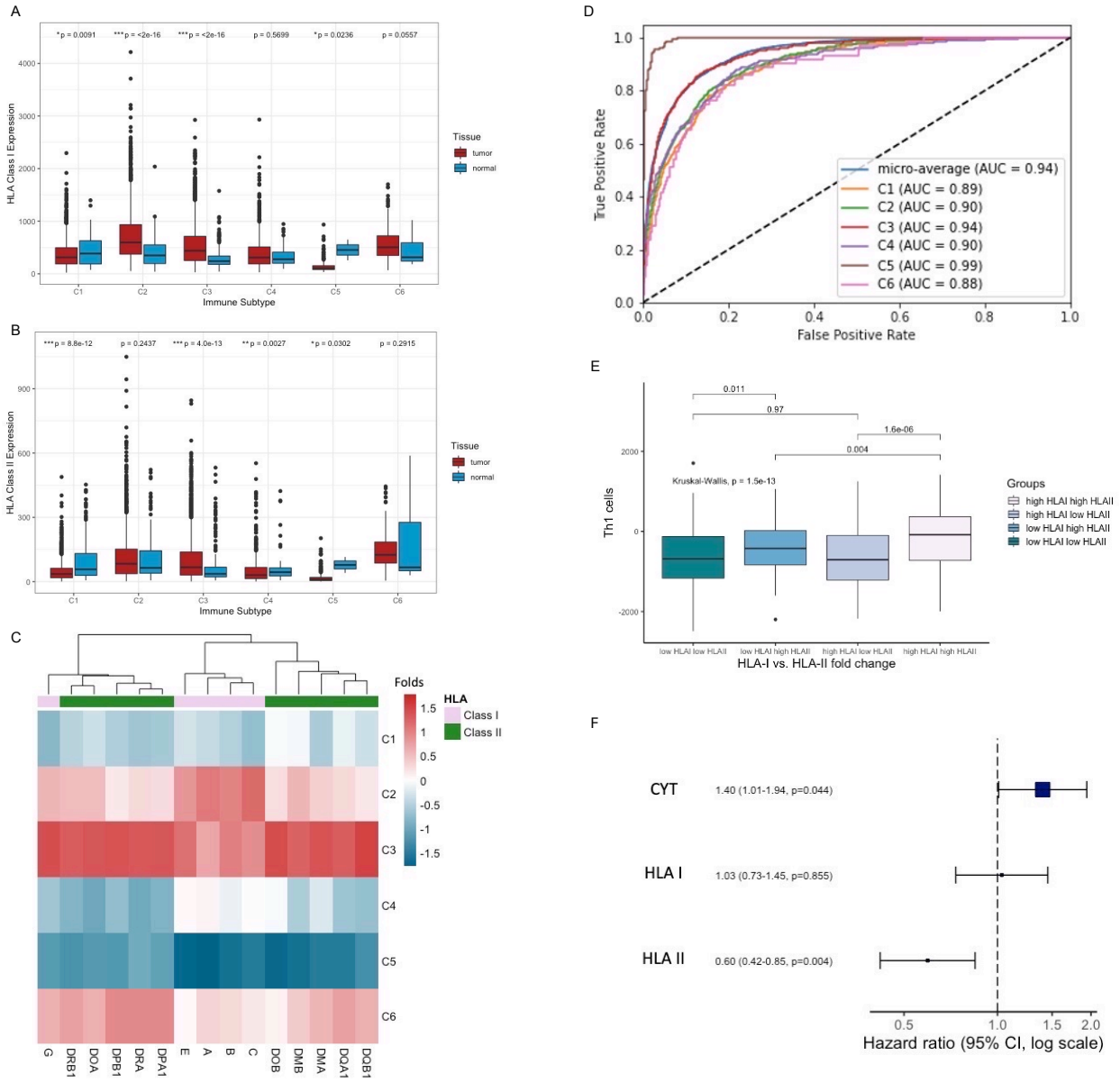


**Figure 4. HLA Expression Predicts the Immune Subtypes and Is Associated with Improved Survival.** (A) HLA class I expression, and (B) HLA class II expression between tumors and matched normal samples across six immune subtypes, C1-C6. Boxes in the box plot represent interquartile ranges. P values between pairwise tumor-normal samples are calculated by Wilcoxon rank-sum test and are unadjusted. Asterisks denote significant FDR-adjusted (Benjamini & Hochberg method) p values (*p<0.05; **p<0.01; ***p<0.001). C1: wound-healing; C2: IFN-dominant; C3: inflammatory; C4: lymphocyte-depleted; C5: immunologically quiet; C6: TGF- dominant. (C) Expression fold change of classical and non-classical HLA class I and class II genes between tumors and matched normal samples across six immune subtypes. The heatmap is scaled by columns. (D) Receiver Operating Characteristic (ROC) curve of a neural network model. The model is trained on log-transformed expression levels for classical and non-classical HLA class I and class II genes and predicts the immune subtypes of tumors. Micro-average area under the curve (AUC) as well as AUC for individual immune subtype predictions are presented. (E) Helper 1 T cells (Th1) signatures across 4 groups based on their HLA class I and class II expression fold change. High HLA I and high HLA II are defined as the top 50% of all tumor samples, respectively. Pairwise p values are calculated by Wilcoxon rank-sum test and are unadjusted. p-value across all 4 groups is calculated by the Kruskal-Wallis test. (F) The hazard ratio (HR) with CYT, HLA I fold change, HLA II fold change as explanatory variables, and PFS as the dependent variable. Squares denote the HR and horizontal bars represent 95% confidence intervals. HR and p-value are calculated from Cox Proportional-hazards Model. CYT: cytolytic activity; HLA I: HLA I expression fold change; HLA II: HLA II expression fold change.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

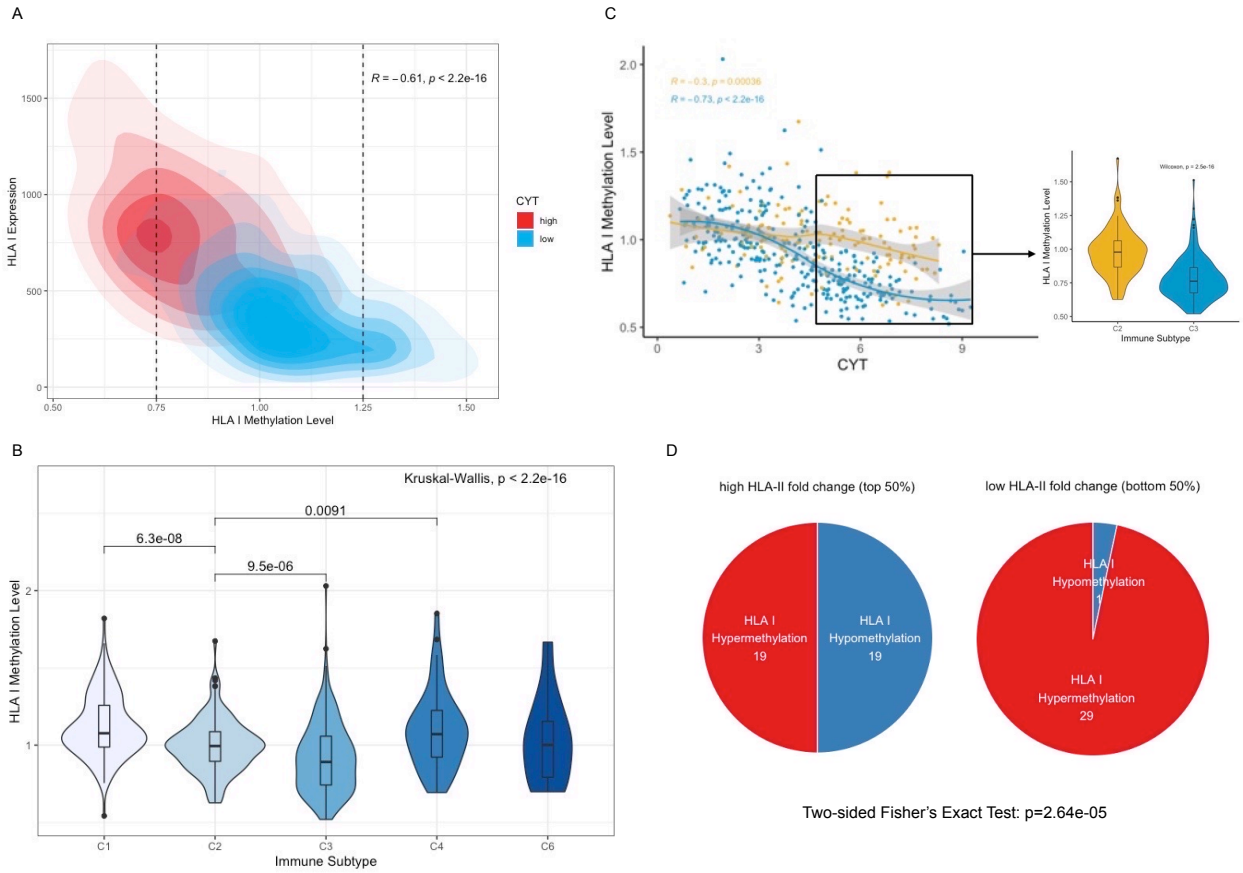


**Figure 5. HLA Class I Methylation Is Favored in A Subset of Hot Tumors and Dampens Immune Activity.** (A) A kernel density plot of HLA class I methylation levels and expression across groups with high (top 75%) or low cytolytic activity. Methylation level was represented as the tumor:normal ratio of mean -scores across all CpG sites near HLA I genes. Correlation coefficient and p values are calculated from Spearman's rank correlation. CYT: cytolytic activity; Hypomethylation: HLA I methylation level < 0.75; Hypermethylation: HLA I methylation level > 1.25. (B) HLA class I methylation levels across five immune subtypes (excluding C5 with no available data). Boxes in the violin plot represent interquartile ranges. p values between pairwise immune subtypes are calculated by Wilcoxon rank-sum test and are unadjusted. p-value across six immune subtypes is calculated by the Kruskal-Wallis test. (C) Left panel: the Spearman correlations between cytolytic activity (CYT) and HLA class I methylation in C2 and C3 tumors. p values within each immune subtype are calculated from Spearman's rank correlation. Right panel: HLA class I methylation in tumors with high cytolytic activity (top 50%) across C2 and C3. p-value is calculated by Wilcoxon rank-sum test and is unadjusted. (D) Association between HLA II fold change and aberrant methylation states near HLA I genes. Two groups are defined based on HLA II fold change in all samples. Numbers on pie charts indicate the number of tumor samples showing hypermethylated or hypomethylated HLA I genes. p-value is calculated from two-sided Fisher's Exact Test.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444511; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

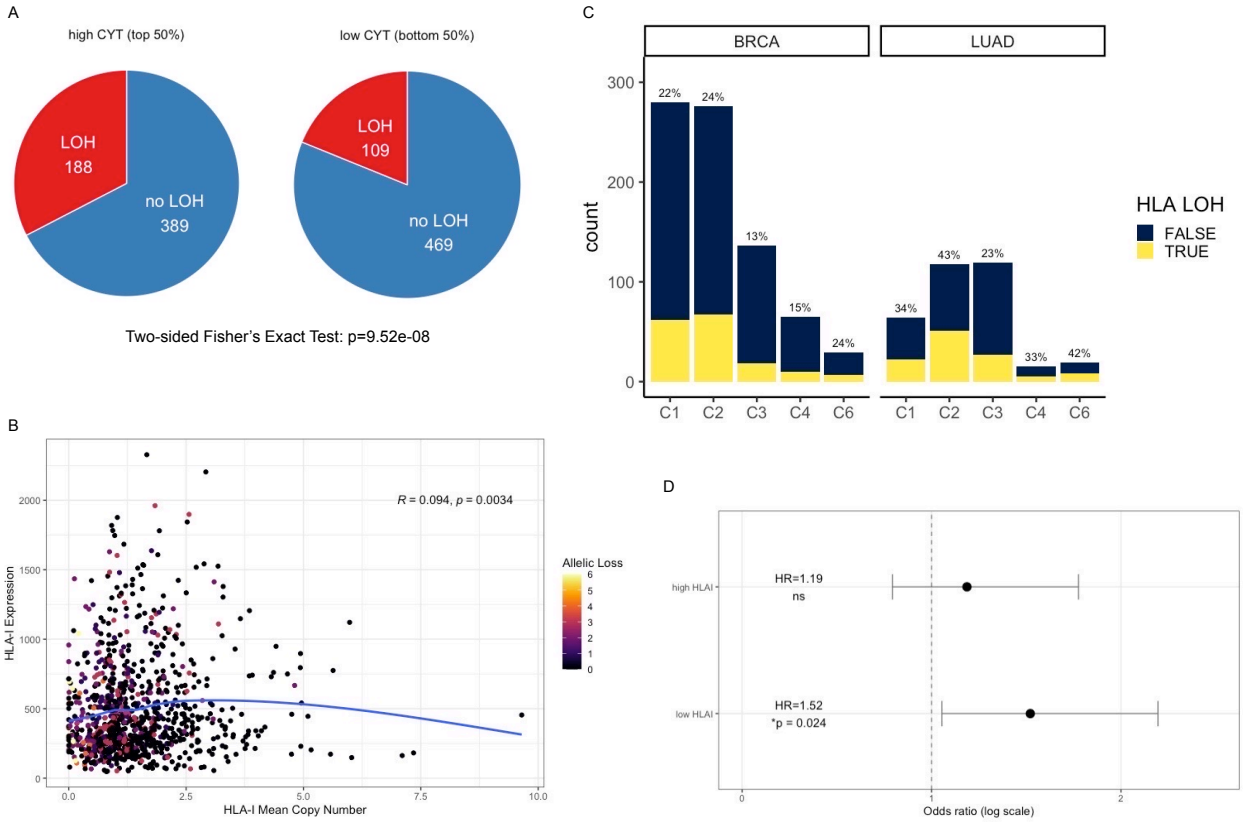


**Figure 6. HLA LOH Is Exploited by Hot Tumors with Worse Patients' Survival.** (A) Association between cytolytic activity and HLA LOH. Two groups are defined based on cytolytic activity in Tumors (high CYT: top 50%). Numbers on pie charts indicate the number of tumors harboring HLA LOH. p-value is calculated from two-sided Fisher's Exact Test. (B) Tumor samples harboring HLA LOH across five immune subtypes (excluding C5 with no available data) in TCGA BRCA and LUAD, respectively. Percentage on top of each bar represents the percent of tumor samples with HLA LOH within the corresponding immune subtype. (C) Spearman correlation between the mean copy number of HLA class I genes (*HLA-A*, *HLA-B*, *HLA-C*) and HLA class I expression. The correlation coefficient and p-value are calculated from Spearman's rank correlation. The allelic loss represents the number of lost HLA I alleles (raw copy number<0.5 and p<0.05 from LOHHLA output). (D) The hazard ratio (HR) of the effect of HLA class I expression on PFS. Circles represent the HR and horizontal bars represent 95% confidence intervals. HR and p values are calculated from the log-rank test. High HLAI: high HLA class I expression (top 50%); low HLAI: low HLA class I expression.