Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity

Gergő Mihály Balogh

PhD Thesis

Szeged

2023

UNIVERSITY OF SZEGED, ALBERT SZENT-GYÖRGYI MEDICAL SCHOOL
DEPARTMENT OF DERMATOLOGY AND ALLERGOLOGY,
DOCTORAL SCHOOL OF CLINICAL MEDICINE

Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity

Gergő Mihály Balogh

PhD thesis

Supervisor:
Dr. Máté Manczinger

Szeged

2023

Publications

Scientific paper related to the thesis

Máté Manczinger, Balázs Koncz, Gergő Mihály Balogh, Benjamin Tamás Papp, Leó Asztalos, Lajos Kemény, Balázs Papp & Csaba Pál. Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity. *Nature Cancer* (2021). DOI: 10.1038/s43018-021-00226-4. IF: 22.7, SCImago Journal Rank: Q1 / D1

Publications not directly related to the thesis

Balázs Koncz, Gergő M. Balogh, Benjamin T. Papp, Leó Asztalos, Lajos Kemény, Máté Manczinger: Self-mediated positive selection of T cells sets an obstacle to the recognition of nonself. *Proceedings of the National Academy of Sciences* (2021). DOI: 10.1073/pnas.2100542118. IF: 11.205, SCImago Journal Rank: Q1 / D1.

Cumulative impact factor: 33,905

Table of Contents

Abbreviations

ANOVA	analysis of variance
BCL-2	B-cell lymphoma 2
BLIMP1 (PRDM1)	B lymphocyte-induced maturation protein-1 (PR domain zinc finger protein 1)
BLOSUM62	blocks substitution matrix 62
BTLA	B- and T-lymphocyte attenuator
CD	cluster of differentiation
COSMIC	Catalogue of Somatic Mutations in Cancer
cTEC	cortical thymic epithelial cell
CTLA-4	cytotoxic T-lymphocyte-associated protein 4
DAI	differential agretopicity index
FDR	false discovery rate
FOXP3	forkhead box P3
GSEA	gene set enrichment analysis
HED	HLA-I evolutionary divergence
HLA	human leukocyte antigen
IC50	jalf maximal inhibitory concentration
ICI	immune checkpoint inhibitor
IDO	indoleamine 2,3-dioxygenase
IEDB	Immune Epitope Database
LAG-3	lymphocyte-activation gene 3
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
mTEC	medullary thymic epithelial cell
NK	natural killer (cell)
NSCLC	non-small cell lung cancer
OR	odds ratio
PD-1, PD-L1	programmed cell death protein 1; programmed death-ligand 1 and 2
pHLA	peptide-HLA
Pr	promiscuity
RECIST	response evaluation criteria in solid tumors
RR	relative risk
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SKCM	skin cutaneous melanoma
T-bet	T-box expressed in T cells
TCGA	The Cancer Genome Atlas
TCR	T-cell receptor
TGF-β	transforming growth factor beta
TIDE	Tumor Immune Dysfunction and Exclusion
TIGIT	T-cell immunoreceptor with Ig and ITIM domains
TIM3	T-cell immunoglobulin and mucin-domain containing-3
TMB	tumor mutational burden
TOX	thymocyte selection-associated high mobility group box factor
T_{reg}	regulatory T-cell

1 INTRODUCTION

1.1 HLA-I are key molecules of the cellular adaptive immune response

Adaptive immunity is a major component of the mammalian immune system, including that of humans. It generates a specific and refined reaction to self and non-self antigens by producing clonal populations of cells that recognize their targets with high affinity. The adaptive immune system comprises two arms, cellular and humoral immunity, which collaborate with the innate immune system to eliminate pathogens.[1]

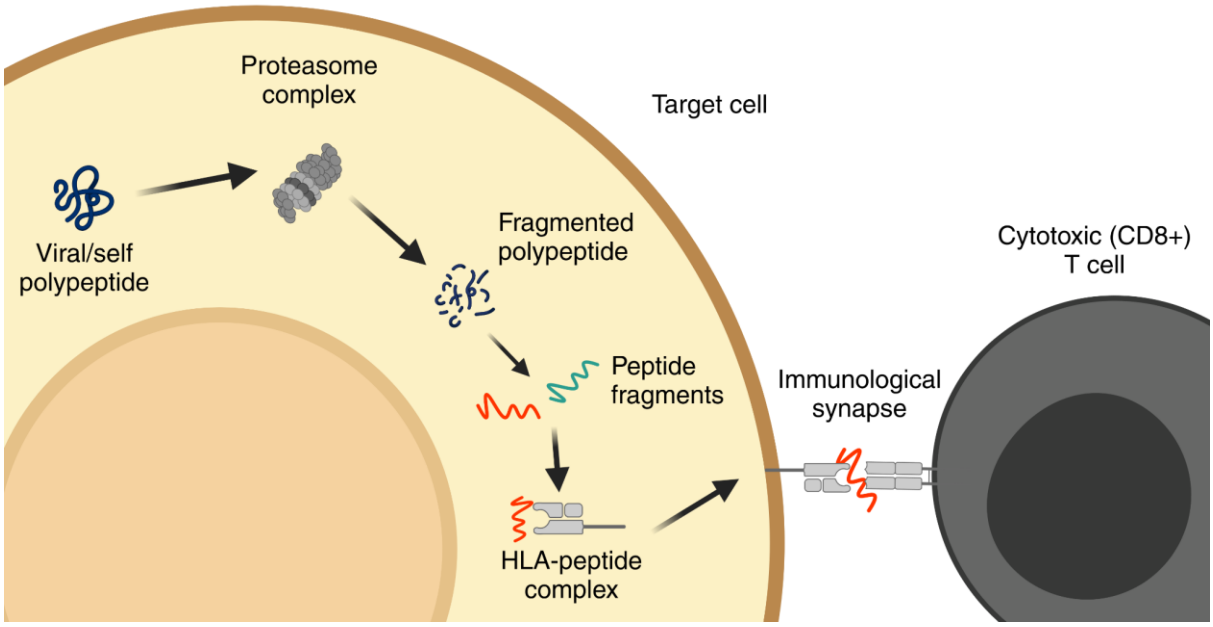


Figure 1: *overview of cellular adaptive immune recognition. The figure is created using Biorender.com.*

The cellular arm of the adaptive immune system targets and eliminates host cells that are infected by intracellular pathogens (such as viruses or some obligate intracellular bacteria like *Mycobacteria*) or display signs of abnormality (such as tumor cells). The Human Leukocyte Antigen class I (HLA-I) molecule is a key player in this process (Figure 1).[1,2] This protein complex is expressed on the surface of almost all nucleated cells in the human body, except for some immune-privileged sites (e.g. testis). In contrast, HLA class II molecules are expressed on the surface of professional antigen-presenting cells only and present peptides derived from extracellular sources (such as most bacterial pathogens).[3–5] Recently, it has been shown that HLA-II molecules also play a role in antitumor immunity, but less directly compared to the HLA-I system.[4,6,7] This work focuses on HLA class I mediated immune recognition exclusively.

HLA-I molecules consist of two polypeptide chains, the α chain and β2-microglobulin, which are noncovalently associated at the α3 domain of the α chain. The α1 and α2 domains of the α chain form a peptide-binding groove on the surface of the molecule, exposed to the extracellular environment. The third component of a functional HLA-I complex is a short peptide (usually 9 amino acids in length) derived from a protein that was synthesized inside the cell. These peptides stabilize the HLA-I molecule by fitting into the peptide binding pocket of the molecular complex.[8] Foreign peptide–HLA-I complexes can act as danger signals for the cellular arm of the adaptive immune system.[1,9]

Cytotoxic T cells are the effector cells that can recognize these specific HLA-I complexes. These cells express a protein structure called T-cell receptor (TCR), which determines the specificity of each cell clone to a distinct subset of target HLA class I–peptide (pHLA) complexes. The interaction between the pHLA on the target cell and the TCR on the T cell forms the immunological synapse. It triggers the activation of the cytotoxic T lymphocytes and the leads to the subsequent elimination of the targeted host cell. The high specificity of TCR–pHLA binding reduces the likelihood of random immune activation.[10,11]

1.2 HLA-I genotype determines what the cellular adaptive immunity responds to

HLA-I proteins are encoded by three genes (A, B, and C), located at the chromosomal position 6p21.[12] The HLA region is the most variable part of our genome: the number of HLA-I alleles registered in the human population exceeds 25,000. Each allele has a characteristic specificity for peptides carrying different amino acid residues in key positions.[13]

In conclusion to the diversity of HLA-I alleles, there is a high chance for a subject to be heterozygous for all 3 HLA-I loci, meaning that a person usually carries 6 different alleles.[14,15] HLA-I heterozygosity has been shown to play a major role in protection against viral infections and their potential severe outcomes, as it increases the number of presentable epitopes from a pathogen. Research has shown that among HIV-infected individuals, heterozygous patients have a slower rate of disease progression compared to those who are homozygous.[16] Similar trends were observed for HLA-II genotype and hepatitis B clearance.[17] Finally, a study has shown that individuals who are homozygous for any HLA class I alleles in case of a SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) infection have a higher risk of earlier mortality.[18]

1.3 Harmful effects of self-reactive T cells are eliminated via tolerance mechanisms

The development of T-cell repertoires, which react to foreign peptide-HLA complexes, is a multi-step process. Hematopoietic stem cells in the bone marrow differentiate into lymphoid progenitor cells called thymocytes in a two-stage process that occurs during fetal development within the thymus. The first step involves the interaction of double-positive T cells expressing CD4 and CD8 (cluster of differentiation) coreceptors with cortical thymic epithelial cells (cTECs), leading to the positive selection of CD8+ cytotoxic T lymphocytes. This encounter eliminates most T cells that are non-specific to self-peptides, presented by cTECs.[19]

The subsequent step, negative selection, predominantly occurs in the medulla, but it already begins in the cortex.[20] The goal of this process is to remove potentially self-reactive T cells that harbor TCRs exhibiting strong binding towards self-pHLA complexes. Medullary thymic epithelial cells (mTECs) express certain tissue-specific genes, resulting in a presented peptide pool with a distinct composition compared to cTECs.[21,22] A significant subset of cells having TCRs binding strongly to peptides presented by mTECs undergo differentiation into regulatory T cells, facilitating immune tolerance response. Meanwhile, the remaining negatively selected cells die as a result of the maturation process. Consequently, this step results in the production of a functional T-cell pool harboring $< 10^8$ unique TCR sequences.[23]. In overall, the resulting cytotoxic T lymphocyte population should confer tolerance towards self-peptides while exhibiting a destructive response towards foreign amino acid sequence motifs. This phenomenon is called central tolerance.

Additionally, a complex mechanism called peripheral tolerance prevents the spontaneous activation of those T cells in the body, which have not been deleted in the thymus. Such T cells carry TCRs, specific to e.g. food antigens and developmental antigens that are not being presented in the thymus. This mechanism aims to minimize the chance for autoimmunity and pathological immune activation during chronic inflammatory processes.[24]

Multiple different paths of T cell inactivation exist as ways of peripheral tolerance. Anergy occurs when the T cell can interact with its cognate antigen, but lacks co-stimulatory signals needed for complete T-cell activation. This results in a state of long-term hyporesponsiveness in T cells, characterized by active repression of TCR signaling and interleukin-2 expression.[25]

Furthermore, some co-stimulatory signal molecules act as inhibitors of immune activation, e.g. programmed cell death 1 (PD-1) receptor and its ligands PD-L1 and PD-L2 (programmed death-ligand 1 and 2).[26] CTLA-4 (cytotoxic T-lymphocyte-associated protein 4) is getting

highly expressed after T-cell activation.[25] These molecules are also called immune checkpoints, and will be discussed in more detail in section 1.5.

Peripheral tolerance can also be mediated by special subsets of antigen-presenting cells, e.g. tolerogenic dendritic cells. [27,28] Finally, the peripheral deletion of self-reactive lymphocytes may also occur, as a result of the activation of apoptotic signaling pathways.[29–31]

1.4 Adaptive cellular immunity targets the altered self in tumor cells

Cancer is a collection of diseases characterized by anomalous cell growth that can invade or metastasize to other parts of the body. Cancer is a primary cause of death globally, responsible for nearly 10 million fatalities in 2020.[32]

One noteworthy characteristic of certain tumor types is the infiltration of particular immune cell subtypes. Cytotoxic T cells play a crucial role in antitumor immune response. T cells can identify peptides presented by HLA-I molecules on the surface of malignant cells.

These cancer rejection epitopes may arise from two classes of antigens.[33] The first group contains non-mutated peptides, against which T-cell tolerance is incomplete due to their limited tissue expression patterns. For example, cancer-testis antigens are expressed by abnormal cells of a wide range of tumors: bladder, lung, and liver carcinoma, as well as melanoma.[34]

Another type of potential cancer rejection antigens, known as neoantigens, is composed of peptides that do not exist in the normal human genome. In the case of a significant portion of human tumors without a viral origin, these neo-epitopes are the results of tumor-specific genomic alterations that lead to the formation of new protein sequences. In the case of virus-associated tumors like cervical cancer and certain types of head and neck cancers, epitopes resulting from viral open reading frames also add to the pool of neoantigens.[33,35]

Distinct types of genetic changes have different frequencies across tumor types and individual tumors. Missense and silent mutations, caused mainly by single nucleotide changes, are dominant in skin cutaneous melanoma (SKCM) and non-small cell lung cancer (NSCLC).[36,37] Meanwhile, insertions and deletions are frequent in microsatellite-unstable cancers, including distinct subtypes of colorectal and endometrial tumors.[38–40]

It is worth mentioning that the tumor immune microenvironment is shaped by various immune cell types of the innate immune system as well. Tumor-associated macrophages are present in two subtypes, M1 and M2. Macrophages initially tend to show M1-polarization during carcinogenesis – a subtype associated with stronger anti-tumor activity.[41–44] M2 macrophages

usually arise later, promoting tumor development and poor survival by immune suppression, increased angiogenesis, and metastatic potential.[45–50] Myeloid-derived suppressor cells confer a similar effect on tumor progression, by enhancing the effects of T_{reg} cells and neovascular invasion, suppressing cytotoxic T cells, dendritic cells, and natural killer (NK) cells.[51,52]

Tumor-associated neutrophils also have a controversial effect on tumor progression, as N2 cells also promote vascularization and tumor development.[53,54] NK cells are capable of destructing tumor cells via natural cytotoxicity or the antibody-dependent cellular cytotoxicity pathway, resulting in the release of cytotoxic granules or inducing apoptosis of the target cells.[55–57]

Dendritic cells are professional antigen-presenting cells. They play a major role in orchestrating immune response, as distinct subtypes of these cells (e.g. cDC1) take up exogenous antigens, and present them via HLA-I and HLA-II molecules to CD8+ and CD4+ T cells respectively.[58] Meanwhile, distinct subsets of dendritic cells are initiating immune tolerance via factors including PD-L1, activating T_{reg} cells, eventually promoting tumor progression.[59,60]

1.5 Checkpoint molecules serve as regulators in adaptive immune response

Binding between pHLA complexes and specific TCRs is a prerequisite for the initiation of cellular adaptive immunity. However, the decision between activation and inhibition of T lymphocytes is dependent on the balance of additional positive and negative signals delivered through auxiliary membrane receptors. Some of these are co-stimulatory factors. For instance, the binding of CD28 receptors with B7 ligands greatly enhances the activation of T cells, initiating signaling events that work synergistically with TCR-mediated activation pathways. It ultimately promotes T-cell survival and production of the cytokine interleukin-2.[61]

Contrarily, distinct checkpoint molecules are playing an inhibitory role in the initiation of an immune response. CTLA-4, expressed by T cells soon after activation, provides a signal that halts T-cell responses, by interfering with TCR- and CD28-mediated signaling.[62,63] This checkpoint molecule has a markedly stronger affinity to B7 proteins compared to CD28, resulting in inhibition via competition.[64–66] The overall effect is a decreased production of interleukin-2 and weaker proliferation of T cells, maintaining peripheral T cell tolerance.[67]

PD-1, another member of the CD28 family, interacts with two other B7 homologs, namely PD-L1 (also called B7-H1) and PD-L2 (also known as B7-DC).[68–70] The expression of PD-1 molecules is activation dependent, but not specific to T cells, as it had been detected on multiple other cell types, e.g. B lymphocytes and myeloid cells as well.[71] Paradoxically, PD-1 molecules suppress immune activation by forming an excessively stable pHLA-TCR complex. The

stabilization prevents the engagement of subsequent pHLA complexes and, consequently, activation of T cell response. This results in T-cell exhaustion, a phenomenon protecting the body from the toxic effects of chronic immune activation.[72]

Additional checkpoint molecules also take part in controlling immune reactions. Some other notables examples include LAG-3 (Lymphocyte-activation gene 3), TIGIT (T cell immunoreceptor with Ig and ITIM domains), TIM3 (T-cell immunoglobulin and mucin-domain containing-3) and BTLA (B- and T-lymphocyte attenuator).[73]

1.6 Checkpoint molecules aid tumor immune evasion and serve as targets for therapy

Tumors consist of cells that have undergone transformation and exhibit high rates of proliferation. The accumulation of somatic DNA mutations is the most important cause of this process. Some of these alterations are considered neutral by not having a direct impact on the fitness of cancer cells (passenger mutations), while others have a major role in tumor evolution and the formation of more successful subclones.[74–79] Basically, clones that exhibit high rates of proliferation have a greater likelihood to dominate a cell mass, but the picture is more complex, as additional phenotypic traits, e.g. capability for migration, and immunogenicity, also take part in the formation of tumor composition.[80] Latter selective pressure results in a phenomenon called cancer immunoediting, an evolutionary process forming the landscape of tumor cells. It enables cell clones harboring less immunogenic mutations and/or displaying less immunogenic antigens to persist, leading to a reduction in the number of targets available to the immune system.[81–83] Furthermore, tumor cells can downregulate mutated genes that encode potentially immunogenic peptides, as well as components of the HLA-I peptide presentation pathway.[82,84]

The cellular and molecular composition of the tumor microenvironment has a complex impact on the overall properties of the immune response. Regulatory T cells and myeloid-derived suppressor cells express inhibitory molecules that can suppress the maturation of dendritic cells, leading to decreased levels of immune presentation via HLA, and subsequently reducing the production of pro-inflammatory cytokines, including interleukin-12.[85–87] Various small molecules, such as kynurenine, a tryptophan metabolite produced by IDO (indoleamine 2,3-dioxygenase) enzymes found in myeloid and cancer cells, are also involved in the transition from destructive T cell phenotypes to the suppressive T_{reg} state.[87,88] As mentioned previously, distinct subtypes of macrophages can also affect the quality of the antitumor immune response.[89]

Immune checkpoint mechanisms also play a crucial role in immune evasion. During the past two decades, the importance of these molecules, particularly members of the PD-1 and CTLA-4 pathways, has gained considerable attention in both basic research and clinical practice. Elevated CTLA-4 expression was associated with poor outcomes in various tumor types, such as nasopharyngeal carcinoma, thymoma, and esophageal carcinoma.[66,90–92] CTLA-4 blocks immune activation via the previously presented mechanisms (see section 1.5), as well as through the induction of IDO enzymes, resulting in tryptophan depletion.[88,93,94] In the case of certain tumor types, e.g. in NSCLC and breast cancer, the association between CTLA-4 expression and clinical prognosis was detected only in distinct subtypes.[66,95–97] High intra-tumor expression levels of the PD-1 receptor and its ligand PD-L1 is also shown to correlate with poor outcome in multiple cancer types.[98–102]

As inhibitory checkpoint molecules are frequently exploited by tumor cells to avoid immune destruction, the idea of targeting them has been a subject of research for several years.[103] In the last decade, monoclonal antibody-based therapeutic agents that target either CTLA-4 or PD-1/PD-L1 molecules have been approved for clinical use. These agents are commonly referred to as immune checkpoint inhibitors (ICIs).[104] Ipilimumab, a human IgG1 (immunoglobulin G1) antibody that inhibits CTLA-4, was the first drug of its kind to be approved for treating metastatic melanoma in 2011.[105–108] In 2014, nivolumab, the first ICI antibody targeting PD-1 was introduced, which was approved as a medication for melanoma, NSCLC, and a series of additional neoplasms.[109] Since then, multiple additional antibody-based treatment options have been approved, including pembrolizumab and cemipilimab targeting PD-1, atezolizumab, avelumab, and durvalumab targeting PD-L1 molecules.[110] In the last few years, other checkpoint molecules have also been considered as targets. A notable example is the case of LAG-3 and relatlimab, a monoclonal antibody binding to this molecule. In previously untreated metastatic melanoma, the inhibition of both LAG-3 and PD-1 using relatlimab and nivolumab respectively led to a greater benefit in progression-free survival than inhibition of PD-1 alone.[111]

1.7 Identifying predictive factors for ICI therapy outcomes

Prior to the approval of ipilimumab, advanced metastatic melanoma patients had extremely low chances of survival. In a randomized clinical trial, it was shown that patients treated with dacarbazine, a widely used chemotherapeutic agent have an 8.8% 5-year survival rate, while this number is 18.2% for patients treated with ipilimumab and dacarbazine together.[112] The numbers were similar for other tumor types, e.g. the application of nivolumab for treating NSCLC.[113,114] Since then, various combinations and novel settings of ICI therapy have been

introduced in clinical practice. For instance, combining ipilimumab with the PD-1 targeting nivolumab significantly improved 5-year survival rates in melanoma to 52%.[106] However, as this result indicates, the number of non-responders is still high.

The significant failure rate of ICI has prompted extensive research into factors that can affect therapy outcomes. Some of these are basic patient characteristics, including age, gender, and clinical history.[115–117] Distinct molecular markers, such as the intra-tumor expression of PD-L1 are predictive for therapy outcome when using pembrolizumab for treating NSCLC or cervical cancer.[36,118–120] Recently, the impact of gut microbiome composition on ICI outcome has also gained significant attention.[121]

Since the introduction of ICIs into clinical practice, the search for germline and tumor genomic factors associated with immunotherapy efficacy has been a key field of research. A shared characteristic among these factors is that they are mostly associated with the process of cellular adaptive immune presentation. Tumor mutational burden (TMB) is considered to be a major prognostic factor determining tumor immunogenicity and ICI outcome.[36,122–126] Due to the characteristics of epitope processing pathways, only a small subset of mutations in a tumor generates peptides that are ultimately presented by HLA-I molecules.[127,128] It is logical to assume that more altered self-proteins increase the probability of forming HLA-I complexes, presenting non-tolerated neopeptides. TMB is one of the few biomarkers recognized by the U.S. Food and Drug Administration to predict tumor immunotherapy outcomes.[129,130] Despite its initial popularity, several studies have since challenged the significance of this metric by itself.[131,132] Further studies showed that not all mutations have the same importance: factors, like the clonality of mutations, could improve TMB as a predictive biomarker.[133]

Various properties of germline HLA alleles/genotypes have also been associated with either positive or detrimental outcomes. In accordance with the studies on infectious diseases in section 1.2, Chowell et al. assumed that if a patient carries a diverse set of HLA molecules, it potentiates the presentation of a broader set of altered peptides. They reported a significant relationship between HLA-homozygosity, certain HLA-I supertypes, and germline HLA-I evolutionary divergence (HED) with the overall survival of patients receiving immunotherapy.[134,135] However, later studies showed the lack of robustness for HLA-I homozygosity and HED as a predictive biomarker, being highly dependent on additional genetic factors and specificities of the treatment.[136,137] Finally, genomic rearrangements may lead to HLA-I copy number loss which has a detrimental effect in lung cancer.[84]

1.8 Binding promiscuity is an inherent property of HLA alleles

TMB, HLA homozygosity, and HED are factors suggested to correlate with the number of presented mutated peptides on the surface of tumor cells, which could eventually increase the probability of successful immune recognition by CD8+ T cells. These factors are properties of the HLA-I genotype of a patient.

HLA-I alleles themselves have an intrinsic characteristic that determines their peptide-binding specificity, and consequently, the breadth of the presentable peptides. Based on the amino acid sequence of the binding pocket, distinct HLA-I proteins are specific for different peptide motifs.[13] The diversity of presentable peptides by an HLA-I molecule is called allele promiscuity. Based on the range of amino acid sequences a given HLA molecule binds efficiently, we can discriminate between generalist and specialist HLA-I alleles.[138] As an example, HLA-B*27:05 prefers to bind nonamer peptides, having arginine (R) residues in position 2. In contrast, HLA-A*30:01 binds a wider range of amino acids, meaning that this molecule is more promiscuous. Sequence specificity motifs for these two alleles are visualized in the sequence logos in Figure 4b indicating a higher level of generality for HLA-A*30:01.[139]

Generalist and specialist HLA alleles may play a critical role in the evolution of the human body's ability to fight against diseases. Previously, we showed the potential association of HLA-II allele promiscuity in distinct human populations with the number of pathogenic microbial species prevalent in the geographical region, indicating a potential protective effect of generalist HLA alleles against infectious diseases. Furthermore, significant trends were found between the intracellular pathogen diversity of a geographical region and the HLA-A promiscuity of local human populations.[140] Conversely, earlier articles proposed a possible detrimental impact of generalist alleles as they may elicit autoimmune responses more frequently by presenting a larger number of self-peptides.[138]

2 AIMS

Here, we hypothesize that promiscuous HLA-I variants are not only capable of binding more diverse peptides of intracellular pathogens in case of an infection but also higher numbers of neopeptides in tumors. We first aimed to form a robust definition of allele-level HLA-I promiscuity, then expand it to create a reliable metric that characterizes the breadth of peptides presented by a patient's HLA-I set. We hypothesized that individuals carrying one or more promiscuous HLA-I alleles would have considerably better clinical outcomes in immune checkpoint blockade therapy.

3 METHODS

3.1 Details of prediction-based promiscuity calculations

We used a dataset listing nucleotide changes identified in ICI-treated patients and generated all possible nonamer peptide fragments, mapped to the affected coding sequences.[141] Next, 10,000 random nonamers carrying a missense mutation were chosen, which was further reduced to 4650 unique peptides after the elimination of highly similar ones using an iterative approach.[140]

We formed the second peptide set based on the COSMIC (Catalogue of Somatic Mutations in Cancer) database, containing somatic mutations from various tumor types. We downloaded the database on the 18th of May 2018. We selected missense mutations that occurred in at least 5 melanoma samples (n = 11,294). Following the methods of Marty et al., we generated all possible unique 8-11-mer neopeptides from the mutated sequences (n = 189,542).[142]

We used the NetMHCpan 4.0 algorithm to predict the binding of these peptide sets to HLA-I molecules.[139] We used a set of 16 HLA-A and 13 HLA-B alleles, collected from a reference set with maximal population coverage from the Immune Epitope Database (IEDB).[143,144] As the list did not include HLA-C, we selected the first four-digit alleles from each two-digit HLA-C allele class (n = 14), overall producing an HLA-I set of 43 alleles.

3.2 Using Kullback-Leibler divergence to calculate allele promiscuity

In an alternative definition of HLA-I allelic promiscuity, we utilized experimental data from the IEDB on peptide-HLA class I interactions, as well as T-cell activation assays, downloaded on the 17th of January 2019.[145] The dataset was filtered for 8-12-mers, and with available positive HLA-binding or T-cell activation assays. To reliably estimate peptide-binding promiscuity, only those HLA-I alleles were included in the analysis, for which at least 400 unique bound peptides were available. The resulting 253,147 peptide-HLA-I interactions gave us an insight into the binding specificities of 21 HLA-A, 31 HLA-B, and 15 HLA-C alleles.
Next, we used Kullback-Leibler divergence, a metric describing the distance between two discrete probability distributions. An overview of the calculation is shown in Figure 2. In order to determine promiscuity for one allele, presented peptides were classified into separate peptide sets by their amino acid lengths. The following steps were performed for each set separately. For each amino acid position, amino acid frequency distribution in the set was compared with that in the complete human proteome (downloaded from the UniProt database) using the Kullback-Leibler divergence (D_{KL}, Figure 2a).[146] We used the Rtreemix R library to calculate

position-specific D_{KL} values. 10^{-7} was added to the frequency of all amino acids if an amino acid had zero frequency in a distinct position. Finally, the mean of position-specific D_{KL} values were calculated for each peptide and peptide set, yielding four peptide-length-specific D_{KL} values. If an HLA-I allele is more selective (i.e., less promiscuous) in peptide binding, the higher deviation we can observe between the peptides presented by the HLA-I molecules and that could be expected based on the human proteome.

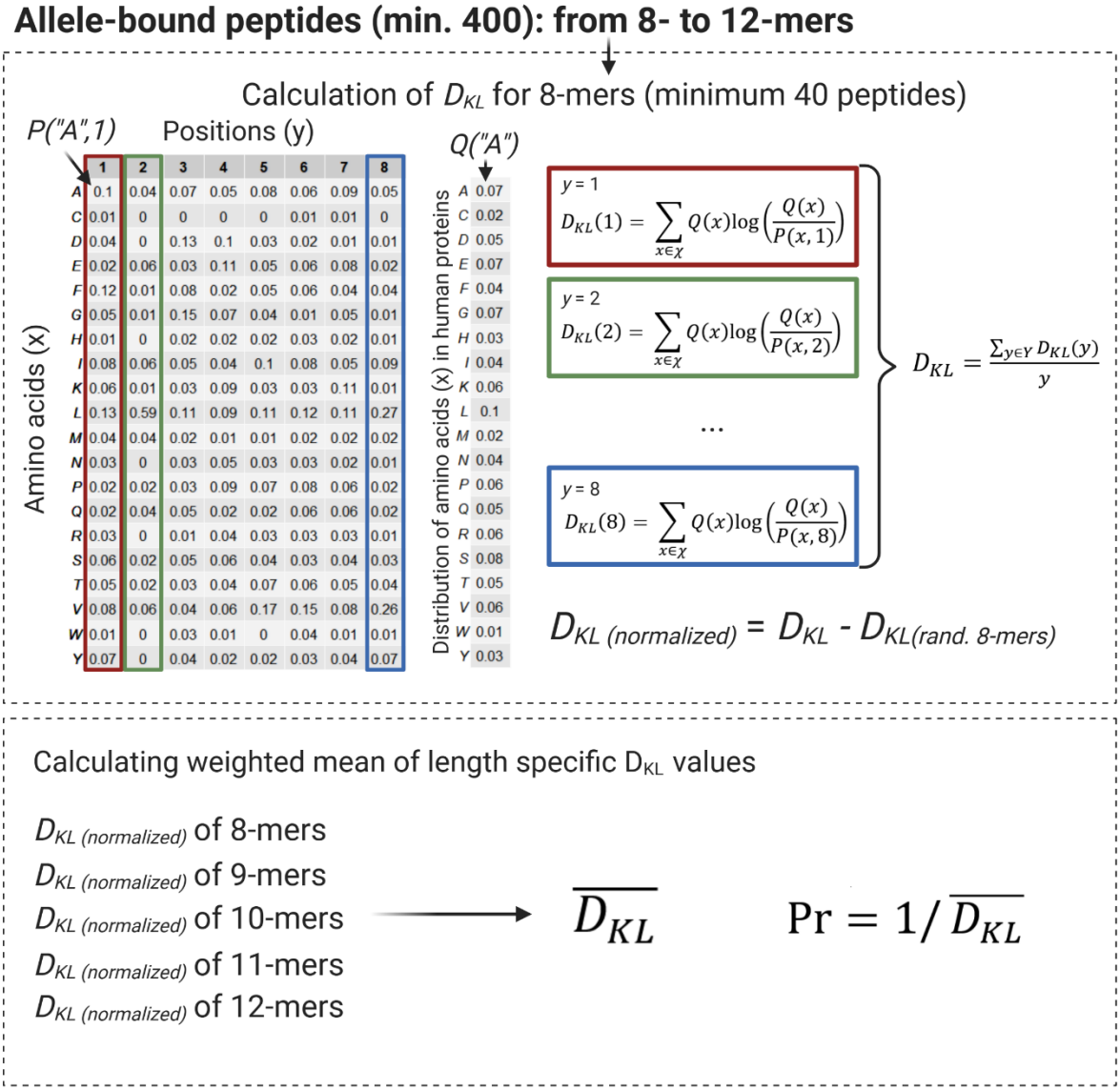


Figure 2**: calculation and reliability of HLA allele promiscuity.* ***Panel a: *$D_{KL(normalized)}$ values were calculated for each peptide length group separately.* ***Panel b:*** *the weighted mean of length-specific values was determined. The figure is created using Biorender.com.*

In order to avoid any potential sample size-related biases, we normalized raw D_{KL} values by subtracting a randomized Kullback-Leibler divergence, $D_{KL(rand)}$. This value is calculated as the mean of the Kullback-Leibler divergence values obtained from 1,000 iterations of randomly selecting peptides with the same length from the human proteome. D_{KL} values after

normalization show no significant correlation with the number of input peptide sequences (Spearman's $\rho = -0.08$, $P = 0.23$).

Subsequently, the mean of peptide length-specific values was calculated, weighted by their relative proportion in the repertoire of peptides belonging to the HLA-I allele of interest, yielding $\overline{D_{KL}}$ (Figure 2b). As the resulting metric is smaller in alleles with high promiscuity (it measures allele specificity), we took the reciprocal value of $\overline{D_{KL}}$ to calculate allele Pr.

3.3 Calculating peptidome diversity of peptides identified by mass spectrometry

To verify the accuracy of the allele Pr estimates, we used mass spectrometry immunopeptidomics data obtained from monoallelic cell lines.[147,148] We calculated the sequence diversity of the peptides using the Shannon entropy index for each HLA-I allele.

3.4 *In vitro* measurement of HLA-I neoepitope binding affinity and complex stability

The TANTIGEN database contains known HLA-I cancer neopeptides.[149] We pre-filtered this dataset by excluding peptides with more than 50% sequence identity using an iterative method, resulting in a set of 29 neoepitopes.[140] Of these, 22 were previously shown to elicit CD8+ T cell response, while the remaining 7 were recognized by tumor-infiltrating lymphocytes *in vivo*.[150] The peptides were synthesized and their binding affinity to pre-selected HLA-I variants was tested using *in vitro* REVEAL binding assays (ProImmune Ltd.). The assay evaluates the ability of candidate peptides to bind to HLA class I alleles by quantifying their capacity to stabilize the HLA-peptide complex. The binding of the tested peptides is measured compared to that of a high-affinity T-cell epitope of the HLA allele of interest.

HLA-I-peptide pairs that exhibited positive results in the REVEAL binding assay were selected for further investigation of complex assembly and stability. Complete rate assays for 66 peptide-HLA complexes were performed (ProImmune Ltd.), measuring the fraction of assembled complexes at six time points in a time frame of 48 hours. The half-lives needed to reach the maximum of assembled complexes were calculated by fitting the values to a one-phase association equation (on-rate values).

Off-rate assays quantified the stability of peptide-HLA complexes by measuring the fraction of denatured complexes at six time points in a duration of 24 hours. The half-lives of the complexes were calculated by fitting these values to a one-phase denaturation equation.

We used the NetMHCstabpan-1.0 algorithm to estimate the stability of HLA-neopeptide complexes.[151] We compiled a dataset of 1,929 experimentally verified neopeptides, all of which

are known to have high binding affinity to certain HLA alleles.[149,152–156] We set the software not to take into account previous information on the exact binding affinity values. Half-life values for a peptide-HLA complex (in hours) were considered as a proxy for complex stability.

3.5 Analyzing cancer outcome in ICI-treated and treatment-naïve patients

Data from multiple sources were used for the analysis of ICI-treatment outcomes, including cohorts of melanoma patients treated with CTLA-4 inhibitors[122,134,157], PD-1 inhibitors[134], and NSCLC treated with PD-1 inhibitors[158]. We excluded patients with zero follow-up time in the anti-PD-1 treated melanoma dataset. We acquired information on HLA genotype, loss of HLA heterozygosity, and cancer mutational burden in these patients from a study by Chowell et al.[134] Information on HED was collected from a later study by the same author.[135] RECIST (Response Evaluation Criteria in Solid Tumors) classification and clinical benefit were compiled from the NCI Genomic Data Commons portal[159] and further independent resources.[122,157,158,160]

Data on TCGA (The Cancer Genome Atlas) patients regarding their treatments and mutational burdens were derived from the NCI Genomic Data Commons portal and cBioPortal, respectively.[159,161] We collected progression-free interval data from a study by Liu et al.[162] HLA genotype data was acquired from Li and colleagues.[163] The additional melanoma cohort, used to evaluate clinical response via RECIST criteria contains ipilimumab-naïve, anti-PD-1-treated patients[160], as well as TCGA melanoma patients, treated with ipilimumab.

For each patient, genotype-level HLA-I promiscuity (termed as genotype Pr) was calculated as the mean of the reciprocal values of $\overline{D_{KL}}$ values across the whole HLA-I genotype of the subject. For patients carrying one HLA-I allele with a missing promiscuity value, we applied an imputation method by replacing the missing promiscuity with the median of existing values for the alleles of the given HLA locus. This technique neither changes the sample median of promiscuity values, nor affects the association between survival and genotype-level HLA-I promiscuity but enlarges the dataset by involving more patients. Subjects having at least two HLA alleles with no known promiscuity values were excluded from the analysis.

The statistical analysis of patient survival was performed using Survminer and Survival R libraries.[164,165] First, patients in each cohort were classified based on genotype Pr. The classification was performed using a fixed cutoff of 2.076, which was determined after merging the three separate cohorts and using a P-value minimization approach from the Survminer R library. The ggsurvplot function of the Survminer R library was applied to perform log-rank tests to evaluate the significance of differences between survival rates of the two groups, formed

based on promiscuity levels. The results were subsequently visualized as Kaplan-Meier curves. In the case of analyses where patients were classified into more than two groups, ordered differences were also tested using the implementation of the Survminer R library.

Cox regression models were built using the coxph function in the Survival R library. To assess the proportional hazard assumption for the Cox regression model fit, we utilized the cox.zph function available in the Survival library. Additionally, the forest plot for multivariate Cox models was generated using the forestmodel R library.

3.6 Calculating Differential Agretopicity Index (DAI)

Differential Agretopicity Index (DAI) is a metric measuring the change in binding affinity of a peptide to an HLA molecule, based on *in silico* predictions. We compiled previously published datasets of experimentally verified neopeptides, binding to certain HLA alleles with high affinity.[149,152–156] As the NetMHCpan-4.0 algorithm showed the highest prediction accuracy for nonamers, we kept only nine amino acid-long peptides in the analysis.[139,166] Peptides with greater than 50% sequence identity were excluded using an iterative method, resulting in 589 neopeptides.[140] To identify the original counterpart of each neopeptide, we aligned them against the reference human proteome using BLAST+ 2.10.0.[167] By retrieving the closest hit, BLOSUM62 (blocks substitution matrix 62) similarities were calculated.[168] We estimated the binding affinity of the original human peptide and its mutated counterpart to the 67 alleles with known allele promiscuity using NetMHCpan 4.0.[139]

We calculated DAI values for each peptide-HLA-I pair as it was previously reported.[166] Median DAI was calculated for each HLA allele, based on all neopeptides that demonstrated significant binding affinity to the specific HLA-I variant. To reduce the likelihood of false-positive results, a binding rank percentile cutoff of 2% was applied to identify neopeptides, ensuring that only the most confidently bound candidate peptides were included in the analysis.[139] In other aspects, DAI calculation was consistent with a publication of Ghorani and colleagues.[166]

In addition, we calculated DAI values for neopeptides based on mutations in 139 melanoma samples from two immunotherapy cohorts.[122,157] Missense mutations were derived from the cBioPortal.[161] We generated the altered peptide sequences using the reference human proteome (downloaded from the UniProt database as of 9th January 2020), collecting all affected 8-12 amino acid long segments in each sample.[146] We calculated binding affinity and DAI values for each HLA-I allele of a patient as described above.

3.7 Analyzing transcriptome data of melanoma samples

We downloaded normalized RNA-sequencing expression data and HLA-I genotypes of anti-CTLA-4-treated melanoma patients from the cBioPortal .[157,161]

For immune deconvolution analysis, we utilized the immunedeconv R library, following the recommendations by Sturm and colleagues.[169] EPIC (Estimating the Proportions of Immune and Cancer cells) algorithm[170] was used for endothelial cells and cancer-associated fibroblasts, MCP-counter[171] (Microenvironment Cell Populations) for macrophages and monocytes, and quanTIseq[172] for regulatory $CD4^+$ T cells. TIDE dysfunction and exclusion score values had already been calculated and were acquired from the TIDE server.[173]

We carried out a gene set enrichment analysis using the GSEA 4.0.3 tool.[174] The software was set to perform 1,000 permutations, a weighted enrichment statistic, and exclusion of gene sets with fewer than five genes. Gene sets of biological processes as reported by the Gene Ontology Consortium were included in the analysis.[175] Gene sets with $P < 0.001$ and false discovery rate (FDR) < 0.1 were considered significant.

3.8 Statistics

In this study, we used retrospective data on cancer patients and validated our results on independent cohorts. As mentioned previously, zero follow-up time patients in the anti-PD-1-treated melanoma cohort could have biased our analysis, so we excluded these subjects from the dataset. Patients having at least two HLA alleles with unknown allelic Pr were also excluded from every cohort in the study since genotype-level Pr values could not be accurately calculated for these patients. Correlations between continuous variables have been determined using Spearman's correlation coefficients and two-sided correlation tests for P-value calculation. On boxplots, vertical lines indicate the median, boxes indicate the interquartile range (IQR) and horizontal lines indicate the first quartile -1.5 × IQR and third quartile +1.5 × IQR. The statistical analyses and visualization were performed using R v.3.6.3 (http://www.r-project.org) and the RStudio 1.2.5033 environment. Smooth curves on plots were fitted with the cubic smoothing spline method.[176]

4 RESULTS

4.1 The road to a proper promiscuity definition

To date, no one has attempted to determine which HLA molecules are generalists and specialists regarding cancer neopeptides. We aimed to find the most proper method to quantify the breadth of presentable mutated peptide fragments by each HLA-I allele. In this section, I present the most significant milestones towards finding the best metrics describing HLA-I promiscuity.

4.1.1 Promiscuity definitions based on HLA-binding predictions

Previously, our group measured promiscuity levels of different alleles by predicting the binding of a representative set of pathogen-derived peptides to HLA molecules.[140] Allele promiscuity was defined as the relative proportion of the peptide set predicted to be bound by the HLA molecule. As an initial approach, we aimed to follow the same logic, by measuring the breadth of neopeptide sequences having predicted binding to distinct HLA alleles.

The NetMHCpan 4.0 algorithm defines the predicted binding of peptides to an HLA-I allele through two main metrics: IC50 (half maximal inhibitory concentration) binding affinity and binding rank. The algorithm classifies peptides with an IC50 value below 500 nM as weak binders, while peptides with values under 50 nM are considered strong binders.[139] However, this threshold may vary depending on the HLA allele.[177] Therefore, binding rank is proposed as a more accurate metric for binding. Binding rank is calculated by comparing the IC50 value of a peptide with those of 200,000 random peptides. The algorithm considers peptides with a binding rank below 2% as weak binders, while the strong binding is defined as rank < 0.5.[139]

Different studies in anti-tumor immunity have used either binding affinity or binding rank to evaluate the binding of peptides to HLA-I alleles. Luksza et al. used binding affinity to build a predictive model for ICI therapy outcomes.[141] Meanwhile, Marty et al. applied binding rank to compute Patient Harmonic mean Best Rank values to describe how presentable the neopeptides derived from certain mutations are by the HLA-I alleles of a cancer patient.[142]

We examined thecorrelations between the two metrics, utilizing the two peptide sets and a reference allele set of 43 alleles (see Methods). We observed strong positive correlations between the binding affinity and binding rank values measured for corresponding allele-peptide combinations, in the case of both peptide sets (Spearman's $\rho = 0.93$, $P < 2.2 \times 10^{-6}$ for both the 4650 neopeptides set and the COSMIC-based peptide set). Next, we calculated the levels of allele promiscuity by determining the proportion of peptides at least weakly or strongly bound

by the 43 alleles. Interestingly, we found no correlation between the affinity-based and rank-based promiscuity scores of the alleles, independently of the peptide set utilized (Figure 3).

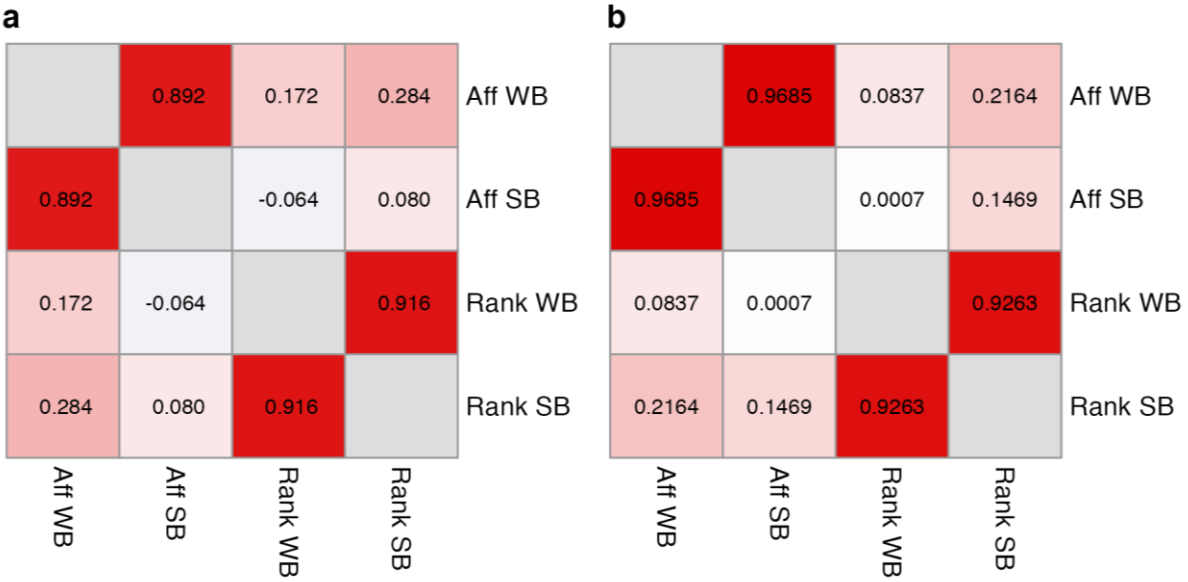


Figure 3: *Spearman's rho values indicating the presence and lack of correlations between certain prediction-based promiscuity values. Aff and Rank indicate that the promiscuity is determined by either predicted IC50 binding affinity or binding rank values, derived from NetMHCpan 4.0 predictions. WB and SB correspond to weak and strong binding.* ***Panels a and b*** *show trends based on the 4650 neopeptide and COSMIC neopeptide sets respectively.*

These results suggest that binding affinity and binding rank are not interchangeable metrics for assessing the peptide binding capacity of HLA-I alleles. Due to its definition, the binding rank may not be able to reliably capture the differences in allele promiscuity levels. A peptide with a binding rank below 2% means that it has a higher binding affinity than 98% of random peptides. But if an allele has a more flexible binding pocket, it would be potentially capable of binding more than 2% of peptides. A fixed 2% rank percentile as binding criterion equalized the peptide binding repertoire sizes of all alleles, thus obscuring the variation in promiscuity. Therefore, the binding rank may reflect more the selectivity than the promiscuity of an allele.

We have previously demonstrated that allele promiscuity is robust to the choice of peptide set and binding cutoff when using affinity measures.[140] Paul et al. have reported that different HLA alleles bind different numbers of pathogen-derived peptides below the predicted 500 nM threshold and that these computational predictions are consistent with experimental data.[177] These findings suggest that affinity binding is a more accurate and reliable basis for promiscuity, compared to binding rank, even though it also means that using the same affinity cutoffs for different alleles to determine bound peptides is not appropriate.

4.1.2 Examining promiscuity using *in vitro* experimental data

To solve these potential issues arising from the usage of binding prediction results, we also aimed to establish a metric purely based on experimental data. To do so, we performed a comprehensive analysis of more than 250,000 peptide–HLA interactions involving 67 HLA-A, -B, or -C alleles with validated *in vitro* assay results from the IEDB database. These alleles are prevalent and represent a substantial proportion of individuals in the human population.[144]

We applied a Kullback-Leibler divergence-based protocol to calculate allele-level promiscuity values, which quantify the amino acid diversity among peptides bound by different HLA-I molecules (see Methods). We observed a significant variation in promiscuity among common HLA-A, -B, and -C alleles (Figure 4a). High promiscuity alleles have more diverse peptide motifs, suggesting that they can bind a broader range of peptides. These sequences are characterized by the presence of either hydrophobic and basic amino acids or tyrosine at anchor positions 2 and/or 9 (e.g. HLA-A*30:01 in Figure 4b). Conclusively, the specificity at these and other positions is low compared to low promiscuity alleles (e.g. HLA-B*27:05).

To validate the reliability of our metric, we compared it with data from two large-scale immunopeptidomics studies.[147,148] They reported naturally eluted self-peptides from the surface of HLA-I monoallelic cell lines, covering 10 HLA-A, 6 HLA-B, and 15 HLA-C alleles. We detected a strong positive association between promiscuity and peptidome diversity on the cell surface (Figure 4c). Our findings were consistent across the two studies (Spearman's $\rho = 0.67$ and 0.56, $P = 0.005$ and 0.03 based on the datasets from Abelin et al. and Di Marco et al., $n = 16$ and 15 alleles respectively). Using multivariate linear regression, we showed that the positive association between promiscuity and cell surface peptide diversity remains even if we control for the locus encoding for the HLA-I allele (coefficient: 0.203, adjusted R-squared: 0.275, P: 0.004, F-test P-value: 0.008).

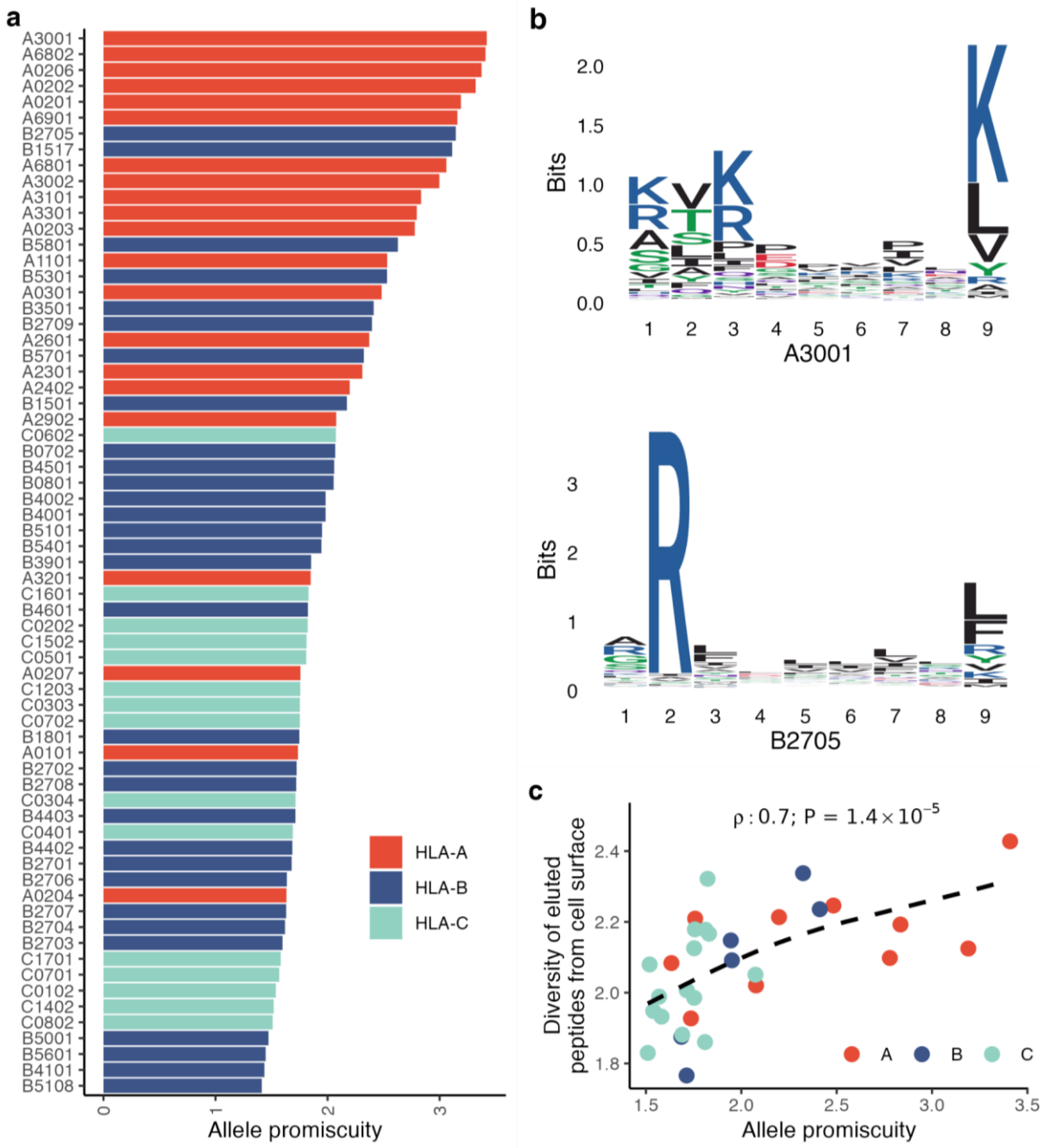


Figure 4: *properties of Kullback-Leibler-based allele promiscuity.* ***Panel a****: the distribution of allele promiscuity values.* ***Panel b****: example binding motifs for an allele with low (HLA-B*27:05) and high (HLA-A*30:01) promiscuity levels, based on immunopeptidomics data by Sarkizova et al.*[178] ***Panel c:*** *the amino acid diversity of self-peptides presented by the given allele on monoallelic cell lines is shown as the function of allele promiscuity. Alleles with higher promiscuity present more diverse peptide sequences (n = 31 alleles).*

4.1.3 Selection of the most appropriate promiscuity definition

To select the most accurate allele Pr definition, we measured the *in vitro* binding affinities of 29 tumor neoepitopes to 11 representative HLA-I alleles using ProImmune REVEAL HLA-peptide binding assays. These alleles are prevalent and belong to a small set of reference HLA

variants that can achieve maximal coverage of human populations.[144] The peptides were selected from the TANTIGEN database,[149] known to induce a cytotoxic immune response *in vitro* or mediate tumor destruction *in vivo*, and have minimal or no sequence overlap (see Methods).[150] We found a significant variation in the numbers of bound peptides (Figure 5).

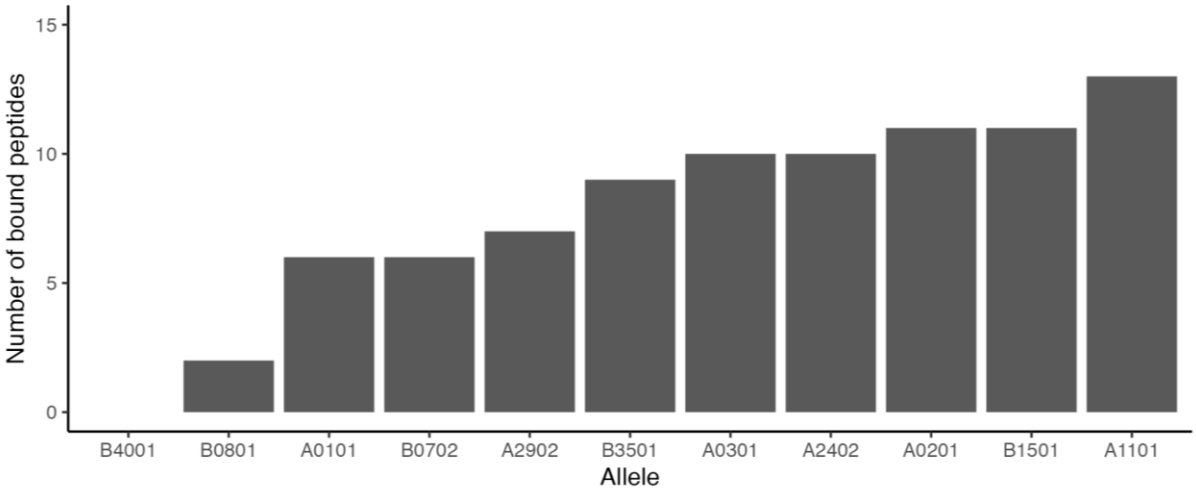


Figure 5*: the numbers of peptides bound of the 29 cancer neopeptides from the TANTIGEN database, according to the in vitro ProImmune REVEAL assays.*

We examined the correlation of different promiscuity values with the number of peptides bound by each allele according to the ProImmune REVEAL binding assays. We found the strongest positive relationship with experimental measurements in the case of the Kullback-Leibler divergence-based promiscuity definition (see Table 1, details on the next page). Conclusively, we used this metric to describe HLA-I allele promiscuity in the subsequent analyses.

Method	**Affinity/rank**	**Binding strength**	**Peptide set**	**Spearman's rho**	**P**
KLD	*irrelevant*	*irrelevant*	IEDB	0.86	0.00068
prediction	affinity	weak	4650	0.82	0.00203
prediction	affinity	strong	COSMIC	0.75	0.00778
prediction	affinity	weak	COSMIC	0.75	0.00778
prediction	affinity	strong	4650	0.54	0.08392
prediction	rank	strong	4650	0.53	0.09368
prediction	rank	weak	4650	0.37	0.26810
prediction	rank	strong	COSMIC	0.18	0.59002
prediction	rank	weak	COSMIC	0.11	0.73764

Table 1: *correlations between promiscuity metrics calculated using different methods and in vitro measured breadth of presentable neopeptides per HLA-I alleles according to the ProImmune REVEAL binding assay. Spearman's rho and two-sided correlation test P values are indicated. KLD represents Kullback-Leibler divergence, prediction refers to NetMHCpan 4.0-based promiscuity metrics. Strong and weak binding is defined in section 4.1.1.*

4.2 HLA-I promiscuity as a prognostic factor for tumor therapy outcomes

4.2.1 Effects of variation in genotype Pr among cancer patients treated with ICI

To identify the effects of genotype-level mean HLA-I promiscuity (gentype Pr) on cancer immunotherapy, we collected publicly available data about patients (n = 316) treated with ICI. We grouped the subjects into three independent cohorts. One cohort consisted of 164 patients who were administered anti-CTLA-4 as a therapeutic intervention for melanoma.[122,134,157] The second cohort, comprising 74 patients, underwent anti-PD-1 therapy for NSCLC.[158] The third group comprised 78 melanoma patients who were treated with anti-PD-1/anti-PD-L1 therapy.[134] These datasets included information on the whole HLA-I genotype of the subjects as well, enabling us to calculate their genotype Pr. The metric showed a substantial variation across patients in the examined cohorts (Figure 6).

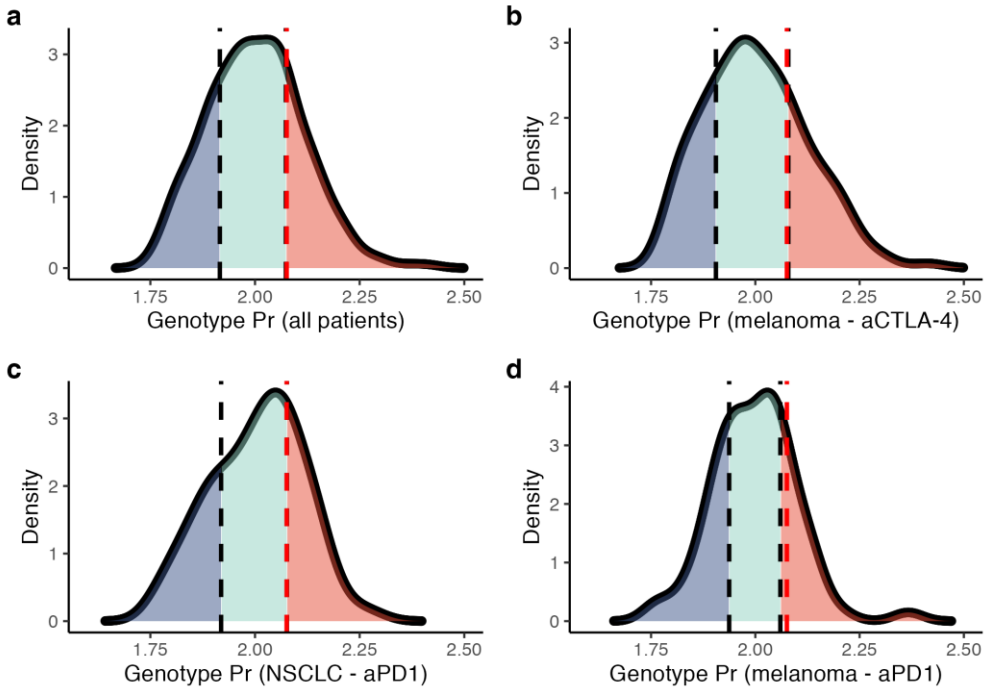


Figure 6: *Distribution of genotype Pr in cancer immunotherapy cohorts. The density of genotype Pr is shown for all patients (**panel a**, n = 316), melanoma patients treated with CTLA-4 inhibitors (**panel b**, n = 164), non-small cell lung cancer (NSCLC) patients treated with PD-1 or PD-L1 inhibitors (**panel c**, n = 74) and melanoma patients treated with PD-1 or PD-L1 inhibitors (**panel d**, n = 78). Dashed black vertical lines represent the 25th and 75th percentile values, while dashed red vertical lines represent the fixed genotype Pr cutoff of 2.076 used for patient classification.*

Next, we examined how genotype Pr affects the overall survival of patients, treated with ICI therapy. To show the robust effect of promiscuity on therapy outcome, we formed low and high genotype Pr groups among the subjects in each cohort using the unified threshold value of 2.076 (see Methods for details). Contrary to our expectations, high genotype Pr was associated with significantly reduced overall survival in all cohorts. Melanoma patients under CTLA-4 inhibitor treatment in the high genotype Pr group showed 44% lower survival rates compared to the low promiscuity group. Among the PD-1-inhibitor-treated NSCLC patients, high genotype Pr subjects were 62% less likely to survive compared to the other group. Finally, high HLA-promiscuity anti-PD-1/anti-PD-L1-treated melanoma patients showed 70% worse survival. All these differences were statistically significant (Figure 7, two-sided log-rank test P = 0.006, P = 0.008, and P = 0.043 respectively).

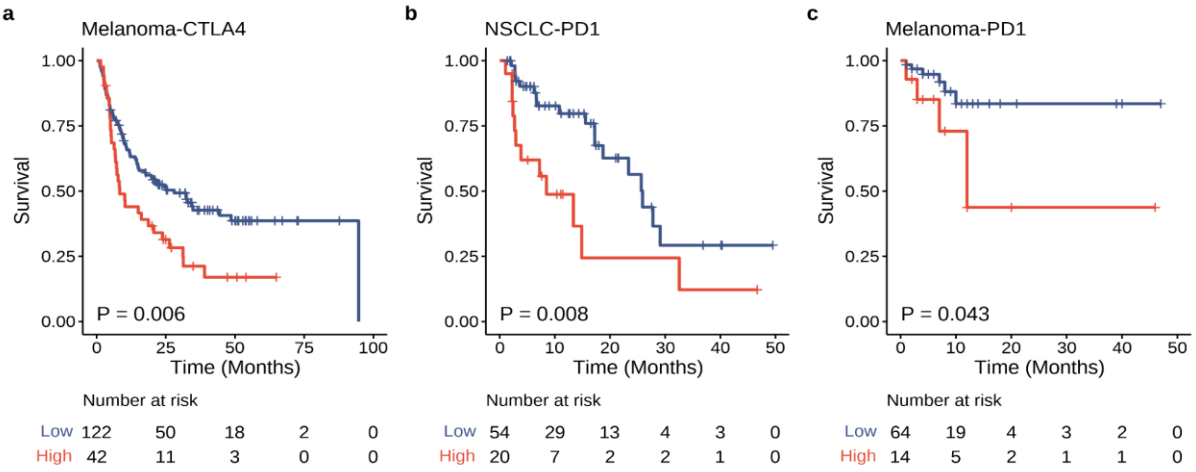


Figure 7: *The association of high HLA-I genotype promiscuity and reduced overall survival. Analysis was carried out for patients with melanoma who were receiving anti-CTLA-4 therapy (**panel a**, n = 164 patients), patients with NSCLC who were receiving anti-PD-1 therapy (**panel b**, n = 74 patients), and patients with melanoma who were receiving anti-PD-1/anti-PD-L1 therapy (**panel c**, n = 78 patients). Higher mean HLA class I promiscuity predisposes to lower overall survival time compared to subjects with low genotype Pr.*

Previous literature suggested that HLA-I molecules encoded by distinct loci may have different weights and roles in adaptive immune recognition.[134,179] Thus, we next examined how the low promiscuity of each HLA-I locus contributes to improved survival rates of immunotherapy patients. We calculated the mean HLA-I promiscuity for HLA-A, -B, and -C alleles of each patient, as well as the combinations thereof (Table 2). Univariate Cox models showed that HLA-B genotype promiscuity was a predictive factor of survival, but this negative association becomes even more significant when HLA-A and -C promiscuity levels are included as well, indicating the importance of all three HLA class loci.

Locus	Coeff.	P (var.)	Log-rank test P	Log-likelihood	ANOVA P (vs. ABC)
ABC	2.380	$8.6 * 10^{-4}$	0.001	-700.915	*NA*
A	0.309	0.164	0.166	-705.346	$< 2 * 10^{-16}$
B	1.100	0.004	0.005	-702.353	$< 2 * 10^{-16}$
C	1.038	0.190	0.196	-705.473	$< 2 * 10^{-16}$

Table 2: *The effect of genotype Pr on patient survival when calculated using individual loci. We examined four univariate Cox models containing the mean promiscuity of individual loci (Models 2 to 4) and all loci (Model 1) as predictors. The effect of variables on patient survival is shown. "P (var.)" values indicate two-sided P values of Z statistics. Two-sided log-rank test P-values and the log-likelihood values for individual models are also indicated. Significant two-sided ANOVA (Analysis of Variance) P-values indicate that the first model fits better than the examined one. NA: not applicable.*

Response Evaluation Criteria in Solid Tumors (RECIST) is a collection of rules defining the behavior of tumors and the direction of disease in cancer patients.[180] The condition of a subject can improve ("respond"), stay the same ("stabilize"), or worsen ("progress") as a result of treatment. In addition to survival rates, we also tested the association of genotype Pr with clinical outcomes using RECIST criteria. In previous publications, treatment was defined as clinically beneficial if it showed a complete or partial response. On the contrary, stable disease or progressive disease was considered as an absence of clinical benefit to the treatment.[181,182] RECIST categorization was available for the anti-CTLA-4 treated melanoma and the anti-PD-1 treated NSCLC patient cohorts, as well as an additional independent cohort composed of TCGA patients treated with ipilimumab and subject data published by Riaz and colleagues.[160] In accordance with the previously presented results on survival rates, patients with no clinical benefit were shown to have significantly higher genotype Pr (Figure 8a and 8b, Wilcoxon test P = 0.007 and 0.008 in the two cohorts respectively). Patients with high genotype Pr were more than twofold less likely to have clinical benefit as a result of ICI treatment in the original cohort (Figure 8a, RR = 2.2, Fisher's Exact Test P = 0.04). Moreover, even stronger trends were observed in the second group of patients, treated with either PD-1- or CTLA-4-inhibitors (Figure 8b, RR = 5.5, Fisher's Exact Test P = 0.01).

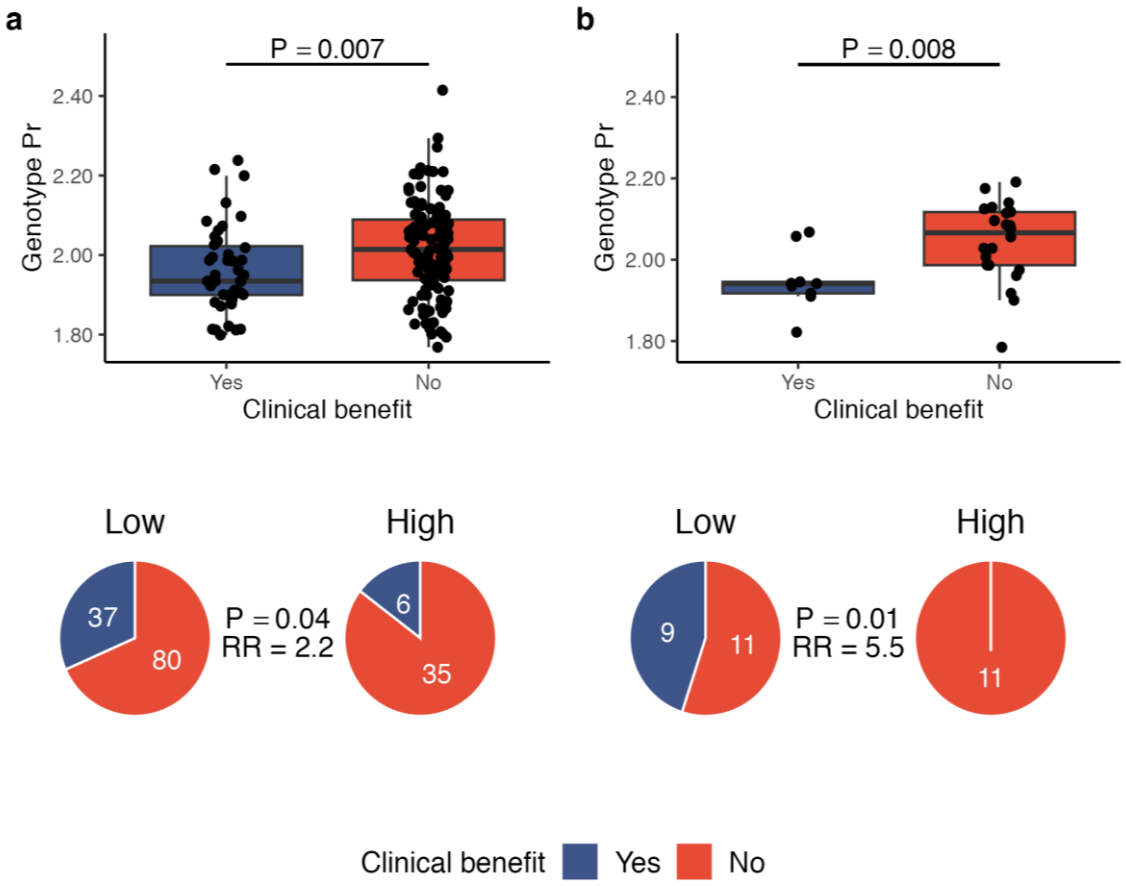


Figure 8: *The association between HLA-I genotype promiscuity and clinical benefit.* ***Panel a:*** *patients with clinical benefit from the anti-CTLA-4 melanoma and anti-PD1 NSCLC cohorts had significantly lower genotype Pr (n = 43 and 115 in patient groups with and without clinical benefit, respectively). Similarly, high genotype Pr was associated with a 2.2-fold reduced chance of clinical benefit after ICI therapy.* ***Panel b:*** *response to ICI therapy in an independent melanoma cohort. Patients were stratified into low (n = 20 patients) and high (n = 11 patients) promiscuity groups using the same cutoff as in* ***panel a.*** *In both* ***panels a and b,*** *P values of a two-sided Wilcoxon's rank-sum test and a two-sided Fisher's exact test are indicated. In the case of* ***panel b****, risk ratios had been calculated by adding one to each element of the contingency table to avoid division by zero (not indicated in the pie chart).*

4.2.2 Effects of genotype Pr variation among non-ICI cancer patients

As ICIs allow the antitumor immune response to freely prevail, we expected the strongest effect of genotype Pr on therapy outcomes in immunotherapy patients. However, as HLA-I-mediated immune presentation plays a key role in antitumor immunity in general, we briefly examined if similar trends are visible in those patients who had never received ICI therapy. The Cancer Genome Atlas is a comprehensive dataset of about 20,000 tumors and matched normal samples from 33 tumor types, maintained by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI).[183] In addition to the previously mentioned immunotherapy cohorts, we analyzed data about melanoma and NSCLC patients.

We divided melanoma samples (n = 301) in the TCGA database into high- and low-mutational-burden (TMB) groups based on the median of mutation count values in the cohort (Figure 9a). In the case of this tumor type, we observed progression-free survival reduced by 38% in patients with high TMB and genotype Pr compared to the low promiscuity group (Two-sided log-rank test P = 0.014). Meanwhile, no significant trends were visible for low TMB samples, suggesting a potential interaction between HLA-I promiscuity and tumor mutational burden (Figure 9b, two-sided log-rank test P = 0.926). In contrast, no impact of HLA-I promiscuity was shown for NSCLC patients, independently of the tumor mutational burden levels (Figure 9c-f). These contradictory findings suggest additional data may be necessary, or the inclusion of other factors, such as tumor mutational signatures should be considered.

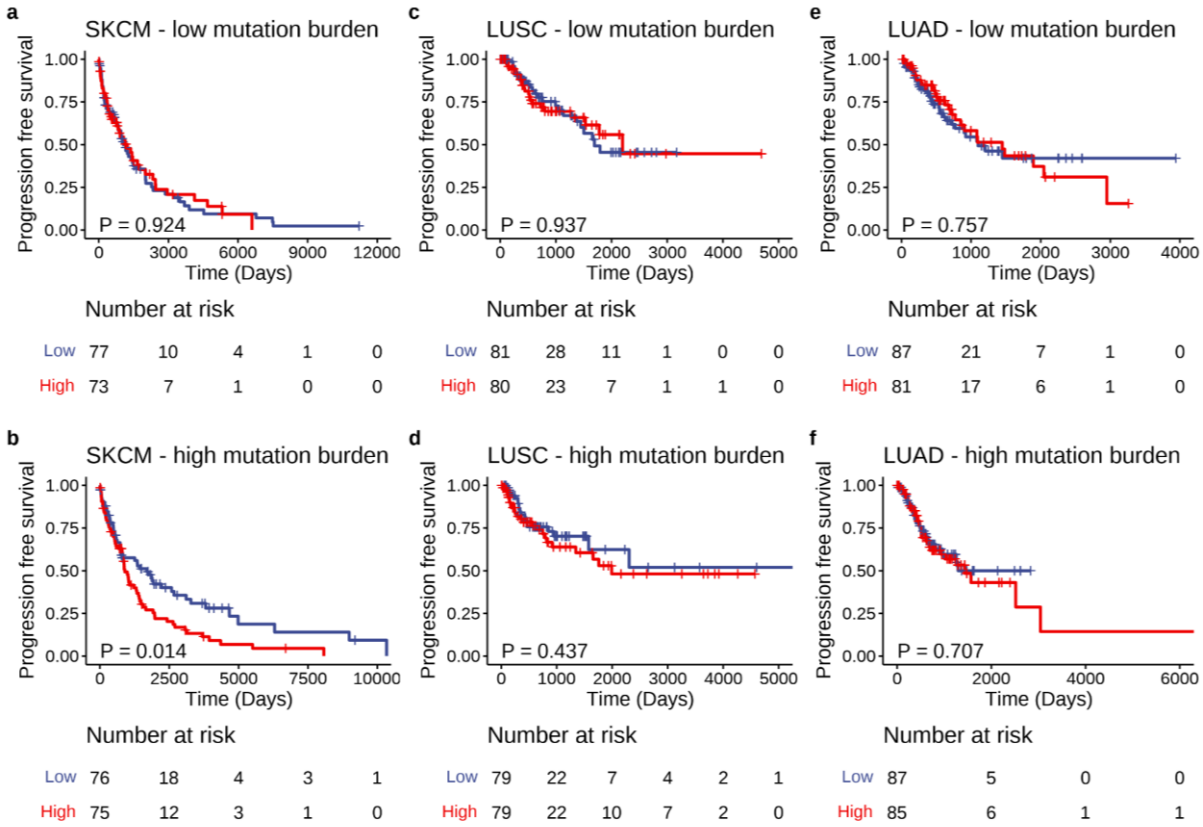


Figure 9: *the relationship between HLA promiscuity and progression-free survival of treatment naïve melanoma* ***(panels a and b)*** *and NSCLC patients* ***(panels c-f)****. Similarly to previous studies*[135]*, patients were classified into low* ***(panels a, c, e)*** *and high* ***(panels b, d, f)*** *mutational burden cohorts using the median as a cutoff. In each cohort, patients were classified into low and high-promiscuity groups using the median as a cutoff. Two-sided log-rank test P values are shown. The vertical axes indicate the probability of progression-free survival. LUSC: lung squamous cell carcinoma; LUAD: lung adenocarcinoma.*

4.3 The effects of HLA-I promiscuity on ICI therapy in context of further factors

As mentioned previously in the *Introduction,* immune checkpoint blockade therapy has variable efficacy among patients, which motivates the search for key factors that determine treatment success. Based on the results described in the previous subchapter, we assumed that genotype Pr could be an additional determinant to consider when deciding on starting an ICI therapy. To evaluate whether HLA-I genotype promiscuity is an independent prognostic factor for immunotherapy outcome, we tested its robustness and compared it to established biomarkers.

4.3.1 Effects of separate HLA alleles on the predictive power of HLA-I genotype Pr values

Arithmetic means are sensitive to outlier values. As HLA-I genotype promiscuity is calculated as the arithmetic mean of values belonging to six alleles, it is important to test if the observed trends are indeed the results of the genotype Pr, and not only the effects of certain alleles with extreme values. We iteratively performed survival analysis for the main ICI cohorts (n = 316), by excluding individuals carrying each specific HLA allele. We included only those alleles in the analyses that were carried by at least 10% of the individuals. We found that the association between HLA-I promiscuity and patient survival remained significant, which indicates that these trends cannot be explained by a single, highly prevalent allele with special peptide-binding properties (Figure 10, details continue on the next page).

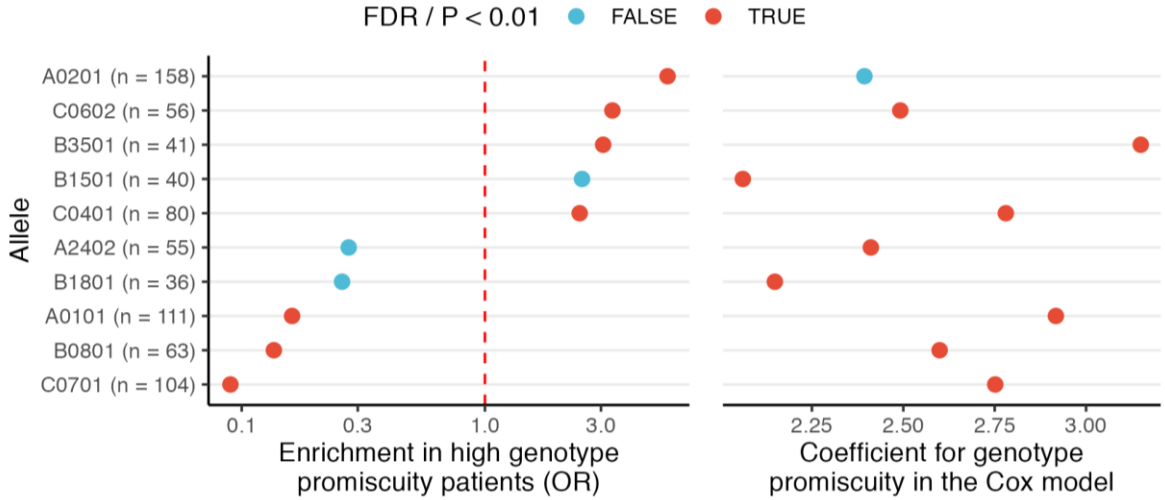


Figure 10: *the effect of HLA allele promiscuity on survival is independent of the presence of a single allele. We analyzed the enrichment of alleles in different promiscuity groups of the merged dataset including all patients from the three immunotherapy cohorts (N = 316). Only those alleles were included in the analysis that were carried by at least 10% of the individuals. We stratified patients into low and high genotype Pr groups using the cutoff described previously for the immunotherapy cohorts (see Methods). Next, we determined the odds ratio (OR) for finding an allele among high genotype Pr subjects and calculated if the enrichment is significant using Fisher's exact tests. The left panel shows data corresponding to FDR < 0.05*

alleles. ***On the left panel,*** *points on the left and right sides of the dashed red line indicate under- or overrepresentation of the allele in the high promiscuity group, respectively.* ***On the right panel,*** *points indicate coefficients of the Cox models that show the effect of genotype Pr on survival, when excluding patients carrying each allele separately. Positive coefficients indicate a lower probability of survival among high genotype Pr subjects. Red points indicate FDR and P values under 0.01 in the two panels, respectively.*

Furthermore, we checked how the number of promiscuous HLA-I alleles in a patient's genotype affects ICI therapy outcome. Patients carrying multiple promiscuous alleles showed decreased overall survival compared to subjects with zero or one promiscuous HLA-I variant (two-sided log-rank test, tested for trend; P = 0.027).

As it was mentioned previously, distinct HLA-I variants significantly differ in their peptide-binding regions. However, genetically related HLA-I alleles can be classified into supertypes based on their similar structural properties and peptide binding capabilities. The most common HLA-A and -B alleles are mostly covered by 12 supertypes.[184] Previously, the positive prognostic effect of the B44 superfamily was reported after ICI treatment, but the reasons remained unclear. Specific variants in the superfamily had been associated with increased patient survival, including B*18:01, B*44:02, B*44:03, B*44:05, and B*50:01.[134] In general, we found that allele-level promiscuity values (which we were able to calculate for all alleles listed above except for B*44:05) are lower for B44 alleles compared to others (Figure 11, Two-sided Wilcoxon's rank-sum test, P = 0.049).

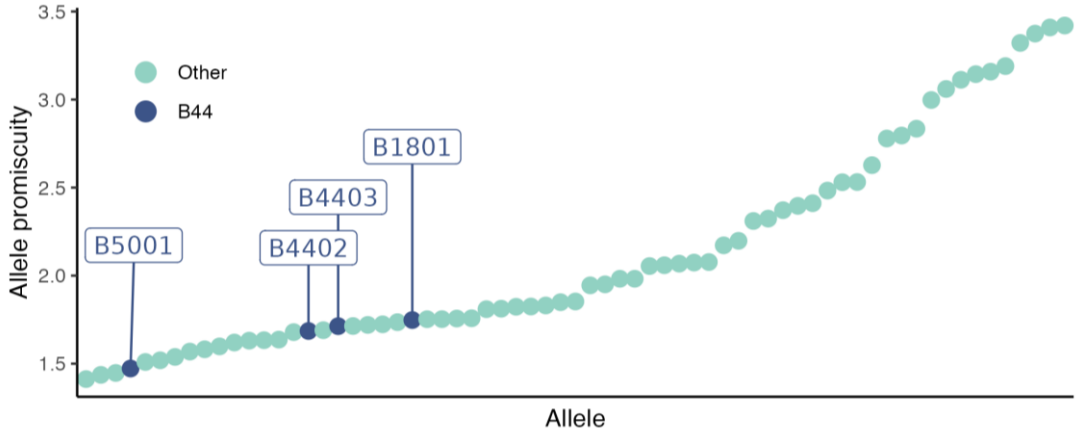


Figure 11: *HLA-I alleles belonging to the B44 (dark blue) supergroup have low promiscuity compared to other alleles (two-sided Wilcoxon's rank-sum test P = 0.049).*

4.3.2 Controlling for tumor mutational burden, HLA-I heterozygosity, and HED

As tumor mutational burden, HLA-I heterozygosity, and HED have previously been suggested as determinants of response to ICI therapy, we decided to examine if HLA-I genotype promiscuity is a predictor for patient survival independently of these established

factors.[126,134,135] To note, all these properties are expected to increase the breadth of presented neopeptides, i.e. the potential targets for adaptive immune recognition of tumor cells.

First, we aimed to determine if any of these characteristics exhibited a statistically significant association with genotype Pr. We could not find a significant difference between genotype promiscuity values of homozygous and heterozygous patients (Figure 12a, Two-sided Wilcoxon's rank-sum test, P = 0.369). Furthermore, there were no correlations of patient-level promiscuity values either with mean HED (Figure 12b, Spearman's ρ: −0.02, P = 0.77, n = 316 patients) or mutation burden (Figure 12c, Spearman's ρ: 0.02, P = 0.77, n = 316 patients).

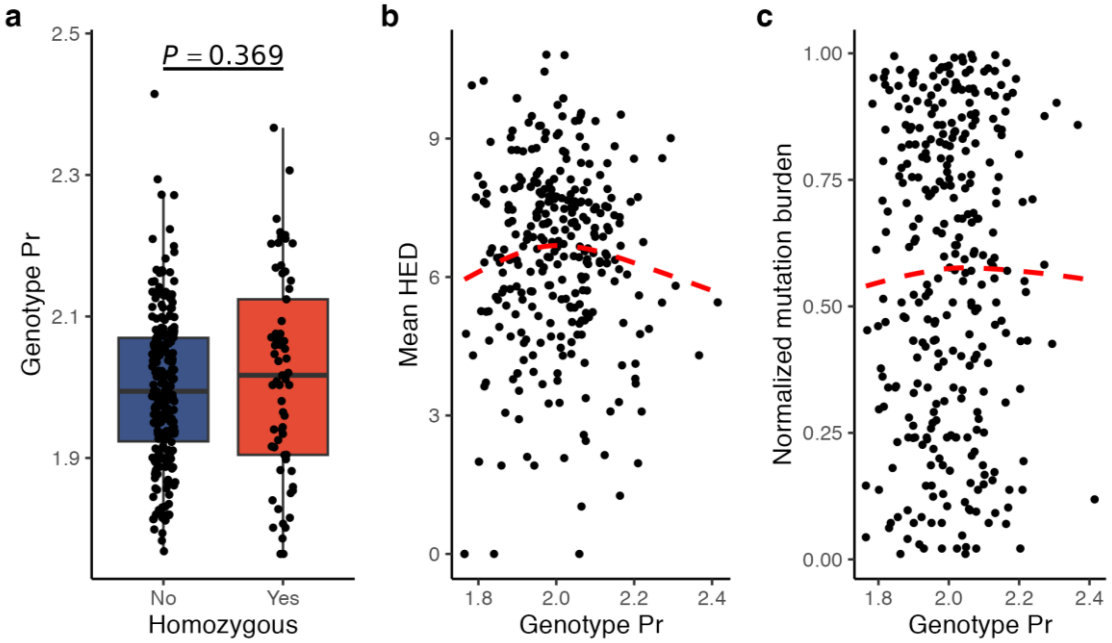


Figure 12: *the lack of significant association between genotype Pr and previously identified factors associated with ICI efficacy. All patients from the three main immunotherapy cohorts were included in the analysis (n = 316).* ***Panel a:*** *there was no significant difference in genotype Pr between fully HLA heterozygous patients (n = 251) and the ones homozygous for at least one HLA-I locus (n = 65 patients).* ***Panel b:*** *similarly, there was no significant association between genotype Pr and mean HED (Spearman's ρ = −0.02, P = 0.77, n = 316 patients).* ***Panel c:*** *there was no significant association between genotype Pr and tumor mutational burden of cancer immunotherapy patients (Spearman's ρ = 0.02, P = 0.77, n = 316 patients).*

Our next objective was to gain insight into the effects of these genomic properties on patient survival. We built a multiple Cox regression model, including genotype Pr, TMB, HLA-I heterozygosity, and HED as well. Tumor mutational burden and HLA-I genotype Pr were found to be the strongest factors in determining overall survival. Meanwhile, HED and HLA-I heterozygosity had no significant effect on therapy outcomes. Notably, not even the presence of B44 and B62 alleles (both supertypes have been associated with better ICI therapy outcomes) as an additional factor in the model showed significant effects on survival time (Figure 13).[134]

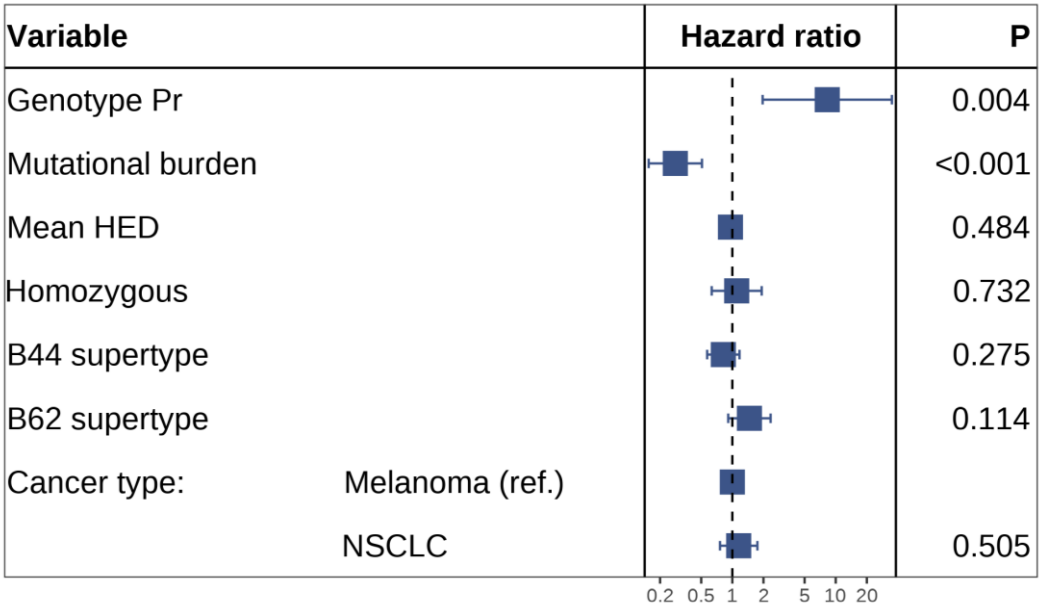


Figure 13: *the effect of genotype Pr on survival remained highly significant in a multiple Cox regression model. The model includes TMB, additional HLA-associated features, and cancer type as covariates. The model involved all patients from the main immunotherapy cohorts (n = 316). Blue squares indicate hazard ratios and horizontal lines indicate 95% confidence intervals. Two-sided P values of z-statistics are shown. The two-sided log-rank test P value of the model was* 5×10^{-6}*. Genotype Pr, TMB and mean HED were treated as continuous variables.*

4.4 Exploring the background of poor prognosis in high genotype Pr patients

Contrary to our preconceptions we found that high genotype Pr is associated with poor prognosis in cancer patients treated with ICI therapy. In the subsequent sections, I present our efforts to identify the underlying causes behind these unexpected trends.

4.4.1 Binding multispecificity does not compromise stability

The stability of peptide–HLA complexes is essential for effective antigen presentation.[185] It is a widely accepted view in enzymology that more promiscuous substrate binding negatively correlates with complex stability.[186] Here, we examined if high HLA-I allele promiscuity is associated with a reduced capacity to form stable protein complexes with neopeptides - assuming that this phenomenon might explain the poor survival rates among high genotype promiscuity patients.

We utilized the *in vitro* ProImmune Complete Rate assays to measure the assembly and dissociation rates of 66 representative neopeptide-HLA class I complexes. Interestingly, we observed a significantly shorter time for complex assembly (Figure 14a and c) and longer half-life of the resulting complexes (Figure 14b and d) in the case of high promiscuity HLA alleles.

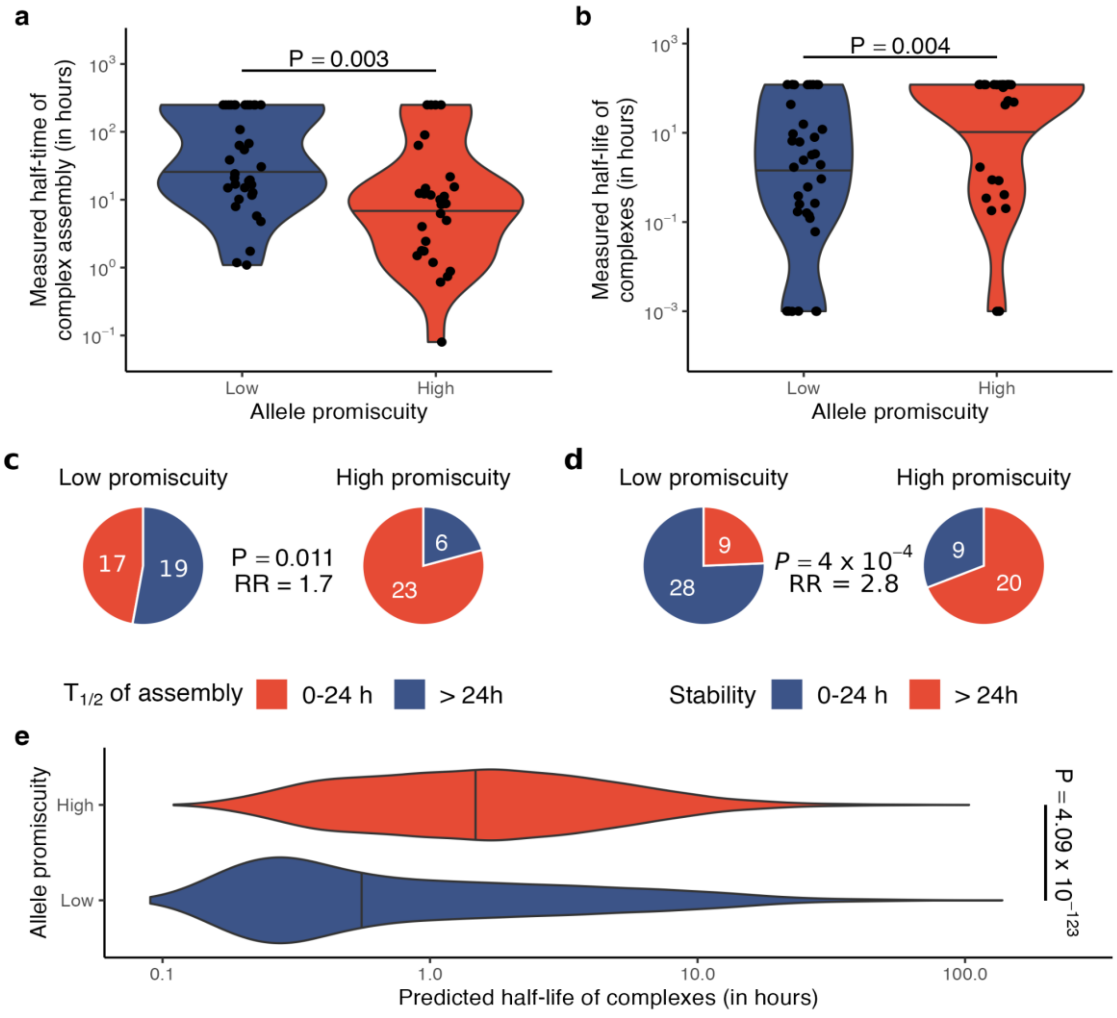


Figure 14: *high allele Pr is associated with higher stability peptide–HLA complexes.* ***Panel a:*** *in vitro measured half-lives of complex assembly for 65 allele-neoepitope pairs. Alleles were stratified into low- (n = 36 complexes) and high- (n = 29 complexes) promiscuity groups based on the median allele Pr. The result for one allele-neopeptide pair is not shown because stable complexes were not detected at the end of the measurement period.* ***Panel b:*** *in vitro measured stability of 66 allele–neoepitope pairs. Peptide-HLA complexes of promiscuous HLA variants are more stable. The vertical axis shows the complex half-life in hours (n = 37 and 29 complexes in low and high allele Pr groups, respectively). For allele-peptide pairs that did not form any stable complexes at the start of the measurement period, the complex half-life was defined as 0.* ***Panel c and d****: promiscuous alleles are 1.7-fold more likely to form complexes with an assembly half-life ($T_{1/2}$) <24 h (**panel c**) and 2.8-fold more likely to form particularly stable complexes with $T_{1/2}$ > 24 h (**panel d**). P values for two-sided Fisher's exact tests and relative risk (RR) values are indicated.* ***Panel e:*** *Predicted half-life of neopeptide–HLA complexes using the NetMHCstabpan algorithm.*[151] *Complexes were identified by predicting the binding affinity of 1,929 neopeptides to 67 HLA-I alleles (Fig. 1d and Methods). Alleles were stratified into low- (n = 3,518 complexes) and high- (n = 4,895 complexes) promiscuity groups using the same cutoff as in panels a and b. For plots in* ***panels a, b, and e****, P values for two-sided Wilcoxon's rank-sum tests are indicated. For* ***panels a and b****, violin plots show the density function of values indicated on vertical axes, and horizontal lines indicate the median value in each group.*

We extended the *in vitro* findings with the results of an *in silico* experiment. Using NetMHCstabpan[151], we predicted the HLA-peptide complex stability for 67 HLA class I alleles and a set of 1,929 experimentally verified neopeptides. Predictions also suggested increased levels of stability among promiscuous HLA-I molecules (Figure 14e).

In overall, in contrast to other proteins, binding multispecificity and binding stability of HLA-I molecules are positively correlated.

4.4.2 High HLA-I promiscuity hampers discrimination between self and altered self

A potential reason for the immunogenicity of a neopeptide is that its original counterpart is not bound by the patient's HLA-I molecules, while the mutated one is presented on the surface of the tumor cells. As previously described in the *Introduction*, there are central tolerance mechanisms remove T cells that are reactive to self-peptides – but it requires HLA-binding.[21,22] Conclusively, it has previously been proposed that tumors carrying more neopeptides with highly increased binding capabilities to the patient's HLA-I molecules should be more susceptible to immune recognition, and, thus, responsive to ICI therapy.[141,166,187,188]

Differential Agretopicity Index (DAI) is an indicator measuring the difference between the predicted affinity values for a neopeptide and its homologous non-mutated pair to a specific HLA-I allele. Therefore, in the case of high DAI neopeptides, the original one had a much lower predicted chance for binding to an HLA-I molecule compared to its mutated counterpart.[166]

In the next step, we examined how certain HLA-I variants discriminate between self and mutated peptides. We applied a set of 589 experimentally verified tumor neopeptides (see Methods), by calculating the median DAI for these sequences to the 67 HLA-I alleles with known allele Pr. In conclusion, we found a strong negative correlation between DAI and the allele Pr of HLA-A and HLA-B alleles, but not for HLA-C alleles (Figure 15).

Conclusively, we expected to see an abundance of high-DAI neopeptides in tumor samples of patients having low HLA-I genotype promiscuity. By analyzing nonsynonymous mutations in melanoma samples of 139 patients treated with CTLA-4 inhibitors, we determined the complete set of neopeptides presented by at least one HLA-I allele of the corresponding subject (see Methods for details). Consistent with the findings for separate alleles, the median DAI for neopeptides showed a significant negative association with genotype Pr as well (Figure 16a, Spearman's $\rho = -0.34$; $P = 3.78 \times 10^{-5}$). In accordance with the literature, higher median DAI was also found to be positively associated with patient overall survival (Figure 16b, two-sided log-rank test, tested for trend, $P = 0.0001$).[166]

Taken together, these results suggest that patients who carry promiscuous HLA-I alleles have a reduced ability to differentiate between self-peptides and the corresponding mutant tumor neopeptides, which may have implications for patient survival.

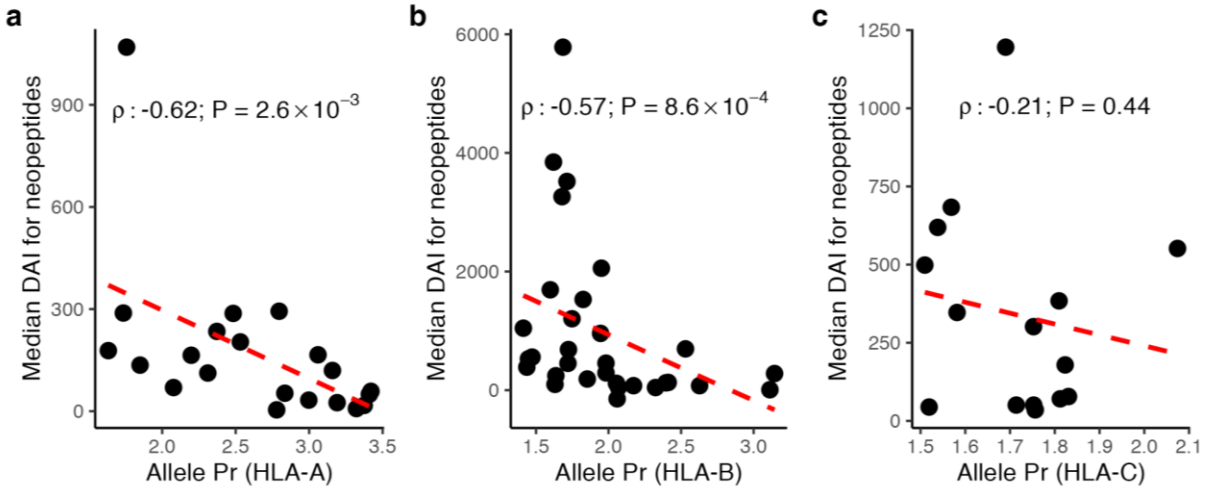


Figure 15: *correlations between allele promiscuity and median DAI for neopeptides. The relationship between HLA-A (**panel a**, n = 32 alleles), HLA-B (**panel b**, n = 50 alleles), and HLA-C (**panel c**, n = 21 alleles) allele promiscuity and median DAI for bound tumor neopeptides are shown. Higher HLA-A and -B promiscuity is associated with lower median DAI values, while there is no significant relationship between the two variables in the case of HLA-C.*

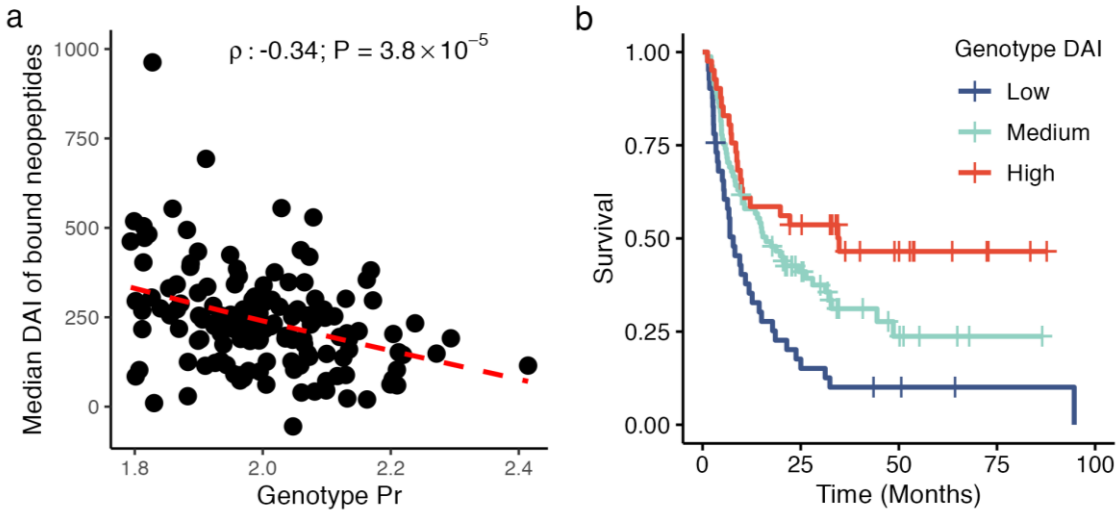


Figure 16: *Correlation of patient-level median DAI and genotype Pr values. **Panel a:** in melanoma samples (n = 139), neopeptides bound by promiscuous HLA alleles have a lower median DAI. Spearman's ρ and the associated P values are shown. **Panel b:** patients from the anti-CTLA-4-treated melanoma cohort were stratified into low- (n = 41 patients), medium- (n = 81 patients), and high- (n = 41 patients) genotype DAI groups based on the first and third quartiles of all values. A higher median DAI was associated with longer survival. The two-sided log-rank test P value is 0.0001. Between-group differences were also tested for trend (Methods). Y axis indicates the probability of survival.*

4.4.3 Association of high HLA-I genotype promiscuity with tolerogenic responses

Our hypothesis on peripheral tolerance

Complementary to central tolerance, CD8+ T cells can also be inhibited through several pathways on the periphery. Peripheral tolerance is initiated when original self-peptides are presented by HLA-I molecules.[141,166] These mechanisms aim to avoid self-reactive T cells from unnecessary activation and subsequent autoimmune reactions.[24] These otherwise important processes can be exploited by tumors, as T-cell tolerance in the tumor microenvironment can enhance uncontrolled cell growth.[189,190] As promiscuous HLA-I alleles are seemingly less able to discriminate between original and altered self-peptides, regulation of T cells may be shifted towards tolerance responses instead of immune destruction.

Here, we aimed to identify molecular and cellular properties of T-cell tolerance in cancer samples. We utilized a dataset about patients with melanoma treated with anti-CTLA-4 antibody ipilimumab (n = 30), as it includes information for each patient's pre-treatment transcriptomes, HLA genotype, and disease progression as well.[157]

Gene set enrichment analysis (GSEA) is a computational method that determines if a set of genes (for example, sharing common biological function, chromosomal location, or regulation) exhibit statistically significant, consistent differences between two biological conditions. RNA expression is one of the most frequently used variables in GSEA.[174] We investigated whether genes with certain functions are more expressed in tumors of high genotype Pr patients compared to the low group. The analysis showed that gene sets associated with the positive regulation of T-cell induction, type 2 immune response, extracellular matrix secretion, and macrophage induction were upregulated in the high genotype Pr group, as well as genes associated with the negative regulation of T helper-1 cell-mediated immune response and CD4+ alpha–beta T-cell differentiation (Figure 17).

All these gene classes are associated with an immunosuppressive tumor microenvironment. As described previously in the *Introduction*, the activity of macrophages may lead to the inhibition of effector immune responses, as well as angiogenesis.[191] Meanwhile, type 2 immune response is reported to be significantly less effective against cancer cells, compared to the T helper cell-mediated type 1 response.[191,192] The transcriptomic markers of CD4+ T-cell differentiation are significantly lower, which is also an indicator of poor prognosis.[189,190] The role of extracellular matrix and immune suppression in cancer progression has also been described in previous literature.[193,194]

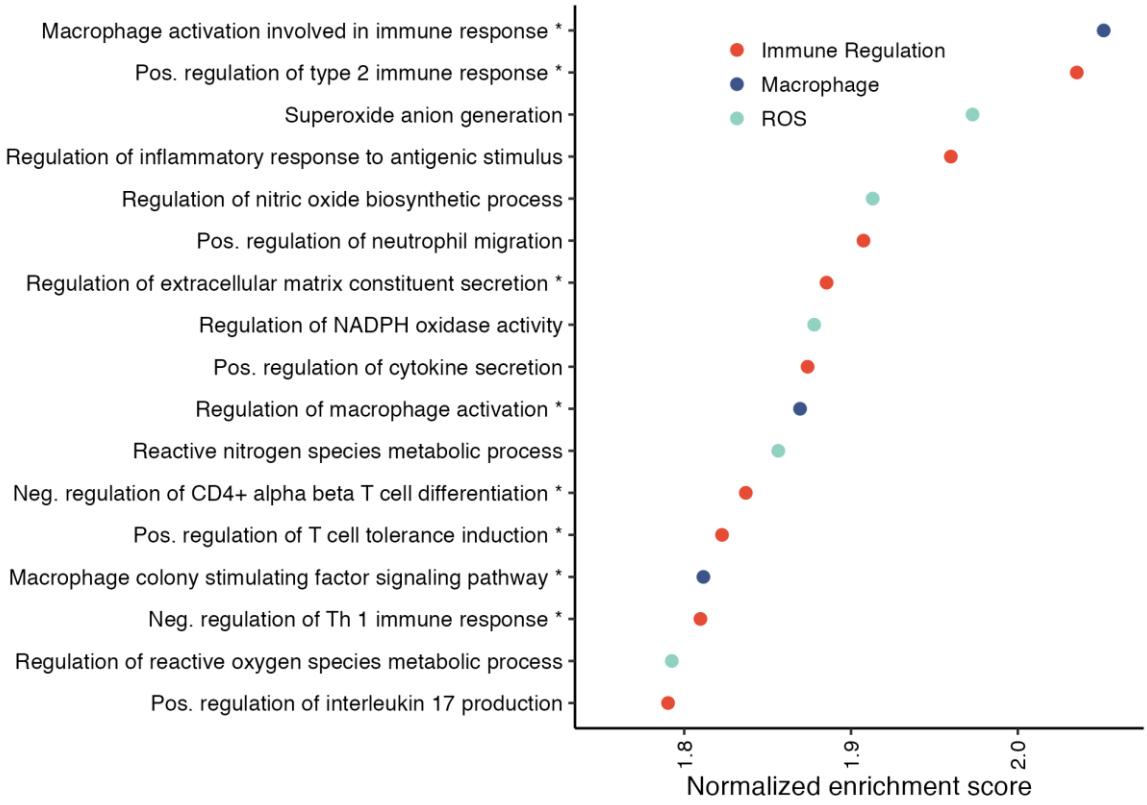


Figure 17: *Results of the gene set enrichment analysis. Patients were stratified into low- and high- genotype Pr groups using the median as a cutoff (n = 15 samples in both groups). Gene sets enriched in high genotype Pr samples with P < 0.001 and FDR < 0.1 are shown. Terms are classified into three main categories, and those associated with the development of an immunosuppressive tumor microenvironment are marked with asterisks.*

Immune deconvolution

Immune deconvolution is a computational method that estimates the proportions of different immune cells in samples, based on gene expression levels measured in a tumor biopsy.[195] We assessed the immune cell composition of the tumor samples utilizing a state-of-the-art software tool and recommendations from previous benchmark studies.[169] We observed an over-representation of T_{reg} cells, macrophages and monocytes, cancer-associated fibroblasts, and endothelial cells in tumor samples with high genotype Pr; indicators of suppressive microenvironment (Figure 18).

Key genes with increased expression in high promiscuity tumor samples include TGF-β (Transforming Growth Factor β), produced by regulatory T cells and macrophages, inhibiting effector immune functions; as well as FOXP3 (forkhead box P3), encoding a master regulator protein in T_{reg} development and function (Figure 19). All these factors had been previously shown to contribute to poor outcomes in cancer patients.[189,196]

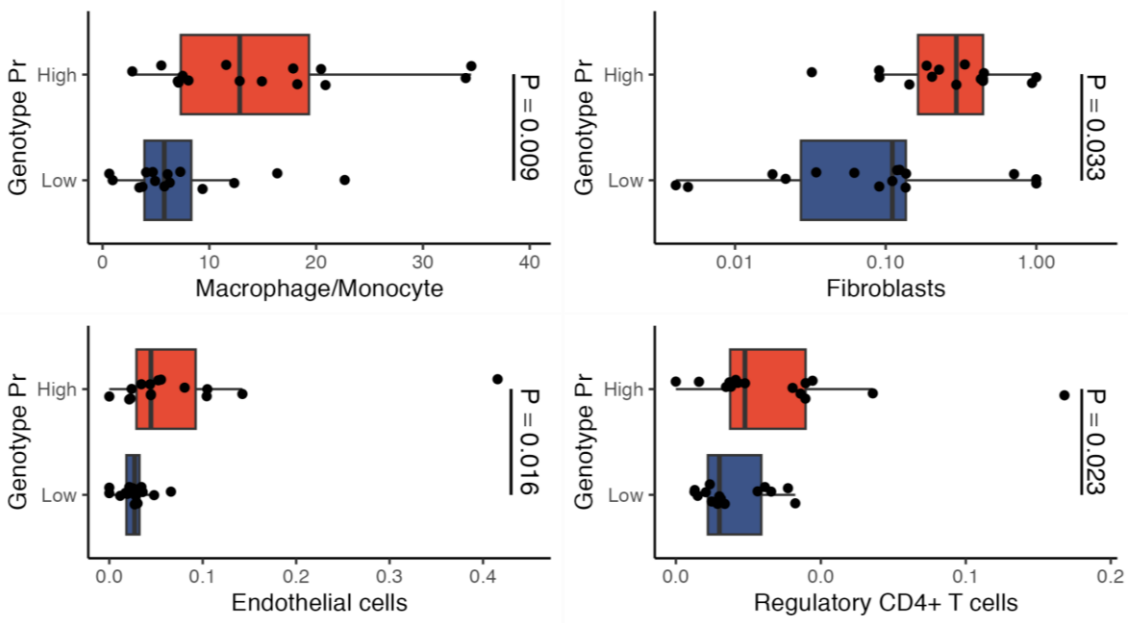


Figure 18: *association of high HLA-I genotype promiscuity with markers of an immunosuppressive cancer microenvironment, using immune deconvolution analysis. Boxplots show approximated cell abundance in samples with low and high genotype Pr. Patients were stratified into low- and high-genotype Pr groups using the median as a cutoff (n = 15 samples in both groups).*

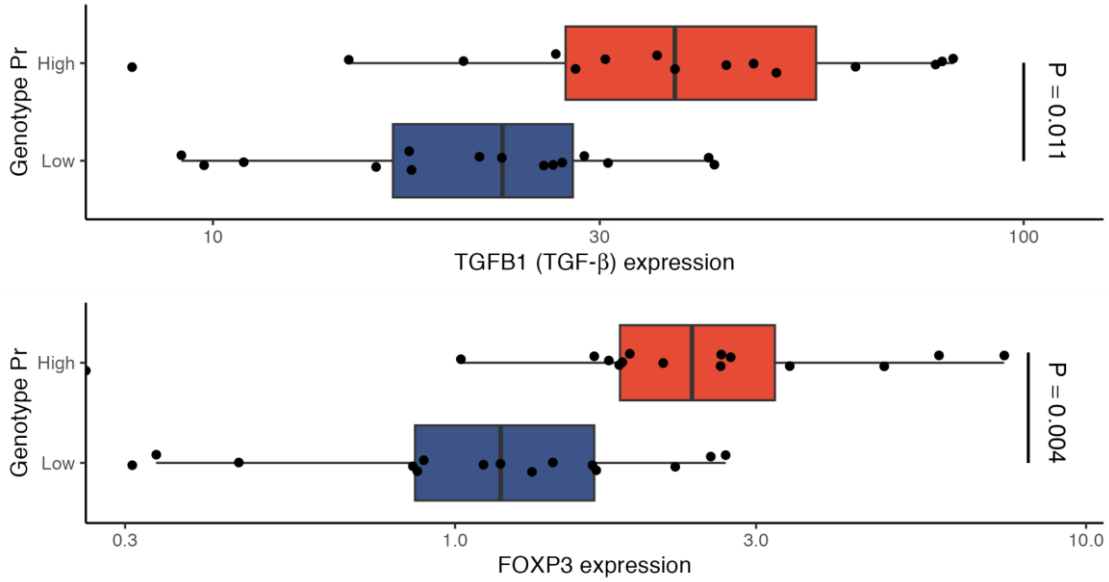


Figure 19: *expression of genes encoding the cytokine TGF-β and the transcription factor FOXP3, molecules associated with T-cell tolerance. Patients were stratified into low- and high-genotype Pr groups using the median as a cutoff (n = 15 samples in both groups).*

The role of immune checkpoint molecules

We investigated the expression levels of genes encoding for a selected set of costimulatory proteins and immune checkpoint receptors. These, including CTLA-4, PD-1, TIGIT, and TIM3 are responsible for the initiation of immune tolerance, and consequently, the formation of an

immunosuppressive tumor microenvironment.[197,198] Importantly, all genes displayed higher expression in cancer samples with high genotype Pr (Figure 20).

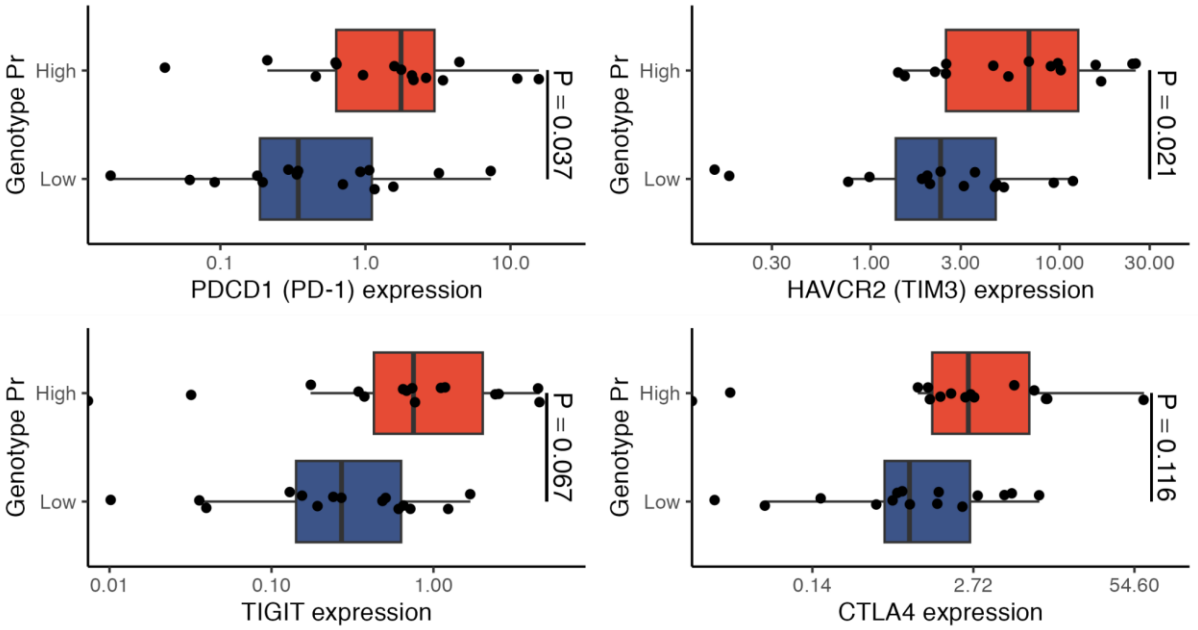


Figure 20: *expression levels of genes encoding immune checkpoint molecules. Patients were stratified into low- and high-genotype Pr groups using the median as a cutoff (n = 15 samples in both groups).*

Irreversible T-cell dysfunction is frequently associated with the formation of an immunosuppressive tumor microenvironment, as well as results in poor response to ICI therapy.[173,190] In this state, T cells have a reduced proliferative capacity and effector function. Next, we aimed to investigate whether high HLA-I genotype promiscuity is linked to T-cell dysfunction. We found that TOX, TOX2, T-bet, and BLIMP1 genes, transcription factors reported to have a major role in the formation and maintenance of T-cell dysfunction, had higher expression levels in cancer samples with high genotype Pr (Figure 21a).[199,200]

We also performed a systematic analysis using Tumor Immune Dysfunction and Exclusion (TIDE) method, a computational tool providing scores that describe T-cell dysfunction and exclusion signatures for tumors, especially the level of late-stage, irreversible dysfunction of T cells. Previously, the output scores had been shown to reliably predict ICI outcomes.[173] We found a positive correlation between genotype Pr and T-cell dysfunction score (Figure 21b, Spearman's $\rho = 0.4$, $P = 0.03$), but significant trends were not observable for T-cell exclusion score (Figure 21c, Spearman's $\rho = -0.1$, $P = 0.6$). This indicates that high genotype Pr does not prevent the infiltration of T cells directly, in contrast, it shapes the activity and functions of T cells in the tumor microenvironment.

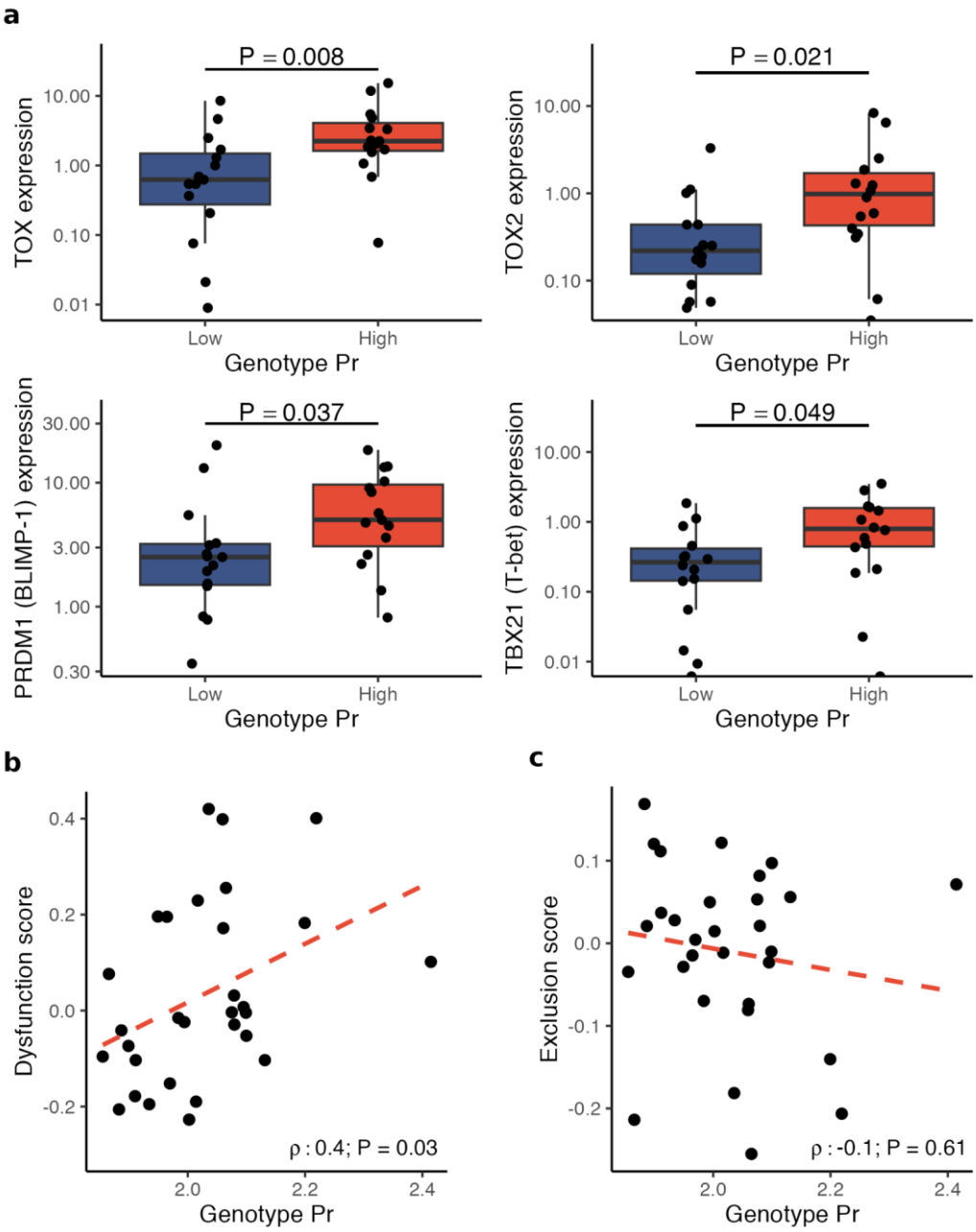


Figure 21: *association between genotype Pr and markers of T cell dysfunction.* ***Panel a:*** *Expression of genes encoding master regulators associated with T-cell dysfunction. Patients were stratified similarly to the comparisons in Figure O. High genotype Pr is associated with an elevated T-cell dysfunction score calculated by the TIDE algorithm (**panel b**). In contrast, no significant trend is observable for the exclusion score (**panel c**). Spearman's ρ and P values of a two-sided correlation test are indicated in the scatter plots of **panels b and c** (n = 30 samples); the dashed red line indicates a linear regression line.*

Furthermore, we investigated whether there are any correlations between T cell dysfunction and HLA-I-associated features, previously suggested to be clinically relevant, including HLA-I heterozygosity, HLA-I HED, and the presence of HLA-encoding gene variants belonging to certain allele supertypes (Figure 22). Importantly, neither of these factors showed a significant correlation with the T cell dysfunction score.

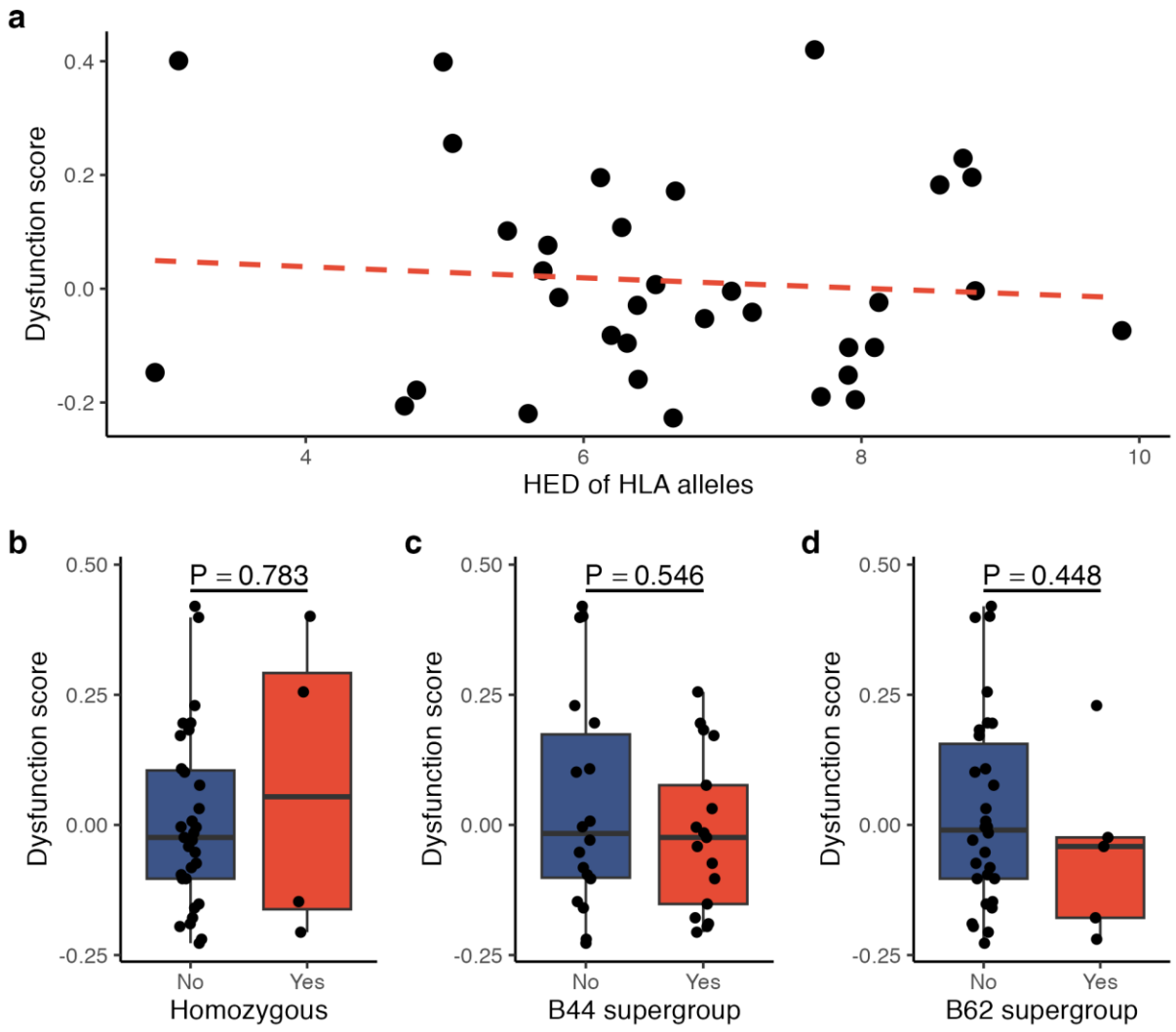


Figure 22: *the TIDE dysfunction score is shown as a function of HLA evolutionary divergence. There was no relationship between the two variables (**panel a**, Spearman's ρ: −0.01, two-sided correlation test P = 0.96, n = 35 samples). Similarly, neither HLA homozygosity **(panel b)** nor the carrier status for B44 (**panel c**, n = 17 and 18 samples in B44 positive and B44 negative groups, respectively) or B62 (**panel d**, n = 5 and 30 samples in B62 positive and negative groups, respectively) supertype alleles affected the TIDE dysfunction score. In **panel a**, the dashed red line indicates a linear regression line.*

5 SUMMARY AND DISCUSSION

HLA-I genes are key components of the human adaptive immune system, and their allelic variants encoded by HLA-A, -B, and -C loci restrict the peptides presented to CD8+ T cells.[1] The peptide-binding region of HLA class I alleles exhibits significant sequence variation, making the peptide repertoire presented by each allele distinct and only partially overlapping with others. This peptide repertoire is called immunopeptidome, and as a result of high allelic diversity in the human population, it shows high variation between individuals.

Moreover, different HLA variants have different peptide-binding capacities, with some being highly promiscuous, while others preferentially bind a narrow set of peptides.[140,201] Although HLA promiscuity is thought to be advantageous in pathogen-rich environments, its role in anticancer immunity has not been fully established yet.[140]

First, we aimed to quantify the potential size of the immunopeptidome for each allele. Initially, we followed the methodology from our previous study.[140] To put it in the context of tumor immunity, we replaced pathogen-derived peptides with neopeptide sequences. The NetMHCpan 4.0 algorithm, utilized for the *in silico* predictions, determines the strength of binding via two distinct metrics, IC50 binding affinity and rank percentile. We showed that for a chosen set of allele-peptide combinations, the two measurements show strong correlations. However, a noticeable disparity was observed in the relative fraction of peptides categorized to be weak binders or strong binders for a particular allele when using the two metrics. As delineated in the Results, these discrepancies stem from differences in the definitions of rank percentile and binding affinity. The former unifies the size of weakly or strongly bound peptide repertoires for all alleles – exactly what we aimed to quantify. On the other hand, prior studies have revealed that distinct HLA-I molecules have different IC50 nanomolar binding thresholds.

To build a robust method to quantify HLA-I promiscuity, we attempted using *in vitro* assay data. We designed a metric, which showed a positive correlation with the diversity of eluted peptides from immunopeptidomic studies. In addition, utilizing *in vitro* ProImmune REVEAL assays, we showed experimentally that this Kullback-Leibler divergence-based metric correlates most strongly with the breadth of peptides bound by HLA-I molecules.

This finding was followed by an investigation focusing on the impacts of HLA-I promiscuity on the response to ICI therapy in melanoma and NSCLC patients. Contrary to our expectations, patients carrying highly promiscuous HLA-I alleles had worse prognosis. We found a prominent effect of HLA-B promiscuity on patient survival. This might be explained by the

relatively high abundance of this molecule on the cell surface relative to that of HLA-A and HLA-C.[134,179] However, based on univariate Cox models, genotype Pr based on all loci is a better predictor of survival, compared to the three loci separately.

To explore the potential mechanisms underlying this trend, a multiple Cox regression model was applied. We found that genotype Pr and TMB were the strongest predictors of patient survival, while HED and HLA-I heterozygosity had no significant effect. The limitations of these metrics had been previously shown in a work investigating the survival of immunotherapy patients with NSCLC.[136] Interestingly, TMB was found to be a factor in significant interaction with genotype Pr only in the case of patients with melanoma, but not in lung cancer. Further research and data on additional factors (e.g. mutational signature composition) is necessary to find the underlying causes of this trend.

Peptide-HLA-I complex stability is a key factor in determining the immunogenicity of an antigen.[185] Surprisingly, we found that promiscuous HLA variants tended to form more stable complexes with bound peptides, challenging previous assumptions. This finding may have important implications for immunology and should be explored further in future studies.

Moreover, we observed a negative correlation between the level of HLA-I promiscuity and the differential binding affinity of mutated neopeptides. This suggests that patients with high genotype Pr have a reduced ability to discriminate between self- and mutated tumor peptides, having important implications for their anticancer immune response. To note, we did not observe significant trends between DAI for HLA-C molecules and their corresponding Pr values. The lack of correlation might be the result of the low variance HLA-C allele promiscuity.

We hypothesized that, the regulation of T cells in the tumor microenvironment is shifted from effector functions to tolerance, yielding a reduced capacity to eliminate cancer cells. In agreement with the expectations, melanoma samples from patients carrying promiscuous HLA-I alleles displayed molecular and cellular signatures of peripheral T-cell tolerance and an immunosuppressive tumor microenvironment. We detected an increased prevalence of T_{reg} cells, as well as elevated expression of genes encoding immune checkpoint receptors (PD-1), suppressive soluble mediators (TGF-β), and master control genes involved in regulatory T-cell development and maintenance (for example, FOXP3). One potential explanation for immune tolerance in individuals with high genotype Pr could be the paradoxical effects of excessively stable pHLA-TCR complexes on immune activation. This concept, as discussed in

the *Introduction* concerning the mechanism of action of PD-1 molecules, suggests that the enhanced stability of these complexes might hinder proper immune activation.[72] In overall, future work should elucidate the precise molecular pathways underlying these patterns. Additionally, it is important to elucidate the impact of HLA-I promiscuity on central tolerance and the potential depletion of T-cells in the repertoires of healthy individuals.[202]

High peptide-binding promiscuity could be a shared characteristic of genetically diverse HLA-class I variants that are associated with decreased survival rates following ICI therapy. Consequently, this metric could potentially serve as a useful target for future clinical trials identifying genetic variables of antitumor immunity. It is important to highlight that HLA-I promiscuity is a conceptually and statistically distinct factor compared to other, previously established determinants of therapy response and the occurrence of adverse events associated with ICI immunotherapy.[134,135,203] An interesting, unresolved issue is the potential interaction between high genotype Pr and high TMB in tumors. This question would be addressed in later analyses, as more immunotherapy cohort data becomes publicly available.

The presented findings raise important questions to be considered in future studies. The importance of HLA class II immune presentation had been described in multiple studies.[204,205] This current work neither investigated the potential effects of HLA-II alleles in ICI therapy nor aimed to quantify the binding promiscuity of these molecules. Measuring HLA-II promiscuity is challenging for several reasons. As a result of a more open structure, the preferred length of peptides bound by HLA-II is not as limited as it is for HLA-I, which adds complexity to the analysis.[145,206] Additionally, there is a limited amount of data available on the binding affinities of peptides to HLA-II molecules, which makes it difficult to accurately measure the levels of allele promiscuity.

Previously it has been shown that the presentation of a broader range of viral peptides could be beneficial in acute viral infections.[140,201] Our results suggest that the presence of high promiscuity alleles capable of binding peptides by less specificity could hamper antitumor immunity. These trends indicate a negative trade-off between the genetic susceptibility for effective antitumor immune responses and the capability to efficiently resolve viral infections. There might be various potential reasons for these opposite trends. Viral peptides usually have a lower sequence similarity to the human proteome (Figure 23), in contrast to mutated self-peptides, which mostly differ only in one amino acid compared to the germline version of the original protein.[207,208] In the case of viral epitopes, HLA-I molecules do not pay the cost of

presenting a higher number of peptides. Meanwhile, in the case of tumor neopeptides the discriminative capabilities of highly promiscuous HLA-I alleles are limited.

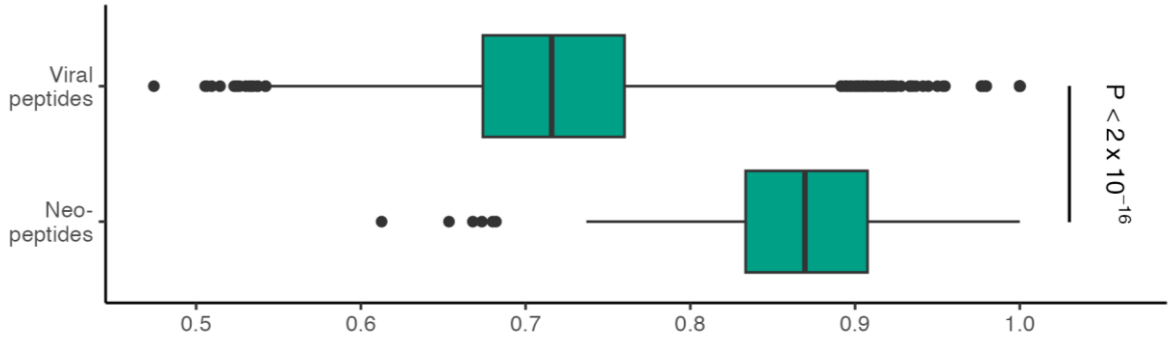


***Figure 23**: the maximum BLOSUM62 sequence similarity of neopeptides (n = 589) and immunogenic viral (n = 1038) peptides to the non-mutated human proteome. Only nine amino acid-long peptides were included in this analysis.*

In a previous study, our group has shown that promiscuous HLA-A and -B variants are highly frequent in South-East Asia, likely as a consequence of selection by high pathogen load in the region.[140] This pattern raises the question if there are geographical and ethnic differences in genetic factors shaping cancer immunotherapy outcomes and antitumor immunity in general. Finally, studies highlight the relevance of neoantigen immunogenicity and quality additionally to neoantigen quantity in patient survival.[209] This work suggests that HLA-I peptide-binding promiscuity is an important factor in shaping neoantigen immunogenicity, with potential implications on cancer vaccine design.[210]

6 ACKNOWLEDGEMENTS

I wish to express my gratitude to Máté Manczinger, whose constant support and guidance helped me master the most important methods in computational biology, including the essentials of programming and data analysis in R. I am thankful for the opportunities to engage in exciting projects, ranging from the investigation of molecular mimicry of tumors to unexpected roles of SARS-CoV-2 mutations.

I am truly grateful to Professor Lajos Kemény for granting me the opportunity to begin my academic journey at the Department of Dermatology and Allergology, University of Szeged. His institute served as a home for our workgroup during the initial years, providing us with a workplace and computational infrastructure to perform our investigations.

I wish to express my gratitude to my colleagues, namely Balázs Koncz, Benjamin Tamás Papp, and Leó Asztalos, for their substantial contributions and assistance. I am thankful for the brainstorming sessions, producing multitudes of novel ideas for further exploration in the forthcoming years.

I also would like to thank Csaba Pál and Balázs Papp for their suggestions and guidance throughout the research process. Their expertise has greatly enriched our work and helped us avoid several pitfalls in study design and publication.

I am profoundly grateful to my parents, that they always supported my childhood desire to discover and invent something new and motivated me to strive for perfection. Finally, I extend my sincerest thanks to my beloved wife, Szilvi, for helping me throughout these years, and in whom I found a companion to believe in ideas that may seem nonsense to others. I am also grateful for our scientific discussions about adaptive immune recognition, even amidst the depths of raspberry bushes while picking berries or during treks in the woods. Furthermore, I am grateful for the boundless joy and strength that our daughter, Luca, brings to our life.

SDG

7 REFERENCES

1. Parham, P. & Janeway, C. *The immune system*. (Garland Science, Taylor & Francis Group, 2015).
2. Alberts, B. *et al. Molecular biology of the cell*. (W. W. Norton & Company, 2022).
3. Couture, A. *et al.* HLA-Class II Artificial Antigen Presenting Cells in CD4+ T Cell-Based Immunotherapy. *Front. Immunol.* **10**, 1081 (2019).
4. Neefjes, J., Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
5. Johnson, D. B. *et al.* Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat. Commun.* **7**, 10582 (2016).
6. Sharonov, G. V., Serebrovskaya, E. O., Yuzhakova, D. V., Britanova, O. V. & Chudakov, D. M. B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nat. Rev. Immunol.* **20**, 294–307 (2020).
7. Papp, B. T. *et al. Synergy of HLA class I and II shapes the timing of antitumor immune response*. http://biorxiv.org/lookup/doi/10.1101/2022.11.16.516740 (2022) doi:10.1101/2022.11.16.516740.
8. Kostyu, D. D. *et al.* HLA Class I Polymorphism: Structure and Function and Still Questions. *Hum. Immunol.* **57**, 1–18 (1997).
9. Matzinger, P. Tolerance, danger, and the extended family. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
10. Dustin, M. L. The Immunological Synapse. *Cancer Immunol. Res.* **2**, 1023–1033 (2014).
11. Grakoui, A. *et al.* The Immunological Synapse: A Molecular Machine Controlling T Cell Activation. *Science* **285**, 221–227 (1999).
12. Shiina, T., Hosomichi, K., Inoko, H. & Kulski, J. K. The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* **54**, 15–39 (2009).
13. Nguyen, A. T., Szeto, C. & Gras, S. The pockets guide to HLA class I molecules. *Biochem. Soc. Trans.* **49**, 2319–2331 (2021).
14. Marsh, S. G. E. *et al.* Nomenclature for factors of the HLA system, 2010. *Tissue Antigens* **75**, 291–455 (2010).
15. Robinson, J. *et al.* The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* **43**, D423–D431 (2015).

16. Carrington, M. *et al.* HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283**, 1748–1752 (1999).
17. Thursz, M. R., Thomas, H. C., Greenwood, B. M. & Hill, A. V. S. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat. Genet.* **17**, 11–12 (1997).
18. Shkurnikov, M. *et al.* Association of HLA Class I Genotypes With Severity of Coronavirus Disease-19. *Front. Immunol.* **12**, (2021).
19. Zlotoff, D. A. & Bhandoola, A. Hematopoietic progenitor migration to the adult thymus: Hematopoietic progenitor migration. *Ann. N. Y. Acad. Sci.* **1217**, 122–138 (2011).
20. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* **14**, 377–391 (2014).
21. Kyewski, B. & Derbinski, J. Self-representation in the thymus: an extended view. *Nat. Rev. Immunol.* **4**, 688–698 (2004).
22. Gallegos, A. M. & Bevan, M. J. Central Tolerance to Tissue-specific Antigens Mediated by Direct and Indirect Antigen Presentation. *J. Exp. Med.* **200**, 1039–1049 (2004).
23. Arstila, T. P. *et al.* A direct estimate of the human alphabeta T cell receptor diversity. *Science* **286**, 958–961 (1999).
24. Xing, Y. & Hogquist, K. A. T-Cell Tolerance: Central and Peripheral. *Cold Spring Harb. Perspect. Biol.* **4**, a006957–a006957 (2012).
25. Schwartz, R. H. T Cell Anergy. *Annu. Rev. Immunol.* **21**, 305–334 (2003).
26. Keir, M. E., Butte, M. J., Freeman, G. J. & Sharpe, A. H. PD-1 and Its Ligands in Tolerance and Immunity. *Annu. Rev. Immunol.* **26**, 677–704 (2008).
27. Hawiger, D. *et al.* Dendritic Cells Induce Peripheral T Cell Unresponsiveness under Steady State Conditions in Vivo. *J. Exp. Med.* **194**, 769–780 (2001).
28. Wölfle, S. J. *et al.* PD-L1 expression on tolerogenic APCs is controlled by STAT-3. *Eur. J. Immunol.* **41**, 413–424 (2011).
29. Strasser, A. & Pellegrini, M. T-lymphocyte death during shutdown of an immune response. *Trends Immunol.* **25**, 610–615 (2004).
30. Pellegrini, M., Belz, G., Bouillet, P. & Strasser, A. Shutdown of an acute T cell immune response to viral infection is mediated by the proapoptotic Bcl-2 homology 3-only protein Bim. *Proc. Natl. Acad. Sci.* **100**, 14175–14180 (2003).
31. Hildeman, D. A. *et al.* Activated T Cell Death In Vivo Mediated by Proapoptotic Bcl-2 Family Member Bim. *Immunity* **16**, 759–767 (2002).

32. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
33. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
34. Simpson, A. J. G., Caballero, O. L., Jungbluth, A., Chen, Y.-T. & Old, L. J. Cancer/testis antigens, gametogenesis and cancer. *Nat. Rev. Cancer* **5**, 615–625 (2005).
35. Sasagawa, T., Takagi, H. & Makinoda, S. Immune responses against human papillomavirus (HPV) infection and evasion of host defense in cervical cancer. *J. Infect. Chemother.* **18**, 807–815 (2012).
36. Chan, T. A. *et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* **30**, 44–56 (2019).
37. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
38. Ballhausen, A. *et al.* The shared frameshift mutation landscape of microsatellite-unstable cancers suggests immunoediting during tumor evolution. *Nat. Commun.* **11**, 4740 (2020).
39. Kloor, M. & Von Knebel Doeberitz, M. The Immune Biology of Microsatellite-Unstable Cancer. *Trends Cancer* **2**, 121–133 (2016).
40. Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* **22**, 1342–1350 (2016).
41. Bingle, L., Brown, N. J. & Lewis, C. E. The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. *J. Pathol.* **196**, 254–265 (2002).
42. Haabeth, O. A. W., Bogen, B. & Corthay, A. A model for cancer-suppressive inflammation. *OncoImmunology* **1**, 1146–1155 (2012).
43. Bruns, H. *et al.* Vitamin D–dependent induction of cathelicidin in human macrophages results in cytotoxicity against high-grade B cell lymphoma. *Sci. Transl. Med.* **7**, (2015).
44. Bernsmeier, C., Van Der Merwe, S. & Périanin, A. Innate immune cells in cirrhosis. *J. Hepatol.* **73**, 186–201 (2020).
45. Linde, N. *et al.* Vascular endothelial growth factor-induced skin carcinogenesis depends on recruitment and alternative activation of macrophages. *J. Pathol.* **227**, 17–28 (2012).

46. Zhu, C. *et al.* Activation of CECR1 in M2-like TAMs promotes paracrine stimulation-mediated glial tumor progression. *Neuro-Oncol.* now251 (2017) doi:10.1093/neuonc/now251.
47. Yin, M. *et al.* Tumor-associated macrophages drive spheroid formation during early transcoelomic metastasis of ovarian cancer. *J. Clin. Invest.* **126**, 4157–4173 (2016).
48. Wei, Y. *et al.* The local immune landscape determines tumor PD-L1 heterogeneity and sensitivity to therapy. *J. Clin. Invest.* **129**, 3347–3360 (2019).
49. Hao, S. *et al.* Macrophage phenotypic mechanomodulation of enhancing bone regeneration by superparamagnetic scaffold upon magnetization. *Biomaterials* **140**, 16–25 (2017).
50. Egawa, M. *et al.* Inflammatory Monocytes Recruited to Allergic Skin Acquire an Anti-inflammatory M2 Phenotype via Basophil-Derived Interleukin-4. *Immunity* **38**, 570–580 (2013).
51. Kolahian, S. *et al.* The emerging role of myeloid-derived suppressor cells in lung diseases. *Eur. Respir. J.* **47**, 967–977 (2016).
52. Qu, X., Zhuang, G., Yu, L., Meng, G. & Ferrara, N. Induction of Bv8 Expression by Granulocyte Colony-stimulating Factor in CD11b+Gr1+ Cells. *J. Biol. Chem.* **287**, 19574–19584 (2012).
53. Ardi, V. C., Kupriyanova, T. A., Deryugina, E. I. & Quigley, J. P. Human neutrophils uniquely release TIMP-free MMP-9 to provide a potent catalytic stimulator of angiogenesis. *Proc. Natl. Acad. Sci.* **104**, 20262–20267 (2007).
54. Bekes, E. M. *et al.* Tumor-Recruited Neutrophils and Neutrophil TIMP-Free MMP-9 Regulate Coordinately the Levels of Tumor Angiogenesis and Efficiency of Malignant Cell Intravasation. *Am. J. Pathol.* **179**, 1455–1470 (2011).
55. Smyth, M. J. *et al.* Activation of NK cell cytotoxicity. *Mol. Immunol.* **42**, 501–510 (2005).
56. Prager, I. & Watzl, C. Mechanisms of natural killer cell-mediated cellular cytotoxicity. *J. Leukoc. Biol.* **105**, 1319–1329 (2019).
57. Lu, C., Liu, Y., Ali, N. M., Zhang, B. & Cui, X. The role of innate immune cells in the tumor microenvironment and research progress in anti-tumor therapy. *Front. Immunol.* **13**, 1039260 (2023).
58. Joffre, O., Nolte, M. A., Spörri, R. & Sousa, C. R. e. Inflammatory signals in dendritic cell activation and the induction of adaptive immunity. *Immunol. Rev.* **227**, 234–247 (2009).

59. Castellaneta, A., Sumpter, T. L., Chen, L., Tokita, D. & Thomson, A. W. NOD2 Ligation Subverts IFN-α Production by Liver Plasmacytoid Dendritic Cells and Inhibits Their T Cell Allostimulatory Activity via B7-H1 Up-Regulation. *J. Immunol.* **183**, 6922–6932 (2009).
60. Jahrsdörfer, B. *et al.* Granzyme B produced by human plasmacytoid dendritic cells suppresses T-cell expansion. *Blood* **115**, 1156–1165 (2010).
61. Lanzavecchia, A., Iezzi, G. & Viola, A. From TCR Engagement to T Cell Activation. *Cell* **96**, 1–4 (1999).
62. Krummel, M. F. & Allison, J. P. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *J. Exp. Med.* **182**, 459–465 (1995).
63. Krummel, M. F. & Allison, J. P. CTLA-4 engagement inhibits IL-2 accumulation and cell cycle progression upon activation of resting T cells. *J. Exp. Med.* **183**, 2533–2540 (1996).
64. Linsley, P. S. *et al.* Human B7-1 (CD80) and B7-2 (CD86) bind with similar avidities but distinct kinetics to CD28 and CTLA-4 receptors. *Immunity* **1**, 793–801 (1994).
65. Van Der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P. & Davis, S. J. CD80 (B7-1) Binds Both CD28 and CTLA-4 with a Low Affinity and Very Fast Kinetics. *J. Exp. Med.* **185**, 393–404 (1997).
66. Oyewole-Said, D. *et al.* Beyond T-Cells: Functional Characterization of CTLA-4 Expression in Immune and Non-Immune Cell Types. *Front. Immunol.* **11**, 608024 (2020).
67. Walunas, T. L. *et al.* CTLA-4 can function as a negative regulator of T cell activation. *Immunity* **1**, 405–413 (1994).
68. Ishida, Y., Agata, Y., Shibahara, K. & Honjo, T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* **11**, 3887–3895 (1992).
69. Freeman, G. J. *et al.* Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J. Exp. Med.* **192**, 1027–1034 (2000).
70. Latchman, Y. *et al.* PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat. Immunol.* **2**, 261–268 (2001).
71. Agata, Y. *et al.* Expression of the PD-1 antigen on the surface of stimulated mouse T and B lymphocytes. *Int. Immunol.* **8**, 765–772 (1996).
72. Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439**, 682–687 (2006).

73. He, X. & Xu, C. Immune checkpoint signaling and cancer immunotherapy. *Cell Res.* **30**, 660–669 (2020).
74. Chowell, D. *et al.* Modeling the Subclonal Evolution of Cancer Cell Populations. *Cancer Res.* **78**, 830–839 (2018).
75. Loeb, L. A. *et al.* Extensive subclonal mutational diversity in human colorectal cancer and its significance. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 26863–26872 (2019).
76. Frankell, A. M. *et al.* The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533 (2023).
77. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
78. Sjöblom, T. *et al.* The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science* **314**, 268–274 (2006).
79. Kinzler, K. W. & Vogelstein, B. Landscaping the Cancer Terrain. *Science* **280**, 1036–1037 (1998).
80. Gallaher, J. A., Brown, J. S. & Anderson, A. R. A. The impact of proliferation-migration tradeoffs on phenotypic evolution in cancer. *Sci. Rep.* **9**, 2425 (2019).
81. Zapata, L. *et al.* Immune selection determines tumor antigenicity and influences response to checkpoint inhibitors. *Nat. Genet.* **55**, 451–460 (2023).
82. Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
83. Castro, A. *et al.* Strength of immune selection in tumors varies with sex and age. *Nat. Commun.* **11**, 4128 (2020).
84. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171**, 1259-1271.e11 (2017).
85. Li, C., Jiang, P., Wei, S., Xu, X. & Wang, J. Regulatory T cells in tumor microenvironment: new mechanisms, potential therapeutic strategies and future prospects. *Mol. Cancer* **19**, 116 (2020).
86. Lindau, D., Gielen, P., Kroesen, M., Wesseling, P. & Adema, G. J. The immunosuppressive tumour network: myeloid-derived suppressor cells, regulatory T cells and natural killer T cells. *Immunology* **138**, 105–115 (2013).
87. Kim, S. K. & Cho, S. W. The Evasion Mechanisms of Cancer Immunity and Drug Intervention in the Tumor Microenvironment. *Front. Pharmacol.* **13**, 868695 (2022).
88. Munn, D. H. & Mellor, A. L. IDO in the Tumor Microenvironment: Inflammation, Counter-Regulation, and Tolerance. *Trends Immunol.* **37**, 193–207 (2016).

89. Pan, Y., Yu, Y., Wang, X. & Zhang, T. Tumor-Associated Macrophages in Tumor Immunity. *Front. Immunol.* **11**, 583084 (2020).
90. Santoni, G. *et al.* High CTLA-4 expression correlates with poor prognosis in thymoma patients. *Oncotarget* **9**, 16665–16677 (2018).
91. Zhang, X.-F. *et al.* Cytotoxic T lymphocyte antigen-4 expression in esophageal carcinoma: implications for prognosis. *Oncotarget* **7**, 26670–26679 (2016).
92. Huang, P.-Y. *et al.* Tumor CTLA-4 overexpression predicts poor survival in patients with nasopharyngeal carcinoma. *Oncotarget* **7**, 13060–13068 (2016).
93. Grohmann, U. *et al.* CTLA-4-Ig regulates tryptophan catabolism in vivo. *Nat. Immunol.* **3**, 1097–1101 (2002).
94. Fallarino, F. *et al.* Modulation of tryptophan catabolism by regulatory T cells. *Nat. Immunol.* **4**, 1206–1212 (2003).
95. Paulsen, E.-E. *et al.* CTLA-4 expression in the non-small cell lung cancer patient tumor microenvironment: diverging prognostic impact in primary tumors and lymph node metastases. *Cancer Immunol. Immunother.* **66**, 1449–1461 (2017).
96. Yu, H. *et al.* Cytotoxic T lymphocyte antigen 4 expression in human breast cancer: implications for prognosis. *Cancer Immunol. Immunother.* **64**, 853–860 (2015).
97. Chen, X. *et al.* CTLA-4 positive breast cancer cells suppress dendritic cells maturation and function. *Oncotarget* **8**, 13703–13715 (2017).
98. Deng, M. *et al.* Relationship between PD-L1 expression, CD8+ T-cell infiltration and prognosis in intrahepatic cholangiocarcinoma patients. *Cancer Cell Int.* **21**, 371 (2021).
99. Wang, X. & Liu, Y. PD-L1 expression in tumor infiltrated lymphocytes predicts survival in triple-negative breast cancer. *Pathol. - Res. Pract.* **216**, 152802 (2020).
100. Fan, P. *et al.* PD-1 Expression Status on CD8+ Tumour Infiltrating Lymphocytes Associates With Survival in Cervical Cancer. *Front. Oncol.* **11**, 678758 (2021).
101. Mu, C.-Y., Huang, J.-A., Chen, Y., Chen, C. & Zhang, X.-G. High expression of PD-L1 in lung cancer may contribute to poor prognosis and tumor cells immune escape through suppressing tumor infiltrating dendritic cells maturation. *Med. Oncol. Northwood Lond. Engl.* **28**, 682–688 (2011).
102. Ohigashi, Y. *et al.* Clinical Significance of Programmed Death-1 Ligand-1 and Programmed Death-1 Ligand-2 Expression in Human Esophageal Cancer. *Clin. Cancer Res.* **11**, 2947–2953 (2005).
103. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).

104. Robert, C. A decade of immune-checkpoint inhibitors in cancer therapy. *Nat. Commun.* **11**, 3801 (2020).
105. Robert, C. *et al.* Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **372**, 2521–2532 (2015).
106. Larkin, J. *et al.* Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* **381**, 1535–1546 (2019).
107. Wolchok, J. D. *et al.* Development of ipilimumab: a novel immunotherapeutic approach for the treatment of advanced melanoma. *Ann. N. Y. Acad. Sci.* **1291**, 1–13 (2013).
108. Topalian, S. L. *et al.* Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **32**, 1020–1030 (2014).
109. Johnson, D. B., Peng, C. & Sosman, J. A. Nivolumab in melanoma: latest evidence and clinical potential. *Ther. Adv. Med. Oncol.* **7**, 97–106 (2015).
110. Jacob, J. B., Jacob, M. K. & Parajuli, P. Review of immune checkpoint inhibitors in immuno-oncology. in *Advances in Pharmacology* vol. 91 111–139 (Elsevier, 2021).
111. Tawbi, H. A. *et al.* Relatlimab and Nivolumab versus Nivolumab in Untreated Advanced Melanoma. *N. Engl. J. Med.* **386**, 24–34 (2022).
112. Maio, M. *et al.* Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **33**, 1191–1196 (2015).
113. Topalian, S. L. *et al.* Five-Year Survival and Correlates Among Patients With Advanced Melanoma, Renal Cell Carcinoma, or Non–Small Cell Lung Cancer Treated With Nivolumab. *JAMA Oncol.* **5**, 1411 (2019).
114. Brahmer, J. R. *et al.* Five-Year Survival Outcomes With Nivolumab Plus Ipilimumab Versus Chemotherapy as First-Line Treatment for Metastatic Non–Small-Cell Lung Cancer in CheckMate 227. *J. Clin. Oncol.* **41**, 1200–1212 (2023).
115. Grassadonia, A. *et al.* Effect of Gender on the Outcome of Patients Receiving Immune Checkpoint Inhibitors for Advanced Cancer: A Systematic Review and Meta-Analysis of Phase III Randomized Clinical Trials. *J. Clin. Med.* **7**, 542 (2018).
116. Ye, Y. *et al.* Sex-associated molecular differences for cancer immunotherapy. *Nat. Commun.* **11**, 1779 (2020).
117. Erbe, R. *et al.* Evaluating the impact of age on immune checkpoint therapy biomarkers. *Cell Rep.* **36**, (2021).

118. Herbst, R. S. *et al.* Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *The Lancet* **387**, 1540–1550 (2016).
119. Reck, M. *et al.* Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **375**, 1823–1833 (2016).
120. Chung, H. C. *et al.* Efficacy and Safety of Pembrolizumab in Previously Treated Advanced Cervical Cancer: Results From the Phase II KEYNOTE-158 Study. *J. Clin. Oncol.* **37**, 1470–1478 (2019).
121. Li, X., Zhang, S., Guo, G., Han, J. & Yu, J. Gut microbiome in modulating immune checkpoint inhibitors. *eBioMedicine* **82**, (2022).
122. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
123. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
124. Hugo, W. *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44 (2016).
125. Carbone, D. P. *et al.* First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2415–2426 (2017).
126. Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
127. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 13404 (2016).
128. Coulie, P. G., Van Den Eynde, B. J., Van Der Bruggen, P. & Boon, T. Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat. Rev. Cancer* **14**, 135–146 (2014).
129. Lemery, S., Keegan, P. & Pazdur, R. First FDA Approval Agnostic of Cancer Site - When a Biomarker Defines the Indication. *N. Engl. J. Med.* **377**, 1409–1412 (2017).
130. Middha, S. *et al.* Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. *JCO Precis. Oncol.* **2017**, PO.17.00084 (2017).
131. Valero, C. *et al.* The association between tumor mutational burden and prognosis is dependent on treatment context. *Nat. Genet.* **53**, 11–15 (2021).

132. Jardim, D. L., Goodman, A., De Melo Gagliato, D. & Kurzrock, R. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. *Cancer Cell* **39**, 154–173 (2021).
133. Gao, Y. *et al.* Integration of the Tumor Mutational Burden and Tumor Heterogeneity Identify an Immunological Subtype of Melanoma With Favorable Survival. *Front. Oncol.* **10**, 571545 (2020).
134. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).
135. Chowell, D. *et al.* Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat. Med.* **25**, 1715–1720 (2019).
136. Negrao, M. V. *et al.* PD-L1 Expression, Tumor Mutational Burden, and Cancer Gene Mutations Are Stronger Predictors of Benefit from Immune Checkpoint Blockade than HLA Class I Genotype in Non-Small Cell Lung Cancer. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **14**, 1021–1031 (2019).
137. Litchfield, K. *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596-614.e14 (2021).
138. Kaufman, J. From Chickens to Humans: The Importance of Peptide Repertoires for MHC Class I Alleles. *Front. Immunol.* **11**, 601089 (2020).
139. Jurtz, V. *et al.* NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
140. Manczinger, M. *et al.* Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLOS Biol.* **17**, e3000131 (2019).
141. Łuksza, M. *et al.* A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).
142. Marty, R. *et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* **171**, 1272-1283.e15 (2017).
143. HLA allele frequencies and reference sets with maximal population coverage. *IEDB Solutions Center* https://help.iedb.org/hc/en-us/articles/114094151851-HLA-allele-frequencies-and-reference-sets-with-maximal-population-coverage (2020).
144. Weiskopf, D. *et al.* Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E2046-2053 (2013).

145. Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
146. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
147. Abelin, J. G. *et al.* Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **46**, 315–326 (2017).
148. Di Marco, M. *et al.* Unveiling the Peptide Motifs of HLA-C and HLA-G from Naturally Presented Peptides and Generation of Binding Prediction Matrices. *J. Immunol. Baltim. Md 1950* **199**, 2639–2651 (2017).
149. Olsen, L. R. *et al.* TANTIGEN: a comprehensive database of tumor T cell antigens. *Cancer Immunol. Immunother. CII* **66**, 731–735 (2017).
150. Robbins, P. F. *et al.* Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat. Med.* **19**, 747–752 (2013).
151. Rasmussen, M. *et al.* Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. *J. Immunol. Baltim. Md 1950* **197**, 1517–1524 (2016).
152. Bjerregaard, A.-M. *et al.* An Analysis of Natural T Cell Responses to Predicted Tumor Neoepitopes. *Front. Immunol.* **8**, 1566 (2017).
153. Gros, A. *et al.* Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* **22**, 433–438 (2016).
154. Zacharakis, N. *et al.* Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat. Med.* **24**, 724–730 (2018).
155. Tran, E. *et al.* Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **350**, 1387–1390 (2015).
156. Carreno, B. M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).
157. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
158. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science* **348**, 124–128 (2015).
159. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).

160. Riaz, N. *et al.* Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **171**, 934-949.e16 (2017).
161. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
162. Liu, J. *et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416.e11 (2018).
163. Li, X. *et al.* Benchmarking HLA genotyping and clarifying HLA impact on survival in tumor immunotherapy. *Mol. Oncol.* **15**, 1764–1782 (2021).
164. Terry M. Therneau & Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. (Springer, 2000).
165. Kassambara, A., Kosinski, M. & Biecek, P. *survminer: Drawing Survival Curves using 'ggplot2'*. (2021).
166. Ghorani, E. *et al.* Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. *Ann. Oncol.* **29**, 271–279 (2018).
167. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
168. Carrasco Pro, S. *et al.* Microbiota epitope similarity either dampens or enhances the immunogenicity of disease-associated antigenic epitopes. *PloS One* **13**, e0196551 (2018).
169. Sturm, G. *et al.* Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, i436–i445 (2019).
170. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **6**, e26476 (2017).
171. Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
172. Finotello, F. *et al.* Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).
173. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
174. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

175. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
176. De Boor, C. *A practical guide to splines: with 32 figures*. (Springer, 2001).
177. Paul, S. *et al.* HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *J. Immunol.* **191**, 5831–5839 (2013).
178. Sarkizova, S. *et al.* A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
179. Kaur, G. *et al.* Structural and regulatory diversity shape HLA-C protein expression levels. *Nat. Commun.* **8**, 15924 (2017).
180. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer Oxf. Engl. 1990* **45**, 228–247 (2009).
181. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based immunotherapy. *Science* **362**, eaar3593 (2018).
182. Mariathasan, S. *et al.* TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
183. Hutter, C. & Zenklusen, J. C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **173**, 283–285 (2018).
184. Sidney, J., Peters, B., Frahm, N., Brander, C. & Sette, A. HLA class I supertypes: a revised and updated classification. *BMC Immunol.* **9**, 1 (2008).
185. Harndahl, M. *et al.* Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur. J. Immunol.* **42**, 1405–1416 (2012).
186. Fromer, M. & Shifman, J. M. Tradeoff between stability and multispecificity in the design of promiscuous proteins. *PLoS Comput. Biol.* **5**, e1000627 (2009).
187. Rech, A. J. *et al.* Tumor Immunity and Survival as a Function of Alternative Neopeptides in Human Cancer. *Cancer Immunol. Res.* **6**, 276–287 (2018).
188. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211**, 2231–2248 (2014).
189. Nurieva, R., Wang, J. & Sahoo, A. T-cell tolerance in cancer. *Immunotherapy* **5**, 513–531 (2013).
190. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
191. Noy, R. & Pollard, J. W. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* **41**, 49–61 (2014).

192. Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. Natural innate and adaptive immunity to cancer. *Annu. Rev. Immunol.* **29**, 235–271 (2011).
193. Henke, E., Nandigama, R. & Ergün, S. Extracellular Matrix in the Tumor Microenvironment and Its Impact on Cancer Therapy. *Front. Mol. Biosci.* **6**, 160 (2019).
194. Walker, C., Mojares, E. & Del Río Hernández, A. Role of Extracellular Matrix in Development and Cancer Progression. *Int. J. Mol. Sci.* **19**, 3028 (2018).
195. Sturm, G., Finotello, F. & List, M. In Silico Cell-Type Deconvolution Methods in Cancer Immunotherapy. *Methods Mol. Biol. Clifton NJ* **2120**, 213–222 (2020).
196. Colak, S. & Ten Dijke, P. Targeting TGF-β Signaling in Cancer. *Trends Cancer* **3**, 56–71 (2017).
197. Das, M., Zhu, C. & Kuchroo, V. K. Tim-3 and its role in regulating anti-tumor immunity. *Immunol. Rev.* **276**, 97–111 (2017).
198. Manieri, N. A., Chiang, E. Y. & Grogan, J. L. TIGIT: A Key Inhibitor of the Cancer Immunity Cycle. *Trends Immunol.* **38**, 20–28 (2017).
199. Scott, A. C. *et al.* TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).
200. Xia, A., Zhang, Y., Xu, J., Yin, T. & Lu, X.-J. T Cell Dysfunction in Cancer Immunity and Immunotherapy. *Front. Immunol.* **10**, (2019).
201. Kaufman, J. Generalists and Specialists: A New View of How MHC Class I Molecules Fight Infectious Pathogens. *Trends Immunol.* **39**, 367–379 (2018).
202. Radwan, J., Babik, W., Kaufman, J., Lenz, T. L. & Winternitz, J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genet. TIG* **36**, 298–311 (2020).
203. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
204. Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
205. Marty Pyke, R. *et al.* Evolutionary Pressure against MHC Class II Binding Cancer Mutations. *Cell* **175**, 416-428.e13 (2018).
206. Heng, Y. *et al.* A simple pan-specific RNN model for predicting HLA-II binding peptides. *Mol. Immunol.* **139**, 177–183 (2021).
207. Rolland, M. *et al.* Recognition of HIV-1 peptides by host CTL is related to HIV-1 similarity to human proteins. *PloS One* **2**, e823 (2007).

208. Richman, L. P., Vonderheide, R. H. & Rech, A. J. Neoantigen Dissimilarity to the Self-Proteome Predicts Immunogenicity and Response to Immune Checkpoint Blockade. *Cell Syst.* **9**, 375-382.e4 (2019).
209. McGranahan, N. & Swanton, C. Neoantigen quality, not quantity. *Sci. Transl. Med.* **11**, eaax7918 (2019).
210. Peng, M. *et al.* Neoantigen vaccine: an emerging tumor immunotherapy. *Mol. Cancer* **18**, 128 (2019).

I

ARTICLES
https://doi.org/10.1038/s43018-021-00226-4

nature cancer

Check for updates

Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity

Máté Manczinger[1,2,3,4,7] ✉, Balázs Koncz[2,7], Gergő Mihály Balogh[2,7], Benjamin Tamás Papp[2,5], Leó Asztalos[2,5], Lajos Kemény[2,3,4], Balázs Papp[1,6] and Csaba Pál[1] ✉

Human leukocyte antigen class I (HLA-I) genes shape our immune response against pathogens and cancer. Certain HLA-I variants can bind a wider range of peptides than others, a feature that could be favorable against a range of viral diseases. However, the implications of this phenomenon on cancer immune response are unknown. Here we quantified peptide repertoire breadth (or promiscuity) of a representative set of HLA-I alleles and found that patients with cancer who were carrying HLA-I alleles with high peptide-binding promiscuity have significantly worse prognosis after immune checkpoint inhibition. This can be explained by a reduced capacity of the immune system to discriminate tumor neopeptides from self-peptides when patients carry highly promiscuous HLA-I variants, shifting the regulation of tumor-infiltrating T cells from activation to tolerance. In summary, HLA-I peptide-binding specificity shapes neopeptide immunogenicity and the self-immunopeptidome repertoire in an antagonistic manner, and could underlie a negative trade-off between antitumor immunity and genetic susceptibility to viral infections.

HLA-I molecules generally reside on the surface of human nucleated cells and present intracellularly derived peptides. HLA-I genes are the most variable in the human genome[1], and this variation shapes susceptibility to infectious diseases, autoimmune disorders and cancers[1].

The adaptive immune recognition of mutated peptides (neopeptides) is essential for effective tumor destruction by effector immune cells[2]. These neopeptides are presented on the surface of cancer cells generally bound by HLA-I molecules, and are subsequently recognized by T cells which can potentially induce an anticancer immune response[3]. Allelic HLA-I variation shapes the efficacy of antitumor immunity in a complex manner. HLA-I homozygosity, certain HLA-I supertypes and germline HLA-I evolutionary divergence (HED) are associated with an altered overall survival in patients receiving immunotherapy[4,5], but these associations may be dependent on the type of cancer, genetic background and specific details of treatment[6,7]. Additionally, the HLA-I genotype shapes the mutational landscape of tumors: common missense mutations are generally bound more weakly by HLA molecules[8] and lung cancer cells frequently lose one copy (or both copies) of their HLA-I alleles, resulting in immune escape of cancer cells[9].

An important feature of HLA-I alleles that has received little attention in cancer immunity is peptide-binding promiscuity. There is a substantial variation in the size of the bound and presented peptide repertoire across HLA-I alleles[10–12]. Certain HLA-I alleles are capable of binding an exceptionally large set of peptide segments. For instance, a bioinformatics analysis has revealed >16-fold variation in the number of peptides bound by common HLA-I alleles from a wide range of dengue-virus-derived peptides[12]. Similarly, the binding capacity of HLA-B molecules to a large set of self-peptides varies extensively, and influences the native repertoire of T-cell clones developing in the thymus[13]. Previous studies suggest that HLA class I and II peptide-binding promiscuity has been positively selected during the evolutionary diversification of human populations, because these provide an enhanced capacity to withstand high pathogen diversity[10,11].

In this work we studied the impact of the peptide-binding repertoire of HLA-I alleles on cancer immune checkpoint inhibitor (ICI) therapy. ICIs target inhibitory receptors on T cells and thereby stimulate antitumor immunity[14]. However, because only a subset of patients responded adequately to ICI-based immunotherapies, understanding the immunological mechanisms underlying therapeutic success is of paramount importance. We found that patients with cancer carrying HLA-I alleles with promiscuous peptide binding were characterized by significantly worse prognosis after ICI immunotherapy. This trend can be explained by a reduced capacity of the immune system to discriminate tumor neopeptides from self-peptides in individuals carrying highly promiscuous HLA-I variants. As a consequence, cancer samples from these patients display signatures of an immunosuppressive tumor microenvironment with prominent T-cell dysfunction. Taken together, these results indicate that HLA-I alleles with low peptide-binding specificity represent a genetic barrier to effective cancer immunotherapy.

Results

Peptide-binding specificity of HLA-I alleles. Our first aim was to estimate the peptide-binding specificity of individual HLA-I alleles by collecting peptide–HLA-I interactions validated experimentally

[1]Biological Research Centre, Institute of Biochemistry, Synthetic and Systems Biology Unit, Eötvös Loránd Research Network (ELKH), Szeged, Hungary. [2]Department of Dermatology and Allergology, University of Szeged, Szeged, Hungary. [3]MTA-SZTE Dermatological Research Group, Eötvös Loránd Research Network (ELKH), University of Szeged, Szeged, Hungary. [4]HCEMM-USZ Skin Research Group, Szeged, Hungary. [5]Szeged Scientist Academy, Szeged, Hungary. [6]HCEMM-BRC Metabolic Systems Biology Lab, Szeged, Hungary. [7]These authors contributed equally: Máté Manczinger, Balázs Koncz, Gergő Mihály Balogh. ✉e-mail: manczinger.mate@brc.hu; cpal@brc.hu

in vitro (Methods). In total, we analyzed >250,000 peptide–HLA interactions covering 67 *HLA-A*, *-B* or *-C* alleles with appropriate in vitro data. These alleles are present at detectable frequencies and cover a large fraction of individuals in numerous human populations. Importantly, all *HLA-A* and *-B* alleles of a well-established reference set with maximal population coverage[15] were included in our analysis.

An established protocol was used to identify the amino acid composition of the 8–12-amino acid-long peptides bound by the peptide-binding domain of each HLA-I allele (Methods). Using these data, we calculated peptide-binding promiscuity (*Pr*), an index that estimates the sequence diversity of the peptides presented by each allele (Methods and Extended Data Fig. 1a,b). Alleles with high *Pr* are capable of binding a wide variety of peptides and can be explained with diverse peptide motifs (Extended Data Fig. 2). Bound peptides are enriched in either hydrophobic and basic amino acids or tyrosine at anchor positions 2 and/or 9 (Extended Data Fig. 2). At the same time, the selectivity at these and other positions is low compared to other alleles (Extended Data Fig. 2).

We found a substantial variation in *Pr* across common *HLA-A*, *-B* and *-C* alleles (Fig. 1a). The reliability of the calculated *Pr* values was first confirmed by analyzing the data of two large-scale immunopeptidomics studies[16,17]. These studies identified naturally eluted self-peptides from the surface of HLA-I monoallelic cell lines, and include peptide data for ten *HLA-A*, six *HLA-B* and 15 *HLA-C* alleles. We found a strong positive correlation between *Pr* and peptidome diversity on the cell surface (Fig. 1b), which remained when data of the two studies were analyzed separately (Extended Data Fig. 3a,b).

Additionally, the in vitro binding affinities of 11 representative HLA-I alleles to 29 tumor neoepitopes were measured with ProImmune REVEAL HLA-peptide binding assays. These alleles are widespread and are consequently part of a small set of reference HLA variants that allows maximal coverage of human populations[15]. The selected peptides were collected from the TANTIGEN database[18]. These peptides elicit either a cytotoxic immune response in vitro or can mediate tumor destruction in vivo[19] (Supplementary Table 1), and display limited or no overlap in sequence (Methods). The analysis revealed a substantial variation in binding affinities to different peptides (Supplementary Table 1). Reassuringly, there was a strong positive correlation between *Pr* and the fraction of neopeptides bound by a given allele (Fig. 1c; Spearman's $\rho = 0.86$, two-sided correlation test $P = 7 \times 10^{-4}$).

To confirm the results reported above, the binding specificities of HLA-I alleles against established sets of 1,929 cancer peptides (Supplementary Table 2) and 9,544 viral peptides derived from diverse pathogens (Supplementary Table 3) were calculated using the NetMHCpan-4.0 algorithm[20]. Additionally, it was also tested whether *Pr* is associated with the diversity of human self-immunopeptidome using a set of 212,090 peptides identified in 25 immunopeptidomics studies (Supplementary Table 4). It was found that HLA-I variants with high *Pr* present a larger diversity of neopeptides, human self-peptides and viral peptides (Fig. 1d–f). These associations remained significant in multivariate linear regression models that included the *HLA-I* locus as a categorical predictor variable (Supplementary Table 5). The relationship between promiscuity and the fraction of bound neopeptides holds for both *HLA-A* and *-B*, but not for *HLA-C*, when these loci were analyzed separately (Extended Data Fig. 3c–e). Together, these results indicate that HLA-I peptide-binding promiscuity shapes the diversity of the presented tumor, human self and viral immunopeptidome. They also verify that *Pr* is a suitable index for estimation of the diversity of peptides presented by HLA-I alleles.

Because the stability of peptide–HLA complexes is essential for effective antigen presentation[21], we finally evaluated whether promiscuous HLA-I alleles have an altered capacity to form stable protein complexes with neoepitopes. The assembly and dissociation rates of 66 representative neoepitope–HLA-I complexes were measured in vitro using ProImmune Complete rate assays (Methods). On average, assembly of the complexes tends to be faster and the resulting complexes more stable with high-*Pr* HLA-I molecules (Fig. 2a–d). These results were further supported by the computational analysis of HLA–peptide complex stability involving 67 HLA alleles and 1,929 neopeptides (Fig. 2e and Methods). We conclude that, in contrast to other proteins[22], there is no evidence for a negative trade-off between binding multispecificity and complex stability in the case of HLA-I molecules.

High level of HLA-I promiscuity yields worse prognosis. To investigate the impact of *Pr* on cancer immunotherapy, we focused on previously published cohorts of patients with cancer who were treated with ICI[4,23–25]. We collected data on patients with melanoma who were receiving anti-CTLA-4 therapy[4,23,24] ($n = 164$), patients with non-small cell lung cancer (NSCL) who were treated with anti-PD-1 therapy[25] ($n = 74$) and patients with melanoma who were treated with anti-PD-1/anti-PD-L1 therapy[4] ($n = 78$). The HLA genotype of patients in these three cohorts was determined previously[4]. For each patient we calculated genotype *Pr* as the mean of six *HLA-A*, *-B* and *-C* allelic *Pr* values, under the assumption that each locus contributes to the presentation of neopeptides. There was a substantial variation in genotype *Pr* across patients in the examined cohorts (Extended Data Fig. 4).

We next examined how genotype *Pr* affects overall survival probability in response to ICI therapy. Patients were classified into high and low genotype *Pr*, based on a fixed cutoff used in all cohorts (Methods). High genotype *Pr* was found to be associated with reduced overall survival in all three cohorts (Fig. 3a–c). For example, survival rates of CTLA-4 inhibitor-treated melanoma patients with high genotype *Pr* were 44% lower than those of patients in the low-*Pr* group, while the high genotype *Pr* group of PD-1-inhibitor-treated patients with NSCLC were 62% less likely to survive than individuals in the low-*Pr* group.

We also tested the impact of genotype *Pr* on clinical outcome using the Response Evaluation Criteria in Solid Tumors (RECIST) criteria. As in previous studies[26,27], clinical benefit to treatment was defined as complete or partial response while absence of clinical benefit was defined as stable disease or progressive disease (Methods). Reassuringly, patients with high genotype *Pr* were >twofold less likely to receive clinical benefit from ICI treatment (Fig. 3g). Moreover, this effect was even stronger in an independent cohort containing patients with melanoma who were treated with either PD-1 (ref. [28]) or CTLA-4 inhibitors (Extended Data Fig. 5a,b and Methods).

Finally, we briefly examined whether HLA-I promiscuity shapes disease progression in those patients who did not receive ICI therapy. Melanoma samples in The Cancer Genome Atlas (TCGA) were classified into high- and low-mutational-burden (TMB) groups. High genotype *Pr* was found to be associated with reduced progression-free survival in patients with high TMB (Extended Data Fig. 5f), while progression-free survival was independent of genotype *Pr* in patients with low TMB (Extended Data Fig. 5c). This suggests that HLA-I promiscuity and TMB may jointly shape disease progression in patients with melanoma. By contrast, HLA-I promiscuity had no impact on the survival of NSCLC patients regardless of the TMB level (Extended Data Fig. 5d,e,g,h). Clearly, additional data will be needed to establish the prognostic effects of HLA-I promiscuity on antitumor immunity.

HLA-I promiscuity is a major determinant of patient survival. The relationship between the level of HLA-I promiscuity and survival remained significant or became even stronger when patients

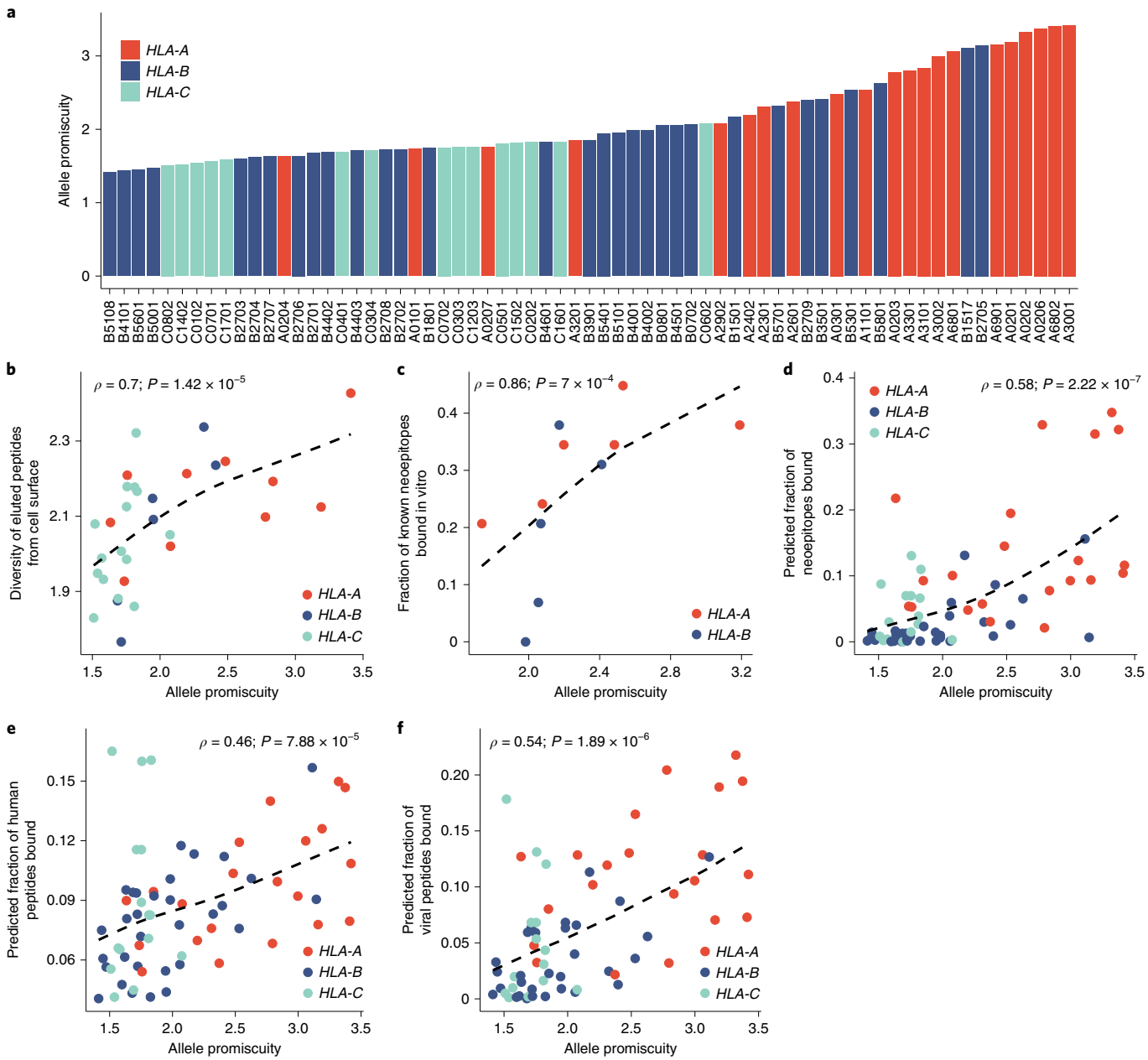


Fig. 1 | Basic properties of HLA-I allelic promiscuity. a, *Pr* values for the 67 HLA class I alleles used in our study are shown in increasing order. **b**, Amino acid diversity of self-peptides presented by the given allele on monoallelic cell lines is shown as the function of allele promiscuity. Alleles with higher promiscuity present more diverse peptide sequences ($n = 31$ alleles). **c**, The fraction of immunogenic neoepitopes bound by the given allele is shown as the function of allele promiscuity. Binding was determined in vitro with ProImmune REVEAL assays ($n = 11$ alleles; see Methods for assay details). **d**–**f**, The predicted fractions of bound neopeptides (**d**), human peptides identified in immunopeptidomics studies (**e**) and viral peptides (**f**) are shown as the function of HLA promiscuity ($n = 67$ alleles in all panels). **b**–**e**, Spearman's coefficients and *P* values of two-sided correlation tests are shown; dashed lines indicate smooth curve fitted using the cubic smoothing spline method in R (Methods).

homozygous on any of the *HLA-I* loci were excluded from the analysis (Fig. 3d–f).

Similarly, the association between genotype *Pr* and patient survival remained significant when the analysis was iteratively repeated with the exclusion of individuals carrying specific HLA alleles (Supplementary Table 6). Hence, the impact of HLA promiscuity on patient survival cannot be explained by a single, highly prevalent allele with peculiar peptide-binding properties. An additional analysis revealed that the overall survival of patients carrying multiple promiscuous HLA-I alleles decreased, while that of patients carrying only one promiscuous HLA-I allele remained unaltered (Extended Data Fig. 6a).

Based on the structural properties of the peptide-binding regions, genetically related HLA-I alleles were previously classified into 12 supertypes that cover the majority of common *HLA-A* and *-B* alleles[29]. Patients with melanoma who were carrying *HLA-B* alleles belonging to the B44 superfamily were reported to have significantly better survival after ICI treatment, but the reasons for this remain unclear[4]. Notably, the association between patient survival and B44 supertype is mainly due to specific *HLA-B* alleles such

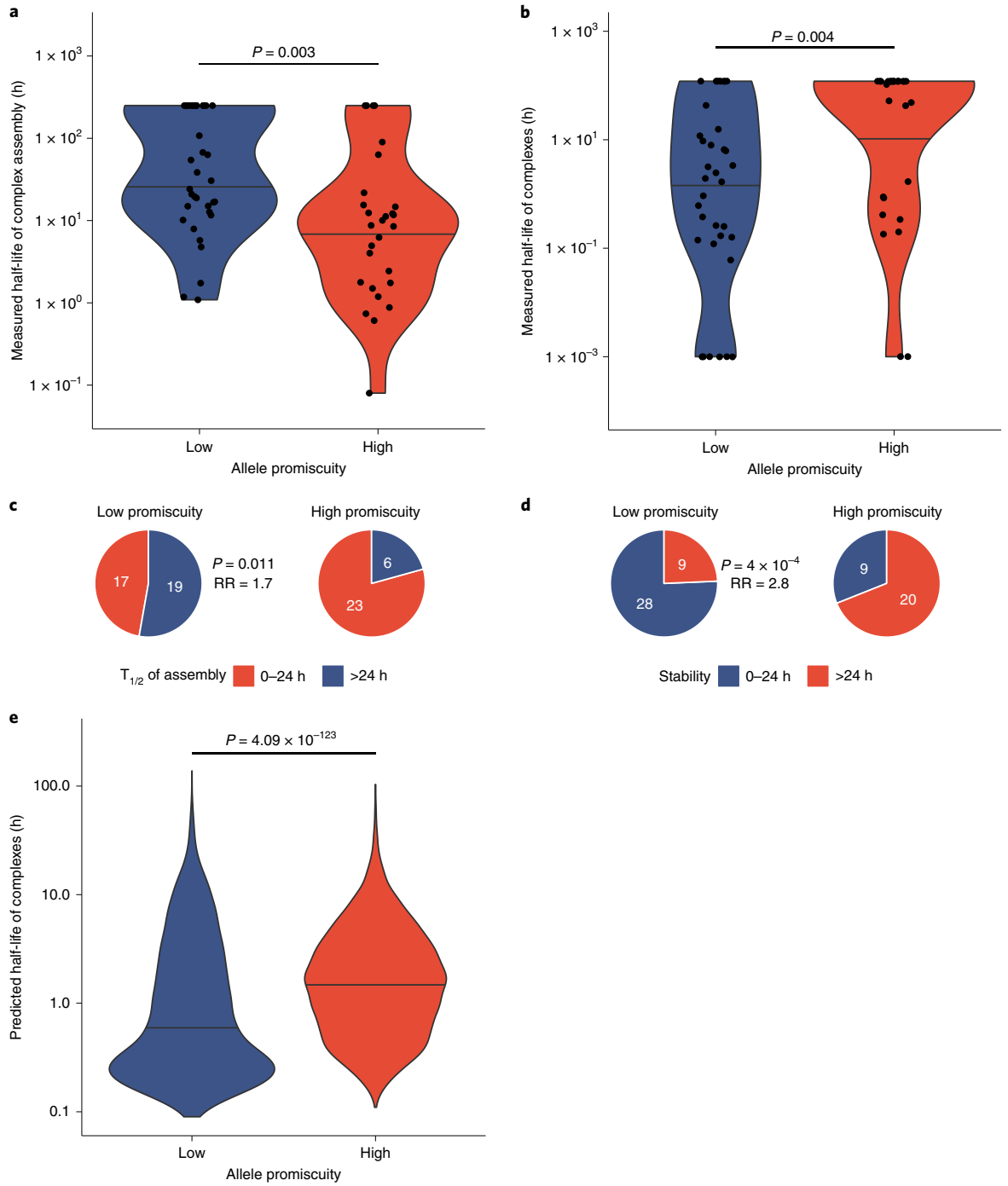


Fig. 2 | High allele promiscuity is associated with more stable peptide–HLA complexes. a, In vitro measured half-life of complex assembly for 65 allele–neoepitope pairs. Alleles were stratified into low- ($n = 36$ complexes) and high- ($n = 29$ complexes) promiscuity groups based on the median. The result for one allele–neopeptide pair is not shown because stable complexes were not detected at the end of the measurement period. **b**, In vitro measured stability of 66 allele–neoepitope pairs. Peptide–HLA complexes of promiscuous HLA variants are more stable. The vertical axis shows the complex half-life in hours ($n = 37$ and 29 complexes in low and high allele promiscuity groups, respectively). For allele–peptide pairs that did not form any stable complexes at the start of the measurement period, complex half-life was defined as 0. **c**,**d**, Promiscuous alleles are 1.7-fold more likely to form complexes with an assembly half-life ($T_{1/2}$) <24 h (**c**) and 2.8-fold more likely to form particularly stable complexes with $T_{1/2} > 24$ h (**d**). **c**,**d**, *P* values for two-sided Fisher's exact tests are indicated. RR, relative risk. **e**, Predicted half-life of neopeptide–HLA complexes using the NetMHCstabpan algorithm[70]. Complexes were identified by predicting the binding affinity of 1,929 neopeptides to 67 HLA-I alleles (Fig. 1d and Methods). Alleles were stratified into low- ($n = 3{,}518$ complexes) and high- ($n = 4895$ complexes) promiscuity groups using the same cutoff as in **a**,**b**. **a**,**b**,**e**, *P* values for two-sided Wilcoxon's rank-sum tests are indicated. Violin plots show the density function of values indicated on vertical axes, and horizontal lines indicate the median value in each group.

as B*18:01, B*44:02, B*44:03, B*44:05 and B*50:01. The four B44 alleles included in our analysis have particularly low *Pr* compared to other alleles (Extended Data Fig. 6b).

Tumor mutational burden[30], HLA-I heterozygosity[4] and HED[5] have previously been suggested as the determinants of response to ICI therapy. These features are expected to increase the diversity of

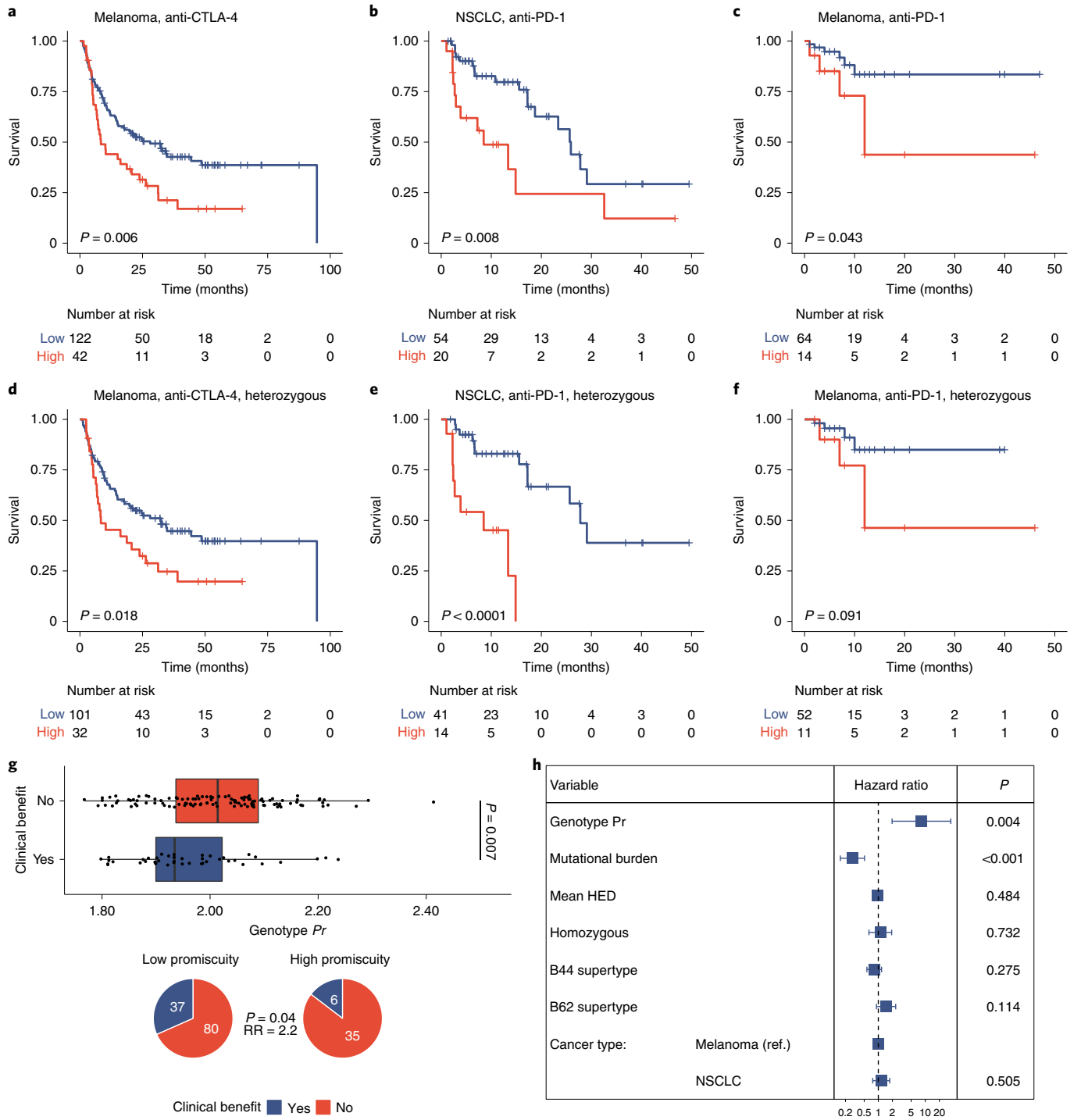


Fig. 3 | High allele promiscuity and survival rate in patient cohorts treated with checkpoint inhibitor therapy. a–c, Analysis was carried out for patients with melanoma who were receiving anti-CTLA-4 therapy (**a**, $n=164$ patients), patients with NSCLC who were receiving anti-PD-1 therapy (**b**, $n=74$ patients) and patients with melanoma who were receiving anti-PD-1/anti-PD-L1 therapy (**c**, $n=78$ patients). **d–f**, The analysis was repeated for cohorts including only those patients fully heterozygous at the *HLA-I* loci ($n=133$ patients in the anti-CTLA-4-treated melanoma cohort (**d**), $n=55$ patients in the anti-PD-1-treated NSCLC cohort (**e**) and $n=63$ patients in the anti-PD-1/anti-PD-L1-treated melanoma cohort (**f**)). Patients were classified into low and high allele promiscuity groups using the same cutoff in all cohorts (Methods). Two-sided log-rank test *P* values are shown. *Y* axes indicate the probability of survival. **g**, Patients with clinical benefit in cohorts from **a**,**b** had significantly lower genotype *Pr* ($n=43$ and 115 in patient groups with and without clinical benefit, respectively). Similarly, high genotype *Pr* was associated with 2.2-fold reduced chance of clinical benefit after ICI therapy. *P* values of a two-sided Wilcoxon's rank-sum test and a two-sided Fisher's exact test are indicated. On the boxplot, vertical lines indicate median, boxes indicate the interquartile range (IQR) and horizontal lines indicate first quartile −1.5 × IQR and third quartile +1.5 × IQR. **h**, The effect of genotype *Pr* on survival remained highly significant in a multiple Cox regression model including mutational burden, other HLA-associated features and cancer type as covariates. The model involved all patients from **a**–**c** ($n=316$ patients). Blue squares indicate hazard ratios and horizontal lines indicate 95% confidence intervals. Two-sided *P* values of *z*-statistics are shown. The two-sided log-rank test *P* value of the model was 5×10^{-6}. Genotype *Pr*, mutational burden and mean HED were treated as continuous variables.

neopeptides that can be presented on the surface of tumor cells[14]. Genotype *Pr* showed no statistically significant relationship with HED, HLA-I heterozygosity or TMB (Extended Data Fig. 6c–e). Moreover, the negative association between genotype *Pr* and patient survival remained highly significant in a multiple Cox regression model after controlling for these features (Fig. 3h). In particular, TMB and genotype *Pr* appeared to be the strongest determinants while HED and HLA-I heterozygosity had no significant effect on survival in this model. This is in line with previous work showing no association between HLA class I heterozygosity and survival of patients with NSCLC who were receiving immunotherapy[6]. Furthermore, the impact of *Pr* on patient survival remained after controlling for the effect of specific *HLA-B* supertypes B44 and B62, which have previously been associated with survival outcome[4] (Fig. 3h).

We next evaluated whether the low *Pr* of each HLA locus contributes to improved survival. For this purpose, the mean *Pr* was calculated for *HLA-A*, *-B* and *-C* loci individually and also for combinations thereof. The promiscuity level of *HLA-B* was predictive of survival outcome, but the negative association between genotype *Pr* and patient survival becomes even stronger by the inclusion of *HLA-A* and *-C* in univariate Cox models (Supplementary Table 7). The prominent effect of *HLA-B* on patient survival might be explained by the relatively high abundance of this molecule on the cell surface relative to that of *HLA-A* and *-C*[4,31]. Finally, genomic analysis of cancer mutations in 165 patients revealed no significant association between genotype *Pr* and the frequency of loss of heterozygosity events at any of the HLA-I loci (two-sided Wilcoxon's rank-sum test, $P=0.64$).

Neopeptide dissimilarity to the self-proteome and promiscuity. We next asked whether the presentation of more diverse peptides by promiscuous HLA-I molecules would result in a reduced capacity to differentiate between human self-peptides and their mutated counterparts (neopeptides). For this purpose we calculated the differential agretopicity index[32] (DAI), which is a broad indicator of neopeptide dissimilarity from self and thereby a feature of immunogenicity[32]. DAI estimates the difference between the predicted affinity for the mutant peptide and the corresponding nonmutated homologous peptide pair[32]. To gain insight into the effect of each HLA-I molecule on discrimination of self and tumor peptides, we employed a set of 589 experimentally verified tumor neopeptides (Methods). For each HLA-I variant we calculated the median DAI for neopeptides bound by the given HLA-I allele (Methods). A strong negative association was found between median DAI and the level of promiscuity for *HLA-A* and *-B* alleles, but not for *HLA-C* alleles (Fig. 4a–c and Extended Data Fig. 7a,b).

It has previously been suggested that tumors carrying high-DAI neopeptides are more susceptible to immune recognition and hence should be responsive to ICI therapy[32–35]. Therefore, the next important question is how HLA-I promiscuity shapes the abundance of potentially immunogenic, high-DAI neopeptides in patients with cancer. For this purpose, we analyzed the cancer genomes of 139 patients with melanoma who had been treated with CTLA-4 inhibitors[23,24]. From the detected set of nonsynonymous mutations in each cancer sample, we determined the complete set of neopeptides bound to at least one of the HLA-I alleles in the corresponding patient (Methods). The median DAI for neopeptides showed a significant negative association with genotype *Pr* (Fig. 4d), indicating that the immunogenicity of neopeptides is contingent upon HLA-I promiscuity level. As previously reported[32,36], we also found that higher median DAI is associated positively with patient survival (Fig. 4e). These results do not simply reflect differences in the set of cancer mutations across patients. For each genotype, DAI was calculated as the mean of allele-specific values determined on a fixed set of 589 experimentally confirmed neopeptides, and the positive association between DAI and survival was found to remain (Extended Data Fig. 7c). Together, these results indicate that patients who carry promiscuous HLA-I alleles have a reduced capacity to differentiate between self-peptides and the corresponding mutant tumor neopeptides, with implications for patient survival[32].

HLA-I promiscuity and T-cell tolerance. Immunogenic neopeptides need to have a strong binding affinity to HLA-I/T-cell receptor complexes. However, when their nonmutated counterparts are also presented by HLA molecules, the corresponding $CD8^+$ T cells reactive to neopeptides are expected either to be removed by central tolerance in the thymus or inhibited by peripheral tolerance[32,35]. Here we focus on the impact of HLA-I promiscuity on peripheral tolerance, a mechanism ensuring that self-reactive T cells that have not been eliminated by central tolerance in the thymus do not recognize self-peptides as foreign and initiate autoimmune reactions[37]. However, an elevated level of T-cell tolerance in the tumor microenvironment contributes to uncontrolled tumor growth[38,39]. We hypothesized that, due to the reduced discrimination capacity of self and tumor peptides in patients carrying high-*Pr* HLA-I alleles, regulation of T cells in the tumor microenvironment is shifted from activation to tolerance, yielding reduced capacity to eliminate cancer cells following ICI therapy.

To identify molecular and cellular signatures of T-cell tolerance in cancer samples, we focused on datasets from patients with melanoma who had been treated with the anti-CTLA-4 antibody ipilimumab[23]. A unique aspect of this dataset is that it combines information on the transcriptome for a sufficient number of pretreatment cancer samples with the HLA genotype and disease progression of each patient.

We first performed a gene set enrichment analysis comparing patients with high and low genotype *Pr*. Most notably, those genes involved in the positive regulation of T-cell tolerance induction, type 2 immune response, extracellular matrix secretion and macrophage induction were found to be upregulated in high-*Pr* cancer samples (Fig. 5a). Similarly, those genes associated with the negative regulation of T helper-1-cell-mediated immune response and $CD4^+$ alpha–beta T-cell differentiation were expressed at higher levels in high-*Pr* samples (Fig. 5a). These cellular processes are linked to the induction of an immunosuppressive tumor microenvironment[39,40]. Specifically, macrophages inhibit effector immune responses and stimulate angiogenesis[41]. Additionally, they skew type 1 immune response (mediated by T helper-1 cells) towards type 2, which is less effective against cancer cells[40,41]. Remodeling of the extracellular matrix is essential for immune suppression and cancer progression[42,43]. Finally, the lower level of $CD4^+$ T-cell differentiation and T-cell tolerance induction prohibits effective tumor destruction[38,39].

Next, by applying state-of-the-art immune deconvolution methods and recommendations from previous benchmark studies[44], we assessed the immune cell composition in these tumor samples[44]. The analysis revealed an over-representation of cells associated with immunosuppression in tumor samples derived from patients with high genotype *Pr* (Fig. 5b), including regulatory T cells (Tregs), cancer-associated fibroblasts, macrophages/monocytes and endothelial cells. Tregs and macrophages produce immunosuppressive mediators such as TGF-β, which downregulates effector immune functions and induces the development and survival of Tregs[45]. As expected, cancer samples from patients with high *Pr* displayed increased expression of *TGFB1* encoding TGF-β (Fig. 5c). This is notable because elevated TGF-β expression is associated with the promotion of tumorigenesis[46]. Similarly, the gene *FOXP3*, encoding a master regulator protein involved in the development and function of Tregs, was also highly expressed in these samples[38] (Fig. 5c).

T-cell costimulation is an essential step in the development of T-cell tolerance or function[38]. Negative costimulatory molecules inhibit T-cell activation at immune checkpoints, leading to T-cell

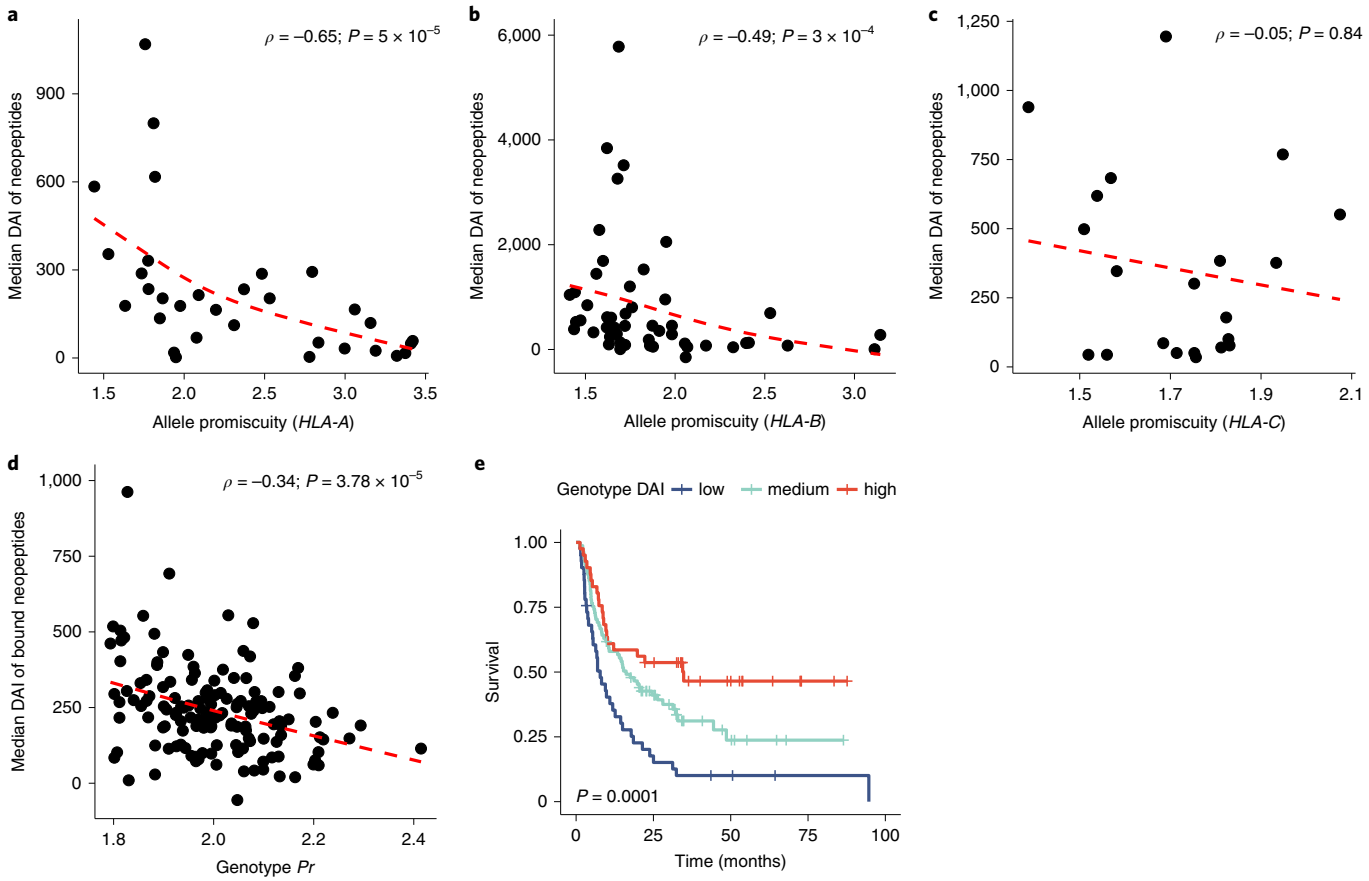


Fig. 4 | HLA-I allelic promiscuity and differentiation capacity between neopeptides and corresponding nonmutated peptides. a–**c**, Relationship between the promiscuity of *HLA-A* (**a**, $n=32$ alleles), *HLA-B* (**b**, $n=50$ alleles) and *HLA-C* (**c**, $n=21$ alleles) alleles and median DAI for bound tumor neopeptides. Higher *HLA-A* and *-B* promiscuity is associated with lower median DAI values, while there is no significant relationship between the two variables in the case of *HLA-C*. **d**, In melanoma samples ($n=139$), neopeptides bound by promiscuous HLA alleles have a lower median DAI. **e**, Patients from Fig. 3a were stratified into low- ($n=41$ patients), medium- ($n=81$ patients) and high- ($n=41$ patients) genotype DAI groups based on the first and third quartiles of all values. A higher median DAI was associated with better survival. A two-sided log-rank test *P* value is shown. Between-group differences were also tested for trend (Methods). *Y* axis indicates the probability of survival. **a**–**d**, Spearman's ρ and the associated *P* values of two-sided correlation tests are shown; dashed lines indicate a smooth curve fitted using the cubic smoothing spline method in R (Methods).

tolerance[38]. Consequently, the inhibitory immune checkpoint receptor PD-1, as well as the coinhibitory receptors TIGIT and TIM3, have established roles in promotion of tumor immune evasion[47,48]. Genes encoding these three molecules displayed significantly elevated expression in cancer samples with high *Pr* (Fig. 5d). It is well known that TIM3 is especially abundant in FOXP3-expressing Treg cells and macrophages[49].

Irreversible dysfunction of T cells is a common cause of nonresponsiveness to immune checkpoint blockade therapy[50], and is linked to the formation of an immunosuppressive tumor microenvironment[39]. Therefore, we next asked whether high-*Pr* HLA-I alleles are associated with signatures of T-cell dysfunction. Remarkably, we found that cancer samples with high *Pr* display high expression levels of genes encoding TOX, TOX2, T-bet and BLIMP1 (Fig. 5e). These transcription factors are critical for the progression and maintenance of T-cell dysfunction[51,52]. To investigate T-cell dysfunction signatures in cancer samples more systematically, we employed the Tumor Immune Dysfunction and Exclusion (TIDE) method[50]. TIDE is a recently developed computational tool providing genome-wide scores of T-cell dysfunction and T-cell exclusion signatures for each tumor, and can reliably predict ICI response. Importantly, the dysfunction score describes the level of late-stage, irreversible dysfunction of T cells as reported by the authors[50]. We found a positive correlation between genotype *Pr* and T-cell dysfunction score (Fig. 5f; Spearman's $\rho=0.4$, two-sided correlation test $P=0.03$), but not between genotype *Pr* and T-cell exclusion score (Spearman's $\rho=-0.1$, two-sided correlation test $P=0.6$). This suggests that HLA-I promiscuity level does not affect T-cell infiltration directly, but rather it shapes the functioning of T cells in the tumor microenvironment. Importantly, other clinically relevant HLA-I-associated features, such as HLA-I heterozygosity, germline HLA-I evolutionary divergence and specific HLA-I supertypes, showed no association with T-cell dysfunction score (Extended Data Fig. 8).

Taken together, these results indicate that cancer samples carrying highly promiscuous HLA-I alleles tend to display signatures of an immunosuppressive tumor microenvironment and T-cell dysfunction, resulting in resistance to ICI therapy.

Discussion

HLA-I genes encode a critical component of the adaptive immune system. The human genome contains six primary allelic variants encoded by *HLA-A*, *-B* and *-C*. Together, these alleles define the set of individual peptides presented to T cells[1]. The resulting peptide repertoire—or immunopeptidome—substantially differs across individuals due to the remarkable sequence variation in the peptide-binding region of these genes. As a consequence, each allele presents a distinct, and only partially overlapping, set of peptides.

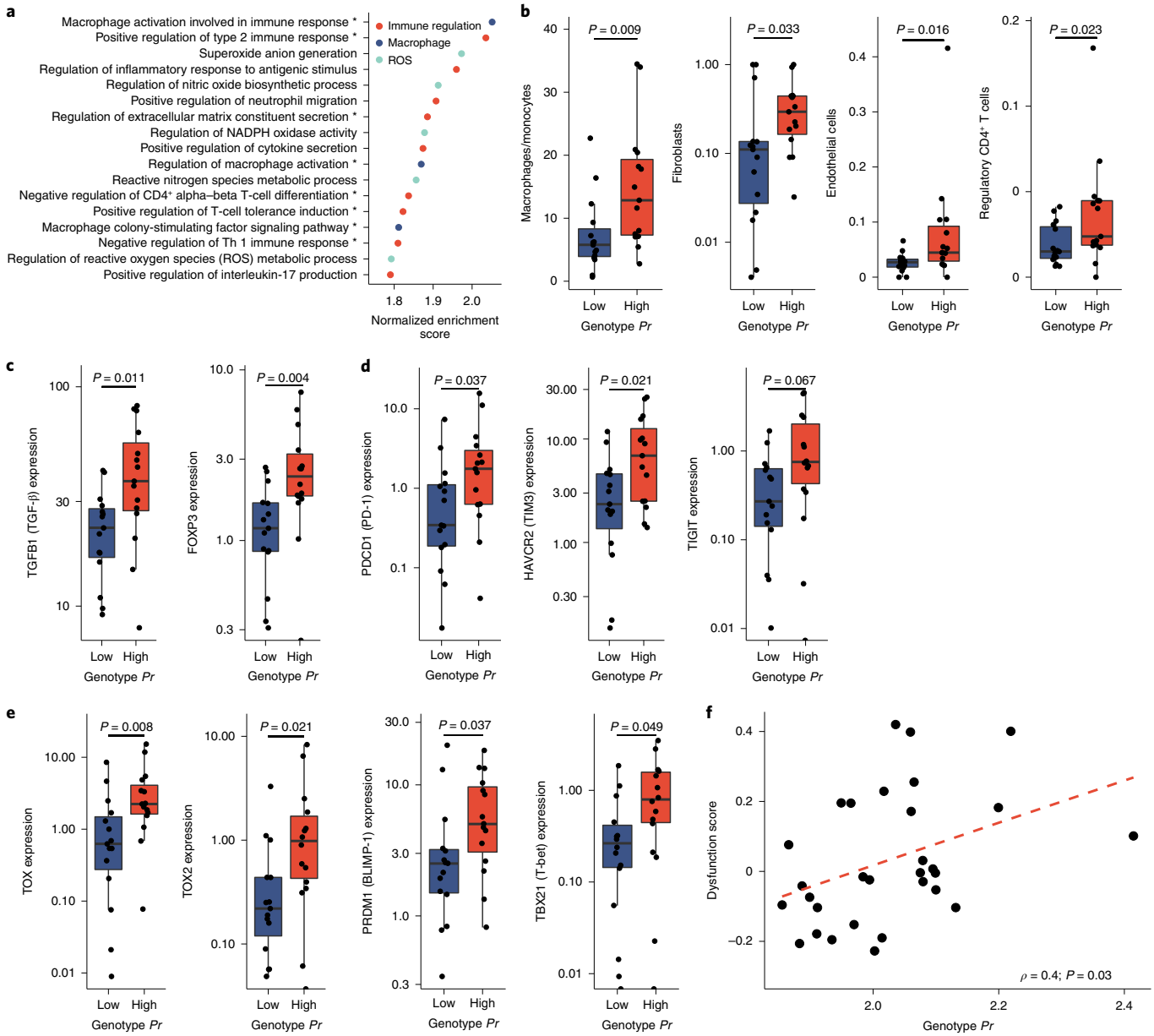


Fig. 5 | Association of high HLA genotype promiscuity with an immunosuppressive cancer microenvironment. a, Results of gene set enrichment analysis. Gene sets enriched in high genotype *Pr* samples with $P < 0.001$ and FDR < 0.1 are shown. Terms are classified into three main categories, and those associated with the development of an immunosuppressive tumor microenvironment are marked with asterisks. **b**, Results of immune deconvolution analysis. Boxplots show approximated cell abundance in samples with low and high genotype *Pr*. **c**, Expression of genes encoding the cytokine TGF-β and the transcription factor FOXP3, molecules associated with T-cell tolerance. **d**, Expression levels of genes encoding checkpoint molecules. **e**, Expression of genes encoding master regulators associated with T-cell dysfunction. **a–e**, Patients were stratified into low- and high -genotype *Pr* groups using the median as a cutoff ($n = 15$ samples in both groups). **f**, High genotype *Pr* is associated with an elevated T-cell dysfunction score calculated by the TIDE algorithm. Spearman's ρ and *P* values of a two-sided correlation test are indicated ($n = 30$ samples). The dashed red line indicates a linear regression line. **b–e**, *P* values for two-sided Wilcoxon's rank-sum tests are shown. **b–e**, In boxplots, horizontal lines indicate median, boxes indicate interquartile range and vertical lines indicate first quartile −1.5 × IQR and third quartile +1.5 × IQR.

Moreover, HLA variants differ by orders of magnitude in the breadth of peptides they can bind[10,11]. HLA-I alleles with promiscuous peptide binding may be favorable in pathogen-rich environments[10], but their role in anticancer immunity remains to be established.

To investigate this issue we first developed a computational tool to systematically explore the sequence diversity of the peptide set presented by different HLA-I molecules, and found that high HLA-I promiscuity is associated also with enhanced diversity of tumor, viral and self-immunopeptidomes (Fig. 1b,d–f). Next, we showed that patients with melanoma or NSCLC cancer and carrying HLA-I alleles with high *Pr* display a significantly worse prognosis after ICI immunotherapy (Fig. 3). This pattern is surprising, because the breadth of bound peptides by HLA-I molecules is expected to underlie a successful immunologic control of cancer. Therefore, we next aimed to establish the deleterious side effects of elevated HLA-I promiscuity level on anticancer immunity. Of note, tumor mutational burden and HLA promiscuity level were the strongest determinants in a multiple Cox regression model, while HED and

HLA-I heterozygosity had no significant effect on overall patient survival (Fig. 3h).

The stability of the peptide–HLA-I complex is a strong predictor of antigen immunogenicity[21]. In contrast to expectations, promiscuous HLA variants tend to form more stable HLA–peptide complexes (Fig. 2). This finding has important immunological implications and should be further investigated in future research. On the other hand, we found a strong negative association between the level of promiscuity and the differential binding affinity of mutated neopeptides (Fig. 4). This suggests that the immune system of patients with high *Pr* has a reduced capacity to discriminate self-peptides from mutated tumor peptides.

We hypothesized that, as a consequence, regulation of T cells in the tumor microenvironment is shifted from activation to tolerance, yielding a reduced capacity to eliminate cancer cells following ICI therapy. In agreement with expectations, cancer samples from patients with melanoma carrying promiscuous HLA-I alleles displayed molecular and cellular signatures of peripheral T-cell tolerance and an immunosuppressive tumor microenvironment (Fig. 5). These signatures include increased prevalence of regulatory T cells and other immunosuppressive cells in the tumor microenvironment and elevated expression of genes encoding immune checkpoint inhibitory receptors (PD-1), suppressive soluble mediators (TGF-β) and master control genes involved in regulatory T-cell development and maintenance (for example, *FOXP3*). Future work should elucidate the precise molecular pathways underlying these patterns. This research will require a series of HLA transgenic mouse studies on the role of promiscuity in T-cell-initiated immune response. The impact of HLA-I promiscuity level on central tolerance and potential depletion of T cells in the repertoire in healthy individuals also remains to be elucidated[53].

In sum, high peptide-binding promiscuity could be a common feature of genetically diverse HLA-I variants associated with reduced survival following ICI therapy. Therefore, this metric could potentially be used for future clinical trials to identify genetic variables that affect antitumor immunity. Importantly, HLA promiscuity is conceptually different from and statistically independent of other, previously identified determinants driving response and adverse effects of ICI immunotherapy[4,5,14] (Extended Data Fig. 6c–e). An important and unresolved issue is whether the impact of HLA-I promiscuity level on clinical benefit is especially strong in tumors with a high mutational burden. A preliminary analysis of mismatch-repair-deficient and mismatch-repair-proficient patients with colorectal cancer who were treated with PD-1 inhibitors[54] indicates that this may indeed be so (Extended Data Fig. 5i,j). As a consequence of defects in mismatch repair, cancer samples in the former group have accumulated an especially large number of somatic mutations, and high TMB was associated with prolonged progression-free survival[54]. To increase the size of the dataset, we focused on one important aspect of clinical benefit—that is, the overall changes in tumor size after treatment. The analysis revealed a significant negative association between genotype HLA-I *Pr* and clinical benefit for patients with defects in mismatch repair (Extended Data Fig. 5i). By contrast, *Pr* had no impact on clinical benefit after treatment in patients with mismatch-repair-proficient cancers (Extended Data Fig. 5j). This may indicate that TMB shapes the association between HLA-I promiscuity and clinical benefit. However, the number of patients in this study was rather limited, and future studies are required to settle this issue.

Our work has several important implications for future studies. First, HLA class II molecules are known to drive antitumor responses and tumor evolution in general[55,56]. Therefore, it is important to establish how the functional diversity of HLA-II molecules shapes the efficacy of ICI immunotherapy. Second, while the presentation of a broad range of viral peptides by promiscuous HLA-I alleles could be favorable in terms of acute viral infections[10,11], our work indicates that such alleles impede antitumor immunity. In other words, *Pr* could underlie a negative trade-off between antitumor immunity and genetic susceptibility to viral infections. Why is this so? Because viral peptides tend to display limited sequence similarity to the human proteome[36,57] (Extended Data Fig. 7d), promiscuous HLA-I molecules present a wider range of viral epitopes (Fig. 1f) without paying the cost of reduced discrimination between self- and virus-derived peptides. Third, and on a related note, promiscuous *HLA-A* and *-B* variants are especially widespread in South-East Asia, possibly as a result of selection to withstand high pathogen load[10]. This pattern raises the possibility that there are geographical differences in the genetic make-up shaping the efficacy of cancer immunotherapy and antitumor immunity in general. Finally, previous studies suggested that neoantigen immunogenicity and quality—and not simply neoantigen quantity—determine patient survival[58]. Our work indicates that HLA-I peptide-binding promiscuity shapes neoantigen immunogenicity, with implications on the design of neoantigen vaccines[59].

Methods

Calculation of peptide-binding specificity. Data on peptide–HLA class I interactions and corresponding T-cell activation assays were downloaded from The Immune Epitope Database[60], as of 17 January 2019. Only peptides of length 8–12 standard amino acids, and with positive major histocompatibility complex binding and/or T-cell activation assays, were considered further. To ensure reliable estimation of peptide-binding promiscuity, only HLA-I alleles with data on the binding of at least 400 different peptides were analyzed further. This procedure resulted in a total of 253,147 peptide–HLA-I interactions across 21 *HLA-A*, 31 *HLA-B* and 15 *HLA-C* alleles. HLA-I allelic promiscuity was calculated as follows. Peptides were classified into discrete peptide sets according to their amino acid length, and each set was analyzed separately (Extended Data Fig. 1a). At each amino acid position, amino acid frequency distribution in the peptide set was compared with that in the complete human proteome (downloaded from UniProt[61]), using Kullback–Leibler divergence (D_{KL}). KL divergence is a standard metric regularly used to measure the distance between two discrete probability distributions (P and Q) and is calculated as follows:

$$D_{KL}(y) = \sum_{x \in \chi} P(x,y) \log\left(\frac{P(x,y)}{Q(x)}\right)$$

where $P(x,y)$ denotes the frequency of amino acid x in the peptide set at amino acid position y, while $Q(x)$ denotes the frequency of the same amino acid in the complete human proteome (χ is the set of 20 amino acids). Position-specific $D_{KL}(y)$ values were calculated by the Rtreemix R library. A constant value of 10^{-7} was added to the frequency of all amino acids when a given amino acid had zero frequency at the examined position. Finally, position-specific $D_{KL}(y)$ values were averaged across the full peptide length and for each peptide set, yielding peptide-length-specific D_{KL} values. The more selective the binding of the HLA-I allele at a given position, the higher the deviation between the two distributions, resulting in higher D_{KL}. To control for any potential sample-size-associated biases, raw D_{KL} values were normalized by subtracting $D_{KL(rand)}$. $D_{KL(rand)}$ is the average of D_{KL} values based on the same number of randomly selected peptides with the same length in the human proteome iterated 1,000 times. Normalized D_{KL} values show no significant correlation with the number of input peptide sequences (Spearman's $\rho = -0.08$, $P = 0.23$). Peptide-length-specific values were then averaged by weighting with their relative proportion in the repertoire, yielding $\overline{D_{KL}}$ (Extended Data Fig. 1b). HLA allelic *Pr* is defined as the reciprocal value of $\overline{D_{KL}}$—that is, $Pr = 1/\overline{D_{KL}}$. *Pr* values do not change significantly when using $Q(x)$ derived from the proteomes of bacterial or viral species rather than from the complete human proteome (Extended Data Fig. 1e,f). Computer simulations also showed that *Pr* can be reliably inferred using the above-defined threshold of peptide number per HLA allele (Extended Data Fig. 1c,d). Of note, KL divergence, which measures the discrepancy between two probability distributions, is conceptually related to the Shannon entropy index[62]. As expected, *Pr* values based on Shannon entropy index and KL divergence show a strong positive correlation with each other (Spearman's $\rho = 0.86$, $P < 2.2 \times 10^{-16}$).

Experimental analysis of HLA-I neoepitope binding affinity. HLA-I cancer neopeptides were collected from the TANTIGEN database[18] (Supplementary Table 1). Epitopes with >50% sequence identity were excluded using an iterative method[10], resulting in 29 peptide neoepitopes. Importantly, 22 peptides were previously shown to elicit a cytotoxic T-cell response in vitro while seven were found to be recognized by tumor-infiltrating lymphocytes in vivo[19]. Each peptide was synthesized and its binding affinity towards a preselected set of 11 HLA-I variants was tested using the in vitro HLA class I REVEAL binding

assay (ProImmune Ltd.). The binding assay determines the capability of each candidate peptide to bind to one or more HLA class I alleles by assessing its ability to stabilize the HLA–peptide complex. For each allele, the binding of the tested peptides is scored relative to that of a corresponding high-affinity T-cell epitope (Supplementary Table 1). Of note, our experimental assay could confirm 21 of 22 HLA–peptide associations reported in previous studies.

Experimental analysis of peptide–HLA complex assembly and stability. All peptide–HLA-I pairs with a positive outcome in the HLA class I REVEAL binding assay were studied further. The assembly and stability of 66 putative peptide–HLA complexes were measured using complete rate assays (ProImmune Ltd). The fraction of assembled complexes was measured at six different time points over 48 h. After fitting the values to a one-phase association equation, the half-life needed to reach the maximum of assembled complexes was calculated and reported as the on-rate of the studied peptide–HLA complex. Stability of the complexes was measured using off-rate assays, which measure the fraction of denaturated complexes at six different time points over 24 h. These values were fitted to a one-phase denaturation equation and the half-life of complexes was then calculated. For on- and off-rate values, see Supplementary Table 8.

Computational prediction of HLA binding affinity and complex stability. Using published information[18,63–67], we compiled datasets of (1) 1,929 experimentally verified neopeptides, all of which are known to have high binding affinity to certain HLA alleles (Supplementary Table 2); (2) 212,090 HLA-I-bound self-peptides identified in a range of immunopeptidomics studies (Supplementary Table 4); and (3) 9,544 viral peptides (Supplementary Table 3) acquired from a dataset published by Ogishi and Yotsuyanagi[68]. The binding affinity of each peptide to the 67 HLA alleles was calculated using the NetMHCpan-4.0 algorithm[20]. Raw binding affinity values substantially vary across alleles[20,69], but can accurately estimate peptide repertoire size[12]. Therefore, as described above, peptide repertoire size—that is, percentage of the predicted binders from a fixed set of tested peptides—was calculated using a binding affinity threshold of 500 nM. The stability of complexes between HLA alleles and bound neopeptides was determined using the NetMHCstabpan-1.0 algorithm[70], with the option of excluding previous information on exact binding affinity values (half-maximal inhibitory concentration). The half-life of the peptide–HLA complex (in hours) was used as a proxy for peptide–HLA complex stability.

Calculation of peptidome diversity from mass spectrometry studies. The accuracy of *Pr* estimates was tested using mass spectrometry immunopeptidomes derived from monoallelic cell lines[16,17]. For each HLA class I allele with appropriate peptidome data, sequence diversity of the identified peptides was measured by the Shannon entropy index.

Analysis of ICI-treated and treatment-naïve patients. Data were collected on published cohorts of patients with melanoma and who were treated with CTLA-4 inhibitors[4,23,24], melanoma treated with PD-1 inhibitors[4] and NSCLC treated with PD-1 inhibitors[25]. Patients with zero follow-up time were excluded from the anti-PD-1-treated melanoma cohort. Information on HLA genotype, loss of HLA heterozygosity and cancer mutational burden in these patients was derived from ref. [4]. Treatment and mutational burden data on patients of TCGA were downloaded from the NCI Genomic Data Commons portal[71] and cBioPortal[72], respectively. Progression-free interval and HLA genotype data were acquired from refs. [73,74], respectively. The independent melanoma cohort in Extended Data Fig. 5a,b contains the ipilimumab-naïve, anti-PD-1-treated patients of ref. [28] and ipilimumab-treated patients with melanoma of TCGA. Genotype-level HLA promiscuity (or genotype *Pr*) was determined by calculating the mean of $\overline{D_{KL}}$ values across all *HLA-A*, *-B* and *-C* alleles of the patient and taking the reciprocal value. For patients with a single HLA allele with no appropriate allelic *Pr* data, a standard imputation technique was involved by replacing the missing *Pr* with the median of that variable for all other cases. This protocol has the benefit of not changing the sample median for *Pr*, and it enlarges the dataset, but it has no impact on the association between survival and *Pr*. Patients having at least two HLA alleles with unknown allelic Pr were excluded from the analysis.

Data on HED were derived from a study by Chowell et al.[5]. Survminer and Survival R libraries were used for statistical analysis of patient survival. The effect of genotype *Pr* on survival was tested with the log-rank test and by the ggsurvplot function of the Survminer R library, and the results were visualized by Kaplan–Meier curves. Patients in each cohort were classified into *Pr* groups using a fixed cutoff. This cutoff was determined after merging the three separate cohorts and with an established *P* value minimization approach implemented in the Survminer R library. Survival curves of cohorts classified into more than two groups were also tested for ordered differences, as implemented by the Survminer R library. Cox regression models were generated using the coxph function of the Survival R library. The proportional hazard assumption for the Cox regression model fit was tested with the cox.zph function of the Survival library. The forest plot for multivariate Cox models was generated with the forestmodel R library. Data on RECIST classification and clinical benefit were acquired from refs. [23–25,28] and the NCI Genomic Data Commons portal[71].

Calculation of DAI. Using published information[18,63–67] we compiled a dataset of experimentally verified neopeptides, all of which are known to have high binding affinity to certain HLA alleles (Supplementary Table 2). To reach the best prediction accuracy using the NetMHCpan-4.0 algorithm[20,32], the analysis focused on peptides with a length of nine amino acids. Using an established method[10], neopeptides with >50% sequence identity were excluded, resulting in 589 neopeptides. Similarly to a previous study[36], each neopeptide was aligned against the nonmutated human proteome using BLAST+ 2.10.0 (ref. [75]). The closest hit was retrieved, and BLOSUM62 sequence similarities were calculated using an established method[76]. The binding affinity of each neopeptide and the corresponding nonmutated human peptide to each of the 32 *HLA-A*, 50 *HLA-B* and 21 *HLA-C* alleles[77] was calculated using NetMHCpan-4.0 (ref. [20]). For each peptide–HLA-I pair, DAI values were calculated as previously reported[32]. For each HLA allele, the median DAI value was calculated across all neopeptides with significant binding to the given HLA variant. We used a 2% binding rank percentile cutoff to identify these neopeptides, because percentile rank minimizes the fraction of false-positive hits[20]. The use of percentile rank was crucial, because we aimed to focus on calculating DAI for the most reliable candidate peptides. In all other respects, our DAI calculation was fully consistent with that of an earlier study[32].

DAI values were also determined for neopeptides in 139 melanoma samples derived from two immunotherapy cohorts[23,24]. Genomic data on missense mutations in protein-coding genes were acquired from cBioPortal[72]. The corresponding mutated proteins were identified using the reference human proteome (downloaded from UniProt database[61], as of 9 January 2020). Using these protein data, all possible mutated peptide segments of length 8–12 amino acids were determined in each sample. Binding affinity and DAI values of each mutated peptide to each HLA class I allele of the patient were calculated as above.

Transcriptome analysis of melanoma cancer samples. Normalized RNA-sequencing expression data of cancer samples derived from patients with anti-CTLA-4-treated melanoma[23] and with established HLA-I genotype were downloaded from cBioPortal[72]. Immune deconvolution analysis was carried out following recommendations by Sturm and colleagues[44] and using the immunedeconv R library[44]. As recommended[44], we used the EPIC algorithm[78] for endothelial cells and cancer-associated fibroblasts, MCP-counter[79] for macrophages/monocytes and quanTIseq[80] for regulatory CD4$^+$ T cells. Precalculated TIDE dysfunction and exclusion score values were acquired from the TIDE server[50]. Gene set enrichment analysis was carried out with the GSEA 4.0.3 (ref. [81]) software using 1,000 permutations, a weighted enrichment statistic and exclusion of gene sets with fewer than five genes. Gene sets of biological processes as reported by the Gene Ontology Consortium[82] were included in the analysis. Gene sets enriched with $P < 0.001$ and false discovery rate (FDR) < 0.1 were treated as significant.

Statistics and reproducibility. Published data on established cohorts of cancer patients were used, and thus statistical analysis was not carried out to predetermine sample size (the number of individuals in each cohort is reported in the text and/or figures). Similarly, randomization of patients has already been carried out in previous studies. We used retrospective cohorts of patients with cancer and validated our findings on independent cohorts. We excluded patients with zero follow-up time from the anti-PD-1-treated melanoma cohort, since these could have biased the analysis. Patients having at least two HLA alleles with unknown allelic *Pr* were also excluded from every cohort in the study, since genome-level *Pr* values could not be accurately calculated for these patients. Statistical analyses and visualization were carried out using R v.3.6.3 (http://www.r-project.org) under the RStudio 1.2.5033 environment. Smooth curves on plots were fitted with the cubic smoothing spline method[83].

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

Data availability

Previously published clinical datasets that were reanalyzed here are available at refs. [4,23–25,28,54]. The following databases were used in the study: the Immune Epitope Database, https://www.iedb.org/; TANTIGEN database, http://projects.met-hilab.org/tadb/; the Cancer Genome Atlas, https://portal.gdc.cancer.gov/; cBioPortal, https://www.cbioportal.org/; UniProt database, https://www.uniprot.org/. Precalculated data on T-cell dysfunction were downloaded from the TIDE server, http://tide.dfci.harvard.edu/. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code availability

The code for HLA allele promiscuity calculation is available at https://github.com/matemanc/hla_promiscuity.

Received: 11 August 2020; Accepted: 19 May 2021;
Published online: 8 July 2021

References

1. Trowsdale, J. The MHC, disease and selection. *Immunol. Lett.* **137**, 1–8 (2011).
2. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
3. Verdegaal, E. M. E. et al. Neoantigen landscape dynamics during human melanoma–T cell interactions. *Nature* **536**, 91–95 (2016).
4. Chowell, D. et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).
5. Chowell, D. et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat. Med.* **25**, 1715–1720 (2019).
6. Negrao, M. V. et al. PD-L1 expression, tumor mutational burden, and cancer gene mutations are stronger predictors of benefit from immune checkpoint blockade than HLA class I genotype in non-small cell lung cancer. *J. Thorac. Oncol.* **14**, 1021–1031 (2019).
7. Litchfield, K. et al. Meta-analysis of tumor and T cell intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614 (2021).
8. Marty, R. et al. MHC-I genotype restricts the oncogenic mutational landscape. *Cell* **171**, 1272–1283 (2017).
9. McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* **171**, 1259–1271 (2017).
10. Manczinger, M. et al. Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol.* **17**, e3000131 (2019).
11. Kaufman, J. Generalists and specialists: a new view of how MHC class I molecules fight infectious pathogens. *Trends Immunol.* **39**, 367–379 (2018).
12. Paul, S. et al. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J. Immunol.* **191**, 5831–5839 (2013).
13. Košmrlj, A. et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* **465**, 350–354 (2010).
14. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).
15. Weiskopf, D. et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for $CD8^+$ T cells. *Proc. Natl Acad. Sci. USA* **110**, E2046–E2053 (2013).
16. Abelin, J. G. et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* **46**, 315–326 (2017).
17. Di Marco, M. et al. Unveiling the peptide motifs of HLA-C and HLA-G from naturally presented peptides and generation of binding prediction matrices. *J. Immunol.* **199**, 2639–2651 (2017).
18. Olsen, L. R. et al. TANTIGEN: a comprehensive database of tumor T cell antigens. *Cancer Immunol. Immunother.* **66**, 731–735 (2017).
19. Robbins, P. F. et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat. Med.* **19**, 747–752 (2013).
20. Jurtz, V. et al. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
21. Harndahl, M. et al. Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity: antigen processing. *Eur. J. Immunol.* **42**, 1405–1416 (2012).
22. Fromer, M. & Shifman, J. M. Tradeoff between stability and multispecificity in the design of promiscuous proteins. *PLoS Comput. Biol.* **5**, e1000627 (2009).
23. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
24. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
25. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
26. Mariathasan, S. et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
27. Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593 (2018).
28. Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934–949 (2017).
29. Sidney, J., Peters, B., Frahm, N., Brander, C. & Sette, A. HLA class I supertypes: a revised and updated classification. *BMC Immunol.* **9**, 1 (2008).
30. Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
31. Kaur, G. et al. Structural and regulatory diversity shape HLA-C protein expression levels. *Nat. Commun.* **8**, 15924 (2017).
32. Ghorani, E. et al. Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. *Ann. Oncol.* **29**, 271–279 (2018).
33. Rech, A. J. et al. Tumor immunity and survival as a function of alternative neopeptides in human cancer. *Cancer Immunol. Res.* **6**, 276–287 (2018).
34. Duan, F. et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211**, 2231–2248 (2014).
35. Łuksza, M. et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).
36. Richman, L. P., Vonderheide, R. H. & Rech, A. J. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Syst.* **9**, 375–382 (2019).
37. Xing, Y. & Hogquist, K. A. T-cell tolerance: central and peripheral. *Cold Spring Harb. Perspect. Biol.* **4**, a006957 (2012).
38. Nurieva, R., Wang, J. & Sahoo, A. T-cell tolerance in cancer. *Immunotherapy* **5**, 513–531 (2013).
39. Binnewies, M. et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
40. Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. Natural innate and adaptive immunity to cancer. *Annu. Rev. Immunol.* **29**, 235–271 (2011).
41. Noy, R. & Pollard, J. W. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* **41**, 49–61 (2014).
42. Walker, C., Mojares, E. & Del Río Hernández, A. Role of extracellular matrix in development and cancer progression. *Int. J. Mol. Sci.* **19**, 3028 (2018).
43. Henke, E., Nandigama, R. & Ergün, S. Extracellular matrix in the tumor microenvironment and its impact on cancer therapy. *Front. Mol. Biosci.* **6**, 160 (2019).
44. Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, i436–i445 (2019).
45. Haque, S. & Morris, J. C. Transforming growth factor-β: a therapeutic target for cancer. *Hum. Vaccin. Immunother.* **13**, 1741–1750 (2017).
46. Colak, S. & ten Dijke, P. Targeting TGF-β signaling in cancer. *Trends Cancer* **3**, 56–71 (2017).
47. Das, M., Zhu, C. & Kuchroo, V. K. Tim-3 and its role in regulating anti-tumor immunity. *Immunol. Rev.* **276**, 97–111 (2017).
48. Manieri, N. A., Chiang, E. Y. & Grogan, J. L. TIGIT: a key inhibitor of the cancer immunity cycle. *Trends Immunol.* **38**, 20–28 (2017).
49. Gautron, A.-S., Dominguez-Villar, M., de Marcken, M. & Hafler, D. A. Enhanced suppressor function of $TIM\text{-}3^+$ $FoxP3^+$ regulatory T cells: immunomodulation. *Eur. J. Immunol.* **44**, 2703–2711 (2014).
50. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
51. Xia, A., Zhang, Y., Xu, J., Yin, T. & Lu, X.-J. T cell dysfunction in cancer immunity and immunotherapy. *Front. Immunol.* **10**, 1719 (2019).
52. Scott, A. C. et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).
53. Radwan, J., Babik, W., Kaufman, J., Lenz, T. L. & Winternitz, J. Advances in the evolutionary understanding of MHC polymorphism. *Trends Genet.* **36**, 298–311 (2020).
54. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
55. Alspach, E. et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
56. Marty Pyke, R. et al. Evolutionary pressure against MHC class II binding cancer mutations. *Cell* **175**, 416–428 (2018).
57. Rolland, M. et al. Recognition of HIV-1 peptides by host CTL is related to HIV-1 similarity to human proteins. *PLoS ONE* **2**, e823 (2007).
58. McGranahan, N. & Swanton, C. Neoantigen quality, not quantity. *Sci. Transl. Med.* **11**, eaax7918 (2019).
59. Peng, M. et al. Neoantigen vaccine: an emerging tumor immunotherapy. *Mol. Cancer* **18**, 128 (2019).
60. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
61. The UniProt Consortium UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
62. Marcon, E., Hérault, B., Baraloto, C. & Lang, G. The decomposition of Shannon's entropy and a confidence interval for beta diversity. *Oikos* **121**, 516–522 (2012).
63. Bjerregaard, A.-M. et al. An analysis of natural T cell responses to predicted tumor neoepitopes. *Front. Immunol.* **8**, 1566 (2017).
64. Gros, A. et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* **22**, 433–438 (2016).
65. Zacharakis, N. et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat. Med.* **24**, 724–730 (2018).
66. Tran, E. et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **350**, 1387–1390 (2015).
67. Carreno, B. M. et al. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).
68. Ogishi, M. & Yotsuyanagi, H. Quantitative prediction of the landscape of T cell epitope immunogenicity in sequence space. *Front. Immunol.* **10**, 827 (2019).

69. Nielsen, M. & Andreatta, M. 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 (2016).
70. Rasmussen, M. et al. Pan-specific prediction of peptide–MHC class I complex stability, a correlate of T cell immunogenicity. *J. Immunol.* **197**, 1517–1524 (2016).
71. Grossman, R. L. et al. Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).
72. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data: figure 1. *Cancer Discov.* **2**, 401–404 (2012).
73. Liu, J. et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416 (2018).
74. Li, X. et al. Benchmarking HLA genotyping and clarifying HLA impact on survival in tumor immunotherapy. *Mol. Oncol.* https://pubmed.ncbi.nlm.nih.gov/33411982/ (2021).
75. Camacho, C. et al. BLAST^{+}: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
76. Carrasco Pro, S. et al. Microbiota epitope similarity either dampens or enhances the immunogenicity of disease-associated antigenic epitopes. *PLoS ONE* **13**, e0196551 (2018).
77. Sarkizova, S. et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
78. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *eLife* **6**, e26476 (2017).
79. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
80. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).
81. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
82. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
83. De Boor, C. *A Practical Guide to Splines* (Springer, 1978).
84. Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).

Acknowledgements

We thank T. Lenz for earlier discussions on this topic and for providing us with the raw data on HED values of HLA-I allele pairs. We also thank the anonymous reviewers for their insightful suggestions on the manuscript, and B. Győrffy. for his suggestions on survival analysis. The results here are in part based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga). We thank H. Ye for providing the HLA genotype data of TCGA patients. Similarly, we thank R. Marty, J. Font-Burgada and H. Carter for sending HLA genotype data of TCGA patients for earlier analyses. The study was supported by the following research grants: The European Research Council (H2020-ERC-2014-CoG, no. 648364 – Resistance Evolution, to C.P.); 'Célzott Lendület' Program of the Hungarian Academy of Sciences (LP-2017–10/2017, to C.P.); 'Frontline' Research Excellence Program of the National Research, Development and Innovation Office, Hungary (KKP-126506, to C.P.); EVOMER (GINOP-2.3.2–15–2016–00014, to C.P. and B.P.); MolMedEx TUMORDNS (GINOP-2.3.2–15–2016–00020 and GINOP-2.3.3–15–2016–00001, to C.P., and GINOP-2.3.2–15–2016–00026, to B.P.); and 'Frontline' Research Excellence Program of the National Research, Development and Innovation Office, Hungary (KKP-129814, to B.P.). The project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 739593 (to B.P., L.K. and M.M). M.M. was supported by the New National Excellence Program of the Ministry of Human Capacities (UNKP-20-5) and by the Bolyai János Research Fellowship of the Hungarian Academy of Sciences. L.A. was supported by the New National Excellence Program of the Ministry for Innovation and Technology (UNKP-20-2). G.M.B. was supported by the New National Excellence Program of the Ministry for Innovation and Technology (UNKP-19-3). B.K. was supported by the New National Excellence Program of the Ministry for Innovation and Technology (UNKP-20-4). B.T.P. was supported by the New National Excellence Program of the Ministry for Innovation and Technology (UNKP-20-3). This research work was conducted with the support of the Szeged Scientists Academy under the sponsorship of the Hungarian Ministry of Innovation and Technology (FEIF/433-4/2020-ITM_SZERZ). We thank D. Bokor for proofreading of the manuscript.

Author contributions

M.M., L.K., B.P. and C.P. undertook conceptualization. M.M., B.K., G.M.B., B.T.P. and L.A. were responsible for methodology. Software was provided by M.M. Formal analysis was performed by M.M., B.K., G.M.B., B.T.P. and L.A. Investigation was carried out by M.M., B.K., G.M.B., B.T.P., B.P. and C.P. Resources were provided by M.M., C.P. and B.P. Data curation was undertaken by M.M., B.K., G.M.B. and B.T.P. Writing of the original draft was done by M.M., B.P. and C.P. Writing, review and editing were performed by M.M., B.P. and C.P. M.M., L.K., C.P. and B.P. supervised the project. Project administration was carried out by M.M. and C.P. Funding acquisition was the responsibility of C.P. and B.P.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43018-021-00226-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43018-021-00226-4.

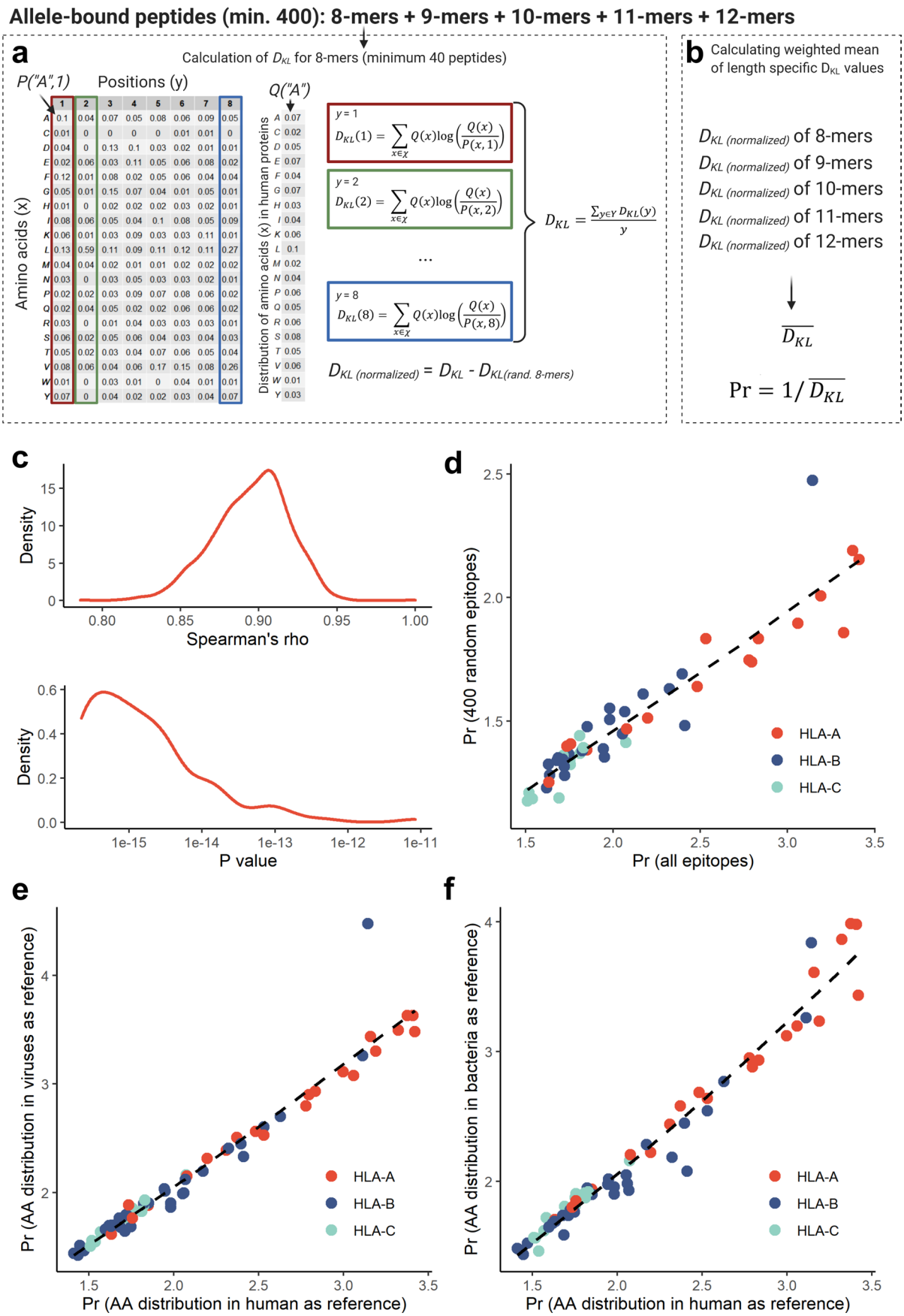
Correspondence and requests for materials should be addressed to M.M. or C.P.

Peer review information *Nature Cancer* thanks Pramod Srivastava and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

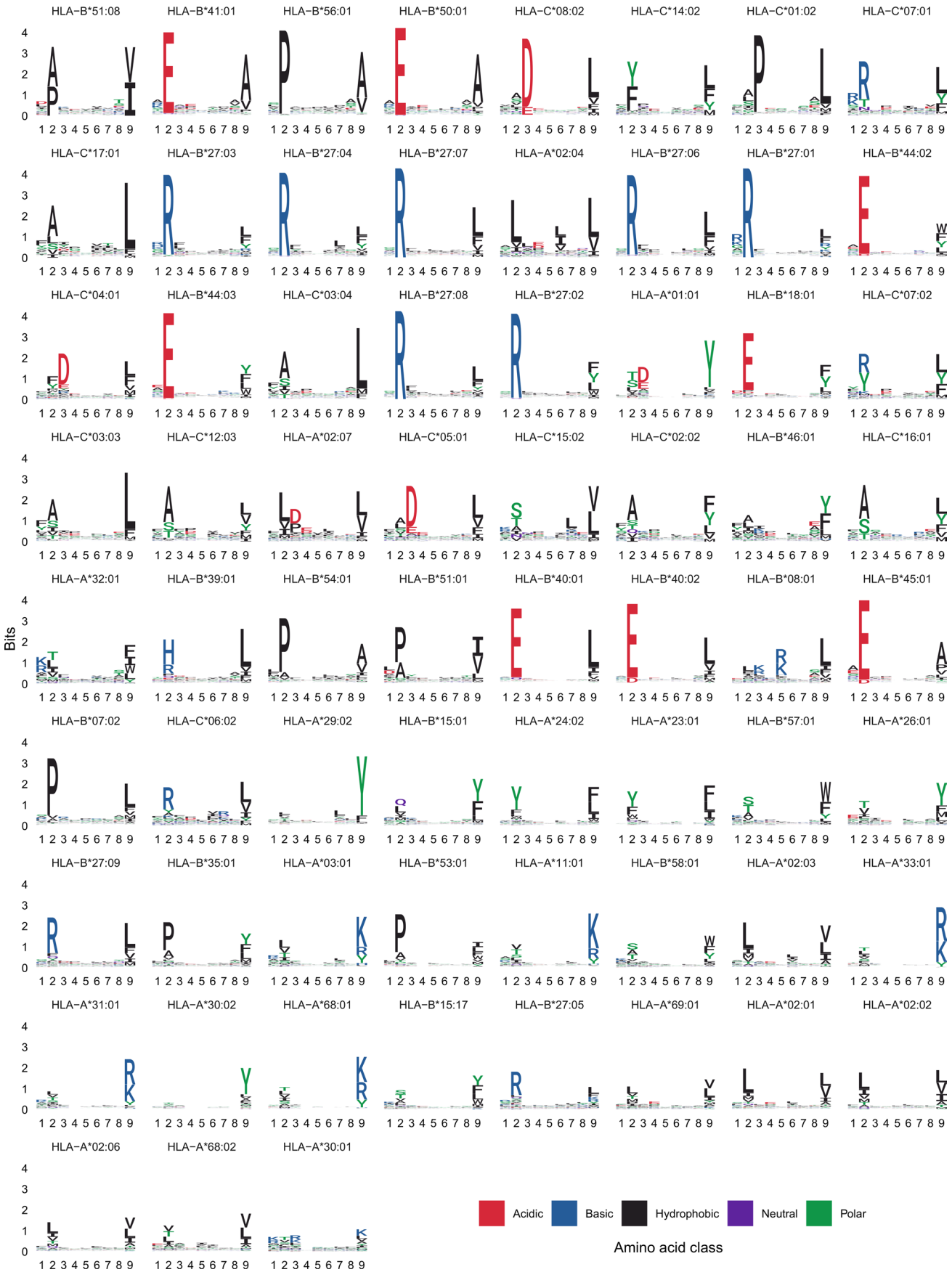
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021



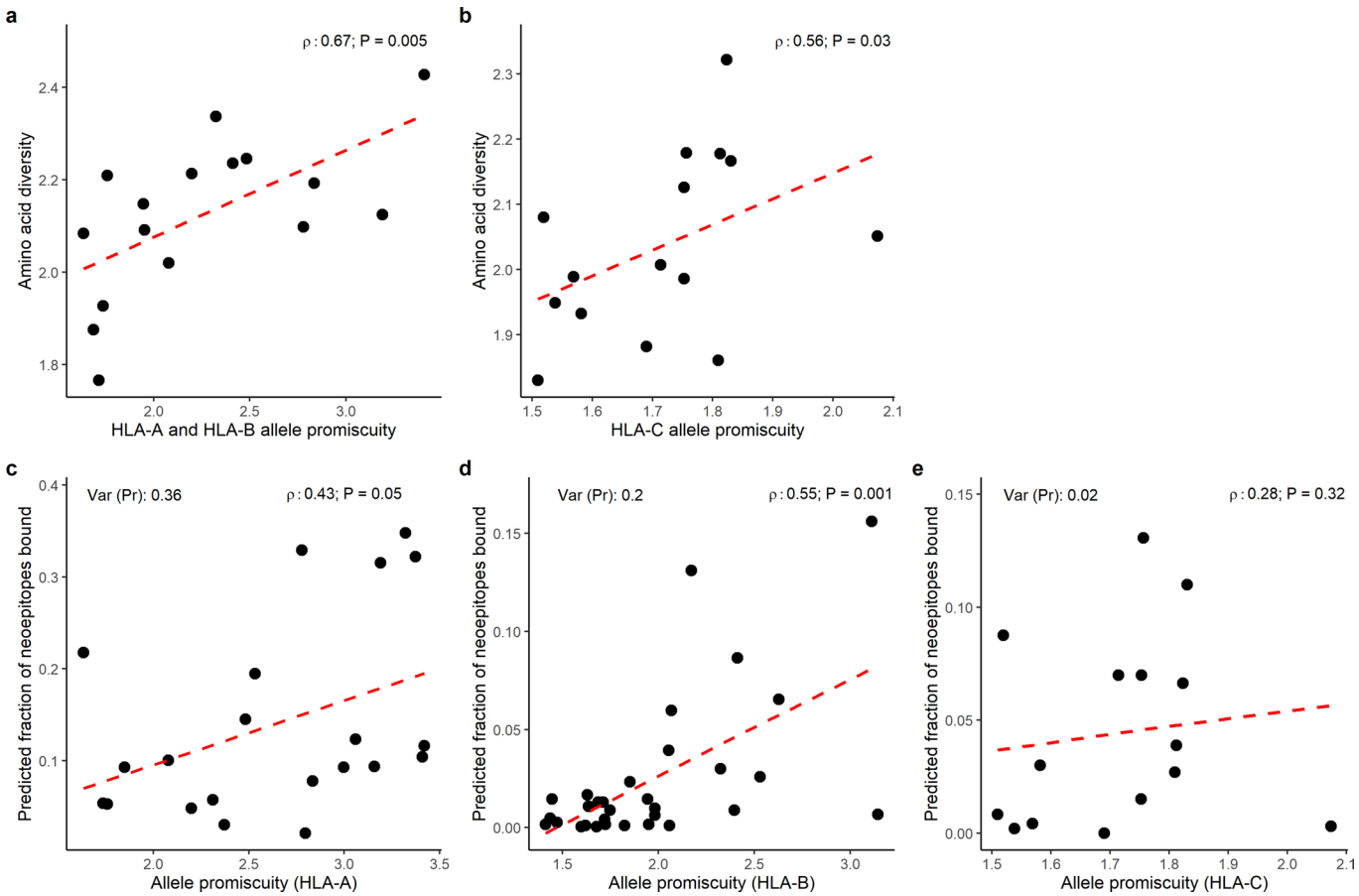
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Calculation and reliability of HLA allele promiscuity. a, D_{KL} values were calculated for each peptide length group separately (see Methods for details) and (**b**) the weighted mean of length-specific values was determined. **c**, **d**, The number of identified peptide sequences bound by a given HLA allele is variable, which may have a significant effect on the calculated Kullback-Leibler (KL) divergence values. To test whether KL divergence can be reliably determined using a minimal peptide sequence number of 400, we first selected HLA alleles with many (more than 1000) identified peptides ($n = 51$ alleles). After randomly selecting 400 sequences from the repertoire of each allele, we calculated KL divergence as explained in the Methods section. We iterated this process 1000 times and compared the *Pr* values with those calculated using the full dataset. Panel c indicates the distribution of Spearman's ρ values and the two-sided correlation test P values. **d**, An example plot indicating the relationship between promiscuity values calculated with either 400 random or all peptides bound by a given allele ($n = 51$ alleles). **e**, **f**, The calculated promiscuity values are independent of the reference amino acid (AA) distribution used during calculation (*Q* on panel **a**). The *Pr* values showed strong correlation when calculated using (**e**) human and viral or (**f**) human and bacterial reference AA distributions (Spearman's ρ: 0.99, two-sided correlation test $P < 2.2 \times 10^{-16}$ for both comparisons, $n = 67$ alleles on both panels). On plots **d** to **f**, dashed lines indicate a smooth curve fitted using cubic smoothing spline method in R (see Methods).

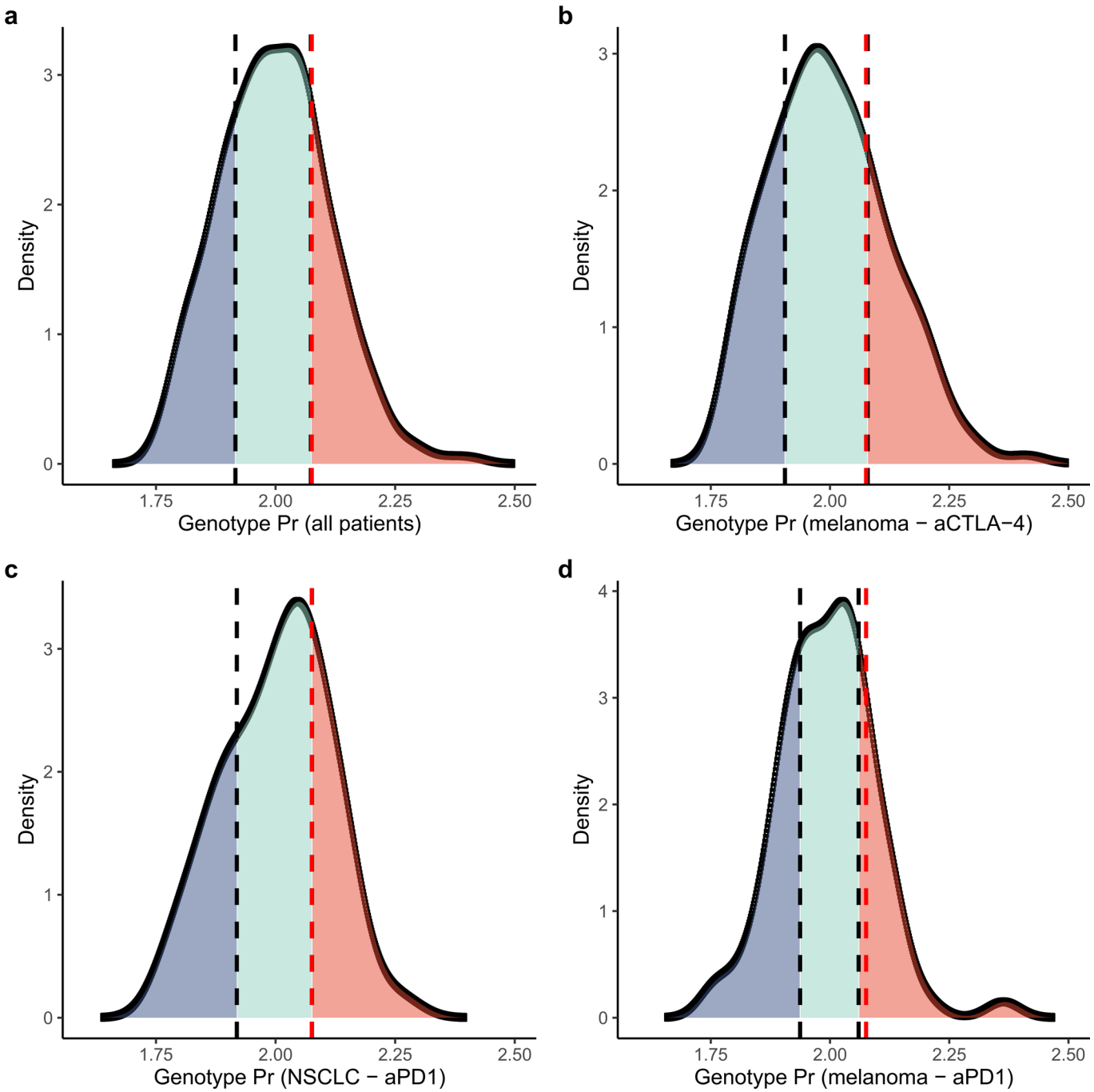


Extended Data Fig. 2 | See next page for caption.

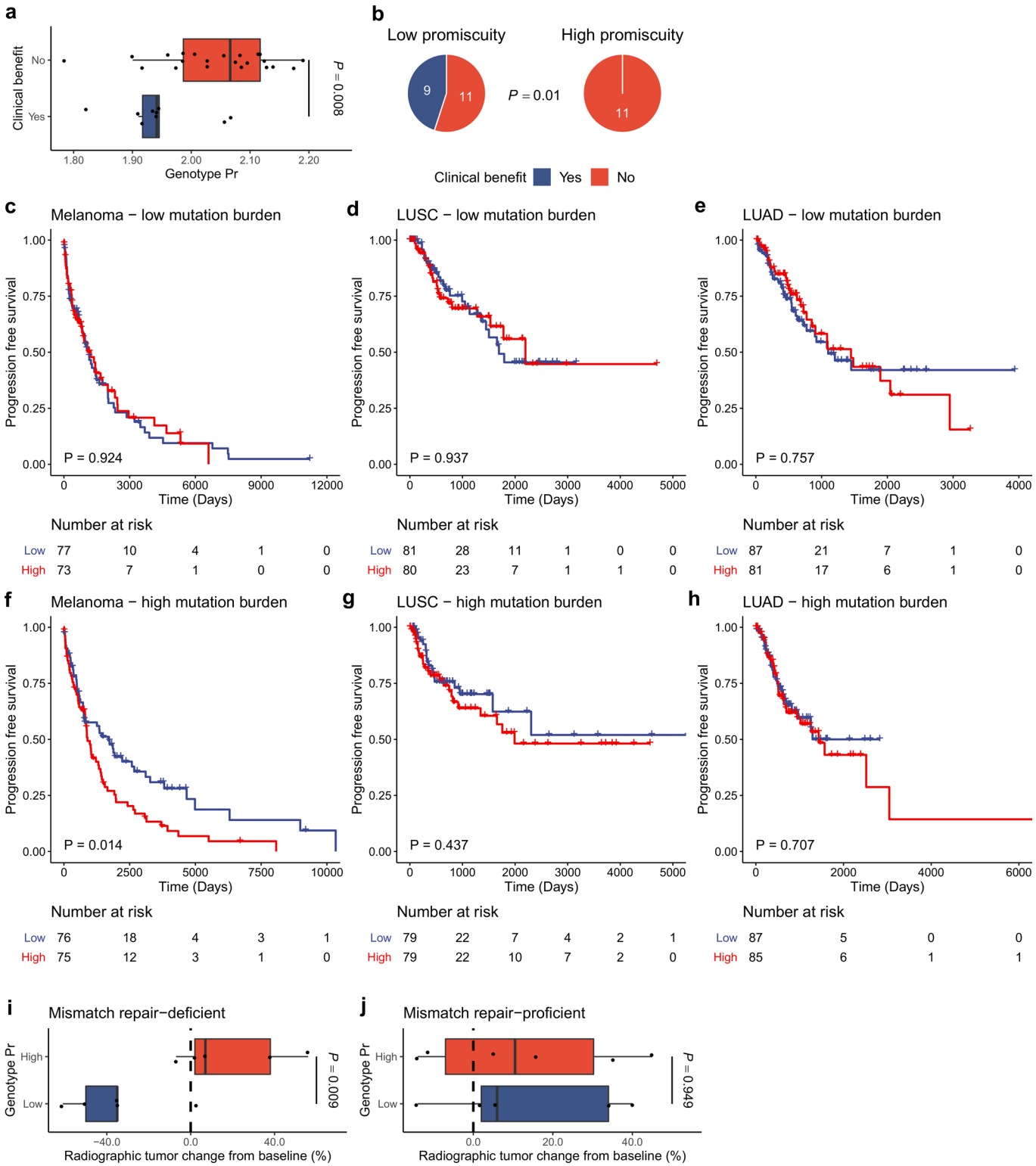
Extended Data Fig. 2 | Sequence diversity of peptides presented by HLA class I alleles. Amino acid preference was determined at different peptide positions bound by each of the 67 investigated HLA class I alleles. Sequence logos were visualized for 9 amino acid long peptides only[84]. Alleles are ordered according to their promiscuity level. The vertical axes indicate the information content of the given position in bits. The relative sizes of characters are proportional to their frequency at the given position.



Extended Data Fig. 3 | The relationship between HLA allele promiscuity and (a, b) the diversity of presented peptides or (c-e) fraction of bound neoepitopes. a, **b**, We found a strong correlation between the HLA allele promiscuity and sequence diversity of peptides eluted from the surface of monoallelic cell lines in studies by (**a**) Abelin et al. ($n=16$ alleles) and (**b**) Di Marco et al. ($n=15$ alleles). **c**-**e**, The relationship between the promiscuity of (**c**) *HLA-A* ($n=21$ alleles) (**d**) *HLA-B* ($n=31$ alleles) and (**e**) *HLA-C* ($n=15$ alleles) alleles and the predicted fraction of bound neopeptides. The variance of allele promiscuity is shown on panels **c** to **e**. On all plots, Spearman's ρ and the corresponding two-sided correlation test P values are indicated and dashed red lines indicate a linear regression line.

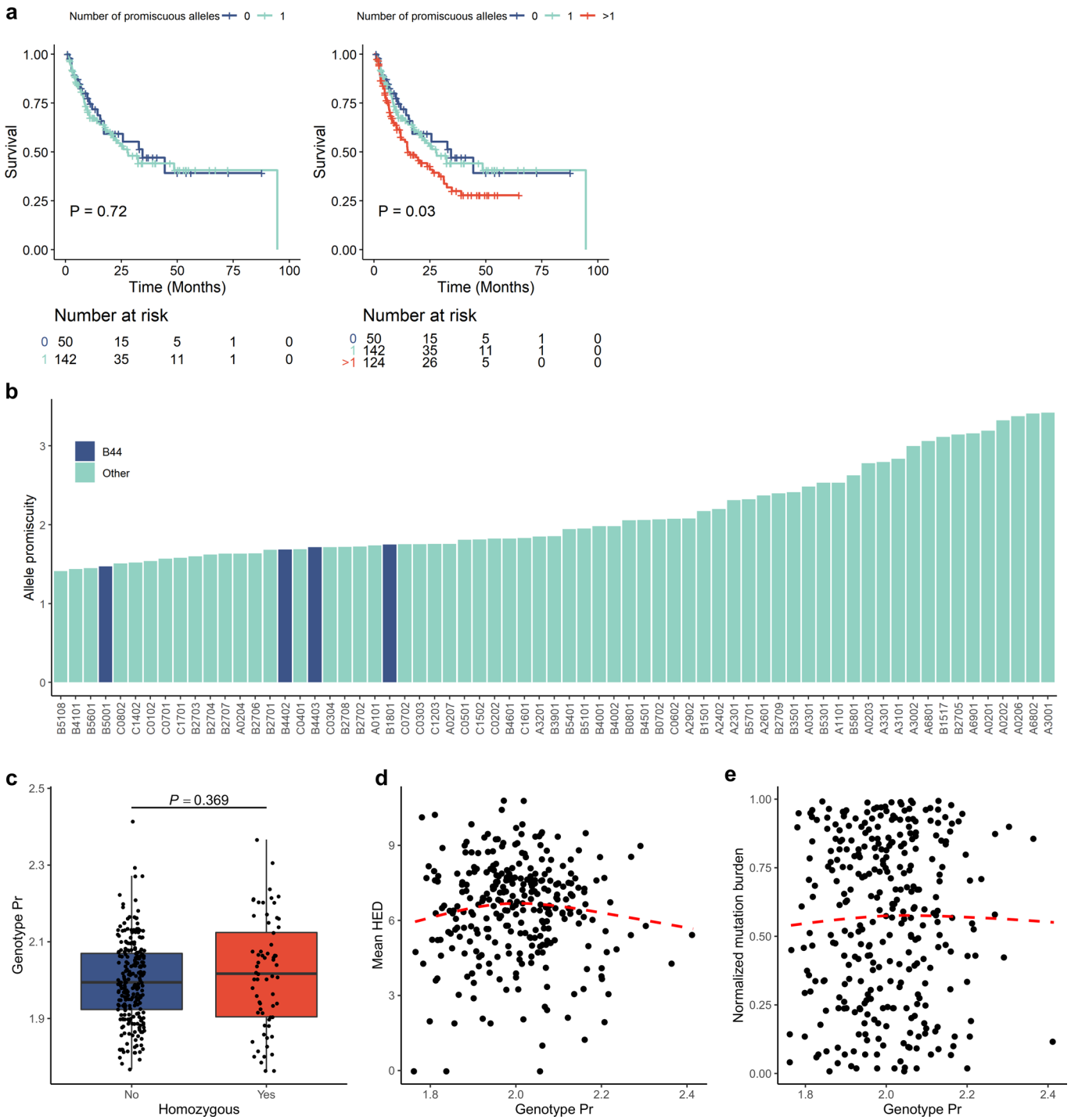


Extended Data Fig. 4 | Distribution of genotype *Pr* in cancer immunotherapy cohorts. The density of genotype *Pr* is shown for (**a**) all patients ($n=316$), (**b**) melanoma patients treated with CTLA-4 inhibitors ($n=164$), (**c**) non-small cell lung cancer (NSCLC) patients treated with PD-1 or PD-L1 inhibitors ($n=74$) and (**d**) melanoma patients treated with PD-1 or PD-L1 inhibitors ($n=78$). Dashed black vertical lines represent the 25th and 75th percentile values, while dashed red vertical lines represent the fixed genotype *Pr* cutoff of 2.076 used for patient classification.

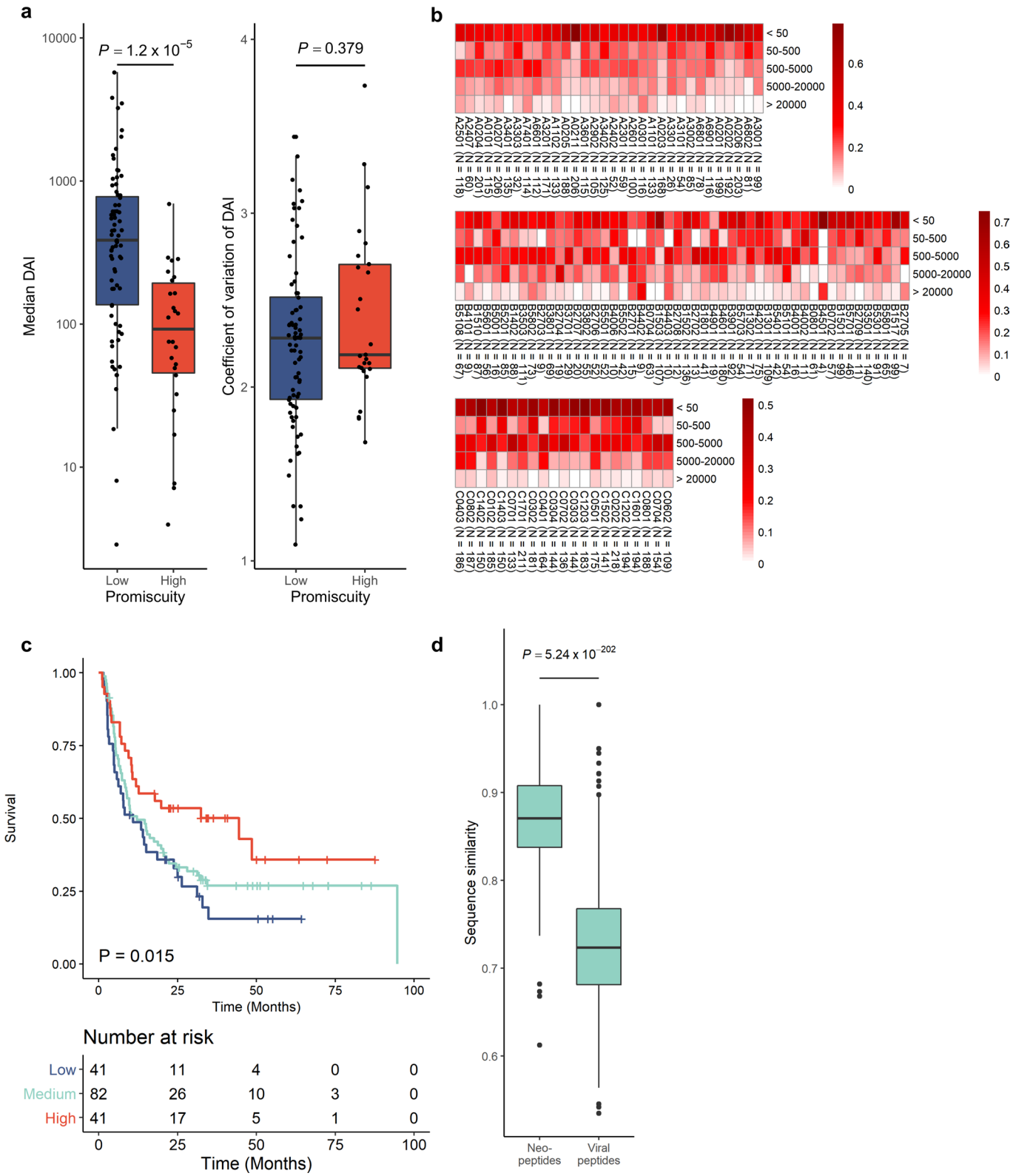


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | The effect of HLA promiscuity on response to ICI therapy and survival of treatment naïve patients. a, **b**, Response to ICI therapy in an independent melanoma cohort. Patients were stratified into low ($n=20$ patients) and high ($n=11$ patients) promiscuity groups using the same cutoff as in Fig. 3. On the boxplot, the P value of a two-sided Wilcoxon's rank-sum test is indicated. Vertical lines indicate median, boxes indicate the interquartile range (IQR), horizontal lines indicate 1st quartile - 1.5 × IQR and 3rd quartile + 1.5 × IQR. On the pie chart, the P value of a two-sided Fisher's exact test is shown. **c-h**, The relationship between HLA promiscuity and progression-free survival of treatment naïve melanoma (**c** and **f**) and NSCLC patients (**d**, **e**, **g** and **h**). Similarly to previous studies[5], patients were classified into low (**c** to **e**) and high (**f** to **h**) mutational burden cohorts using the median as a cutoff. In each cohort, patients were classified into low and high promiscuity groups using the median as a cutoff. Two-sided log-rank test P values are shown. The vertical axes indicate the probability of progression-free survival. LUSC: lung squamous cell carcinoma; LUAD: lung adenocarcinoma. **i**, **j**, The relationship between HLA-I genotype *Pr* and changes in tumor size after anti-PD-1 treatment in mismatch repair deficient (**i**, $n=10$) and proficient (**j**, $n=11$) cancer patients. Patients were stratified into low and high genotype *Pr* groups using the median value as a cutoff. P values indicate the outcome of two-sided two-sample T-tests. On boxplots, vertical lines indicate median, boxes indicate the interquartile range (IQR), horizontal lines indicate 1st quartile - 1.5 × IQR and 3rd quartile + 1.5 × IQR.

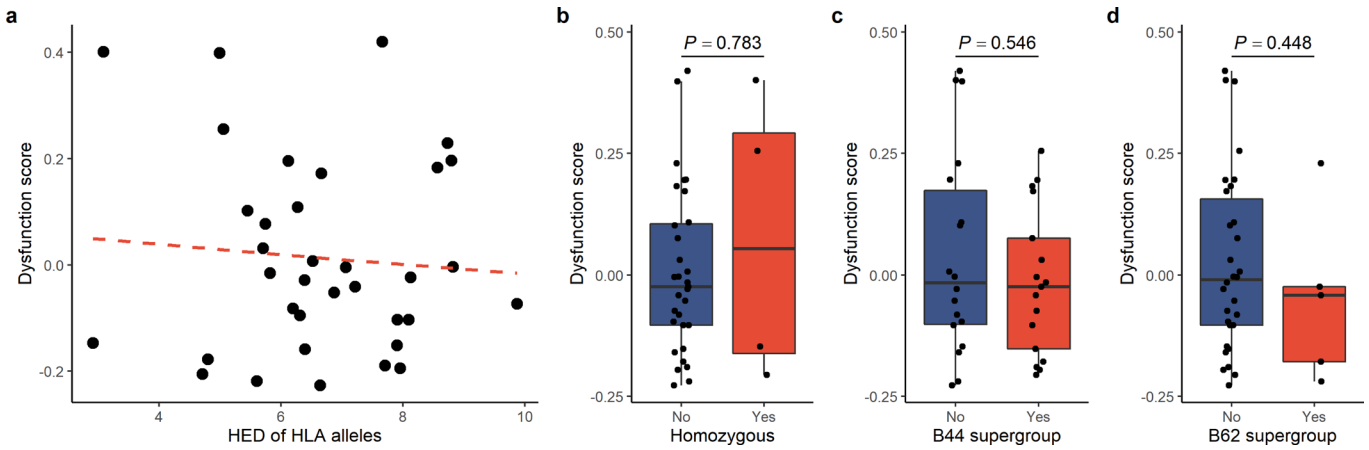


Extended Data Fig. 6 | HLA-I promiscuity is a major determinant of patient survival. a, Carrying a single promiscuous HLA-I allele has no impact on survival. The patients in Fig. 1a to c were classified based on the number of promiscuous alleles in their genotypes. Alleles with promiscuity levels in the top quartile were considered to be promiscuous. Two-sided log-rank test P values are shown. In the second analysis, between-group differences were also tested for trend (see Methods). The vertical axes indicate the probability of survival. **b**, HLA alleles belonging to the B44 supergroup have low promiscuity. B44 alleles are highlighted in blue and have significantly lower promiscuity than alleles not belonging to this group (two-sided Wilcoxon's rank-sum test P: 0.049). **c**-**e**, No significant association between HLA-I genotype *Pr* and (**c**) HLA homozygosity, (**d**) HLA-I evolutionary divergence (HED), (**e**) mutational burden. All patients from Fig. 3a to c were included in these analyses. **c**, There was no significant difference in genotype *Pr* between fully HLA heterozygous patients ($n = 251$) and the ones homozygous for at least one HLA-I locus ($n = 65$ patients). The P value for a two-sided Wilcoxon's rank-sum test is indicated. Similarly, there was no significant association between (**d**) genotype *Pr* and mean HED (Spearman's ρ: −0.02, two-sided correlation test P = 0.77, $n = 316$ patients) and (**e**) genotype *Pr* and tumor mutational burden of cancer immunotherapy patients (Spearman's ρ: 0.02, two-sided correlation test P = 0.77, $n = 316$ patients). Dashed red lines indicate a smooth curve fitted using the cubic smoothing spline method in R (see Methods). On the boxplot, horizontal lines indicate median, boxes indicate the interquartile range (IQR), vertical lines indicate 1st quartile - 1.5 × IQR and 3rd quartile + 1.5 × IQR.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | The relationship between genotype *Pr* and DAI. a, The median and variance of DAI in different allele promiscuity groups. Alleles were grouped based on the 3rd quartile as a cutoff ($n = 26$ and 77 alleles in the high and low promiscuity group, respectively). High allele promiscuity was associated with lower median DAI, but it had no significant effect on the coefficient of variation. P values of two-sided Wilcoxon's rank-sum tests are shown. On the boxplots, horizontal lines indicate median, boxes indicate the interquartile range (IQR), vertical lines indicate 1st quartile - 1.5 × IQR and 3rd quartile + 1.5 × IQR. **b**, The number and fraction of HLA-bound neopeptides in different DAI value ranges. The binding of 589 neopeptides to 103 HLA alleles was determined and DAI values were calculated for each neopeptide-human self-peptide pair. Alleles are shown in increasing order of promiscuity. For each allele, the fraction of peptide pairs belonging to different DAI categories is shown color-coded. **c**, The effect of genotype DAI on survival when determined as the average of allele-specific DAI values. Patients were stratified into groups using the 25th and 75th percentile of all values as cutoffs. A two-sided log-rank test P value is shown. Between-group differences were also tested for trend (see Methods). The vertical axis indicates the probability of survival. Higher genotype DAI value was associated with better survival. **d**, The maximum BLOSUM62 sequence similarity of neopeptides ($n = 589$) and immunogenic viral ($n = 1038$) peptides to the non-mutated human proteome. Only nine amino acid-long peptides were included in this analysis. P value of a two-sided Wilcoxon's rank-sum test is indicated. On boxplots, horizontal lines indicate median, boxes indicate the interquartile range (IQR), vertical lines indicate 1st quartile - 1.5 × IQR and 3rd quartile + 1.5 × IQR.



Extended Data Fig. 8 | The effect of HLA-associated features on TIDE dysfunction score. a, The TIDE dysfunction score is shown as a function of HLA evolutionary divergence. There was no relationship between the two variables (Spearman's ρ: −0.01, two-sided correlation test P = 0.96, *n* = 35 samples). Similarly, neither HLA homozygosity (**b**) nor the carrier status for B44 (**c**, *n* = 17 and 18 samples in B44 positive and B44 negative groups, respectively) or B62 (**d**, *n* = 5 and 30 samples in B62 positive and negative groups, respectively) supertype alleles affected the TIDE dysfunction score. P values indicate two-sided Wilcoxon's rank-sum test results. On plot **a**, the dashed red line indicates a linear regression line. On boxplots, horizontal lines indicate median, boxes indicate the interquartile range (IQR), vertical lines indicate 1st quartile - 1.5 × IQR and 3rd quartile + 1.5 × IQR.