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

INTRODUCTION

Key players of adaptive cellular immunity

The Human Leukocyte Antigen class I (HLA-I) molecule is a key player in the adaptive immune system. Its role is to bind peptides that are being produced in the cell. This protein complex is expressed on the surface of almost all nucleated cells in the human body.

Cytotoxic T cells can recognize HLA-I complexes binding foreign peptides. 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. This leads to the activation of the cytotoxic T cell and the subsequent elimination of the targeted host cell. The high specificity of TCR–pHLA binding reduces the likelihood of random immune activation.

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. The HLA region is the most variable part of our genome. Each allele has a characteristic specificity for peptides carrying different amino acid residues in key positions.

In conclusion to the diversity of HLA-I loci, 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. The importance of HLA-I heterozygosity has been shown to play a major role in protection against several viral infections, as it increases the number of presentable epitopes from a pathogen. Similar trends were reported for the adaptive immune recognition of tumors as well.

Adaptive cellular immunity targets the altered self in tumor cells

Cancer is a collection of diseases characterized by anomalous cell growth that has the capacity to invade or metastasize. One noteworthy characteristic of numerous tumor types is the infiltration of immune cells. Cytotoxic T can identify peptides presented by HLA-I molecules on malignant cells, and subsequently eliminate them.

A significant set of potential cancer rejection antigens, known as neoantigens, is composed of peptides that do not exist in the normal human genome. In several tumor types, neoantigens are generated by genomic alterations, including missense point mutations, deletions, and insertions.

Checkpoint molecules serve as regulators in the adaptive immune response and targets for therapy

Binding between pHLA complexes and specific TCRs is a prerequisite for the initiation of cellular adaptive immunity. However, the final decision between activation and inhibition of T lymphocytes is

dependent on the balance of additional positive and negative signals delivered through auxillary membrane receptors.

Immune checkpoint molecules, playing inhibitory roles in the initiation of an immune response have an importance in forming the tumor microenvironment and consequently, the outcome of cancer. CTLA-4, expressed by T cells soon after activation, provides an inhibitory signal that halts T cell responses, by interfering with TCR- and CD28-mediated signaling. The overall effect is a weaker proliferation of T cells, maintaining peripheral T cell tolerance.

PD-1, another member of the CD28 family, interacts with its ligands PD-L1 and PD-L2. The expression of PD-1 molecules is activation dependent, but not specific to T cells, as it has been detected on multiple other cell types, e.g., B lymphocytes and myeloid cells as well. Paradoxically, PD-1 molecules suppress immune activation by forming an excessively stable pHLA-TCR complex during chronic viral infection and cancer. This stabilization prevents the engagement of subsequent pHLA complexes and, consequently, activation of T cell response. The process results in a T cell exhaustion, aiming to protect the body from the toxicity associated with chronic immune activation. These immune checkpoint molecules have relevance in the clinical outcomes of cancer patients. Elevated CTLA-4 expression was associated with poor survival in various tumor types, such as nasopharyngeal carcinoma, thymoma, esophageal carcinoma and certain subtypes of non-small cell lung cancer (NSCLC) and breast

cancer. The high intra-tumor expression levels of PD-1 receptor, and its ligand PD-L1 is also shown to correlate with outcome in several cancer types.

In the last decade, monoclonal antibody-based medications targeting either CTLA-4 or PD-1/PD-L1 molecules have been approved. These are commonly referred to as immune checkpoint inhibitors (ICI). Ipilimumab, a human IgG1 antibody that inhibits CTLA-4, was the first drug of its kind to treat metastatic melanoma. It significantly improved the survival rates of patients who faced a high risk of death. In 2014, nivolumab, the first ICI antibody targeting PD-1 was introduced, which has become approved as a medication for melanoma, NSCLC, and a series of additional neoplasms.

Predictive factors determining ICI therapy efficacy

The number of non-responders to ICI therapy is still high, prompting extensive research into factors that can affect disease outcomes. Some of these are basic patient characteristics, including age, gender, and clinical history. Distinct molecular markers, such as the expression of PD-L1 by tumor or immune cells are predictive for therapy outcomes in certain cancer types. Recently, the impact of gut microbiome composition on ICI outcome has also gained attention.

Since the introduction of ICIs into clinical practice, the impact of genomic factors on therapy efficacy has become a key field of research. A shared characteristic among these factors is that they are mostly associated with the process of immune presentation.

Tumor mutation burden (TMB) is a major prognostic factor potentially affecting immunogenicity and ICI outcome. It is logical to assume that more altered self-proteins lead to higher probabilities for the formation of HLA-I complexes, presenting non-tolerated neoepitopes. TMB is one of the few biomarkers recognized by the FDA to predict tumor immunotherapy outcomes.

Various properties of germline HLA alleles/genotypes have also been associated with either positive or detrimental outcomes. *Chowell et al.* assumed that if a patient carries a diverse set of HLA molecules, it potentiates for 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. Importantly, later studies showed the lack of robustness for HLA-I homozygosity, HED and TMB as predictive biomarkers. Finally, genomic rearrangements may lead to HLA-I copy number loss which has a detrimental effect in lung cancer.

Binding promiscuity is an inherent property of HLA alleles

Based on the amino acid sequence of the binding pocket, distinct HLA-I proteins are specific for different peptide motifs. HLA-I variants have an intrinsic characteristic determining their peptide-binding specificity, discriminating between generalist and specialist HLA-I alleles. The metric, describing the breadth of peptides presented by an HLA molecule is called *peptide binding promiscuity*.

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 HLA-A promiscuity of local human populations. 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.

AIMS

In this work, we hypothesized 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 expected an increased chance for adaptive immune recognition in the case of more presentable neopeptides, thus enhancing the efficacy of antitumor immunity. According to our hypothesis, higher HLA-I genotype promiscuity (genotype Pr) should result in considerably better clinical outcomes in immune checkpoint inhibitor therapy.

RESULTS

Development of a valid HLA-I promiscuity definition

In a previous study, our group measured promiscuity levels of different alleles by measuring the relative fraction of a representative set of pathogen-derived peptides predicted to be bound by certain HLA molecules. We used the NetMHCpan 4.0 algorithm to predict the binding of neopeptides to a set of frequent alleles (prediction-based promiscuity). We utilized two sets of altered self-peptides based on mutations from an ICI-treated cohort ("4650 set") and the COSMIC database. The algorithm defines the binding strength by two metrics: IC50 binding affinity and binding rank. In the case of both, the developers published recommendation on the thresholds for weak and strong binding.

We found strong correlations between the promiscuity metrics measured for the two peptide sets in the same settings. Surprisingly, there were only weak correlations between promiscuity values calculated based on binding affinity and rank percentile values. These results suggest that binding affinity and binding rank may not be interchangeable metrics for assessing the peptide binding capacity of HLA-I alleles. Due to its definition, binding rank unifies the size of bound peptide repertoires, thus it is not able to reliably capture the differences in the breadths of presentable peptides. This suggests that affinity binding is a more accurate and reliable measure of promiscuity, compared to binding rank, even though based on a study

by Paul et al., using the same affinity cutoffs for several alleles to determine bound peptides is not appropriate.

To solve this problem, we aimed to establish a metric purely based on experimental data (Kullback-Leibler divergence-based promiscuity). 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. We calculated allele-level promiscuity values using Kullback-Leibler divergence, quantifying the amino acid diversity of peptides bound by HLA-I molecules. High promiscuity alleles have more diverse peptide motifs, suggesting that they can bind a broader range of peptides.

We validated our metric using data on naturally eluted self-peptides detected on the surface of HLA-I monoallelic cell lines. We found a strong positive association between promiscuity and peptidome diversity on the cell surface.

To select the most accurate promiscuity definition, we measured the *in vitro* binding affinities of 11 representative HLA-I alleles to 29 tumor neoepitopes using ProImmune REVEAL HLA-peptide binding assays. We found the strongest positive correlation in the case of the Kullback-Leibler divergence-based promiscuity definition. Conclusively, we used this metric to describe HLA-I allele promiscuity in the subsequent analyses.

High HLA-I genotype Pr is associated with worse survival among ICI patients

To identify the effects of mean genotype-level HLA-I promiscuity (genotype Pr) on cancer immunotherapy, we collected publicly available data about patients (in overall n = 316) treated with ICI.

In contrast to our expectations, high genotype Pr (defined using a unified threshold) was associated with reduced overall survival. Based on RECIST criteria patients with no clinical benefit were shown to have significantly higher genotype Pr.

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.

As genotype Pr is calculated as the arithmetic mean of values belonging to six alleles, it is important to test if the observed trends are not the effects of certain alleles with extreme values. We iteratively performed survival analysis for the main ICI cohorts, by excluding individuals carrying each specific HLA allele. We found that the association between HLA-I promiscuity and patient survival remained significant, which indicates that these trends cannot be explained by the effects of single alleles. Furthermore, patients carrying multiple promiscuous alleles showed decreased overall survival compared to subjects with zero or one promiscuous HLA-I variant.

Previously, the positive prognostic effect of the B44 HLA superfamily was reported after ICI treatment. In general, we found that allele-level promiscuity values are lower for B44 alleles compared to other alleles. Tumor mutational burden (TMB), HLA-I heterozygosity and HED have previously been suggested as determinants of response to ICI therapy. Thus, we decided to examine if genotype Pr is an independent predictive factor. We could not find a significant difference between genotype Pr values of homozygous and heterozygous patients. Furthermore, there were no correlations of patient-level promiscuity values either with mean HED or TMB. We also built a multiple Cox regression model, including genotype Pr, TMB, HLA-I heterozygosity and HED, to examine the effects of these variables on patient survival. Tumor mutation burden and HLA-I genotype Pr were the strongest factors in determining overall survival. Meanwhile, HED, HLA-I heterozygosity and the presence of B44 and B62 alleles had no significant effect on therapy outcome.

High HLA-I genotype Pr is associated with worse survival for non-ICI cancer patients as well

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 (TCGA) is a comprehensive dataset about 20,000 tumor and matched normal samples from 33 tumor types. Here, we analyzed data about melanoma and NSCLC patients.

We divided melanoma samples in the TCGA database into high- and low-mutational-burden (TMB) groups. In the case of this tumor type, we observed progression-free survival reduced by 38% in patients with high TMB and high genotype Pr group compared to the low Pr group. Meanwhile, no significant trends were visible for low TMB samples, suggesting a potential interaction between HLA-I promiscuity and TMB. In contrast, no impact of genotype Pr was shown for NSCLC patients.

Binding multispecificity does not compromise stability

Next, we aimed to further investigate the potential causes of reduced survival rates in patients with high genotype Pr. The stability of peptide–HLA complexes is essential for effective antigen presentation. It is a widely accepted view in enzymology that more promiscuous substrate binding negatively correlates with complex stability. Here, we examined if high HLA-I allele promiscuity is associated with a reduced capacity to form stable protein complexes with neopeptides

We utilized the *in vitro* ProImmune Complete rate assays to measure the assembly and dissociation rates of 66 representative pHLA complexes. Interestingly, we observed a significantly shorter time for complex assembly and a longer half-life of the resulting complexes in the case of high promiscuity HLA alleles. Using NetMHCstabpan, we predicted the complex stability for 67 HLA-I alleles and 1,929

neopeptides. *In silico* results also suggested increased levels of stability among promiscuous HLA-I molecules.

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

High allele Pr hampers discrimination between self and altered self

Tolerance might form towards those self-peptides which are presented by the HLA-I molecules of a subject. 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 tumor cells. 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.

Differential Agretopicity Index (DAI) measures the change in the predicted binding of a peptide as a result of an amino acid change (higher DAI corresponds to increased HLA-binding of the peptide). Based on a set of 589 experimentally verified neopeptides we calculated the median DAI for 67 HLA-I alleles. 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. We also observed similar trends on a patient level, for HLA-I genotypes and mutation-affected peptides of 139, anti-CTLA-4-treated melanoma subjects as well.

Higher median DAI was found to be positively associated with patient overall survival in this cohort.

Based on these results, the adaptive immune systems of patients with highly promiscuous HLA-I alleles have difficulties in discriminating between self and mutated self peptides.

High genotype Pr is associated with tolerogenic immune responses

CD8+ T cells can be inhibited through several pathways on the periphery. As highly promiscuous HLA-I alleles are less able to discriminate between original and altered self peptides, regulation of T cells may be shifted towards tolerogenic responses.

To identify the molecular and cellular properties of peripheral T cell tolerance in tumors, in the following analyses we utilized a dataset about patients with melanoma treated with ipilimumab (n = 30).

A gene set enrichment analysis showed that genes associated with the positive regulation of T-cell induction, type 2 immune response, extracellular matrix secretion, macrophage induction, genes associated with the negative regulation of T helper-1 cell-mediated immune response and CD4+ alpha–beta T-cell differentiation were upregulated in the high genotype Pr group. All these gene classes are associated with an immunosuppressive tumor microenvironment.

Immune deconvolution estimates the fractions of immune cells in a tumor biopsy using transcriptome data. We found an enrichment of regulatory T cells (T_{reg}), macrophages and monocytes, cancer-

associated fibroblasts, and endothelial cells in high genotype Pr tumors. These are indicators of an immunosuppressive environment. Additionally, two key genes related to immune tolerance, TGFB1 and FOXP3 had elevated expression levels in tumors of high genotype Pr patients. Similar trends were observable for immune checkpoint molecules PDCD-1 (PD-1), HAVCR2 (TIM3), TIGIT and CTLA-4.

T cells in tumors frequently become dysfunctional, which means they have a reduced proliferative capacity and effector function. We examined the expression of the most important markers for irreversible T cell dysfunction in the tumor samples. TOX, TOX2, T-bet and BLIMP1 genes had significantly higher expression in cancer samples with high genotype Pr. Using the more systematic TIDE (Tumor Immune Dysfunction and Exclusion) method, we showed that high genotype Pr is associated with increased T-cell dysfunction, but not with T-cell exclusion. This indicates that high HLA-I promiscuity does not prevent the infiltration of T cells directly but shapes the activity and functions of immune cells in the tumor microenvironment.

SUMMARY

In this work, we developed a metric to measure the diversity of peptides, presentable by HLA class I molecules. In contrast to the definition of HLA promiscuity from our previous publication, this metric is based solely on experimental data, not relying on HLA binding prediction results. We validated our promiscuity definition utilizing an *in vitro* HLA binding assays.

In contrast to our initial hypothesis, we found that high genotype Pr is associated with worse survival in both ICI therapy and ICI treatment-naïve patients. The results remain significant even when controlling for additional genomic properties, including HLA-I HED, homozygosity and TMB.

Next, we investigated the potential causes of these unexpected trends. First, we examined if high HLA-I allele promiscuity is associated with decreased complex stability. Surprisingly, peptide-HLA complexes involving promiscuous alleles show significantly higher stability.

We showed that the high promiscuity HLA molecules are less able to discriminate between self and altered self peptides, binding both of them. The low capacity of differentiation between original and mutated peptides is potentially associated with a response skewed towards immune tolerance. This finding has been confirmed through gene set enrichment analysis results, immune deconvolution methods and the examination of T cell dysfunction markers.

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 and my grandmother, 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