University of Szeged, Albert Szent-Györgyi Medical School
Department of Dermatology and Allergology
Doctoral School of Clinical Medicine

Interacting Mutational and Antigen Presentation Constraints Shape Tumor Immunogenicity

Dr. Benjamin Tamás Papp

Ph.D. thesis

Supervisor: Dr. Máté Manczinger

Szeged,

2026

Table of contents

Introduction 3

Research objectives 5

Materials and methods 6

Results 8

Amino acid substitution signatures define protein-level mutational landscapes. 8

Substitution landscapes predict tumor immune microenvironment. 9

Summary and discussion 11

Acknowledgement 12

Introduction

Cancer develops through the gradual accumulation of somatic mutations generated by diverse endogenous and environmental processes. These processes leave characteristic patterns in cancer genomes, commonly referred to as mutational signatures. Such signatures reflect the activity of specific DNA damage sources and repair mechanisms, including ultraviolet radiation, oxidative stress, and defects in DNA repair pathways. Over the past decade, the systematic characterization of mutational signatures has provided important insights into the origins of many cancers and has helped link particular exposures or biological mechanisms to observed genomic alterations. However, most mutational signature analyses operate at the nucleotide level and describe patterns of base substitutions without directly addressing the functional consequences of these mutations at the protein level.

In parallel with advances in cancer genomics, tumor immunology has increasingly highlighted the importance of tumor-derived neopeptides in shaping anti-tumor immune responses. Neopeptides arise from somatic mutations that alter protein sequences and generate peptides that can be presented on the cell surface by human leukocyte antigen (HLA) molecules. When recognized by T cells, these peptides can trigger immune responses against tumor cells. This concept has gained particular importance in the era of immune checkpoint blockade therapies, where the presence and characteristics of neopeptides are considered important determinants of treatment response. Despite this progress, the relationship between mutational processes that generate genomic alterations and the immunological properties of the resulting neoantigens remains incompletely understood.

A key limitation of current approaches is the conceptual gap between DNA-level mutational signatures and the biological properties of peptides that determine immune recognition. While nucleotide-based signatures capture the mechanisms generating mutations, immune recognition occurs at the level of amino acid sequences presented as peptides by HLA molecules. Consequently, mutations that appear similar at the nucleotide level may produce distinct amino acid substitutions with different biochemical characteristics and immunological consequences.

Different mutational processes can generate distinct substitution landscapes, potentially influencing the qualitative properties of neoapeptides produced by tumors. Certain substitutions may introduce substantial physicochemical changes or alter peptide motifs that affect HLA binding.

As a result, mutational processes may influence tumor immunogenicity not only by increasing the number of mutations, but also by biasing the types of amino acid changes that occur.

Importantly, immune recognition of tumor-derived peptides depends not only on the mutational landscape of the tumor but also on host genetic factors, most notably the HLA genotype of the patient. HLA molecules determine which peptides can be presented on the cell surface, thereby shaping the repertoire of potential neoantigens available for T cell recognition. Because different HLA alleles exhibit distinct peptide-binding preferences, the immunological impact of a given substitution landscape may vary depending on the patient's antigen presentation machinery. Consequently, the relationship between mutational processes and tumor immunogenicity cannot be fully understood without considering the interaction between tumor-derived peptide space and host-specific HLA constraints.

This thesis investigates tumor immunogenicity from this interaction-based perspective. It proposes that immune recognition emerges from the combined effects of tumor-intrinsic mutational processes and host-mediated antigen presentation. By introducing amino acid substitution signatures (AASs) as a protein-level representation of mutagenesis, the work establishes a framework linking mutational processes to neoantigen properties and immune phenotypes. Through integrative computational analyses and experimental validation, the thesis explores how substitution landscapes influence the immunopeptidome and immune recognition in cancer. Together, these findings support a model in which tumor immunogenicity is shaped by the compatibility between mutationally generated peptide repertoires and the antigen presentation capacity defined by the patient's HLA genotype.

Research objectives

The overall aim of this thesis was to investigate how the interaction between tumor-intrinsic mutational processes and host-specific antigen presentation shapes tumor immunogenicity. The work focuses on the qualitative properties of amino acid substitutions generated by different mutational processes and their compatibility with the patient's HLA genotype. To address this overarching goal, the specific objectives of the thesis were:

1. **To define protein-level representations of mutagenesis.** The first objective was to characterize recurrent and interpretable amino acid substitution signatures across human cancers, reflecting the cumulative outcomes of environmental exposures, endogenous mutational mechanisms, and DNA repair deficiencies.

2. **To link amino acid substitution patterns to neoantigen quality.** The thesis aimed to determine whether distinct substitution signatures bias the biophysical properties of amino acid changes, thereby influencing predicted neoantigen binding and immunogenicity. We seek to determine whether these associations manifest at the tumor sample level as distinct immune phenotypes reflecting altered antigen presentation and tumor–immune interaction.

3. **To assess the clinical relevance of amino acid substitution signatures.** The thesis sought to investigate whether substitution-defined mutational landscapes are associated with clinical outcome and response to immune checkpoint blockade, and whether they provide complementary information to established genomic and immunological biomarkers.

4. **To examine host–tumor interactions mediated by HLA genotype.** An important objective was to evaluate whether the immunological impact of tumor-derived amino acid substitution patterns depends on patient-specific HLA alleles, and whether certain HLA variants are preferentially suited to present peptides generated by specific substitution landscapes.

5. **To support the interaction model with experimental evidence.** Finally, the thesis aimed to provide experimental proof-of-concept that mutagen-induced amino acid substitution landscapes can elicit HLA-restricted T cell responses under compatible antigen presentation contexts, thereby supporting a causal link between mutational processes, HLA genotype, and immune recognition.

Materials and methods

To investigate the relationship between mutational processes, amino acid substitution patterns, and tumor immunogenicity, this study combined large-scale computational analyses of cancer genomics data with validation using external datasets and controlled experimental systems. The overall analytical framework integrated tumor mutational data, predicted antigen presentation, immune phenotype characterization, and functional validation.

Somatic mutation data from human tumors were primarily obtained from The Cancer Genome Atlas (TCGA). Only missense mutations were retained for downstream analyses, as these directly alter protein sequences and generate amino acid substitutions that may influence antigen presentation. For each tumor sample, nucleotide-level mutations were translated into corresponding amino acid substitutions to create a protein-level representation of mutational landscapes across cancer types. Clinical data and additional immune-related annotations associated with TCGA samples were also incorporated to enable the analysis of tumor immune characteristics and clinical outcomes.

To identify recurrent patterns in amino acid substitutions, amino acid substitution signatures (AASs) were extracted using non-negative matrix factorization (NMF). This approach decomposes the observed substitution frequencies across tumor samples into a limited set of signatures that represent distinct substitution patterns. The resulting AAS profiles were then used to characterize the substitution landscapes of individual tumors and to investigate their associations with known mutational processes, tumor types, and biological features.

To evaluate the potential immunological consequences of these substitution landscapes, computational analyses were performed to examine how AAS profiles influence tumor-derived peptide repertoires and predicted antigen presentation. The relationship between substitution patterns and predicted peptide–HLA binding properties was analyzed to assess whether specific substitution landscapes favor the generation of peptides compatible with HLA presentation. In addition, associations between AAS profiles and tumor immune microenvironment characteristics were examined using available immune cell infiltration estimates and gene expression–based immune signatures. Clinical relevance was further evaluated by analyzing the relationship between substitution landscapes and outcomes in cohorts of patients treated with immune checkpoint blockade.

Independent validation of the computational findings was performed using publicly available immunopeptidomics datasets containing experimentally identified HLA-bound peptides from cancer samples with matched genomic information. These datasets allowed the comparison of observed neoantigen repertoires with predicted substitution patterns, providing empirical support for the influence of amino acid substitution landscapes on peptides presented by HLA molecules.

Finally, experimental validation was conducted using in vitro models to test whether mutagen-induced substitution landscapes can generate HLA-restricted immune responses. Human lung cancer A549 cell lines expressing defined HLA alleles were exposed to chemical mutagenesis to induce somatic mutations and generate novel amino acid substitutions. Whole-genome sequencing was used to identify the resulting mutations, and peripheral blood mononuclear cells (PBMCs) from donors with defined HLA genotypes were co-cultured with the mutagenized tumor cells. T cell proliferation assays were used to measure immune activation, allowing assessment of whether mutagen-derived neoantigens could induce HLA-restricted T cell responses under compatible antigen presentation contexts.

Results

The results of this thesis establish a framework linking tumor mutational processes to immune recognition through amino acid substitution patterns. First, recurrent amino acid substitution signatures were identified across human cancers, providing a protein-level representation of mutational landscapes. Second, these substitution landscapes were shown to associate with tumor immune phenotypes and clinical characteristics. Finally, experimental validation demonstrated that mutagen-induced substitution patterns can generate HLA-restricted immune responses under compatible antigen presentation contexts.

Amino acid substitution signatures define protein-level mutational landscapes.

To investigate how mutational processes shape tumor-derived peptides, somatic mutations from cancer genomes were translated into amino acid substitutions and analyzed at the protein level. Using non-negative matrix factorization, five recurrent amino acid substitution signatures (AASs) were identified across human cancers. Each signature represents a characteristic pattern of amino acid changes reflecting the combined effects of mutational processes acting in tumors.

The identified AASs displayed distinct substitution profiles and were associated with specific tumor types and mutational contexts. Certain signatures were enriched in cancers exposed to well-known mutagenic factors, such as ultraviolet radiation, while others were linked to endogenous mutational mechanisms including oxidative damage or DNA repair deficiencies. These associations demonstrate that mutational processes leave recognizable footprints not only at the nucleotide level but also in the spectrum of amino acid substitutions produced in tumor proteins.

Importantly, the substitution patterns represented by individual AASs differed in their biochemical properties. Some signatures were characterized by substitutions that introduce substantial physicochemical changes in amino acid side chains, while others consisted of more conservative replacements. These differences have potential implications for the structure and immunological properties of peptides derived from mutated proteins. As a result, the substitution landscape of a tumor can influence the characteristics of the peptide repertoire generated from its mutations.

By summarizing tumor mutations at the level of amino acid substitutions, AASs provide an interpretable intermediate representation linking mutational processes to downstream biological

effects. This framework enables the investigation of how distinct mutational mechanisms shape the qualitative properties of tumor-derived peptides that are relevant for immune recognition.

Substitution landscapes predict tumor immune microenvironment.

Having identified amino acid substitution signatures across cancers, the next step was to investigate whether these substitution landscapes are associated with tumor immune phenotypes. Computational analyses revealed that tumors characterized by different AAS profiles exhibit systematic differences in predicted neoantigen properties and immune microenvironment features.

Distinct substitution landscapes were associated with variations in the predicted compatibility of tumor-derived peptides with HLA presentation. Certain substitution patterns generated amino acid changes more likely to produce peptides with favorable binding properties for HLA molecules, whereas others resulted in substitution profiles less compatible with antigen presentation. These differences suggest that mutational processes may influence tumor immunogenicity by shaping the types of peptides available for presentation to the immune system.

Consistent with this hypothesis, tumors enriched for specific substitution signatures displayed distinct immune microenvironment characteristics. Associations were observed between AAS profiles and indicators of immune activity, including estimated immune cell infiltration and gene expression–based immune signatures. These results indicate that the qualitative structure of tumor mutational landscapes can influence the immune context in which tumors develop.

The clinical relevance of these findings was further supported by analyses of cohorts of patients treated with immune checkpoint blockade. Tumors characterized by different substitution landscapes exhibited differences in clinical outcomes, suggesting that the properties of mutation-derived peptides may contribute to variability in therapeutic response. Together, these observations demonstrate that amino acid substitution signatures capture biologically meaningful features of tumor mutations that are linked to immune phenotypes and clinical behavior.

Amino acid substitution signature and HLA genotype jointly shape anticancer immune response.

To experimentally test whether mutagen-induced substitution landscapes can generate immunogenic peptides, in vitro experiments were performed using mutagenized tumor cells and T

cell proliferation assays. Human A549 lung cancer cell lines (HLA-Knockout, HLA-B*07:02, HLA-A*03:01) were exposed to the mutagen ethylnitrosourea (ENU), which induces somatic mutations and corresponding amino acid substitutions.

Whole-genome sequencing of mutagenized clones confirmed the presence of numerous ENU-induced missense mutations, generating novel amino acid substitutions in tumor proteins. These mutated cells were then co-cultured with peripheral blood mononuclear cells (PBMCs) obtained from donors with known HLA genotypes in order to assess immune activation.

T cell proliferation assays demonstrated that ENU-mutagenized tumor cells expressing the HLA-B07:02 allele induced robust T cell responses when co-cultured with PBMCs derived from donors carrying the same HLA allele. In contrast, no comparable proliferation was observed when either the antigen-presenting tumor cells or the responding lymphocytes lacked the corresponding HLA allele. These results indicate that the observed immune activation was both mutation-dependent and HLA-restricted.

This experimental system provides proof of principle that mutagen-induced amino acid substitution landscapes can generate neoantigens capable of triggering immune responses when presented in a compatible HLA context. The findings therefore support the interaction-based model proposed in this thesis, in which tumor immunogenicity emerges from the combined effects of mutational processes generating peptide diversity and host-specific antigen presentation constraints determining which peptides can be effectively recognized by the immune system.

.

Summary and discussion

This thesis introduces amino acid substitution signatures (AASs) as a conceptual and analytical framework linking tumor mutational processes to immune recognition. While previous studies have primarily characterized mutagenesis at the nucleotide level, immune recognition occurs at the level of peptides derived from proteins. By translating somatic mutations into amino acid substitution patterns, AASs provide a protein-level representation of mutational processes that is directly relevant to the generation of tumor-derived peptides. The identification of recurrent substitution signatures across cancers demonstrates that mutational processes produce structured amino acid substitution landscapes that extend beyond simple mutation counts.

These findings have important implications for the interpretation of tumor immunogenicity. Tumor mutational burden (TMB) has emerged as a widely used biomarker for predicting response to immune checkpoint blockade, based on the assumption that higher numbers of mutations increase the probability of generating neoantigens. However, the results presented here indicate that the qualitative properties of mutations also play a critical role. Different mutational processes produce distinct substitution patterns that vary in their biochemical characteristics and in their compatibility with antigen presentation. As a result, tumors with similar mutation burdens may differ substantially in the types of peptides generated from their mutations and therefore in their capacity to stimulate immune responses.

The results further support an interaction-based model of tumor immunogenicity in which immune recognition emerges from the combined effects of tumor-derived peptide diversity and host-specific antigen presentation constraints. Mutational processes shape the repertoire of amino acid substitutions available in tumor proteins, while the patient's HLA genotype determines which peptides from this repertoire can be presented to T cells. The experimental findings presented in this thesis provide proof of principle that mutagen-induced substitution landscapes can generate immunogenic peptides capable of triggering HLA-restricted immune responses in compatible antigen presentation contexts.

Several limitations should be considered when interpreting these findings. The analyses primarily focus on missense mutations and therefore capture only one class of genomic alterations contributing to neoantigen formation. In addition, the computational analyses represent static snapshots of tumor genomes and immune environments, whereas tumor–immune interactions are

dynamic processes that evolve during tumor progression and treatment. Despite these limitations, the framework developed in this work provides a mechanistic perspective on how mutational processes influence tumor immunogenicity and highlights the importance of integrating tumor mutational landscapes with host antigen presentation genetics when studying immune responses to cancer.

Acknowledgement

I would like to express my deepest gratitude to my mentor, Dr. Máté Manczinger, whose scientific vision, intellectual rigor, and unwavering support made this work possible. His ability to combine conceptual clarity with critical thinking has profoundly shaped my development as a researcher. I am especially grateful for his patience throughout this journey, during which my excessive enthusiasm undoubtedly tested its limits on more than one occasion.

I am equally thankful to Prof. Dr. Lajos Kemény for his vocational guidance and for providing the academic environment in which this work could flourish. His trust and support were essential foundations for this thesis.

I would like to thank Szilvia Juhász, co–first author of our original article, whose wet-lab expertise was indispensable in translating computational hypotheses into biological insights. I am grateful to my colleagues — Anna Tácia Fülöp, Dávid Kókai, Franciska Dóra Spekhardt, Balázs Koncz, Gergő Mihály Balogh, Balázs Szücs, and Csaba Pál — for their collaboration, discussions, and contributions that strengthened this work in many ways.