Interacting Mutational and Antigen Presentation Constraints Shape Tumor Immunogenicity

Dr. Benjamin Tamás Papp

Ph.D. Thesis

Szeged

2026

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

Co-author certification

I, myself as a corresponding author of the following publication, declare that the authors have no conflict of interest, and Dr. Benjamin Tamás Papp, Ph.D. candidate, had a significant contribution to the jointly published research. The results discussed in his thesis were not used and not intended to be used in any other qualification process for obtaining a PhD degree.

Szeged, 2026.03.10. ....

Dr. Máté Manczinger

The publication relevant to the applicant's thesis:

Juhász, S., Papp, B.T., et al. Five dominant amino acid substitution signatures shape tumour immunity. Mol Syst Biol (2026). https://doi.org/10.1038/s44320-026-00193-x

- IF: 7.7
- SJR Scopus - Agricultural and Biological Sciences (miscellaneous): D1

Table of contents

Introduction

Mutational processes leave interpretable footprints in cancer genomes.

Cancer development is driven by the gradual accumulation of somatic mutations arising from diverse endogenous and exogenous mutational processes. These processes include environmental exposures such as ultraviolet radiation and tobacco-derived carcinogens, as well as endogenous sources of DNA damage and ageing that occur during normal cellular physiology[1–4]. In addition, defects in DNA repair pathways can substantially modify both the rate and spectrum of mutations, amplifying or reshaping the mutational landscape of tumor genomes[5,6].

A key insight from large-scale cancer genome sequencing efforts has been that mutational processes leave characteristic and reproducible patterns in cancer genomes, commonly referred to as mutational signatures[1,3,4,7]. These signatures are defined by specific combinations of nucleotide substitutions and sequence contexts and can often be linked to known biological mechanisms or environmental exposures. For example, ultraviolet radiation induces a characteristic enrichment of C>T substitutions at dipyrimidine sites in skin cancers, while tobacco smoking is associated with a distinct pattern dominated by C>A transversions in lung tumors[8]. Similarly, endogenous enzymatic activities and DNA repair deficiencies give rise to their own signature patterns, reflecting the underlying biochemical processes that generate or fail to repair DNA damage[9,10].

The systematic identification and cataloguing of mutational signatures across cancer types has provided valuable insights into tumor etiology and evolution. By decomposing complex mutation profiles into constituent signatures, these approaches have enabled the reconstruction of mutational histories, the identification of dominant carcinogenic exposures, and the stratification of tumors based on underlying biological processes[11–13]. Importantly, mutational signatures have also been linked to therapeutic vulnerabilities, most notably in the context of DNA repair deficiencies, where tumors with defective mismatch repair or homologous recombination pathways exhibit characteristic mutational patterns and distinct clinical behavior[14,15].

Despite these advances, the biological consequences of different mutational processes extend beyond their contribution to mutation accumulation and genome instability. In particular, how distinct mutational processes influence downstream cellular phenotypes remains incompletely understood. While mutational signatures provide a descriptive framework at the DNA level, they do not directly address how different types of mutations affect protein composition, cellular function, or interactions with the tumor microenvironment.

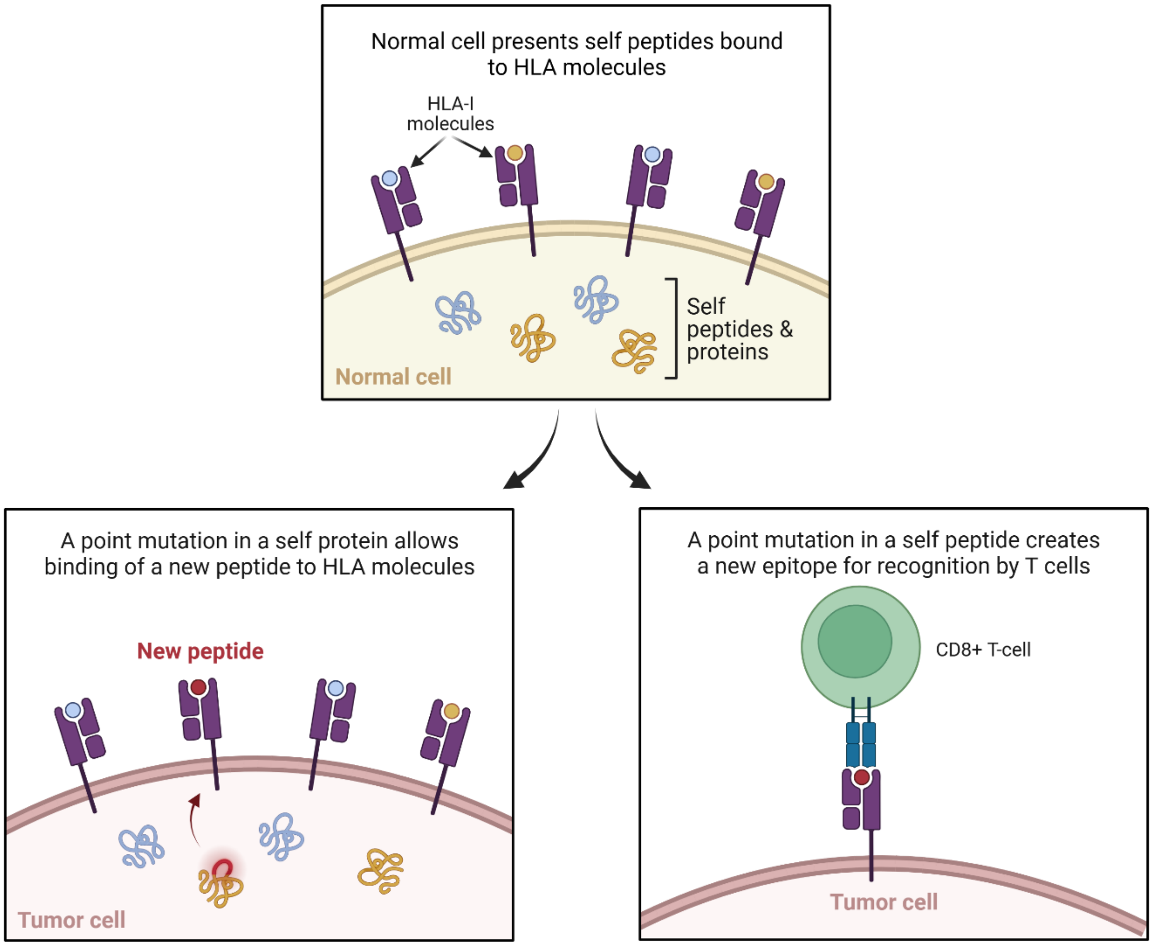
One area where this limitation is particularly evident is tumor immunity. Emerging evidence suggests that certain mutational processes, for example, APOBEC3-related signatures, may be associated with differences in immune infiltration or response to immunotherapy. Yet, these associations have been explored only in a limited number of specific contexts. [8,16–18] Moreover, the mechanistic basis underlying such observations remains unclear, as mutation burden alone does not fully explain the heterogeneity of immune phenotypes observed across tumors with similar numbers of somatic mutations.

Together, these observations highlight a gap between the descriptive power of mutational signatures and the need for biologically interpretable representations of mutagenesis that can be directly linked to functional and immunological consequences. Addressing this gap requires frameworks that connect mutational processes to protein-level changes and, ultimately, to tumor–immune interactions.

From mutation to immune recognition.

Somatic mutations in tumor cells can alter protein-coding sequences, leading to amino acid changes absent from the normal proteome. When such altered proteins are expressed, they can be processed into peptide fragments that are displayed on the tumor cell surface in complex with human leukocyte antigen (HLA) molecules[19]. The mutation-derived neopeptides that the immune system can recognize are commonly referred to as neoantigens. Neoantigens play a central role in antitumor immune recognition and constitute key targets of T cell–mediated immune responses in cancer[20].

The generation of neoantigens is a multistep process that links genomic alterations to immune surveillance. Intracellular proteins are continuously degraded, predominantly by the proteasome, generating a pool of peptides that can be transported into the endoplasmic reticulum and loaded onto HLA molecules. Only a subset of these peptides is stably presented on the cell surface, where T cell receptors on cytotoxic and helper T cells can recognize them[21,22]. Each step of this pathway imposes biochemical constraints on which mutation-derived peptides ultimately become visible to the immune system (Introduction Figure 1).


Normal cell presents self peptides bound to HLA molecules
HLA-I molecules
Self peptides & proteins
Normal cell
A point mutation in a self protein allows binding of a new peptide to HLA molecules
New peptide
Tumor cell
A point mutation in a self peptide creates a new epitope for recognition by T cells
CD8+ T-cell
Tumor cell

Introduction Figure 1. The immunological recognition of neopeptides. In normal cells, HLA class I molecules present self-derived peptides at the cell surface. In tumor cells, somatic mutations can generate altered peptides (neopeptides) that bind to HLA molecules and are displayed for immune surveillance on the cell surface. If recognized as non-self by CD8$^+$ T cells, these neopeptide–HLA complexes can initiate antitumor immune responses.

A critical determinant of neoantigen presentation is the binding specificity of HLA molecules. HLA class I and class II proteins possess highly polymorphic peptide-binding grooves that impose allele-specific preferences for peptide length and amino acid composition. As a consequence, different HLA alleles favor distinct sets of peptides, and only peptides with sufficient binding affinity can be effectively presented at the cell surface[21–24]. This allele specificity introduces a strong host-dependent component into tumor antigen presentation, such that identical tumor-derived mutations may lead to immunogenic peptides in some individuals but not in others.

The extensive polymorphism of HLA genes across human populations results in substantial inter-individual variability in antigen presentation capacity. This diversity has important implications for antitumor immunity, as it shapes the repertoire of neoantigens that can be displayed and recognized in each patient. Consistent with this notion, multiple studies have demonstrated associations between germline HLA genotype and immune-related tumor features, including response to immune checkpoint blockade and overall survival across cancer types[19,25–29]. These observations highlight that effective immune recognition depends not only on the presence of tumor-derived mutations but also on their compatibility with the host's antigen presentation machinery.

Tumor mutational burden (TMB) is widely used in the literature as a surrogate marker of the potential immunogenicity of tumors, based on the premise that a higher number of somatic mutations increases the likelihood of generating neoantigens capable of eliciting T-cell responses[30,31]. However, TMB alone does not fully explain the heterogeneity of immune responses observed across tumors. Tumors with comparable numbers of somatic mutations can exhibit markedly different immune phenotypes, ranging from highly inflamed to immune-depleted states. This suggests that qualitative features of neoantigens, such as peptide composition, binding properties, and presentation efficiency, play a crucial role in shaping antitumor immunity[19,20,26] Consequently, understanding immune recognition requires consideration of both tumor-intrinsic factors that generate neoantigens and host-specific factors that determine which of these antigens are presented.

Together, these principles establish neopeptide processing and HLA-mediated presentation as a central mechanistic link between tumor mutational landscapes and immune recognition. They provide a biological framework for investigating how different types of mutations, beyond their number alone, may influence the immunogenic potential of tumors.

Mutational processes influence neoantigen quality, not only quantity.

The immunogenic potential of tumors has frequently been approximated by the total number of somatic mutations or predicted neoantigens[31–33]. Tumor mutational burden has been shown to correlate with response to immune checkpoint blockade in several cancer types, supporting the notion that increased mutation counts raise the likelihood of generating immunogenic targets for T cells[34–36] However, mutation burden alone does not fully account for the heterogeneity of immune phenotypes and clinical outcomes observed across tumors.

Notably, tumors with comparable mutational burdens can exhibit markedly different levels of immune infiltration, immune activation, and therapeutic response. This discrepancy suggests that factors beyond mutation quantity contribute to effective immune recognition. Indeed, multiple studies have highlighted that the relationship between mutation burden and immunogenicity is context-dependent and influenced by additional tumor-intrinsic and host-related variables[37–39].

One explanation for this variability is that distinct mutational processes generate different types of nucleotide substitutions, which in turn translate into systematic biases in amino acid changes. Environmental exposures, endogenous mutagenic activities, and DNA repair deficiencies do not introduce random mutations but rather favor specific substitution patterns, shaping the resulting protein sequences in characteristic ways[4,7]. These biases imply that tumors with similar mutation counts may differ substantially in the composition of their mutation-derived peptides.

Such differences are particularly relevant for immune recognition, as peptide composition strongly influences antigen processing, HLA binding, and T cell receptor engagement[40–42]. The physicochemical properties of amino acid substitutions, including charge, hydrophobicity, and size, affect whether mutation-derived peptides can be stably presented by HLA molecules and recognized by T cells[20,21]. Consequently, not all mutations have equal potential to generate immunogenic neoantigens, even when they occur at similar frequencies.

Evidence supporting process-specific effects on tumor immunity has begun to emerge. For example, APOBEC-associated mutagenesis has been linked to immune activation and favorable responses to immune checkpoint blockade in several cancer types, although the underlying mechanisms remain debated[8,43,44]. In contrast, other mutational processes have been associated with immune-depleted tumor microenvironments and poorer clinical outcomes, despite high overall mutation burdens[45]. These observations indicate that the qualitative nature of mutations may be as important as their quantity in shaping antitumor immune responses.

Together, these findings motivate a shift from purely quantitative metrics to frameworks that explicitly account for how different mutational processes bias the properties of mutation-derived peptides. By focusing on the qualitative features of neoantigens, such approaches offer the potential to better explain immune heterogeneity across tumors and to provide a more mechanistic understanding of how mutational landscapes influence tumor–immune interactions.

Data resources and integrative frameworks for studying tumor immunogenicity.

Progress in tumor immunology has been closely linked to the availability of large, harmonized datasets that enable systematic analyses across cancer types. A central example is The Cancer Genome Atlas (TCGA), which comprises almost 10,000 tumor samples spanning 33 cancer types and provides multiple complementary data modalities, including somatic mutation profiles, transcriptomic measurements, copy-number alterations, and detailed clinical annotations (https://www.cancer.gov/tcga). The breadth and consistency of these data allow population-level investigation of mutational processes, immune phenotypes, and clinical outcomes within a unified analytical framework.

Interpreting somatic mutations observed in such large cohorts requires curated reference resources that connect genomic alterations to underlying biological mechanisms. The Catalogue Of Somatic Mutations In Cancer (COSMIC) provides a comprehensive and continuously updated knowledge base of somatic mutations and mutational signatures across human cancers[46] By cataloguing characteristic mutation patterns and their associated etiologies, COSMIC enables mechanistic interpretation of tumor mutational landscapes and serves as a foundation for analyses that link observed mutations to specific mutagenic processes. In the context of this thesis, COSMIC-derived mutational signatures provide the basis for connecting nucleotide-level mutation biases to downstream protein-level consequences.

In parallel with advances in cancer genomics, immunopeptidomics has emerged as a powerful experimental approach for directly characterizing the peptides presented by HLA molecules on the cell surface. Using mass spectrometry–based workflows, immunopeptidomics enables the identification of naturally processed and presented peptides, providing empirical insight into antigen presentation beyond *in silico* prediction[47]. Recent reviews have outlined the principles, strengths, and limitations of immunopeptidomics for studying tumor antigens and neoantigens, highlighting its role in bridging genomic variation and immune recognition[48].

Importantly, immunopeptidomics has progressed from small-scale proof-of-concept studies to larger and more systematic investigations. Large-scale analyses have demonstrated that HLA-presented peptide repertoires can be characterized across diverse biological contexts, revealing substantial variability in peptide composition and presentation patterns[49]. Methodological advances have further improved the depth, reproducibility, and scalability of immunopeptidomics workflows, supporting their integration into broader immunogenomic studies[50].

Despite major advances in cancer genomics and tumor immunology, a conceptual gap remains between the characterization of mutational processes at the DNA level and the immune recognition of tumor-derived peptides at the protein level. Mutational signatures describe the origins of genomic alterations, while antigen presentation and T-cell recognition operate on the biochemical properties of peptides shaped by those mutations. However, a systematic framework that links these two levels of organization at the population scale remains lacking. In particular, it remains unclear how the structure of mutational processes translates into qualitative features of neoantigens and how these features interact with host-specific antigen presentation constraints. Addressing this gap requires a representation of tumor mutations that preserves information about underlying mutational mechanisms while capturing properties directly relevant to immune recognition. The work presented in this thesis is motivated by this challenge. It aims to establish such a framework, thereby enabling integrative investigation of how mutational structure and host genotype jointly shape tumor immunogenicity.

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

Collecting and processing mutational data of TCGA cancer samples.

We downloaded data on mutations in 10549 cancer samples of TCGA from the Genomic Data Commons (GDC) portal of the NCBI using the TCGAbiolinks R library[51]. (download date: 27th February 2024, all projects, "Single Nucleotide Variation" data category, "Masked Somatic Mutation" data type). In each sample, we kept only missense mutation data for further analysis. We discarded samples with fewer than ten missense mutations and determined the fraction of each amino acid substitution in each sample. We included samples with relatively low mutation counts (<50) since recent evidence indicates that even cancers with modest mutational burdens can generate immunogenic neoantigens[52]. To exclude potential bias arising from the inclusion of samples with low mutation counts, we repeated the non-negative matrix factorization (NMF) analysis using alternative mutation count thresholds of 25 (n = 7,171 samples) and 50 (n = 4,641 samples). In both cases, the resulting amino acid substitution (AAS) profiles exhibited very high cosine similarity across all substitution signatures. These findings demonstrate that the identified AAS signatures are highly robust and largely insensitive to the chosen mutation count threshold.

Carrying out non-negative matrix factorization (NMF).

Non-negative matrix factorization (NMF) was applied to the complete dataset using the NMF package in R[53]. We generated a matrix containing the relative frequencies of each amino acid substitution for all tumor samples. To identify the optimal number of latent factors—representing distinct amino acid substitution signatures—we conducted 30 independent NMF runs for each factorization rank between 2 and 100. Model quality was assessed using several complementary metrics, including the cophenetic correlation coefficient, silhouette width, dispersion, explained variance, and residual sum of squares. Based on these criteria, a rank of five was selected as the optimal solution. In particular, the cophenetic correlation coefficient remained stable up to five factors and decreased at higher ranks, suggesting reduced stability beyond this point. Consistently, silhouette width and dispersion values also supported a five-factor model, indicating robust clustering and internal consistency. The contribution of individual AASs to each tumor sample was quantified by extracting the basis components from the original amino acid substitution matrix. Tumor samples were classified as dominated by a given AAS if the corresponding basis component exceeded 0.6. In instances where more than one AAS surpassed this threshold (approximately 3%

of samples), the AAS with the highest contribution was designated as dominant. Samples in which no basis component exceeded 0.6 were categorized as mixed.

Retrieving PubMed hits for tumors and mutagens.

We constructed all possible combinations between the full names of tumor types listed in TCGA and the names of environmental mutagens. Using the RISmed package in R, we queried PubMed to obtain the number of publications mentioning both a given tumor type and a specific mutagen. A mutagen was considered to be associated with a particular AAS if the cosine similarity between the mutagen-induced substitution profile and the corresponding AAS exceeded 0.7. For each AAS–tumor type pair, we then summed the PubMed hit counts for the tumor type together with all mutagens assigned to that AAS. To quantify the association between AASs and tumor types, we calculated, for each tumor type, the median value of the corresponding basis component across samples, separately for AAS1 through AAS5.

Simulating the effects of mutagenic processes on amino acid substitutions.

We obtained single-base substitution (SBS) signatures from the COSMIC database[46] and the SIGNAL database[54]. COSMIC provides mutation signatures identified from cancer genomes, whereas the SIGNAL database contains signatures derived from controlled mutagenesis and gene knockout experiments. Coding sequence data were retrieved from Ensembl (Homo_sapiens.GRCh38.cds.all.fa; downloaded on April 3, 2021). For each mutational signature, we simulated 1,000 random missense mutations across the coding genome. During these simulations, single-base substitutions were introduced probabilistically, weighted by the nucleotide substitution frequencies defined for each specific 5′ and 3′ sequence context within the corresponding signature. For COSMIC-derived signatures, transcriptional strand bias was incorporated into the simulation framework; strand bias information was not available for signatures obtained from the SIGNAL database.

Associating nucleotide mutations with AASs.

To quantify the strength of association between nucleotide mutations defined by their 5′ and 3′ sequence contexts and the AASs, we first simulated 1,000 random missense mutations for each of the 192 mutation profiles. These profiles were defined by the identity of the 5′ nucleotide (n = 4), the 3′ nucleotide (n = 4), and the substituted base (n = 12; 4 × 3). Each simulated missense mutation was then randomly assigned to an AAS, with probabilities weighted by the association strength between the corresponding amino acid substitution and the AAS. Based on these assignments, an information count matrix (ICM) motif was constructed using the relative contribution of each AAS across mutation profiles. Motif generation was performed using the universalmotif R package[55].

Analyzing ICB cohort and treatment-naive melanoma patients.

Mutational and clinical data for patients treated with immune checkpoint blockade (ICB) were obtained from Miao *et al*[43] through cBioPortal[56]. For each tumor sample, we calculated the relative frequency of all amino acid substitutions. To quantify the contribution of a given AAS within a sample, we computed the Pearson correlation coefficient between the sample-specific amino acid substitution frequencies and the corresponding AAS association values derived from the NMF coefficient matrix.

Validating the effect of AASs on presented neopeptides using immunopeptidomics data.

We assembled immunopeptidomics datasets from 17 cancer samples (tumor tissues or cell lines) spanning five tumor types (Appendix Table S7) with publicly available somatic mutation profiles and corresponding mutated peptides. For all possible combinations of sample s and AAS k, we computed the cosine similarity $c_{s,k}$_ between the frequency of the observed amino acid substitutions in sample s and AAS k. For each sample, all possible amino acid substitutions (i = 1,…,166) were then classified as either detected or not detected in HLA-I–bound neopeptides. Let $a_{i,k}$ denote the association strength of substitution i with AAS k. For each substitution i in sample s, we defined a weighted sum association strength

$$A_{s,i} = \sum_{k=1}^{5} c_{s,k} a_{i,k}$$

which integrates the contribution of all five AASs according to how similar the sample's substitution profile is to each AAS. Using substitution detection status across samples (detected vs. undetected, n = 17x166) as the response variable and $A_{s,\,i}$ as the predictor, we constructed receiver operating characteristic (ROC) curves using the pROC R package.

Cell lines and cell culture.

All A549 cell lines were obtained from Merck. In particular, we used the HLA Subtype Panel – A549 (HLA003-1KT, Sigma-Aldrich), which included the following lines: B2M knockout (control; HLA003A-1VL; SLCD9117), HLA-A*03:01 (HLA003D-1VL; SLCD9183), and HLA-B*07:02 (HLA003H-1VL; SLCD9187). Cell line identity was authenticated by short tandem repeat (STR) profiling conducted by Sigma-Aldrich. The A549 cell line is extensively annotated at the genomic level and has been widely applied in mutagenesis studies as well as investigations of tumor immunogenicity[57–61]. Owing to these characteristics, A549 represents a suitable model for studying chemically induced somatic mutations and the resulting amino acid substitution signatures.

Cells were routinely screened for mycoplasma contamination using the MycoAlert™ Mycoplasma Detection Kit (Lonza), and all assays were negative. No blinding was applied; investigators were aware of treatment conditions throughout cell culture, clonal isolation, sequencing sample submission, and flow cytometry data acquisition and analysis. Cells were maintained in DMEM:F12 medium (21127030, Gibco) supplemented with 10% fetal bovine serum (FBS-HI-12A, Capricorn), penicillin (100 µg/ml), and streptomycin (100 U/ml) (PS-B, Capricorn), and cultured at 37 °C in a humidified incubator with 5% CO_2.

Treatment with DNA-damaging agents.

The application of DNA-damaging agents followed previously established protocols as described by Kucab *et al*[1]. All compounds used in this study were prepared by dissolving them in suitable solvents and diluting them in culture medium immediately before treatment. For agents requiring metabolic activation, cells were incubated for 3 hours in treatment medium supplemented with an S9 mix, after which the medium was replaced with fresh growth medium. The S9 mix contained 0.25% S9 fraction derived from Aroclor-1254–induced male Golden Syrian hamster liver (#15-03SL.5, Moltox), 3 mM NADP (10128031001, Roche), and 15 mM DL-isocitric acid trisodium salt hydrate (Sigma), prepared in culture medium.

Treatment doses were selected to induce 40–60% cytotoxicity, as previously reported. The concentrations applied in this study were 400 µM N-ethyl-N-nitrosourea (Sigma-Aldrich, N3385-1G) and 0.39 µM benzo[a]pyrene (Sigma-Aldrich, B1760-250MG) administered in the presence of S9 mix. Cells were first subcloned and subsequently exposed to the indicated chemical agents or radiation in culture medium for up to 24 hours. Following treatment, culture medium was replaced and refreshed daily as needed to preserve cell viability and to minimize residual treatment effects. For non-chemical exposures, cells were subjected to simulated solar radiation (SSR) using a Vilber Lourmat VL-6.LM system. The SSR spectrum consisted of 10% UVB (295–315 nm) and 90% UVA (315–400 nm), with a total delivered dose of 1.25 J.

Clonal isolation, genomic DNA extraction, and whole-genome sequencing.

To obtain single-cell–derived clones, treated cell populations were first dissociated into single-cell suspensions using trypsin (TRY-1B, Capricorn). The resulting suspensions were plated at limiting dilution onto 10 cm culture dishes. Culture medium was replaced daily, and individual clones were allowed to form over a period of 10–12 days. After clonal outgrowth, clones from each treatment group were sequentially expanded by passaging into 24-well plates and subsequently into 6-well plates. Cell pellets were collected from each clone, and aliquots were prepared for cryopreservation. Genomic DNA (gDNA) was extracted from the harvested cell pellets using a gDNA purification kit (Lonza) according to the manufacturer's instructions. Purified gDNA samples were then submitted to Novogene for sequencing at an average depth of 30×.

Peripheral blood mononuclear cell (PBMC) sample preparation.

Peripheral blood mononuclear cells (PBMCs) were obtained from CTL Europe GmbH (custom lot: CTL_HP1). Cells were washed with 5 ml of 1× phosphate-buffered saline (PBS) and pelleted by centrifugation at 350 × g for 7 min at room temperature. To eliminate residual red blood cells, the cell pellet was resuspended in 2 ml of ACK lysis buffer (0.15 M NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2EDTA, pH 7.2–7.4) and incubated for 2 min at room temperature. The suspension was then centrifuged at 350 × g for 7 min at room temperature, followed by washing with 12 ml of complete RPMI medium and an additional centrifugation step under identical conditions.

T-cell proliferation assay.

T cell proliferation assays were performed using a protocol adapted and further developed from Mojtaba *et al*[62]. Cells were labeled with carboxyfluorescein succinimidyl ester (CFSE; C34554, Invitrogen) to track cell division. Following centrifugation, cells were resuspended in 1× PBS at a density of 1×10^6 cells/ml. CFSE was added to the suspension to a final concentration of 5 µM, and cells were incubated for 10 min at room temperature with gentle agitation. The labeling reaction was terminated by washing the cells three times with complete RPMI medium (RPMI-A, Capricorn).

After washing, PBMCs were resuspended in culture medium consisting of RPMI supplemented with 10% fetal bovine serum (FBS), 1% antibiotic–antimycotic solution, and 1% MEM non-essential amino acids (NEAA-B). For proliferation measurements, PBMCs were co-cultured with mutagen-treated A549 cells in 96-well plates, with each well containing 200,000 PBMCs and 4,000 A549 cells. As a positive control for stimulation, Dynabeads™ Human T-Activator CD3/CD28 (25 µl per 1 × 10^6 cells; Gibco, 11161D) were included to induce T cell activation and expansion. Interleukin-2 (IL-2; PHC0027, Invitrogen) was added the following day at a final concentration of 1 ng/ml. On day 4 of culture, an additional 50 µl of complete RPMI medium was added to each well. Cells were maintained in culture and analyzed by flow cytometry on day 7.

Immunostaining of PBMCs and flow cytometry analysis.

Prior to flow cytometric analysis, PBMCs were harvested and stained to enable T cell identification and assessment of cell viability. For T cell labeling, 3 µl of Alexa Fluor® 647–conjugated anti-human CD3 antibody (BioLegend, 317312) was added to each well, followed by incubation for 45 min at room temperature. To exclude non-viable cells, propidium iodide (PI; P4170-10MG, Sigma-Aldrich) was added immediately before acquisition at a final concentration of 10 µg/ml in PBS. Samples were analyzed on a Beckman Coulter CytoFlex S flow cytometer, and data were processed using CytExpert software. T cell proliferation was assessed by measuring CFSE dye dilution, which enables discrimination of successive cell division cycles. Cell viability was determined by PI staining, while CD3 expression was used to identify T cells within the PBMC population.

In the proliferation assay, non-dividing cells were defined as the G0 population, characterized by stable CFSE intensity and comparable coefficients of variation (CV) across samples. Dividing cells displayed stepwise CFSE dilution and formed the G1 population. By applying sequential gating strategies for PI (to exclude dead cells), anti-CD3 (to identify T cells), and CFSE intensity (to assess proliferation), G0 and G1 populations were clearly resolved. Based on the relative distribution of cells within the G0 and G1 peaks, the division index was calculated as follows:

$$DI = \frac{\sum_0^i i\frac{N_i}{2^i}}{\sum_0^i \frac{N_i}{2^i}}$$

,

where i is the generation number, N_i is the number of cells in generation i.

Processing WGS data with IsoMut and Annovar.

Raw sequencing reads were processed using Novogene’s proprietary bioinformatics pipeline. In brief, sequencing quality control was performed to ensure that at least 80% of bases achieved a Q30 score or higher. Reads were aligned to the human reference genome using the Burrows–Wheeler Aligner (BWA)[63] v0.7.17. Resulting BAM files were sorted with Sambamba v0.7.1[64], and duplicate reads were identified and marked using Picard v2.18.9[65]. Somatic mutations were detected using the IsoMut[66] algorithm, which is specifically designed to identify experimentally induced mutations across multiple isogenic samples. Only samples containing at least 10,000 detected mutations were retained for downstream analyses. Identified variants were annotated using Annovar[67] (version 2020-06-07).

Results

Characterizing mutational heterogeneity at the amino acid level.

To assess intertumoral variation in the protein-altering mutational landscape, we analyzed 9,374 tumors spanning 33 cancer types in The Cancer Genome Atlas (TCGA). We quantified the frequency of all possible nonsynonymous amino acid substitutions (n = 166) in each sample. These substitution profiles revealed considerable heterogeneity, which was decomposed into five distinct amino acid substitution signatures (AAS1–AAS5) using non-negative matrix factorization. This tool has been proven to be effective for uncovering mutational patterns in cancer[7].

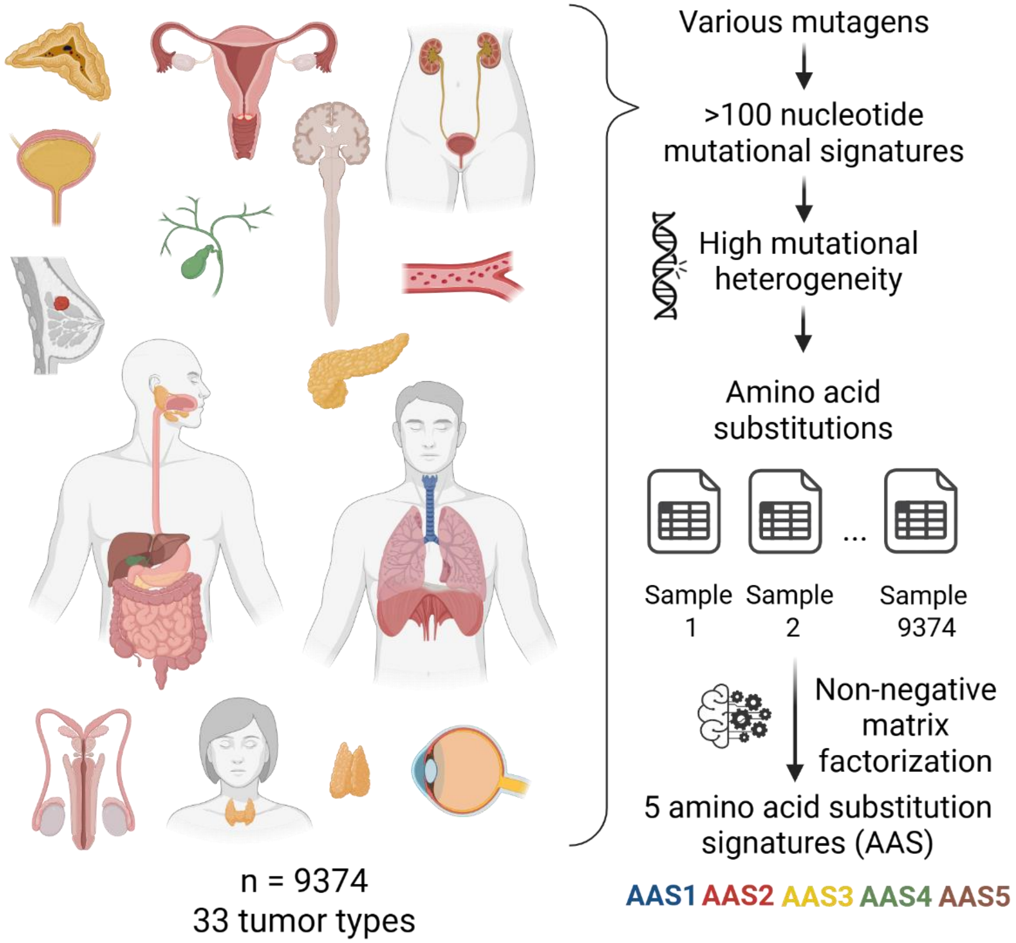

Various mutagens
>100 nucleotide mutational signatures
High mutational heterogeneity
Amino acid substitutions
Sample 1
Sample 2
...
Sample 9374
Non-negative matrix factorization
5 amino acid substitution signatures (AAS)
AAS1 AAS2 AAS3 AAS4 AAS5
n = 9374
33 tumor types

Figure 1. Amino acid substitutions in tumor samples can be categorized into five distinct signatures. Overview of the dataset composition and methodological framework. See text of the Results and Methods sections for details.

Each AAS exhibited a characteristic distribution of substitution types and a distinct prevalence across tumor types. Most tumors were dominated by a single AAS, although 26% of samples displayed a mixed composition with contributions from multiple signatures. To map each AAS to its likely etiological sources, associations with environmental mutagens and DNA repair deficiencies were systematically identified using mutational signatures from the COSMIC[46] database and data from prior experimental studies[1,6]. (See methods, Figure 2A) We additionally assessed associations between AASs, deficiencies in distinct DNA repair genes, and a range of environmental mutagens.

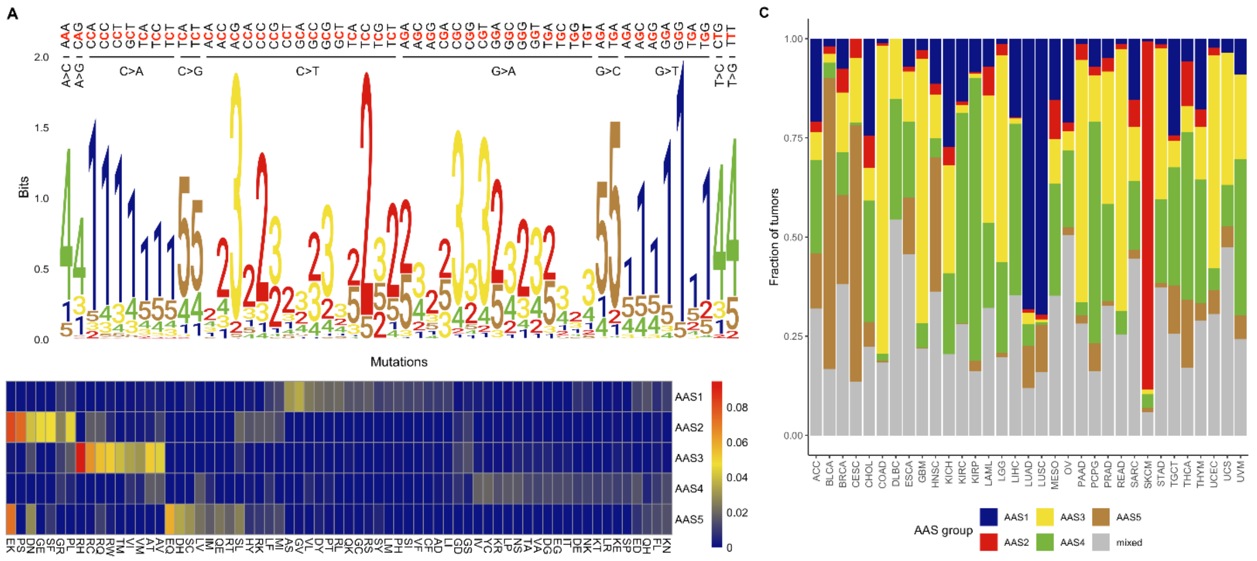


Figure 2. Tumor-derived amino acid substitutions are grouped into five characteristic signatures. A) Each amino acid substitution signature (AAS) is associated with distinct nucleotide mutation patterns. The sequence logos display how specific base substitutions within varying 5' and 3' sequence contexts contribute to different AASs. For clarity, only nucleotide mutations occurring in at least 0.5% of TCGA tumor samples are included (see Methods). **B)** Each AAS shows a unique profile of amino acid substitutions. Association strengths between individual substitutions and the five AASs were inferred from the coefficient matrix produced by non-negative matrix factorization (NMF), and are visualized as a heatmap. Only substitution types with an association value exceeding 0.01 in at least one AAS are shown. **C)** Most tumors are characterized by a dominant AAS. The plot depicts the distribution of dominant signatures across tumor types. Samples without a single prevailing AAS are labeled as ‘mixed.

AAS1 is dominated by C>A substitutions (corresponding to G>T on the complementary strand; Figure 2A) and is most frequently observed in lung cancer types, including lung adenocarcinoma (LUAD) and lung squamous-cell carcinoma (LUSC) (Figure 2C and Table 1). A pronounced enrichment of AAS1 is detected in tumors associated with tobacco-related carcinogens, such as benzo[a]pyrene (Figure 3A), and is reflected by the presence of mutational signatures SBS4 and SBS29 (Figure 3A). These observations indicate that AAS1 captures mutations arising from tobacco smoke-induced DNA damage. In addition, AAS1 is linked to other environmental mutagens, including platinum-based chemotherapeutic agents such as cisplatin (SBS35) and aflatoxin exposure (SBS24). The association between aflatoxin B1, liver hepatocellular carcinoma (LIHC), and AAS1 is particularly notable, as dietary aflatoxin B1 exposure is known to increase the risk of hepatocellular carcinoma[68,69]. Moreover, AAS1-associated mutations are connected to reactive oxygen species (ROS)–mediated DNA damage (SBS18) and to potassium bromate ($KBrO_3$) exposure. Potassium bromate generates hydroxyl radicals that cause multiple DNA lesions, most prominently 8-oxo-dG (Kawanishi & Murata, 2006), which are typically repaired via base-excision repair pathways involving OGG1 and MUTYH[6].

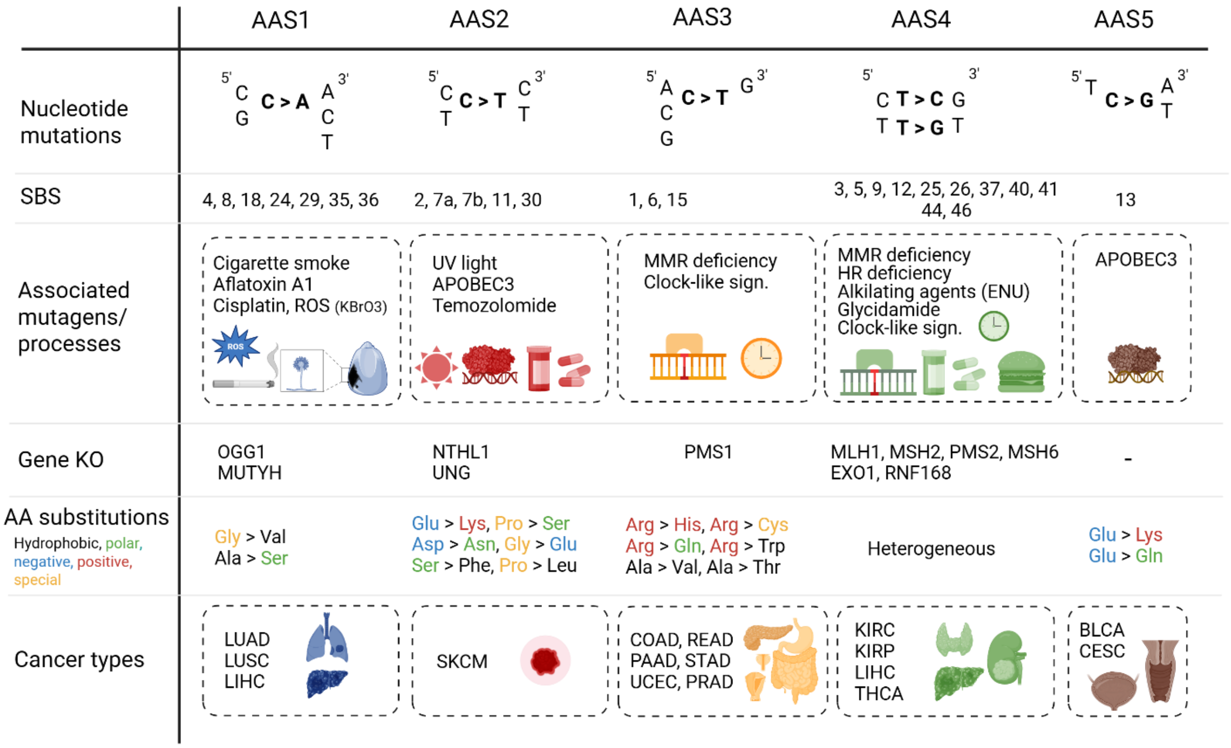
AAS2 is partially defined by C>T substitutions (Figure 2A) and is associated with glutamine to lysine (E>K), proline to serine (P>S), aspartate to asparagine (D>N), glycine to glutamate (G>E) substitutions (Figure 2B). This signature reflects multiple mutational processes, including ultraviolet light–induced damage (SBS7a and SBS7b) and deficiencies in base-excision repair glycosylases such as NTHL1 and UNG. Consistent with these mechanisms, AAS2 is frequently detected in UV-driven melanoma (SKCM; Figure 2C). In addition, AAS2 is associated with SBS2, a mutational signature linked to the activity of AID/APOBEC cytidine deaminases. APOBEC signatures coupled to SBS2 are characterized by an increased frequency of C>T transitions at TpC motifs, a pattern that differs from SBS13-associated APOBEC activity, which predominantly generates C>G substitutions at TpN sites[17].

AAS3 is strongly enriched for arginine substitutions to histidine, cysteine, glutamine, or tryptophan (R>H, R>C, R>Q, R>W; Figure 2B). This signature is prevalent in gastrointestinal malignancies, including colorectal (COAD, READ), pancreatic (PAAD), and gastric adenocarcinoma (STAD) (Figure 2C), and is also observed in uterine corpus endometrial carcinoma (UCEC) and prostate adenocarcinoma (PRAD) (Figure 2C). Tumors exhibiting AAS3 frequently show defects in DNA mismatch repair, as indicated by the presence of SBS15 or SBS6. Notably, loss of the mismatch repair gene PMS1 shows a particularly strong association with AAS3, producing a substitution pattern related to SBS6 that is distinct from those generated by other mismatch repair deficiencies (see AAS4 for comparison). Additionally, AAS3 is linked to the clock-like mutational signature SBS1.

AAS4 is characterized by T>C/A>G and T>G/A>C substitutions (Figure 2A). It is commonly observed in renal cancers, including kidney renal papillary carcinoma (KIRP) and kidney renal clear cell carcinoma (KIRC) (Figure 2C). This signature is associated with defects in several mismatch repair genes, including PMS2 (SBS26) and MLH1, MSH2, and MSH6 (SBS44) (Figure 3A; Appendix Figure S5). In addition, AAS4 can be traced to homologous recombination deficiency–related mutational processes, including loss of EXO1 and RNF168. Alkylating agents, such as N-ethyl-N-nitrosourea (ENU), also generate mutational signatures linked to AAS4. Consistent with this, multiple alkylating compounds induce T>C transitions, which represent substrates for the aforementioned DNA repair pathways[9].

AAS5 is predominantly defined by C>G substitutions (Figure 2A), frequently resulting in glutamate-to-lysine or glutamine substitutions (E>K, E>Q; Figure 2B). This signature is primarily associated with SBS13-coupled APOBEC activity and shows high prevalence in bladder cancer (BLCA) and cervical squamous cell carcinoma (CESC). Consistent with these observations, APOBEC-mediated mutagenesis is known to play a major role in shaping the mutational landscapes of these tumor types[8].

Together, these results define five major classes of amino acid substitution patterns in human cancers, each associated with distinct nucleotide substitution biases, mutagenic exposures, and cancer types. The signatures represent an interpretable layer of protein-level mutational heterogeneity with potential implications for tumor phenotype.



	AAS1	AAS2	AAS3	AAS4	AAS5
Nucleotide mutations	5' C **C > A** A 3' G C T	5' C **C > T** C 3' T T	5' A **C > T** G 3' C G	5' C **T > C** G 3' T **T > G** T	5' T **C > G** A 3' T
SBS	4, 8, 18, 24, 29, 35, 36	2, 7a, 7b, 11, 30	1, 6, 15	3, 5, 9, 12, 25, 26, 37, 40, 41 44, 46	13
Associated mutagens/ processes	Cigarette smoke Aflatoxin A1 Cisplatin, ROS (KBrO3) ROS	UV light APOBEC3 Temozolomide	MMR deficiency Clock-like sign.	MMR deficiency HR deficiency Alkilating agents (ENU) Glycidamide Clock-like sign.	APOBEC3
Gene KO	OGG1 MUTYH	NTHL1 UNG	PMS1	MLH1, MSH2, PMS2, MSH6 EXO1, RNF168	-
AA substitutions Hydrophobic, polar, negative, positive, special	Gly > Val Ala > Ser	Glu > Lys, Pro > Ser Asp > Asn, Gly > Glu Ser > Phe, Pro > Leu	Arg > His, Arg > Cys Arg > Gln, Arg > Trp Ala > Val, Ala > Thr	Heterogeneous	Glu > Lys Glu > Gln
Cancer types	LUAD LUSC LIHC	SKCM	COAD, READ PAAD, STAD UCEC, PRAD	KIRC KIRP LIHC THCA	BLCA CESC

Table 1. Overview of the defining features of amino acid substitution signatures (AASs). This table summarizes the five identified AAS classes (AAS1 through AAS5), detailing their associated nucleotide mutation types, corresponding single-base substitution (SBS) signatures, related mutagens or endogenous processes, relevant gene knockouts (KO), characteristic amino acid changes, and prevalent cancer types. Abbreviations: ROS – reactive oxygen species; MMR – mismatch repair; HR – homologous recombination; ENU – N-ethyl-N-nitrosourea

Amino acid substitution signatures (AASs) shape the cancer immunopeptidome.

Although AASs are defined from the catalogue of missense mutations, their biological relevance ultimately depends on whether the same substitution biases propagate to the neopeptides that reach the tumor cell surface and become visible to T cells. In other words, if a given AAS dominates a tumor's genome, we expect this signature to leave a recognizable imprint on the HLA-bound immunopeptidome as well, because the set of amino acid changes created by the underlying mutational processes constrains which mutant residues can appear in presented peptides. We therefore sought to evaluate, using publicly available immunopeptidomics datasets with matched genomic data, whether the AAS context of a tumor can nonetheless predict which amino acid substitutions are represented among experimentally detected HLA-I–bound neopeptides.

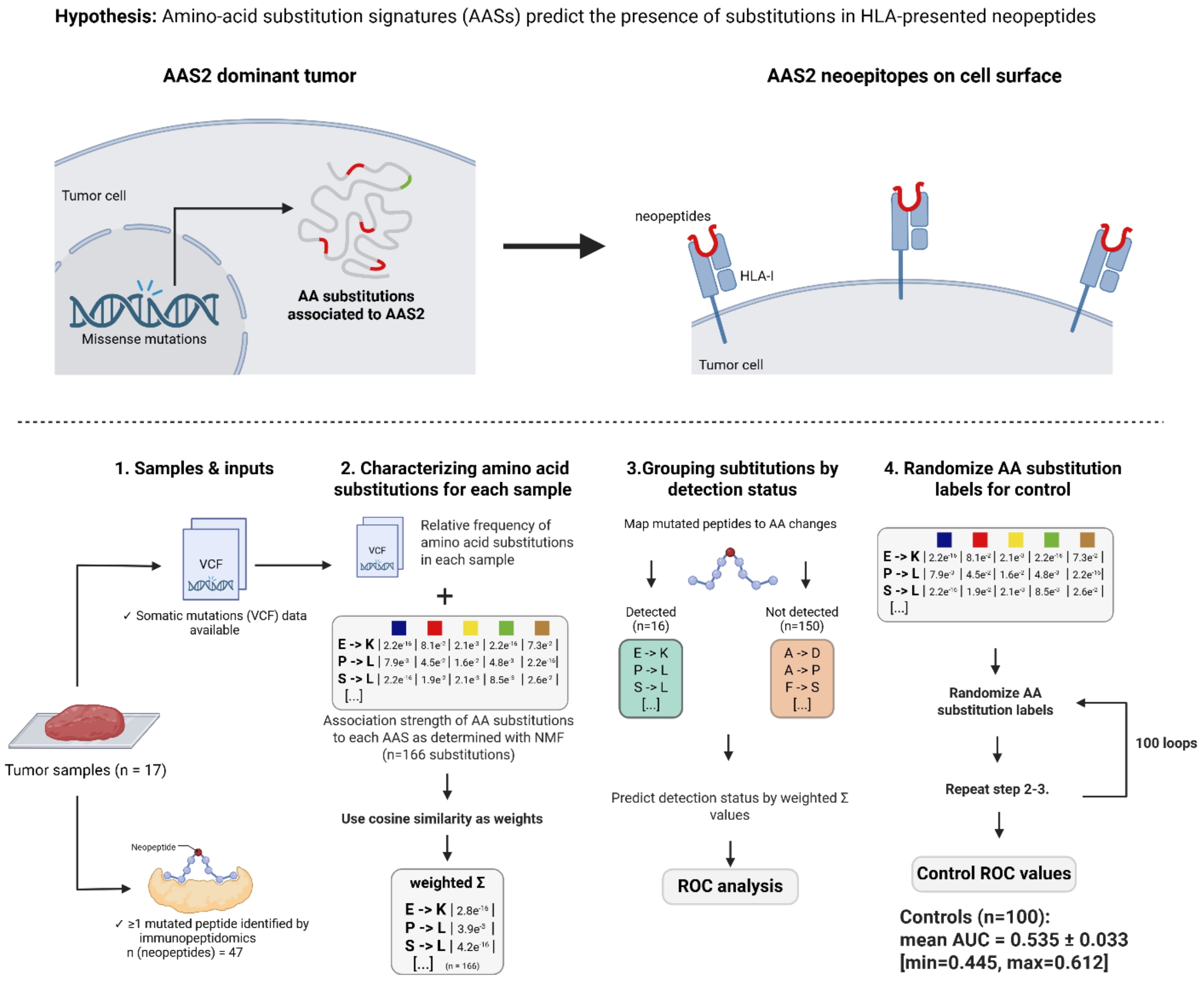

Hypothesis: Amino-acid substitution signatures (AASs) predict the presence of substitutions in HLA-presented neopeptides
AAS2 dominant tumor
AAS2 neoepitopes on cell surface
Tumor cell
Missense mutations
AA substitutions associated to AAS2
neopeptides
HLA-I
Tumor cell
1. Samples & inputs
VCF
✓ Somatic mutations (VCF) data available
Tumor samples (n = 17)
Neopeptide
✓ ≥1 mutated peptide identified by immunopeptidomics
n (neopeptides) = 47
2. Characterizing amino acid substitutions for each sample
VCF
Relative frequency of amino acid substitutions in each sample
+
🟦 🟥 🟨 🟩 🟫
E -> K | $2.2e^{-16}$ | $8.1e^{-2}$ | $2.1e^{-3}$ | $2.2e^{-16}$ | $7.3e^{-2}$ |
P -> L | $7.9e^{-3}$ | $4.5e^{-2}$ | $1.6e^{-2}$ | $4.8e^{-3}$ | $2.2e^{-16}$|
S -> L | $2.2e^{-16}$ | $1.9e^{-2}$ | $2.1e^{-3}$ | $8.5e^{-3}$ | $2.6e^{-2}$ |
[...]
Association strength of AA substitutions to each AAS as determined with NMF (n=166 substitutions)
Use cosine similarity as weights
weighted Σ
E -> K | $2.8e^{-16}$ |
P -> L | $3.9e^{-3}$ |
S -> L | $4.2e^{-16}$ |
[...] (n = 166)
3.Grouping subtitutions by detection status
Map mutated peptides to AA changes
Detected (n=16)
E -> K
P -> L
S -> L
[...]
Not detected (n=150)
A -> D
A -> P
F -> S
[...]
Predict detection status by weighted Σ values
ROC analysis
4. Randomize AA substitution labels for control
🟦 🟥 🟨 🟩 🟫
E -> K | $2.2e^{-16}$ | $8.1e^{-2}$ | $2.1e^{-3}$ | $2.2e^{-16}$ | $7.3e^{-2}$ |
P -> L | $7.9e^{-3}$ | $4.5e^{-2}$ | $1.6e^{-2}$ | $4.8e^{-3}$ | $2.2e^{-16}$|
S -> L | $2.2e^{-16}$ | $1.9e^{-2}$ | $2.1e^{-3}$ | $8.5e^{-3}$ | $2.6e^{-2}$ |
[...]
Randomize AA substitution labels
100 loops
Repeat step 2-3.
Control ROC values
Controls (n=100):
mean AUC = 0.535 ± 0.033
[min=0.445, max=0.612]

Figure 3. Conceptual overview illustrating the rationale and workflow of the analysis performed to validate AAS signatures using immunopeptidomics data. See the main text and Methods for details

In this analysis, we aimed to validate the influence of amino acid substitution signatures (AASs) on the neopeptide repertoire presented by human leukocyte antigen (HLA) molecules on the surfaces of cancer cells. To this end, we analyzed previously published systematic immunopeptidomics datasets profiling HLA-I–bound neopeptides from 17 cancer samples[25,70–78] with matched genomic data describing their somatic mutations. Only datasets meeting two criteria were included: (i) annotated missense variants from whole-exome sequencing or targeted analysis, and (ii) at least one mutant peptide identified by mass spectrometry–based immunopeptidomics. In total, 47 mutated peptides were retained for downstream analysis.

As throughout the paper, the analysis was restricted to missense mutations. We first computed the cosine similarity between the relative frequency of observed amino acid substitutions in each sample and each AAS (Figure 3). For each sample, all possible amino acid substitutions (n = 166) were then classified as either detected or not detected in HLA-I–bound neopeptides. For each substitution, we calculated an association score to the five AASs by summing their association strengths, weighted by the cosine similarity between the given sample and each AAS (see Methods).

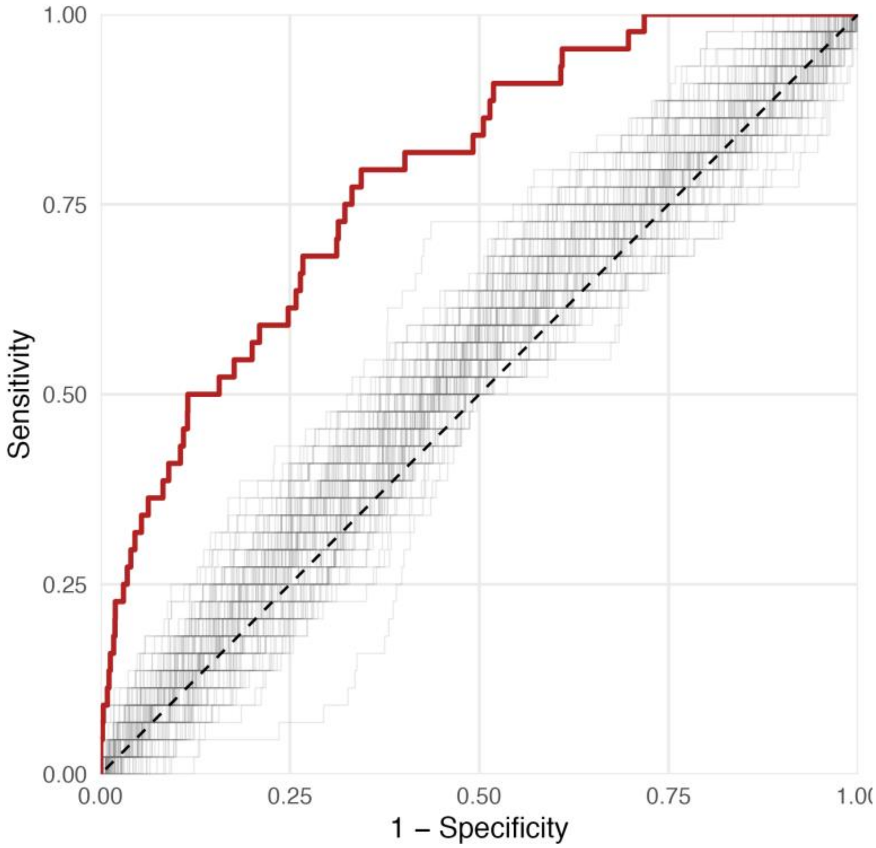


Figure 4. Influence of AASs on the immunopeptidome. The ROC curve showing how well the calculated association score values in tumor samples predict the presence of amino acid substitutions in immunopeptidomics data (AUC = 0.785). For comparison, 100 ROC curves obtained after randomizing amino acid substitution labels in the AAS coefficient matrix are shown in grey (with increased transparency for clarity, AUC min: 0.445, max: 0.612, mean: 0.535, SD: 0.033).

Using substitution detection status across samples (detected vs. undetected, n = 17 × 166) as the response variable and the weighted association scores as predictors, we constructed receiver operating characteristic (ROC) curves. This analysis yielded an area under the ROC curve (AUC) of 0.79 (Figure 4). As a control, we repeated the analysis 100 times after randomizing amino acid substitution labels, obtaining ROC AUC values ranging from 0.445 to 0.612 (mean = 0.535, SD = 0.033). Together, these results indicate that amino acid substitution context, as captured by AASs, shapes the cancer immunopeptidome.

Amino acid substitution profiles predict tumor immune microenvironment.

To infer the functional consequences of the characteristic amino acid substitutions, we sought to describe the physicochemical properties associated with each signature. Accordingly, the 20 standard amino acids were classified according to basic physico-chemical features, including hydrophobicity and charge. Based on these properties, we quantified how each amino acid substitution altered these attributes across the five AASs.

We observed that tumor samples characterized by AAS4-type substitutions show a reduced tendency to accumulate hydrophobic amino acids compared with samples associated with the other AASs. In contrast, tumors linked to AAS5-type substitutions preferentially exhibit more radical amino acid changes, often involving pronounced alterations in charge (Figure 4A, B). For instance, glutamate-to-lysine or glutamine substitutions (E>K, E>Q) occur at a median frequency of 3.8% across all cancer genomes, yet are enriched by a factor of 4.4 in AAS5-associated tumors, reaching a median frequency of 17%. These substitutions replace a negatively charged residue with either a positively charged or neutral one, which may have important functional implications. Such systematic shifts in the physico-chemical properties of amino acids across AASs may influence the immunogenicity of resulting neopeptides[41,79]. First, immunogenic neopeptides have been shown to preferentially contain increased hydrophobicity at T cell receptor contact positions[80,81]. Second, they may arise from radical amino acid substitutions that modify charge and facilitate HLA binding[11]. Consistently, presentation of peptides containing negatively charged anchor residues has been linked to improved survival in patients receiving immune checkpoint blockade therapy. We next asked whether these biophysical biases translate into measurable immune consequences, without assuming causality.

As outlined above, amino acid substitutions associated with AAS4 are characterized by reduced accumulation of hydrophobic residues and only minor alterations in charge. These features suggest that tumor genomes dominated by AAS4 signatures are likely to generate neopeptides with lower immunogenic potential. In contrast, cancer samples linked to AAS5 are expected to accumulate a comparatively higher number of immunogenic neopeptides. To evaluate this hypothesis, we employed the PRIME[81] algorithm to predict immunogenic epitopes. PRIME estimates peptide binding to HLA class I (HLA-I) molecules as well as their likelihood of recognition by T cell receptors. Specifically, PRIME was used to quantify both HLA-binding affinity and predicted immunogenicity of neopeptides derived from cancer samples associated with each AAS. In line with our expectations, neopeptides originating from AAS4-dominant tumors exhibited significantly lower HLA-binding affinities and reduced immunogenicity scores compared with other groups (Figures 5C and 5D). Conversely, neopeptides from AAS5-dominant samples showed substantially higher predicted HLA-binding values (Figure 5D), likely reflecting the pronounced charge-altering amino acid substitutions characteristic of AAS5 (Figure 5B).

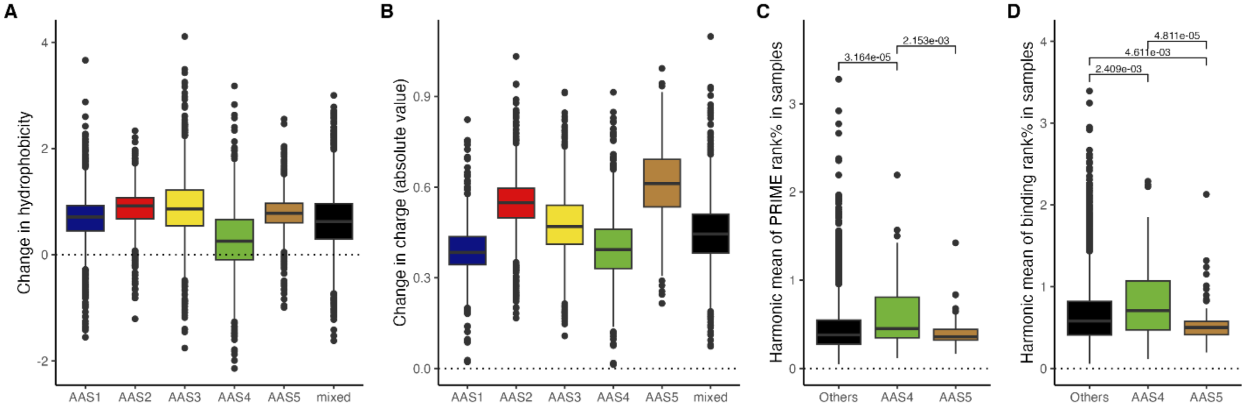


Figure 5. Amino acid substitution signatures influence the immunogenicity of neopeptides. A-B) AASs are linked to diverse effects on the biophysical properties of amino acids. The effect of amino acid substitutions on hydrophobicity (A) and charge (B) in tumor samples dominated by different AASs (n = 1388, 642, 2139, 1513, 1155, and 2537 in AAS1 to 5 and mixed sample groups, respectively). Kruskal-Wallis rank sum test p-values are lower than 2.2 × 10^-16 in both cases. FDR-corrected P values of Dunn's test are not indicated for visualization purposes; they can be found in Appendix Table S2. C-D) AAS4 and 5 are associated with the generation of low and high-immunogenicity neopeptides, respectively. The harmonic mean of HLA binding rank% (C) and PRIME rank% (D) values is shown in tumor samples dominated by AAS4 (n = 1379 samples), AAS5 (n = 1086 samples), and other signatures (n = 6116 samples). Lower rank% values indicate stronger HLA binding (C) and higher immunogenicity (D). Kruskal-Wallis test p values are 5.91 × 10^-5 and 8.7 × 10^-5 for panels C and D, respectively. FDR-corrected p values of Dunn's tests are indicated above horizontal segments. On boxplots, vertical lines indicate the median, boxes indicate the interquartile range (IQR), and horizontal lines indicate the first quartile -1.5 × IQR and the third quartile +1.5 × IQR.

We next examined how these amino acid substitution signatures relate to broader tumor immune phenotypes within the tumor microenvironment. We hypothesized that cancer samples associated with AAS4 would be characterized by a lower prevalence of immune-enriched tumor microenvironments. In contrast, those linked to AAS5 would show an increased frequency of such phenotypes. To test these assumptions, we applied a transcriptome-based classification of tumor microenvironments[82] across all TCGA cancer samples. Previous studies have demonstrated that patients with immune-enriched tumor microenvironments derive the greatest benefit from cancer immunotherapy[82,83]. Consistent with these expectations, tumors associated with AAS5 were significantly enriched for immune-infiltrated microenvironments (Figure 6A, Fisher's exact test P = 0.002), while AAS4-linked tumors showed a marked depletion of these environments (Figure 6A, Fisher's exact test P = 5.4 × 10^{-4}). Importantly, we sought to confirm that the association between AAS classes and tumor microenvironment composition was independent of overall mutational burden and tumor type. To this end, we fitted multivariable logistic regression models incorporating AAS class, tumor mutational burden (TMB), and tumor type as covariates. Notably, the effects of both AAS4 and AAS5 on immune-enriched tumor microenvironments remained statistically significant.

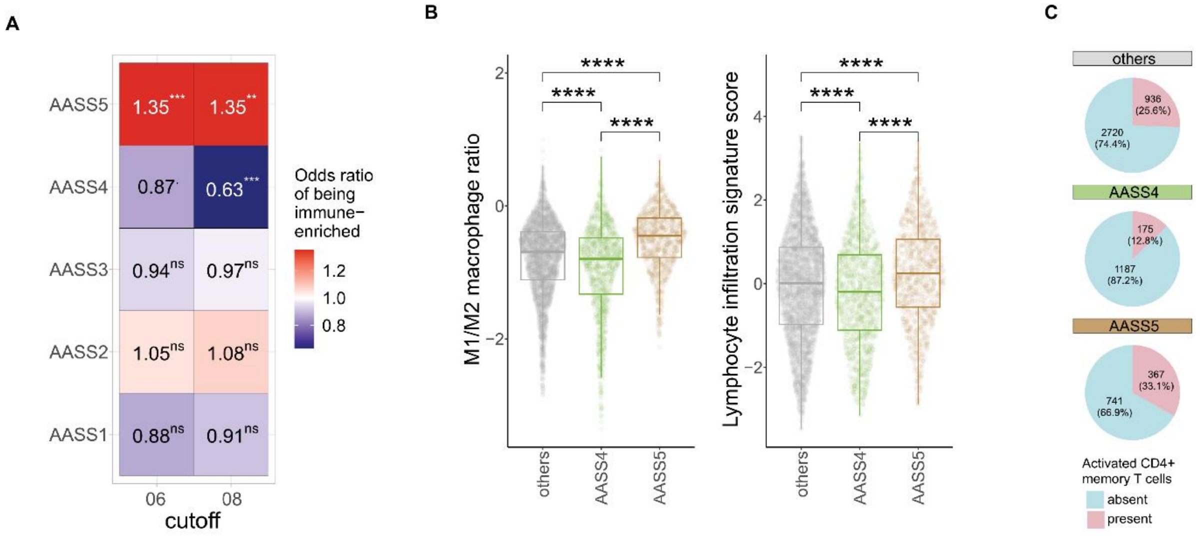


Figure 6. Amino acid substitution signatures influence the tumor microenvironment. A) Immune-enriched tumor microenvironment is differentially distributed across AAS-dominant groups. Odds ratios (ORs) for finding immune-enriched tumors are shown for samples dominated by each AAS, using two cutoff values (0.6 and 0.8) to define AAS dominance. Statistical significance was determined using two-sided Fisher's exact tests. Symbol legend: = $p < 0.1$, * = $p < 0.001$, ** = $p < 0.01$, ns = not significant. **B)** Immune infiltration metrics differ among AAS4-, AAS5-, and other-sample groups. The left panel shows the M1/M2 macrophage ratio; the right panel presents the lymphocyte infiltration score. Statistically significant differences were observed (Kruskal-Wallis test $p < 2.2\times10^{-16}$; Dunn's post hoc test ****$p < 2.2\times10^{-16}$). **C)** Prevalence of activated memory $CD4^+$ T cells across tumors dominated by AAS4, AAS5, or other signatures. OR and p values from Fisher's exact tests are as follows: AAS4 vs. others: OR = 0.43, $p = 9.37\times10^{-24}$; AAS5 vs. others: OR = 1.44, $p = 1.22\times10^{-6}$; AAS5 vs. AAS4: OR = 3.36, $p = 1.01\times10^{-33}$

To further delineate the intratumoral immune landscapes associated with AAS4 and AAS5, we leveraged tumor immune phenotype features defined in a previous study[23]. This analysis showed that tumors characterized by AAS4 generally exhibit lower overall levels of lymphocyte infiltration, whereas AAS5-associated tumors display increased infiltration (Figure 6B), a feature commonly linked to effective antitumor immune activity. More recently, the balance between M1 and M2 tumor-associated macrophages (TAMs) within the tumor microenvironment has emerged as an important biomarker, with a higher M1/M2 ratio being associated with improved prognosis in cancer patients[84]. Consistent with this notion, AAS5-dominant tumors demonstrated elevated M1/M2 ratios compared with AAS4-dominant tumors (Figure 6B). In addition, activated CD4$^+$ memory T cells are critical for coordinating and maintaining antitumor immune responses, as increased numbers of these cells enhance immune recognition and elimination of cancer cells[85,86]. In agreement with these observations, tumors associated with AAS4 exhibited particularly low levels of activated CD4$^+$ memory T cells (Figure 6C).

It is important to note that AAS classification is independent of tumor mutational burden (TMB) and tumor type. To control for these variables, we employed multiparametric logistic regression models, including AAS class, TMB, and tumor type as covariates. The associations of AAS4 and AAS5 with immune-enriched TME remained significant in these models, supporting the robustness of our initial findings (Figure 7).

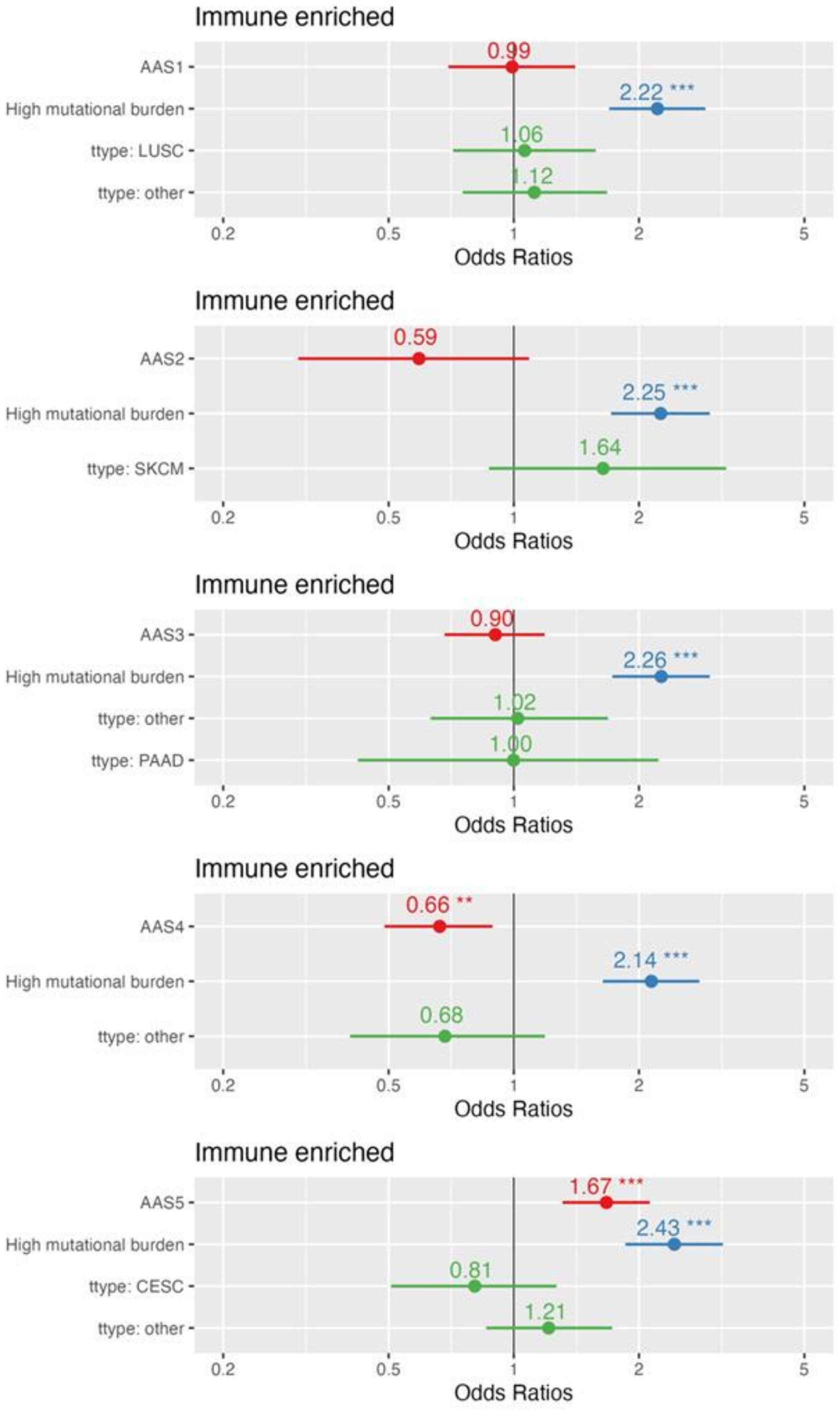


Figure 7. The effect of AAS on immune phenotype is independent of TMB and tumor type. The forest plots display the coefficients and 95% confidence intervals for the predictor variables in the logistic regression models. Predictors with statistically significant effects are marked with asterisks.

Clinical relevance of AASs in ICB response and prognosis.

Based on the above results, we next investigated the impact of amino acid substitution patterns on the efficacy of cancer immunotherapy. We focused on a previously published cohort of 249 cancer patients treated with immune checkpoint blockade (ICB) across multiple tumor types, including melanoma, non-small-cell lung cancer (NSCLC), and bladder cancer, among others[43]. Most patients in this cohort received anti-PD-1, anti-PD-L1, anti-CTLA-4 therapies, or a combination thereof. Specifically, we analyzed mutational data from the corresponding tumor samples and assessed their clinically annotated outcomes following ICB treatment.

To evaluate how amino acid substitution signatures influence therapeutic outcomes, we applied the Response Evaluation Criteria in Solid Tumors (RECIST). In line with previous studies[87]Clinical benefit was defined as a complete or partial response, or stable disease lasting ≥6 months. Conversely, no clinical benefit was defined as stable disease for <6 months or progressive disease. We found that the presence of AAS4 in tumor genomes was negatively associated with clinical benefit from immune checkpoint treatment. This association remained significant in a multivariate logistic regression model that included tumor type, mutational burden, and sex as covariates (Figure 8A).

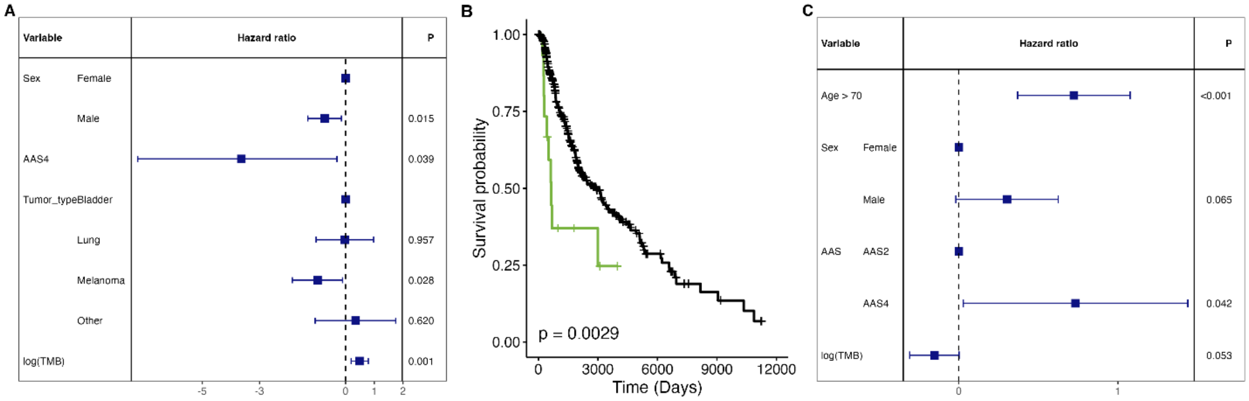


Figure 8. Patient outcomes and immunotherapy responses are influenced by amino acid substitution signatures. A) Dominance of AAS4 is linked to reduced likelihood of clinical benefit from immune checkpoint blockade (ICB) therapy. The logistic regression model incorporates tumor type, sex, tumor mutational burden (TMB), and AAS4 classification. Hazard ratios for each variable are represented by blue squares; corresponding 95% confidence intervals are shown as horizontal blue lines. **B)** Melanoma patients with tumors dominated by AAS4 exhibit poorer overall survival compared to those dominated by AAS2. Kaplan–Meier survival curves display TCGA melanoma samples stratified by dominant AAS (green: AAS4; black: AAS2). The p-value from the log-rank test is indicated. **C)** The association between AAS4 dominance and reduced survival in melanoma remains significant after adjusting for patient age, sex, and TMB. The panel summarizes a multivariable Cox regression model, with hazard ratios shown as blue squares and 95% confidence intervals as blue lines.

Next, we examined whether the prevalence of AAS4 also influences disease progression in patients who did not receive immune checkpoint therapy. The mutational landscape of melanoma is shaped by diverse carcinogenic processes across its subtypes, some of which are unrelated to UV exposure.[88]. Patients with non-UV-associated melanomas generally have a poorer prognosis than those with UV-associated subtypes[88], although the underlying reasons remain unclear. We hypothesized that the observed prognostic differences could be partly explained by the enrichment of AAS4 in non-UV-associated melanoma genomes. In The Cancer Genome Atlas (TCGA), the majority of melanoma genomes are dominated by AAS2 (n = 407), likely reflecting UV-induced mutagenesis, whereas AAS4-dominant samples are relatively rare (n = 16). As anticipated, AAS4-dominant tumors were associated with significantly worse survival compared to AAS2-dominant tumors (Figure 8B). This effect remained significant in a multivariate Cox regression model after adjusting for age and tumor mutational burden (Figure 8C). While these findings are consistent with our hypothesis, additional data will be required to fully establish the prognostic implications of AAS4 in antitumor immunity.

The specificity of HLA-I alleles is quantifiable on amino acid level.

We hypothesized that certain HLA alleles would exhibit a binding preference for the amino acid changes characteristic of individual AASs. Ideally, these patterns would correspond to the five major AASs previously identified, each associated with distinct environmental or intrinsic mutational processes. To test this hypothesis, we established a stepwise analytical framework that combined structural immunopeptidomics data with mutational profiles from tumor genomes.

Our first task was to establish a quantitative representation of HLA-I specificity at the level of individual amino acid substitutions. This level of resolution is necessary to link the mutational biases captured in AASs to the peptide presentation capacity of HLA variants. To achieve this, we analyzed over 185,000 experimentally validated HLA-peptide interactions derived from a large-scale immunopeptidomics dataset[47] encompassing 92 common HLA-A, -B, and -C alleles. Notably, our analysis included all HLA-A and B alleles from a widely recognized reference set that offers maximal population coverage[89].

For each allele, we calculated amino acid preferences using a position-weighted entropy correction method described below, yielding a unique specificity score for each amino acid–allele pair (Figure 9-B).

Determining the amino acid specificity of HLA-I alleles.

We obtained a large immunopeptidomic dataset from Sarkizova et al[47], which includes data for 92 HLA-I alleles. We focused on alleles with at least 400 reported peptides between 8 and 12 amino acids in length. Peptides of different lengths were analyzed separately, and only lengths with at least 100 peptide sequences were analyzed for each allele. For each subset of peptides, we calculated the relative frequency of each amino acid at every position. These frequencies were then weighted by the positional importance, following established methods[90]. Specifically, we calculated the amino acid entropy at each position, reducing the weight of positions with high entropy and emphasizing those with more specificity. Amino acid–specificity values were obtained by summing the weighted frequencies across all positions, resulting in subset-specific specificity scores. An overall specificity value for each amino acid–allele pair was then calculated by averaging these subset-specific scores, weighted by the relative abundance of each peptide length subset. This yielded a comprehensive amino acid specificity profile for every HLA allele. To determine the specificity of a given HLA allele toward mutated amino acids associated with AAS1–5, coefficient values of each AAS were first aggregated by mutated residue, such that substitution-specific coefficients producing the same amino acid change were summed. These aggregated values were multiplied by the corresponding HLA amino acid specificity values and subsequently summed. For analyses at the patient level, the six allele-specific values per individual were averaged. Information count matrix–based binding logos shown in Figure 10B were generated using the universalmotif R package, based on the sequences of 9–amino-acid–long peptides from Sarkizova *et al*[47].

Validating amino acid specificity values of HLA-I alleles.

We next assessed whether the derived amino acid specificity values accurately predict peptide presentation on the cell surface. For this validation, we utilized an independent large-scale immunopeptidomics dataset[91]. In this study, HLA-bound peptides were identified from peripheral blood mononuclear cells (PBMCs) obtained from healthy donors. Our analysis focused on 9-mer peptides, and the predictive performance of the amino acid specificity values was evaluated as follows. All 9-mer peptides detected across samples were pooled. For each sample, HLA-specific scores were calculated for every 9-mer peptide. For each HLA allele present in a given sample, position-specific amino acid specificity values were averaged across the peptide sequence, and the highest resulting average value was selected to represent the strongest predicted HLA binding within that sample.

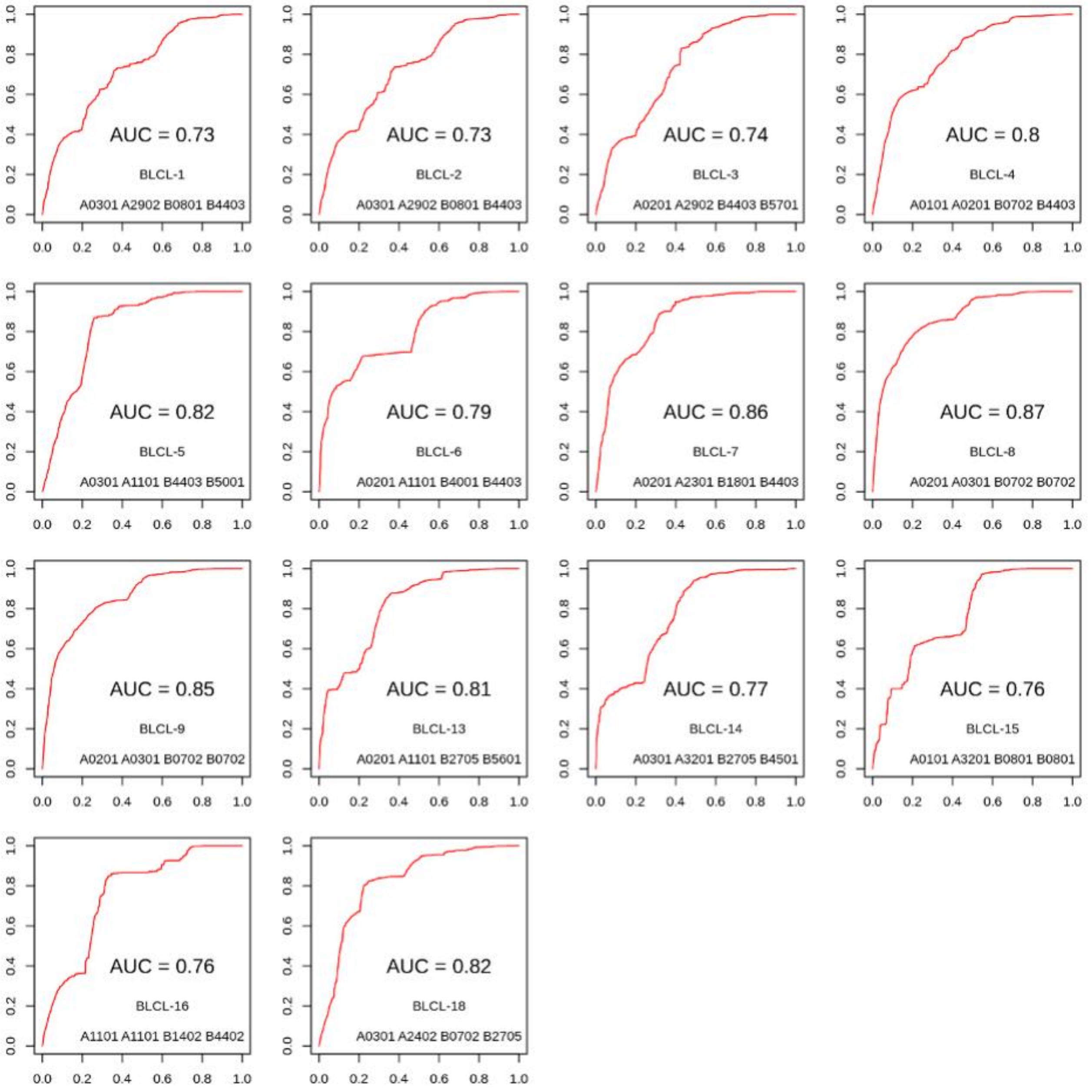


Figure 9-A. The calculated amino acid specificity values are accurate. Area under the ROC curve is shown for different samples. The HLA-A and B alleles of each sample are also indicated.

Peptides were then classified as negative (not detected on the cell surface of the sample) or positive (identified as HLA-bound in the sample). Receiver operating characteristic (ROC) analyses were performed separately for each sample using the calculated scores and peptide classification. The resulting areas under the curve (AUC) ranged from 0.73 to 0.87, indicating strong predictive performance of the HLA amino acid specificity values (Figure 9-A). This approach allowed us to capture the biochemical biases encoded in the peptide-binding grooves of HLA molecules.

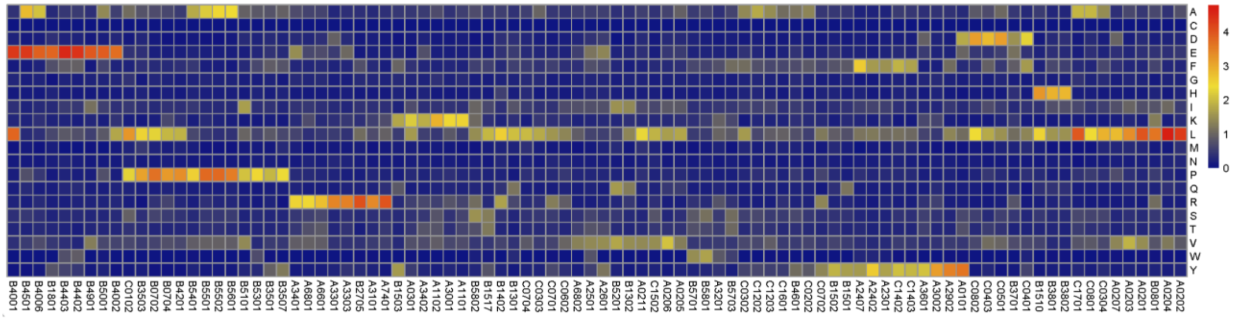


Figure 9-B. The affinity of HLA alleles for the 20 amino acids. The values representing amino acid preferences of HLA alleles are shown color-coded. The columns (alleles) are clustered using the Euclidean distance and the Ward D2 hierarchical clustering method.

Allele-specific antigen presentation explains immune heterogeneity in AAS4-dominant tumors.

Having defined the specificity landscape of HLA-I molecules, we next asked whether certain alleles show enhanced affinity for amino acid substitutions characteristic of specific AASs. We focused this analysis on AAS4, a mutational signature previously shown to be associated with immunologically "cold" tumors and poor clinical outcomes. Interestingly, despite the general immunosuppressive environment of AAS4-dominant tumors, a subset of these samples displayed enriched lymphocyte infiltration. We reasoned that the presence of HLA alleles capable of efficiently presenting AAS4-specific neopeptides could explain this discrepancy.

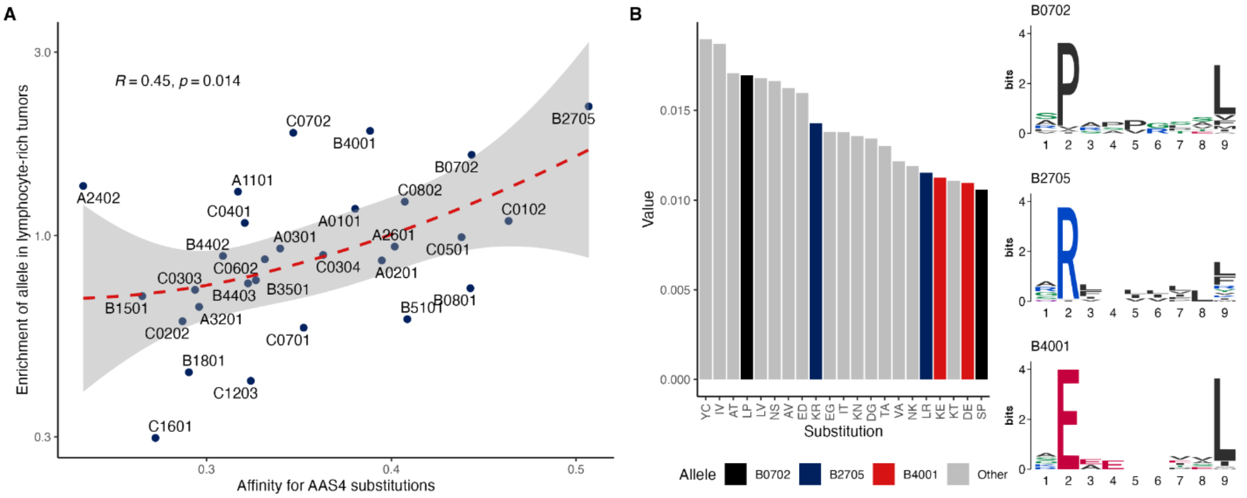


Figure 10. Amino acid substitution signatures modulate antitumor immunity through HLA allele-specific interactions. A) HLA alleles with higher predicted binding affinity for AAS4-associated substitutions are more frequently detected in tumors with high lymphocyte infiltration. For each allele, the odds ratio of occurrence in lymphocyte-rich versus lymphocyte-poor samples is plotted against its predicted specificity for substitutions enriched in AAS4-dominant tumors. Spearman correlation coefficient (rho) and the two-sided p-value are reported. **B)** HLA alleles enriched in lymphocyte-infiltrated, AAS4-dominant tumors preferentially bind amino acid substitutions characteristic of AAS4. Substitutions are ranked by their degree of association with AAS4, and binding motifs (logos) are shown for three representative HLA-B alleles commonly found in lymphocyte-rich tumors.

Mutated amino acids associated with AAS4, such as valine (V), threonine (T), and serine (S), are generally poorly presented by most HLA-I variants (Figure 10), which may contribute to the immunologically cold phenotype observed in many AAS4-dominant tumors. However, HLA-B*27:05, HLA-B*07:02, and HLA-B*40:01 exhibit increased specificity toward the AAS4-associated substitutions, distinguishing them structurally and functionally from the broader HLA-I repertoire. (Figure 10B) Furthermore, we showed that these alleles were indeed more frequent in the immune-infiltrated subset of AAS4 tumors (Figure 10A). Among these candidates, HLA-B*07:02, which is present in up to 20% of the human population[92] emerged as the most frequently observed allele in tumor samples (Figure 10A), making it a logical and robust choice for subsequent experimental validation.

Mutagen exposure generates distinct and predictable amino acid substitution profiles.

To experimentally test whether specific HLA alleles can indeed present neopeptides arising from distinct mutational processes, we designed an in vitro system using chemical mutagens with known signatures. Our aim was to mimic the amino acid substitution patterns associated with

AAS1, AAS2, and AAS4 in a controlled setting and then assess the immunogenicity of resulting neopeptides in the context of different HLA genotypes.

We selected three mutagens, each linked to a different AAS: benzo[a]pyrene (BAP), a component of tobacco smoke, associated with AAS1-type substitutions; simulated solar UV light, representative of AAS2; and N-ethyl-N-nitrosourea (ENU), a potent alkylating agent associated with AAS4 based on our *in silico* analyses. A549 monoallelic cells were subjected to repeated exposure to each mutagen, undergoing up to 70 cycles of treatment. Following mutagenesis, whole-genome sequencing and somatic variant calling were performed to identify the resulting amino acid substitutions (Figure 11, see Methods).

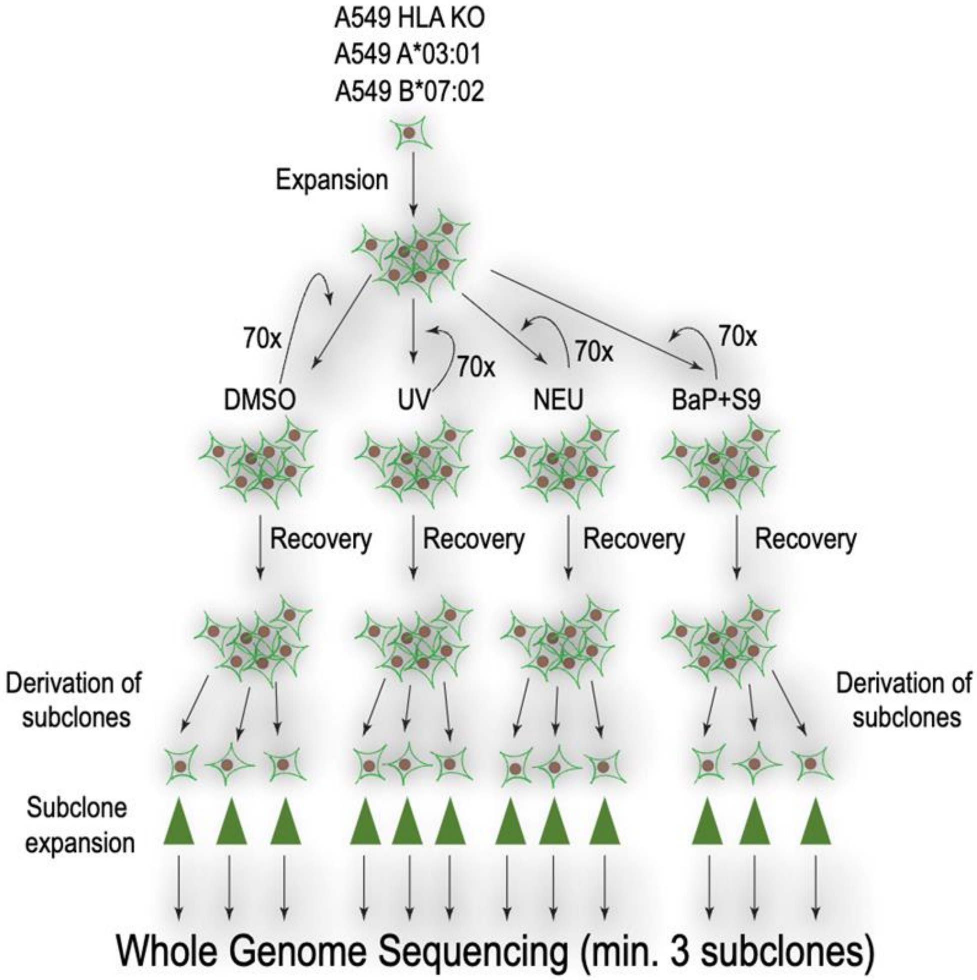


Figure 11. Schematic of the experimental pipeline for mutation induction and subclone analysis. The diagram outlines the procedure used to generate and sequence subclones from A549 cells engineered to express HLA-A**03:01 and HLA-B**07:02 or lacking HLA expression (KO). After cell expansion, four mutagenic conditions were applied over 70 cycles: DMSO as a solvent control; UV irradiation (10% UVB, 90% UVA; total dose 1.25 J); N-ethyl-N-nitrosourea (NEU, 400 µM); and benzo[a]pyrene (0.39 µM) with S9 metabolic activation (BaP+S9). Following exposure and recovery, subclones were isolated and expanded. At least three subclones per treatment group underwent whole-genome sequencing to identify mutational patterns induced by the respective mutagens.

The mutation profiles differed markedly across the three treatments. While the number of amino acid substitutions varied widely, from a few dozen to several thousand depending on the mutagen, the substitution patterns themselves showed strong alignment with our computational expectations. Specifically, there was a high concordance between the experimentally observed amino acid substitutions and those predicted by *in silico* simulations (Figure 12). These simulations were based on DNA-level mutational signatures catalogued in the COSMIC[46] database. By projecting DNA mutations onto coding sequences and translating them into protein-level changes, we established a reference for the expected substitution spectra of each mutagen. The consistency between experimental results and simulations confirmed that our model could accurately capture mutagen-driven changes at the amino acid level.

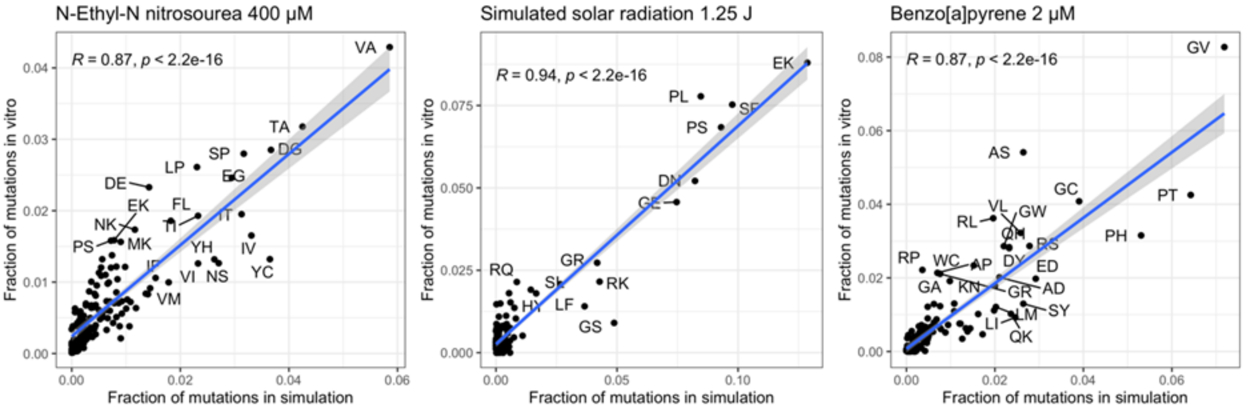


Figure 12. Simulated mutational profiles accurately reflect experimentally observed amino acid substitution patterns. The plots display the correlation between the average frequency of amino acid substitutions identified in experimental datasets and those predicted by computational simulations for three commonly used mutagens. Pearson correlation coefficients and corresponding two-sided p-values are indicated. Blue lines represent fitted linear regression models; shaded grey areas denote the 95% confidence intervals.

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

To test whether the recognition of specific mutational signatures by HLA molecules translates into a functional immune response, we designed a proof-of-concept experiment using chemically induced mutagenesis and in vitro immune assays. Our aim was to evaluate whether neoantigens generated by a mutagen associated with AAS4 could trigger T cell proliferation in an HLA-restricted manner.

We selected the N-ethyl-N-nitrosourea (ENU) treated cell line from our experiment, as it best represents the AAS4 mutational signature. A549 monoallelic lung carcinoma cells were used as the experimental model. Three isogenic cell lines were prepared: one expressing HLA-B*07:02, our candidate allele predicted to recognize AAS4-derived substitutions; one expressing HLA-A*03:01, a common but functionally unrelated control allele; and one knockout line lacking HLA class I expression (B2M KO), serving as a negative control (Figure 11).

Each cell line underwent repeated ENU exposure, up to 70 treatment cycles, mimicking chronic mutagenic stress. Following mutagenesis, whole-genome sequencing was performed on clonal isolates. All cell lines accumulated substantial numbers of missense mutations, but the total mutational burden differed slightly across conditions. Interestingly, the B*07:02-expressing line carried fewer missense mutations (n = 1414) than the control (n = 2715) and knockout (n = 1653) lines, suggesting that any immune effects would not simply reflect a higher neoantigen load. This setup allowed us to test whether specific HLA–mutation combinations drive immune activation.

To assess T cell responses, we co-cultured the ENU-treated A549 clones with CFSE-labelled peripheral blood mononuclear cells (PBMCs) obtained from healthy donors (Figure 13). The experimental setup included ENU-treated A549 cells expressing either HLA-A*03:01, HLA-B*07:02, or lacking HLA class-I alleles, alongside PBMCs with HLA-A*03:01 (control, n = 4), HLA-B*07:02 (n = 4), or both (n = 4). Each co-culture well contained 200,000 PBMCs and 4,000 ENU-treated A549 cells, and interleukin-2 was added on the second day to support T cell growth. After seven days, we used flow cytometry to assess CD3+ T cell proliferation by tracking CFSE dye dilution and excluding dead cells with propidium iodide staining[62,93]

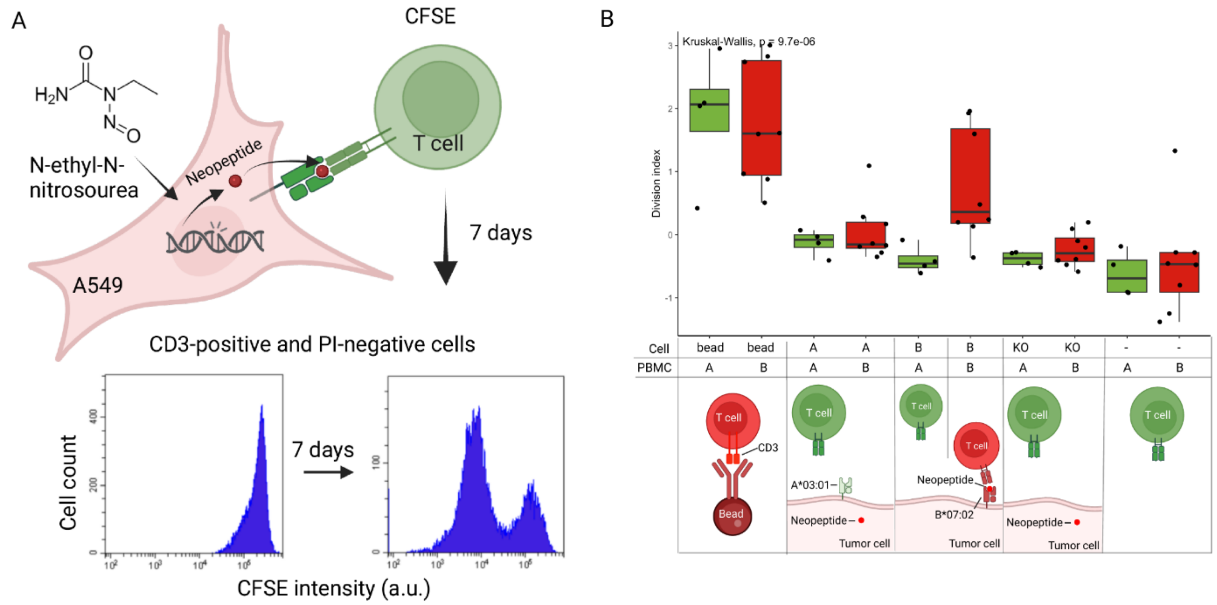


Figure 13. Amino acid substitution signatures modulate T cell responses in an HLA-restricted manner. A) Schematic overview of the experimental setup used to assess T cell proliferation following exposure to N-ethyl-N-nitrosourea (ENU)-treated tumor cells. Peripheral blood mononuclear cells (PBMCs) were labeled with CFSE (carboxyfluorescein diacetate succinimidyl ester) and co-cultured with ENU-treated A549 cells that expressed either HLA-A*03:01, HLA-B*07:02, or no HLA-I alleles (KO). PBMCs were derived from donors carrying A*03:01, B*07:02, or both. After seven days, T cell proliferation was evaluated by flow cytometry using CD3 for T cell gating

and propidium iodide (PI) for exclusion of non-viable cells. Proliferation was quantified based on CFSE dilution, with divided (G1 peak) and undivided (G0 peak) cells identified. The Division Index, representing total cell divisions normalized to the starting population, was used to estimate the magnitude of T cell activation in response to tumor-specific ENU-induced neoantigens. **B)** ENU-treated A549 cells expressing HLA-B*07:02 stimulate proliferation in PBMCs from B*07:02-positive donors. Z-transformed Division Index values are shown across groups. A549 cell types include: B (red) = HLA-B*07:02-expressing, A (green) = HLA-A*03:01-expressing, KO = HLA-null. PBMCs were categorized similarly. A Kruskal–Wallis test p-value is indicated; adjusted p-values from Dunn's post-hoc test were omitted for visualization clarity.

The results aligned with our expectations. ENU-mutagenized A549 cells expressing HLA-B*07:02 induced robust T cell proliferation, but only when co-cultured with PBMCs that also carried the B*07:02 allele (Figure 13B). This effect was not observed in any of the control conditions: neither the HLA-A*03:01-expressing cells nor the HLA knockout line stimulated significant proliferation, regardless of PBMC genotype (Figure 13B). These findings indicate that the response was both HLA-restricted and mutation-dependent. The presence of B*07:02 in both the antigen-presenting cells and the responding lymphocytes was necessary to achieve meaningful immune activation.

Summary and discussion

This thesis examines how tumor immunogenicity arises from the interplay between tumor-intrinsic mutational processes and host-mediated antigen presentation. Rather than focusing solely on mutation counts or individual genomic alterations, the work introduces amino acid substitution signatures (AASs) as an intermediate, protein-level representation of mutagenesis that is directly relevant to immune recognition. By capturing recurrent patterns in the types of amino acid changes generated by different mutational processes, AASs provide a biologically interpretable link between genomic alterations and the properties of resulting neopeptides.

The results presented in this thesis are consistent with a coherent causal framework. First, distinct AASs are shown to arise from specific combinations of environmental exposures and endogenous DNA repair defects. These signatures are then associated with systematic differences in the biophysical properties of amino acid substitutions, which influence the predicted quality of tumor-derived peptides. At the tumor level, these biases translate into differences in immune microenvironment characteristics and clinical outcomes following immune checkpoint blockade. Importantly, the data support a model in which these effects are not uniform but depend on the compatibility between the tumor's substitution landscape and the patient’s HLA genotype.

Together, the findings motivate an interaction-based view of tumor immunogenicity, in which both mutational processes constrain immune recognition and host genetic factors, rather than determined by either in isolation.

From mutational processes to constraints on immune recognition.

Tumor mutational burden (TMB) has become one of the most widely used biomarkers for predicting benefit from immune checkpoint inhibitors[94] (ICIs), largely because it translates a central immunological premise into a quantifiable metric: tumors with more coding mutations are statistically more likely to generate neoantigens capable of eliciting T-cell responses[95]. This mutation-count-based framework has shaped both translational research and clinical decision-making, culminating in tumor-agnostic biomarker strategies built around “TMB-high” classifications. The conceptual appeal of TMB lies in its simplicity. Mutation quantity is measurable, scalable, and broadly comparable across cohorts.

However, accumulating evidence indicates that TMB is an imperfect surrogate for tumor immunogenicity[30]. Not all mutations give rise to peptides that are efficiently processed, presented, and recognized by T cells, and host antigen-presentation properties can significantly modulate this process. Indeed, characteristics of HLA binding have been shown to reshape the relationship between mutation burden and clinical outcome, emphasizing that mutation counts do not translate uniformly into immunogenic potential across individuals[19]. Furthermore, neoantigen "quality" frameworks demonstrate that a limited number of highly immunogenic neoantigens can drive effective immune control, underscoring that qualitative features may outweigh sheer mutation quantity[37,39]. At the clinical level, universal TMB cutoffs perform inconsistently across tumor types and cohorts, and TMB has not consistently achieved robust validation as a predictor of overall survival benefit[30].

Consistent with this evolving perspective, our results show that amino acid substitution signatures (AASs) are associated with distinct tumor immune microenvironment states defined by the pan-cancer immune subtypes of Thorsson et al.[23], independently of both TMB and tumor type. These findings suggest that mutational structure at the protein level encodes immunologically relevant information beyond mutation counts alone, providing a mechanistic layer that may help explain immune heterogeneity among tumors with comparable TMB.

Tumor–host interactions as the foundation of precision immunotherapy.

In parallel with mutation-based approaches, precision immuno-oncology has developed a broad spectrum of biomarkers to predict benefit from immune checkpoint inhibitors (ICIs). These include tumor-intrinsic markers such as PD-L1 expression on tumor or immune cells, microsatellite instability (MSI) and mismatch repair deficiency (dMMR), oncogenic driver alterations; tumor–immune microenvironment features such as CD8^{+} T-cell infiltration, interferon-γ–related gene-expression signatures, and T-cell–inflamed transcriptomic profiles; as well as emerging blood-based markers, including circulating tumor DNA dynamics and peripheral immune-cell phenotypes[96,97]. While several of these biomarkers have demonstrated predictive value in specific tumor types, their performance is often context-dependent, and no single parameter consistently stratifies responders across cancers. This has led to increasing interest in composite or integrative predictors that combine tumor-intrinsic and immune-context features[96].

Within this landscape, HLA genetics occupies a distinct conceptual position. Unlike PD-L1 or gene-expression signatures, which reflect dynamic tumor–immune interactions, HLA molecules constitute the host-intrinsic machinery that determines which tumor-derived peptides can be presented to T cells. Large-scale clinical analyses have demonstrated that patient HLA genotype can influence ICI outcomes. Chowell *et al.* showed that HLA class I genotype features, including heterozygosity, are associated with improved survival in patients treated with checkpoint blockade[98]. Naranbhai *et al.* further identified specific alleles, such as HLA-A*03, that stratify ICI benefit across pooled clinical-trial cohorts[99]. Complementing class I findings, Shao *et al.* reported that an HLA class II–restricted immunogenic mutation burden identifies melanoma and NSCLC patients with longer survival after ICB, suggesting that class II–mediated antigen presentation contributes independently to therapeutic efficacy[100]. Additional studies have linked HLA class II expression and related immune phenotypes to inflamed tumor microenvironments and checkpoint responsiveness[101,102]. Moreover, our previous work indicated that HLA class I peptide-binding promiscuity is linked to differences in antitumor immune responses and outcomes following immune checkpoint blockade, highlighting the importance of host antigen-presentation characteristics[19]. Collectively, these data underscore that HLA-dependent antigen presentation is not a peripheral variable, but a central determinant of tumor immune recognition.

Yet, despite this accumulating evidence, a mechanistically interpretable link between host HLA variation and tumor mutational structure has remained insufficiently defined. Our results address this gap by demonstrating explicit AAS–HLA interaction effects: specific combinations of amino acid substitution signatures and HLA presentation properties associate with distinct tumor immune phenotypes. By integrating tumor-derived mutational structure with host-specific antigen presentation constraints, our framework advances a precision-medicine paradigm in which both tumor and host characteristics jointly determine immunogenicity and clinical behaviour.

Limitations and conceptual boundaries.

Several limitations should be considered when interpreting the findings of this thesis. First, the analyses focus primarily on missense mutations and the resulting amino acid substitutions. Pan-cancer genomic analyses have consistently shown that missense substitutions constitute the majority of coding somatic mutations in solid tumors[103,104]. Nevertheless, other mutation types, particularly frameshift mutations, insertions and deletions, can generate highly immunogenic neoantigens despite their lower frequency[105,106]. Consequently, the conclusions of this thesis pertain to a quantitatively dominant but not exhaustive subset of tumor-derived alterations.

Second, the study adopts a largely static view of the tumor mutational landscape and immune microenvironment. Tumors undergo continuous clonal evolution, generating spatial and temporal heterogeneity in both mutational composition and neoantigen repertoires[107,108]. Immune pressure and therapeutic interventions can further reshape tumor genomes and antigen presentation pathways, as illustrated by longitudinal analyses of patients treated with immune checkpoint inhibitors[109,110]. Moreover, mutational processes themselves may shift during disease progression or in response to treatment, leading to dynamic changes in substitution signatures[7,110]. Therefore, the AAS-based analyses presented here capture snapshots of tumor biology rather than the full evolutionary trajectories that characterize tumor–immune interactions over time.

Additional constraints arise from data availability and modeling assumptions inherent to large-scale immunogenomic analyses. The study relies on predicted peptide–HLA binding, yet mass spectrometry–based immunopeptidomics has shown that binding affinity alone does not fully determine which peptides are naturally presented[78,111]. Antigen processing efficiency, proteasomal cleavage, TAP transport, and intracellular peptide loading introduce additional regulatory layers that are not explicitly modeled in binding-based approaches[21,112]. Likewise, immune phenotypes are inferred from bulk transcriptomic data, which capture aggregate immune states but do not resolve the spatial organization of immune cells within tumors[23,113,114]. These limitations reflect common constraints in computational immunogenomics rather than specific shortcomings of the present work. Together, they define the biological and methodological boundaries within which the proposed interaction-based model of tumor immunogenicity should be interpreted.

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.

I am deeply grateful to Dr. Zsuzsánna Réthi-Nagy, Adrienn Kobl, Alexandra Gyémánt, and Márton Czikkely for the kind of support that extends far beyond professional roles. Zsuzsánna's shared vision for the future of our work and her unwavering commitment to our common goals have been a constant source of inspiration, and from her I learned invaluable lessons about perseverance and the refusal to give up. Adrienn's sense of realism and remarkable communication skills repeatedly helped bring my impulsive, world-changing ambitions closer to what can be built in practice. Alexandra's unwavering emotional support and true friendship carried me through the most demanding periods of this journey. I am also grateful to Márton, whose example as a scientist has served as a constant reminder of why remaining in this field is both meaningful and worthwhile. Among them, I found what, in many ways, is the closest thing I may call a family.

References

1. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**, 821-836.e16 (2019).
2. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).
3. Nik-Zainal, S. *et al.* The genome as a record of environmental exposure. *Mutagenesis* gev073 (2015) doi:10.1093/mutage/gev073.
4. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
5. Volkova, N. V. *et al.* Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).
6. Zou, X. *et al.* A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 (2021).
7. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Rep.* **3**, 246–259 (2013).
8. Butler, K. & Banday, A. R. APOBEC3-mediated mutagenesis in cancer: causes, clinical significance and therapeutic potential. *J. Hematol. Oncol.J Hematol Oncol* **16**, 31 (2023).
9. Kondo, N., Takahashi, A., Ono, K. & Ohnishi, T. DNA Damage Induced by Alkylating Agents and Repair Pathways. *J. Nucleic Acids* **2010**, 543531 (2010).
10. Westcott, P. M. K. *et al.* Mismatch repair deficiency is not sufficient to elicit tumor immunogenicity. *Nat. Genet.* **55**, 1686–1695 (2023).

11. Cummings, A. L. *et al.* Mutational landscape influences immunotherapy outcomes among patients with non-small-cell lung cancer with human leukocyte antigen supertype B44. *Nat. Cancer* **1**, 1167–1175 (2020).

12. Chong, W. *et al.* Association of clock-like mutational signature with immune checkpoint inhibitor outcome in patients with melanoma and NSCLC. *Mol. Ther. - Nucleic Acids* **23**, 89–100 (2021).

13. Szpiech, Z. A. *et al.* Prominent features of the amino acid mutation landscape in cancer. *PLOS ONE* **12**, e0183273 (2017).

14. Gulhan, D. C. *et al.* Predicting response to immune checkpoint blockade therapy among mismatch repair-deficient patients using mutational signatures. Preprint at https://doi.org/10.1101/2024.01.19.24301236 (2024).

15. Boiarsky, D. *et al.* A Panel-Based Mutational Signature of Mismatch Repair Deficiency is Associated With Durable Response to Pembrolizumab in Metastatic Castration-Resistant Prostate Cancer. *Clin. Genitourin. Cancer* **22**, 558-568.e3 (2024).

16. Boichard, A. *et al.* APOBEC-related mutagenesis and neo-peptide hydrophobicity: implications for response to immunotherapy. *OncoImmunology* **8**, 1550341 (2019).

17. DiMarco, A. V. *et al.* APOBEC Mutagenesis Inhibits Breast Cancer Growth through Induction of T cell–Mediated Antitumor Immune Responses. *Cancer Immunol. Res.* **10**, 70–86 (2022).

18. Driscoll, C. B. *et al.* APOBEC3B-mediated corruption of the tumor cell immunopeptidome induces heteroclitic neoepitopes for cancer immunotherapy. *Nat. Commun.* **11**, 790 (2020).

19. Manczinger, M. *et al.* Negative trade-off between neoantigen repertoire breadth and the specificity of HLA-I molecules shapes antitumor immunity. *Nat. Cancer* **2**, 950–961 (2021).

20. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).

21. 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).

22. 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).

23. Thorsson, V. *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e14 (2018).

24. 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).

25. Marty, R. *et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell* **171**, 1272-1283.e15 (2017).

26. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).

27. Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).

28. Kreiter, S. *et al.* Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* **520**, 692–696 (2015).

29. Marty Pyke, R. *et al.* Evolutionary Pressure against MHC Class II Binding Cancer Mutations. *Cell* **175**, 416-428.e13 (2018).

30. Addeo, A., Friedlaender, A., Banna, G. L. & Weiss, G. J. TMB or not TMB as a biomarker: That is the question. *Crit. Rev. Oncol. Hematol.* **163**, 103374 (2021).

31. Sun, S. *et al.* The role of neoantigens and tumor mutational burden in cancer immunotherapy: advances, mechanisms, and perspectives. *J. Hematol. Oncol.J Hematol Oncol* **18**, 84 (2025).

32. Budczies, J. *et al.* Tumour mutational burden: clinical utility, challenges and emerging improvements. *Nat. Rev. Clin. Oncol.* **21**, 725–742 (2024).

33. Jhunjhunwala, S., Hammer, C. & Delamarre, L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat. Rev. Cancer* **21**, 298–312 (2021).

34. Snyder, A. *et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).

35. Ribas, A. *et al.* Association of Pembrolizumab With Tumor Response and Survival Among Patients With Advanced Melanoma. *JAMA* **315**, 1600 (2016).

36. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).

37. McGranahan, N. & Swanton, C. Neoantigen quality, not quantity. *Sci. Transl. Med.* **11**, eaax7918 (2019).

38. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).

39. Łuksza, M. *et al.* Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).

40. Abelin, J. G. *et al.* Defining HLA-II Ligand Processing and Binding Rules with Mass Spectrometry Enhances Cancer Epitope Prediction. *Immunity* **51**, 766-779.e17 (2019).

41. Fritsch, E. F. *et al.* HLA-Binding Properties of Tumor Neoepitopes in Humans. *Cancer Immunol. Res.* **2**, 522–529 (2014).

42. Bell, M. J. *et al.* The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. *Mol. Immunol.* **46**, 1911–1917 (2009).

43. Miao, D. *et al.* Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat. Genet.* **50**, 1271–1281 (2018).

44. Wang, Y. *et al.* APOBEC mutagenesis is a common process in normal human small intestine. *Nat. Genet.* **55**, 246–254 (2023).

45. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* **160**, 48–61 (2015).

46. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).

47. 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).

48. Olsson, N. *et al.* An Integrated Genomic, Proteomic, and Immunopeptidomic Approach to Discover Treatment-Induced Neoantigens. *Front. Immunol.* **12**, 662443 (2021).

49. Cai, Y. *et al.* Immunopeptidomics-guided discovery and characterization of neoantigens for personalized cancer immunotherapy. *Sci. Adv.* **11**, eadv6445 (2025).

50. Pyke, R. M. *et al.* Precision Neoantigen Discovery Using Large-Scale Immunopeptidomes and Composite Modeling of MHC Peptide Presentation. *Mol. Cell. Proteomics* **22**, 100506 (2023).

51. Mounir, M. *et al.* New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLOS Comput. Biol.* **15**, e1006701 (2019).

52. Sethna, Z. *et al.* RNA neoantigen vaccines prime long-lived CD8+ T cells in pancreatic cancer. *Nature* **639**, 1042–1051 (2025).

53. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).

54. Degasperi, A. *et al.* A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. *Nat. Cancer* **1**, 249–263 (2020).

55. Jean-Marie, B. universalmotif. Bioconductor https://doi.org/10.18129/B9.BIOC.UNIVERSALMOTIF (2018).

56. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6**, (2013).

57. Delhomme, T. M. *et al.* Proton and alpha radiation-induced mutational profiles in human cells. *Sci. Rep.* **13**, 9791 (2023).

58. Muradyan, A. *et al.* Acute High-Dose X-Radiation-Induced Genomic Changes in A549 Cells. *Radiat. Res.* **175**, 700 (2011).

59. Zhao, H., Traganos, F. & Darzynkiewicz, Z. Kinetics of the UV-induced DNA damage response in relation to cell cycle phase. Correlation with DNA replication. *Cytometry A* **77A**, 285–293 (2010).

60. Jiang, Y. *et al.* Methyl methanesulfonate induces necroptosis in human lung adenoma A549 cells through the PIG-3-reactive oxygen species pathway. *Tumor Biol.* **37**, 3785–3795 (2016).

61. Maser, E. *et al.* In vitro and in vivo genotoxicity investigations of differently sized amorphous SiO2 nanomaterials. *Mutat. Res. Toxicol. Environ. Mutagen.* **794**, 57–74 (2015).

62. Habib-Agahi, M., Phan, T. T. & Searle, P. F. Co-stimulation with 4-1BB ligand allows extended T-cell proliferation, synergizes with CD80/CD86 and can reactivate anergic T cells. *Int. Immunol.* **19**, 1383–1394 (2007).

63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

64. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

65. Broad Institute. Picard tools. *Broad Institute, GitHub repository* http://broadinstitute.github.io/picard/ (2018).

66. Pipek, O. *et al.* Fast and accurate mutation detection in whole genome sequences of multiple isogenic samples with IsoMut. *BMC Bioinformatics* **18**, 73 (2017).

67. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).

68. Liu, Y. & Wu, F. Global Burden of Aflatoxin-Induced Hepatocellular Carcinoma: A Risk Assessment. *Environ. Health Perspect.* **118**, 818–824 (2010).

69. Wu, H.-C. & Santella, R. The Role of Aflatoxins in Hepatocellular Carcinoma. *Hepat. Mon.* **12**, (2012).

70. Nicholas, B. *et al.* Identification of neoantigens in oesophageal adenocarcinoma. *Immunology* **168**, 420–431 (2023).

71. Shapiro, I. E., Huber, F., Michaux, J. & Bassani-Sternberg, M. Sensitive neoantigen discovery by real-time mutanome-guided immunopeptidomics. *Nat. Commun.* **16**, 7269 (2025).

72. Hirama, T. *et al.* Proteogenomic identification of an immunogenic HLA class I neoantigen in mismatch repair–deficient colorectal cancer tissue. *JCI Insight* **6**, e146356 (2021).

73. Newey, A. *et al.* Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. *J. Immunother. Cancer* **7**, 309 (2019).

74. Gloger, A., Ritz, D., Fugmann, T. & Neri, D. Mass spectrometric analysis of the HLA class I peptidome of melanoma cell lines as a promising tool for the identification of putative tumor-associated HLA epitopes. *Cancer Immunol. Immunother.* **65**, 1377–1393 (2016).

75. Koumantou, D. *et al.* Editing the immunopeptidome of melanoma cells using a potent inhibitor of endoplasmic reticulum aminopeptidase 1 (ERAP1). *Cancer Immunol. Immunother.* **68**, 1245–1261 (2019).

76. Chen, R. *et al.* Chemical Derivatization Strategy for Extending the Identification of MHC Class I Immunopeptides. *Anal. Chem.* **90**, 11409–11416 (2018).

77. Kalaora, S. *et al.* Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. *Oncotarget* **7**, 5110–5117 (2016).

78. Bassani-Sternberg, M. *et al.* Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *PLOS Comput. Biol.* **13**, e1005725 (2017).

79. Sim, M. J. W. & Sun, P. D. T Cell Recognition of Tumor Neoantigens and Insights Into T Cell Immunotherapy. *Front. Immunol.* **13**, 833017 (2022).

80. Chowell, D. *et al.* TCR contact residue hydrophobicity is a hallmark of immunogenic CD8$^+$ T cell epitopes. *Proc. Natl. Acad. Sci.* **112**, (2015).

81. Schmidt, J. *et al.* Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. *Cell Rep. Med.* **2**, 100194 (2021).

82. Bagaev, A. *et al.* Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. *Cancer Cell* **39**, 845-865.e7 (2021).

83. Ji, R.-R. *et al.* An immune-active tumor microenvironment favors clinical response to ipilimumab. *Cancer Immunol. Immunother.* **61**, 1019–1031 (2012).

84. Jayasingam, S. D. *et al.* Evaluating the Polarization of Tumor-Associated Macrophages Into M1 and M2 Phenotypes in Human Cancer Tissue: Technicalities and Challenges in Routine Clinical Practice. *Front. Oncol.* **9**, 1512 (2020).

85. Ning, Z.-K. *et al.* Molecular Subtypes and CD4+ Memory T Cell-Based Signature Associated With Clinical Outcomes in Gastric Cancer. *Front. Oncol.* **10**, 626912 (2021).

86. Beckhove, P. *et al.* Specifically activated memory T cell subsets from cancer patients recognize and reject xenotransplanted autologous tumors. *J. Clin. Invest.* **114**, 67–76 (2004).

87. Roh, W. *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* **9**, eaah3560 (2017).

88. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).

89. 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.* **110**, (2013).

90. Henikoff, S. & Henikoff, J. G. Position-based sequence weights. *J. Mol. Biol.* **243**, 574–578 (1994).

91. Pearson, H. *et al.* MHC class I–associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* **126**, 4690–4701 (2016).

92. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* gkz1029 (2019) doi:10.1093/nar/gkz1029.

93. Terrén, I., Orrantia, A., Vitallé, J., Zenarruzabeitia, O. & Borrego, F. CFSE dilution to study human T and NK cell proliferation in vitro. in *Methods in Enzymology* vol. 631 239–255 (Elsevier, 2020).

94. 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).

95. Klempner, S. J. *et al.* Tumor Mutational Burden as a Predictive Biomarker for Response to Immune Checkpoint Inhibitors: A Review of Current Evidence. *The Oncologist* **25**, e147–e159 (2020).

96. Jiang, Z., Zhou, Y. & Huang, J. A Combination of Biomarkers Predict Response to Immune Checkpoint Blockade Therapy in Non-Small Cell Lung Cancer. *Front. Immunol.* **12**, 813331 (2021).

97. Isaacs, J., Anders, C., McArthur, H. & Force, J. Biomarkers of Immune Checkpoint Blockade Response in Triple-Negative Breast Cancer. *Curr. Treat. Options Oncol.* **22**, 38 (2021).

98. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).

99. Naranbhai, V. *et al.* HLA-A*03 and response to immune checkpoint blockade in cancer: an epidemiological biomarker study. *Lancet Oncol.* **23**, 172–184 (2022).

100. Shao, X. M. *et al.* HLA class II immunogenic mutation burden predicts response to immune checkpoint blockade. *Ann. Oncol.* **33**, 728–738 (2022).

101. Mei, J. *et al.* HLA class II molecule HLA-DRA identifies immuno-hot tumors and predicts the therapeutic response to anti-PD-1 immunotherapy in NSCLC. *BMC Cancer* **22**, 738 (2022).

102. Wang, X. & Mao, N. Tumor-cell HLA-DR expression as a potential biomarker of immunotherapy response in hepatocellular carcinoma. *Front. Oncol.* **15**, 1700181 (2025).

103. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).

104. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).

105. Turajlic, S. *et al.* Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* **18**, 1009–1021 (2017).

106. Wang, S., He, Z., Wang, X., Li, H. & Liu, X.-S. Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *eLife* **8**, e49020 (2019).

107. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).

108. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

109. Australian Pancreatic Cancer Genome Initiative *et al.* Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).

110. Zaretsky, J. M. *et al.* Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N. Engl. J. Med.* **375**, 819–829 (2016).

111. 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).

112. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* **37**, 724–737 (2016).

113. Fridman, W. H., Pagès, F., Sautès-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306 (2012).

114. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e19 (2018).