immuno

MDPI

Article

Identification and Prioritization of Neoantigens Derived from Non-Synonymous Mutations in Melanoma Through HLA Class I Binding Prediction

Karina Trejo-Vázquez [1], Carlos H. Espino-Salinas [1,*], Jorge I. Galván-Tejada [1,*], Karen E. Villagrana-Bañuelos [1], Valeria Maeda-Gutiérrez [1], Carlos E. Galván-Tejada [1], Gloria V. Cerrillo-Rojas [2], Hans C. Correa-Aguado [3] and Manuel A. Soto-Murillo [1]

[1] Centro de Investigación e Innovación Biomédica e Informática, Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Zacatecas 98000, Mexico; ktrejov@uaz.edu.mx (K.T.-V.)
[2] Unidad Académica de Ciencias Químicas, Universidad Autónoma Zacatecas, Zacatecas 98000, Mexico
[3] Unidad Profesional Interdisciplinaria de Ingeniería Campus Zacatecas, Instituto Politécnico Nacional, Zacatecas 98000, Mexico; hcorreaa@ipn.mx
* Correspondence: carlosespino@uaz.edu.mx (C.H.E.-S.); gatejo@uaz.edu.mx (J.I.G.-T.); Tel.: +52-492-113-9268 (C.H.E.-S.)

Abstract

Melanoma is characterized by a high mutational burden making it an established model for studying tumor neoantigens and developing strategies for personalized immunotherapy. In this study, a reproducible bioinformatics pipeline was developed and implemented for the identification and prioritization of candidate neoantigens derived from non-synonymous somatic mutations in melanoma, using genomic data from the MSK-IMPACT cohort (mel-mskimpact-2020; n = 696) and comparative reference information from TCGA-SKCM. From the somatic mutation annotation file (MAF), 16,311 non-synonymous mutations were filtered, from which 50,480 mutant 8–11-mer peptides were generated using a sliding-window approach centered on the mutated position. Peptide–HLA class I binding affinity was predicted using MHCflurry 2.0 across six representative alleles (HLA-A*02:01, HLA-A*24:02, HLA-B*35:01, HLA-B*39:05, HLA-C*04:01, and HLA-C*07:02). Candidate prioritization was initially based on predicted binding percentile (rank $\leq$ 2), identifying 12,209 peptide–HLA combinations with high predicted binding affinity. To refine candidate selection, additional computational analyses were incorporated, including proteasomal cleavage prediction using NetChop 3.1 and estimation of T-cell epitope immunogenicity using the Immune Epitope Database (IEDB) immunogenicity predictor. Furthermore, a direct comparison between mutant (MUT) and corresponding wild-type (WT) peptides was performed using Δaffinity and Δrank metrics to evaluate the predicted impact of somatic mutations on HLA binding. The analysis revealed a predominance of peptides associated with the HLA-B locus, particularly the allele HLA-B*35:01, among the interactions with the lowest predicted binding percentiles. Several high-ranking peptide candidates were derived from genes with known roles in melanoma biology, including PLCG2, GATA3, AKT1, PTEN, PTCH1, and SMO. Overall, the integrative computational framework implemented in this study enables the systematic prioritization of candidate neoantigens derived from non-synonymous mutations in melanoma. This pipeline provides a reproducible strategy for exploring tumor neoantigen repertoires and may serve as a foundation for subsequent experimental validation and for studies related to neoantigen-based immunotherapies and immunopeptidomics.

Academic Editor: Supriya Mahajan

Received: 17 February 2026
Revised: 24 March 2026
Accepted: 25 March 2026
Published: 27 March 2026

Keywords: melanoma; neoantigens; HLA class I; MHCflurry; immunogenicity prediction; bioinformatics

1. Introduction

The adaptive immune response to solid tumors involves the recognition of tumor-derived peptides presented by major histocompatibility complex class I molecules (MHC-I/HLA-I). Among these peptides, neoantigens generated from non-synonymous somatic mutations represent a class of tumor-associated epitopes that may be recognized by $CD8^+$ T cells [1,2]. Because these mutated sequences are absent from normal tissues during central tolerance, they can introduce antigenic determinants that differ from those derived from self-proteins [3].

Structural and functional studies of the T cell receptor–peptide–MHC (TCR–pMHC) complex have shown that even single amino-acid substitutions may influence the structural configuration modify the interface available for T-cell receptor engagement. Such structural changes may affect peptide–MHC complex stability and influence the quality of immune recognition. Therefore, the biological relevance of a neoantigen depends not only on the presence of a somatic mutation, but also on how that mutation affects peptide presentation and T-cell recognition [1,2].

However, the biological relevance of a neoantigen does not depend exclusively on its intrinsic ability to bind HLA-I but also on its evolutionary context within the tumor. Clonal neoantigens, which are present across all tumor subpopulations, have been associated with greater immunoreactivity and improved response to immune checkpoint blockade compared with subclonal neoantigens [4]. In parallel, PD-1 blockade has been shown to remodel the tumor antigen repertoire through immune editing, antigen loss, and alterations in antigen presentation pathways [5]. These findings indicate that the identification of clinically relevant neoantigens should consider not only peptide–HLA binding features but also clonal stability and tumor evolutionary dynamics [4,5].

Cutaneous melanoma represents a suitable model for neoantigen research because of its high tumor mutational burden, well-characterized genomic landscape, and recognized immunogenicity [6,7]. Its genomic architecture includes recurrent alterations in genes such as BRAF, NRAS, NF1, and TERT, as well as mutational processes associated with ultraviolet radiation exposure [6]. In addition, elevated tumor mutational burden has been associated with improved response to immune checkpoint inhibitors, particularly PD-1 blockade, further supporting the relevance of mutation-derived epitopes in this tumor type [7].

Large-scale genomic and immunological studies have provided valuable resources for the characterization of melanoma and its antitumor immune landscape [6,8]. In parallel, studies of tumor-infiltrating lymphocytes have identified $CD8^+$ and $CD4^+$ T-cell populations capable of recognizing mutation-derived epitopes in melanoma, underscoring the relevance of neoantigens in shaping antitumor immune responses [8,9].

The therapeutic relevance of neoantigens has also been explored through personalized immunotherapy strategies. Personalized vaccines based on tumor-specific peptides have been shown to induce neoantigen-specific T-cell responses and immunological memory in patients with melanoma [8]. More recently, personalized mRNA-based neoantigen vaccines administered in combination with immune checkpoint inhibitors have shown encouraging clinical results in patients with resected melanoma [9]. In parallel, tumor antigen-specific T-cell therapies, including T cell receptor-engineered approaches directed against mutation-derived epitopes, have further strengthened the translational relevance of neoantigen identification in cancer immunotherapy [10].

From a computational perspective, the identification of candidate neoantigens generally follows a multi-step analytical framework that includes the detection of somatic mutations, generation of mutant peptide sequences, and prediction of peptide binding to HLA class I molecules [11–13]. Although peptide–HLA binding affinity is an important parameter in this process, antigen presentation also depends on additional biological steps, including proteasomal processing, peptide transport, and peptide–MHC complex stability [14]. For this reason, recent methodological studies recommend the integration of multiple computational predictors to improve the prioritization of candidate neoantigens [11,13].

Machine learning-based predictors such as MHCflurry 2.0 have been developed to estimate peptide–HLA class I binding affinity across multiple alleles using models trained on large experimental datasets [15]. Additional computational tools can provide complementary information for neoantigen prioritization, including proteasomal cleavage prediction using NetChop and sequence-based immunogenicity estimation. In addition, comparison between mutant peptides and their corresponding wild-type sequences can help estimate the impact of somatic mutations on peptide–HLA interactions through metrics such as differences in predicted binding affinity or percentile rank.

In this context, melanoma provides a suitable system for implementing and evaluating computational pipelines for neoantigen identification. Therefore, the aim of the present study was to develop and implement a reproducible bioinformatics pipeline for the identification and prioritization of candidate neoantigens derived from non-synonymous somatic mutations in melanoma. This framework is based on in silico prediction of peptide–HLA class I interactions and integrates multiple computational features to support biological prioritization rather than experimental validation.

2. Scientific Basis

The accurate prioritization of tumor neoantigens represents a central requirement for the development of personalized immunotherapies, particularly in tumors characterized by a high tumor mutational burden such as melanoma. Despite advances in established bioinformatic predictors, only a fraction of coding mutations ultimately generate peptides capable of eliciting detectable immune responses. This limitation highlights the importance of implementing controlled and methodologically robust computational pipelines for the systematic prioritization of candidate neoantigens. Current clinical strategies—including personalized vaccines and T cell receptor-engineered T cell (TCR-T) therapies—depend critically on the quality of the initial set of computationally prioritized candidate peptides [5,6,9].

Structural evidence indicates that peptide–MHC (pMHC) binding affinity, although essential, is not sufficient to determine immunogenicity. Mutations affecting both anchor and non-anchor positions can modify the topology of the pMHC complex and alter the surface recognized by the TCR without substantial changes in HLA–peptide affinity [1,2]. Furthermore, tumor evolutionary dynamics determine the functional relevance of each epitope, with clonal neoantigens being more likely to generate clinically effective T cell responses than subclonal ones [16].

The predictive performance of computational neoantigen identification frameworks improves when additional biological variables are considered alongside peptide–HLA binding predictions. These factors include peptide–MHC complex stability, gene expression levels, antigen processing features, and physicochemical properties of peptide sequences [11–14,17]. However, the incorporation of these variables requires harmonized datasets and reproducible computational frameworks capable of integrating heterogeneous biological evidence.

In this context, melanoma provides a suitable system for evaluating neoantigen prediction pipelines because of its high mutational burden, extensive genomic characterization, and well-documented T cell-mediated immune responses [6,7]. Publicly available genomic resources have further strengthened its value as a reference model for computational immuno-oncology studies [18].

The mel-mskimpact-2020 dataset provides a suitable platform for implementing a reproducible computational workflow focused on the systematic generation and prioritization of neoantigens derived from non-synonymous mutations. This dataset includes detailed mutation annotations generated through clinical sequencing assays, enabling consistent extraction of coding variants for downstream computational analyses. The integration of mutation annotation, peptide generation, peptide–HLA binding prediction using machine-learning models such as MHCflurry 2.0, and complementary prioritization criteria—including proteasomal cleavage prediction and immunogenicity estimation—follows current methodological recommendations for computational neoantigen discovery in personalized immunotherapy research [9,12,19].

This approach is intended for the computational prioritization of candidate neoantigens based on in silico analyses and should be interpreted within the context of hypothesis generation for subsequent experimental validation.

3. Materials and Methods

This study implements a computational prioritization framework based on in silico immunogenomic analyses to identify and rank candidate neoantigens derived from somatic mutations reported in melanoma genomic datasets. The analytical workflow integrates publicly available genomic data with established bioinformatic prediction tools to evaluate peptide–HLA class I binding affinity together with additional features related to antigen processing and predicted T-cell immunogenicity. A schematic representation of the computational workflow used to identify and prioritize candidate neoantigens derived from somatic mutations in melanoma is shown in Figure 1.

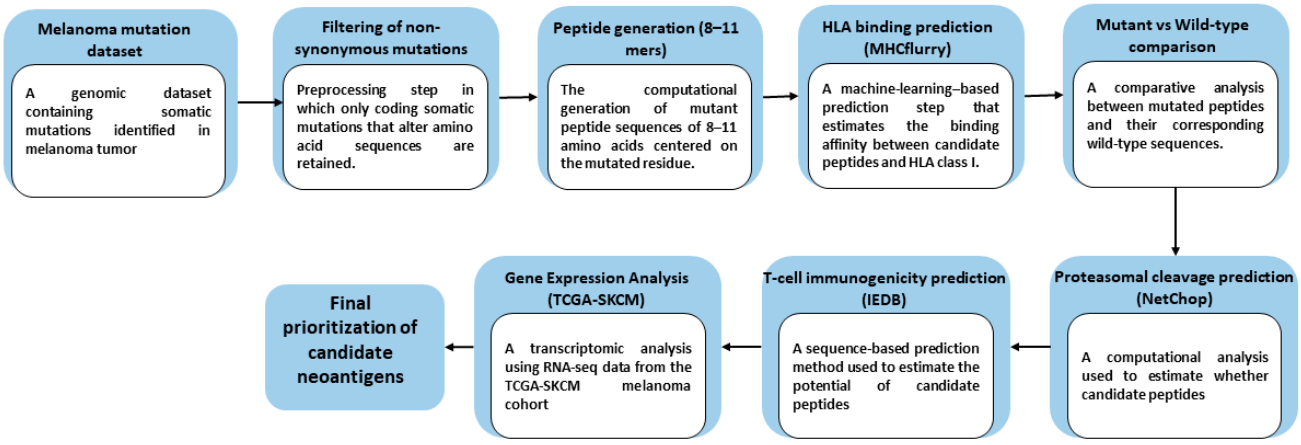

Figure 1. Pipeline overview.

The pipeline includes mutation filtering, generation of mutant peptides (8–11-mers), prediction of peptide–HLA class I binding using MHCflurry, comparison between mutant and wild-type peptides, proteasomal cleavage prediction using NetChop, T-cell immunogenicity prediction using the IEDB tool, and gene expression analysis using TCGA-SKCM transcriptomic data to prioritize biologically relevant neoantigen candidates. The bioinformatic tools and computational resources used throughout the pipeline are summarized in Table 1.

Table 1. Computational pipeline for neoantigen prediction from melanoma somatic mutations.

Pipeline Step	Tool/Resource	Version	Purpose
Somatic mutation dataset retrieval	cBioPortal (mel-skcmimpact-2020)	–	Access and retrieval of melanoma somatic mutation data used as input for the analysis
Mutation filtering and annotation	Genomic dataset processing	–	Selection of non-synonymous somatic mutations suitable for downstream neoantigen prediction
Mutant peptide generation	Custom sliding-window approach	–	Generation of mutant peptides of 8–11 amino acids centered on the mutated position
Peptide–HLA binding prediction	MHCflurry	2.0	Prediction of peptide binding affinity to HLA class I molecules using IC50 and percentile rank metrics
Strong binder selection	Binding affinity criteria	–	Identification of candidate peptides with predicted strong binding (percentile rank $\leq$ 2%)
Proteasomal cleavage prediction	NetChop	3.1	Prediction of proteasomal cleavage likelihood to assess antigen processing compatibility
Immunogenicity prediction	IEDB Epitope Immunogenicity Predictor	–	Estimation of predicted T-cell immunogenicity of candidate peptides
Mutant vs. wild-type comparison	Comparative binding analysis	–	Evaluation of Δaffinity and Δrank to estimate the impact of somatic mutations on HLA binding

All analyses were performed exclusively using computational approaches. No biological samples, cell-based assays, or experimental validation procedures were included in this study.

3.1. Somatic Mutation Data Collection

Somatic mutation data associated with melanoma were retrieved from publicly available cancer genomics resources through the cBioPortal for Cancer Genomics [20,21], including the mel-mskimpact-2020 melanoma cohort generated using the MSK-IMPACT clinical sequencing assay [22].

The MSK-IMPACT program represents a large clinical sequencing initiative that performed targeted next-generation sequencing across hundreds of cancer-related genes in more than 10,000 tumor samples [22].

Only non-synonymous somatic mutations located in protein-coding regions were retained for downstream analysis, as these variants may generate altered peptide sequences suitable for neoantigen prediction. Mutation annotations were extracted to preserve traceability between genes, genomic variants, and the peptide sequences generated in subsequent steps of the computational pipeline.

3.2. Generation of Mutant and Wild-Type Peptides

For each selected somatic mutation, mutant peptide sequences were generated to represent potential HLA class I epitopes derived from the altered proteins. Peptides were

generated using a sliding-window approach centered on the mutated amino acid position, producing peptide lengths compatible with HLA class I presentation (8–11 amino acids).

For each mutant peptide (MUT), the corresponding wild-type (WT) peptide sequence was also generated to enable direct comparison between mutated and non-mutated peptide variants during binding affinity prediction.

3.3. HLA Allele Selection

Six HLA class I alleles were included in the analysis: HLA-A*02:01, HLA-A*24:02, HLA-B*35:01, HLA-B*39:05, HLA-C*04:01, and HLA-C*07:02. These alleles were selected to represent the three classical HLA class I loci (HLA-A, HLA-B, and HLA-C) within the neoantigen prediction pipeline.

The selected panel was used as a representative set to evaluate peptide-binding predictions across multiple HLA loci while maintaining computational feasibility given the large number of candidate peptides generated. This design allowed the inclusion of alleles from each of the three classical HLA class I loci while keeping the computational workflow tractable and enabling locus-level comparison of predicted peptide-binding patterns, consistent with current computational neoantigen prediction frameworks [11–13].

3.4. Peptide–HLA Binding Affinity Prediction

Predicted binding affinity between candidate peptides and HLA class I molecules was evaluated using the machine learning-based prediction tool MHCflurry 2.0 [15]. This predictor estimates peptide–HLA binding using models trained on large experimental peptide–MHC binding datasets and reports two principal metrics: predicted binding affinity (IC50, expressed in nanomolar units) and percentile rank relative to a reference set of random peptides [15].

In this study, percentile rank values were used as the primary metric for peptide prioritization, as this measure provides a normalized estimate of binding strength across different HLA alleles. Peptides with predicted percentile ranks $\leq$ 2% were retained for downstream prioritization analyses as candidate binders. This threshold is commonly used in computational neoantigen discovery workflows to identify peptides with predicted HLA binding potential [11–13,15].

3.5. Proteasomal Cleavage Prediction

To incorporate an additional layer of analysis related to antigen processing, proteasomal cleavage prediction was performed using NetChop 3.1 with the C-terminal model [23]. This analysis was applied to the subset of peptides previously prioritized on the basis of peptide–HLA binding affinity predictions. Predicted cleavage sites were identified using the default threshold recommended by the NetChop algorithm.

The objective of this analysis was to evaluate whether candidate peptide sequences exhibited predicted cleavage patterns compatible with intracellular proteasomal processing, a key step in the HLA class I antigen presentation pathway.

3.6. Predicted T-Cell Immunogenicity Analysis

To incorporate an additional prioritization parameter, predicted T-cell immunogenicity scores were calculated using the epitope immunogenicity prediction tool available through the Immune Epitope Database (IEDB) [24]. This tool estimates the likelihood of T-cell recognition on the basis of amino acid sequence features derived from experimentally validated epitopes.

Immunogenicity predictions were applied to the subset of peptides previously prioritized based on peptide–HLA binding affinity. The resulting scores were used as com-

plementary information to rank candidate neoantigens together with binding affinity and antigen processing predictions.

3.7. Mutant vs. Wild-Type Binding Comparison

To evaluate the predicted impact of somatic mutations on HLA binding affinity, mutant peptides (MUT) were compared with their corresponding wild-type (WT) peptide sequences.

Two comparative metrics were calculated for each peptide pair:

$$\Delta\text{affinity} = \text{affinity\_}MUT - \text{affinity\_}WT \quad (1)$$

$$\Delta\text{rank} = \text{affinity\_}WT - \text{rank\_}MUT \quad (2)$$

These metrics allow direct assessment of the predicted effect of somatic mutations on peptide–HLA binding properties.

- Negative Δaffinity values indicate stronger predicted binding for the mutant peptide.
- Positive Δrank values indicate improved percentile rank for the mutant peptide relative to the corresponding wild-type sequence.

3.8. Neoantigen Candidate Prioritization

Candidate neoantigens were prioritized through an integrated computational framework that considered predicted peptide–HLA binding affinity (IC50 and percentile rank), predicted proteasomal cleavage patterns, predicted T-cell immunogenicity scores, and comparative evaluation of mutant and wild-type peptide sequences.

This framework was used to rank candidate peptides for further investigation and biological interpretation. The resulting candidates represent in silico prioritized neoantigens and should not be interpreted as experimental confirmation of antigen processing, peptide presentation, or immunogenicity.

3.9. Gene Expression Analysis

To provide transcriptomic context for the genes associated with prioritized neoantigen candidates, gene expression analysis was performed using RNA-seq data from the TCGA-SKCM cohort [25]. Expression values were obtained in transcripts per million (TPM) using the TCGAbiolinks package in R-Studio Versión: 4.3.0 [3].

For each gene associated with prioritized neoantigen candidates, the median expression level and the percentage of samples with detectable expression (TPM > 1) were calculated across the cohort. Because mutation data were derived from the MSK-IMPACT dataset and expression data from TCGA-SKCM, this analysis provides cohort-level transcriptomic information rather than matched mutation–expression measurements.

4. Results

This section presents the findings derived from the systematic computational analysis performed in this study. The results are organized according to the main stages of the workflow, including somatic mutation profiling, recurrent gene identification, and the prioritization of candidate neoantigens. Quantitative findings are emphasized in relation to their biological relevance and potential implications for personalized immunotherapy.

4.1. Cohort Characterization and somaTic Mutation Landscape

The analysis was conducted using the mel-mskimpact-2020 cohort [22], which comprises 696 melanoma tumor samples with paired tumor-normal sequencing profiles. As an

external genomic reference, the TCGA-SKCM cohort (n = 472 samples) was used to provide contextual information for the mutational landscape observed in the primary dataset [25].

The descriptive characteristics of the MSK-IMPACT cohort were consistent with the clinical and epidemiological profile commonly reported for advanced cutaneous melanoma, including a predominance of male patients, a mean patient age close to 60 years, and a high proportion of metastatic tumors.

At the genomic level, recurrent alterations were identified in well-characterized melanoma-associated genes, including BRAF, NRAS, NF1, TP53, and TERT, which are frequently reported in melanoma genomic studies [6,7].

After applying filtering criteria to retain only coding, non-synonymous somatic variants, a total of 16,311 mutations were selected for downstream analysis. These variants included missense substitutions, nonsense mutations, in-frame insertions or deletions, and splice-associated coding alterations. The distribution of these recurrent driver genes across both datasets is summarized in Table 2.

Table 2. Comparison of recurrent melanoma driver genes between the MSK-IMPACT cohort and the TCGA-SKCM dataset.

Gene	MSK-IMPACT (*n*)	TCGA-SKCM (*n*)
BRAF	329	281
NRAS	211	131
NF1	258	78
TP53	196	67
TERT	76	13

4.2. Generation of Mutant Peptide Repertoire

From the filtered set of non-synonymous somatic mutations, mutant peptide sequences were generated to represent potential neoantigenic epitopes derived from altered proteins.

Peptides were generated using a sliding-window approach centered on the mutated amino acid residue, ensuring that the altered position was included in each candidate peptide. Peptide lengths of 8–11 amino acids were considered, corresponding to the canonical size range associated with HLA class I presentation.

This procedure generated a total of 50,480 unique mutant peptide sequences, which constituted the initial set of candidate peptides derived from the mutational landscape of melanoma.

Quality control procedures were applied to verify the correspondence between each annotated mutation and the generated peptide sequence, ensure preservation of the correct reading frame, and exclude peptide sequences containing non-canonical amino acid residues.

The resulting peptide repertoire was subsequently used as input for peptide–HLA class I binding prediction and downstream neoantigen prioritization analyses.

4.3. Peptide–HLA Class I Binding Prediction

MHCflurry 2.0 [15] was used to predict binding affinity between the generated mutant peptides and HLA class I molecules. This machine learning-based algorithm is widely used for peptide–HLA binding prediction.

Binding predictions were performed for six representative HLA class I alleles: HLA-A*02:01, HLA-A*24:02, HLA-B*35:01, HLA-B*39:05, HLA-C*04:01, and HLA-C*07:02. These alleles were selected to represent the three classical HLA class I loci (HLA-A, HLA-B, and HLA-C) while maintaining computational feasibility given the large number of candidate peptides generated.

For each peptide–HLA combination, MHCflurry estimated both binding affinity (nM) and percentile rank, which reflects the relative strength of binding compared with a large reference set of random peptides.

Candidate neoantigens were initially prioritized using a percentile rank threshold $\leq 2\%$, a criterion commonly used in computational neoantigen prediction workflows to identify peptides with predicted HLA binding potential [11–13,15].

These peptide–HLA binding predictions formed the basis for subsequent neoantigen prioritization and downstream analyses.

4.4. Prioritization of Candidate Neoantigens

From the complete set of predicted peptide–HLA interactions, candidate neoantigens were prioritized on the basis of their predicted binding affinity and percentile rank obtained from the MHCflurry 2.0 analysis [15].

Peptides with a percentile rank $\leq 2\%$ were considered candidate binders and retained for further evaluation. This threshold is commonly used in computational neoantigen prediction workflows to identify peptides with predicted HLA binding potential [11–13,15].

Using this filtering criterion, a subset of candidate neoantigens with favorable predicted binding properties was identified from the initial peptide repertoire. These candidates correspond to peptides derived from somatic mutations with predicted capacity to bind HLA class I molecules.

To further refine candidate prioritization, predicted T-cell epitope immunogenicity scores were obtained for the top-ranked peptides using the IEDB epitope immunogenicity predictor [24].

Table 3 summarizes the top-ranked candidate neoantigens identified in the analysis, including the associated gene, peptide sequence, HLA allele, predicted binding affinity (nM), and predicted IEDB immunogenicity score.

Table 3. Integrated prioritization of top candidate neoantigens based on predicted peptide–HLA binding affinity (MHCflurry 2.0) and IEDB immunogenicity scores.

Gene	Peptide	HLA Allele	Binding Affinity (nM)	IEDB Immunogenicity Score
PLCG2	FPNDYTLSF	HLA-B*35:01	20.48	−0.053
GATA3	HASPHLFTF	HLA-B*35:01	20.49	0.078
SMO	NRPYAVILF	HLA-C*07:02	29.43	0.170
PDGFRB	YMAPDYNYV	HLA-A*02:01	10.40	0.011
PTEN	DHNPPQLEL	HLA-B*39:05	39.69	−0.084
AKT1	HPFLTALKY	HLA-B*35:01	23.17	−0.034
PTCH1	YPNGYPFLF	HLA-B*35:01	23.36	0.110
INPP4A	LFDALPREI	HLA-C*04:01	27.43	0.128
PTPRD	AYDHSRVLL	HLA-C*04:01	28.12	−0.044
ASXL1	LPAEIPPVF	HLA-B*35:01	24.32	0.247

4.5. Comparison Between Mutant and Wild-Type Peptides

To evaluate the effect of somatic mutations on predicted peptide-HLA binding, mutant (MUT) peptides were compared with their corresponding wild-type (WT) peptide sequences. Predicted binding affinity and percentile rank values were obtained for each peptide pair using MHCflurry 2.0 [15].

Comparison between mutant and wild-type peptides was assessed using Δaffinity and Δrank, as described in the methodological section. Negative Δaffinity and Positive Δrank values indicate improved predicted binding properties for the mutant peptide relative to the corresponding wild-type sequence.

Several mutant peptides exhibited lower predicted binding affinity values and improved percentile ranks relative to their corresponding wild-type counterparts, suggesting that specific amino acid substitutions may influence predicted peptide–HLA interactions. This comparative framework is commonly used in computational neoantigen prioritization workflows to estimate the impact of somatic mutations on peptide–HLA interactions and to identify mutation-derived peptides with enhanced binding potential relative to their wild-type counterparts [11–13,15] (Table 4).

Table 4. Comparison of Mutant vs. Wild-Type Peptide Binding Affinity for top candidate neoantigens with improved mutant binding properties.

Gene	Mutation	HLA Allele	Mutant Peptide	Wild-Type Peptide	Mut Affinity (nM)	WT Affinity (nM)	ΔAffinity	ΔRank
CREBBP	p.E3D	HLA-C*04:01	MADNLLDGP	MAENLLDGP	14,735.927	25,524.938	−10,789.011	15.233
RHOA	p.R5W	HLA-A*02:01	MAIAVWKLW	MAIARRKLV	6942.337	13,834.949	−6892.612	3.063
MITF	p.S3F	HLA-A*02:01	MQFESGVP	MQSESGVP	8379.331	15,010.266	−6630.935	3.109
DNAJB1	p.G2S	HLA-A*02:01	MSKDYQYIL	MGKDYQYIL	3938.190	9769.711	−5831.521	2.246
CREBBP	p.E3D	HLA-A*02:01	MADNLLDGP	MAENLLDGP	19,244.983	24,847.932	−5602.950	8.614
CREBBP	p.E3D	HLA-C*07:02	MADNLLDGP	MAENLLDGP	21,894.959	27,353.451	−5458.492	23.430
PAX5	p.D2N	HLA-A*02:01	MNELKYNPI	MDELKYNPI	12,537.777	17,461.802	−4924.025	2.977
MITF	p.S3F	HLA-C*07:02	MQFESGVP	MQSESGVP	10,690.248	15,265.057	−4574.809	4.938
PBRM1	p.S3F	HLA-C*07:02	MGFKRRRAT	MGSKRRRAT	18,998.191	23,555.303	−4557.112	11.552
SMAD3	p.S3F	HLA-A*02:01	MSIFLPTP	MSSILPTP	14,941.168	18,767.922	−3826.754	2.817

Binding affinity values were predicted using MHCflurry 2.0. Δaffinity was defined as the difference between mutant and wild-type binding affinity values (affinity_MUT–affinity_WT), whereas Δrank was defined as the difference between the percentile ranks of the wild-type and mutant peptides (rank_WT–rank_MUT). Negative Δaffinity values and positive Δrank values indicate improved predicted binding properties for the mutant peptide, reflected by lower predicted binding affinity values (IC50) and a better percentile rank relative to the corresponding wild-type sequence.

4.6. Proteasomal Cleavage Prediction

To evaluate whether prioritized peptide sequences were compatible with intracellular antigen processing, proteasomal cleavage prediction was performed using NetChop 3.1 (C-terminal model) [23]. Predicted cleavage sites were interpreted in the context of generating the C-terminal ends of candidate epitopes, a critical step in antigen processing prior to HLA class I presentation.

This analysis was applied to the subset of top-ranked candidate neoantigens identified during the peptide–HLA binding prediction stage. In total, 179 amino acid positions across the analyzed peptide sequences were evaluated. NetChop predicted 70 potential cleavage sites, corresponding to an estimated cleavage proportion of 39.1% across the analyzed sequences.

These results indicate that several candidate peptides exhibit sequence patterns compatible with predicted proteasomal processing, supporting their potential biological relevance within the antigen presentation pathway.

The results of the proteasomal cleavage analysis are summarized in Table 5.

Proteasomal cleavage prediction was performed using the NetChop 3.1 C-terminal model with default parameters. Predicted cleavage sites were defined as sequence positions assigned a positive cleavage prediction by the NetChop algorithm, corresponding to scores above the recommended threshold.

Table 5. Proteasomal cleavage analysis of top neoantigen candidates (NetChop 3.1, C-term model).

Analysis	Result
Total amino acids analyzed	179.0
Predicted cleavage sites	70.0
Cleavage proportion (%)	39.1

4.7. Predicted T-Cell Immunogenicity Analysis

To further characterize prioritized neoantigen candidates, predicted T-cell immunogenicity scores were calculated using the IEDB epitope immunogenicity prediction tool [24].

This analysis was applied to the subset of peptides previously prioritized on the basis of predicted peptide–HLA class I binding affinity. The IEDB predictor estimates immunogenicity scores using sequence-derived features associated with experimentally validated epitopes, thereby providing an additional computational layer for evaluating the likelihood of T-cell recognition [24].

The resulting scores were used as a complementary parameter for the characterization and prioritization of candidate peptides, together with predicted binding affinity and antigen processing features.

Among the analyzed peptides, several candidates exhibited positive predicted immunogenicity scores, including peptides derived from SMO, PTCH1, INPP4A, and ASXL1, suggesting greater potential for T-cell recognition.

The predicted immunogenicity scores for the prioritized candidate peptides are presented in Table 3.

4.8. Gene Expression Analysis of Neoantigen-Associated Genes

To provide additional transcriptomic context for genes associated with prioritized neoantigen candidates, gene expression levels were analyzed using RNA-seq data from the TCGA-SKCM melanoma cohort [25].

Expression values were quantified as transcripts per million (TPM) and summarized as the median expression across samples for each gene associated with prioritized candidate peptides. The use of TCGA transcriptomic data enabled evaluation of gene expression patterns at the cohort level, thereby providing relevant biological context for the potential impact of mutation-derived peptides [25].

Several genes linked to prioritized neoantigen candidates, including AKT1, SMO, ASXL1, PDGFRB, and PTEN, exhibited higher median expression levels than the remaining genes included in the analysis. In contrast, genes such as INPP4A, PLCG2, PTCH1, GATA3, and PTPRD showed lower median expression across the cohort.

In addition to expression magnitude, the proportion of samples with detectable expression (TPM > 1) was evaluated to describe the frequency of gene expression across the TCGA-SKCM dataset. Notably, genes such as PDGFRB, ASXL1, AKT1, and PTEN were expressed in nearly all samples, indicating consistent transcriptional activity across the cohort. In contrast, genes such as PTPRD and GATA3 were detected in a smaller proportion of samples, suggesting more heterogeneous expression patterns.

These expression patterns are illustrated in Figure 2A and Figure 2B, which summarize the median TPM values and the percentage of samples with detectable expression for genes associated with prioritized neoantigen candidates, respectively.

Because mutation data were obtained from the MSK-IMPACT dataset, whereas expression data were derived from TCGA-SKCM, this analysis provides cohort-level transcriptomic context rather than matched mutation–expression relationships.

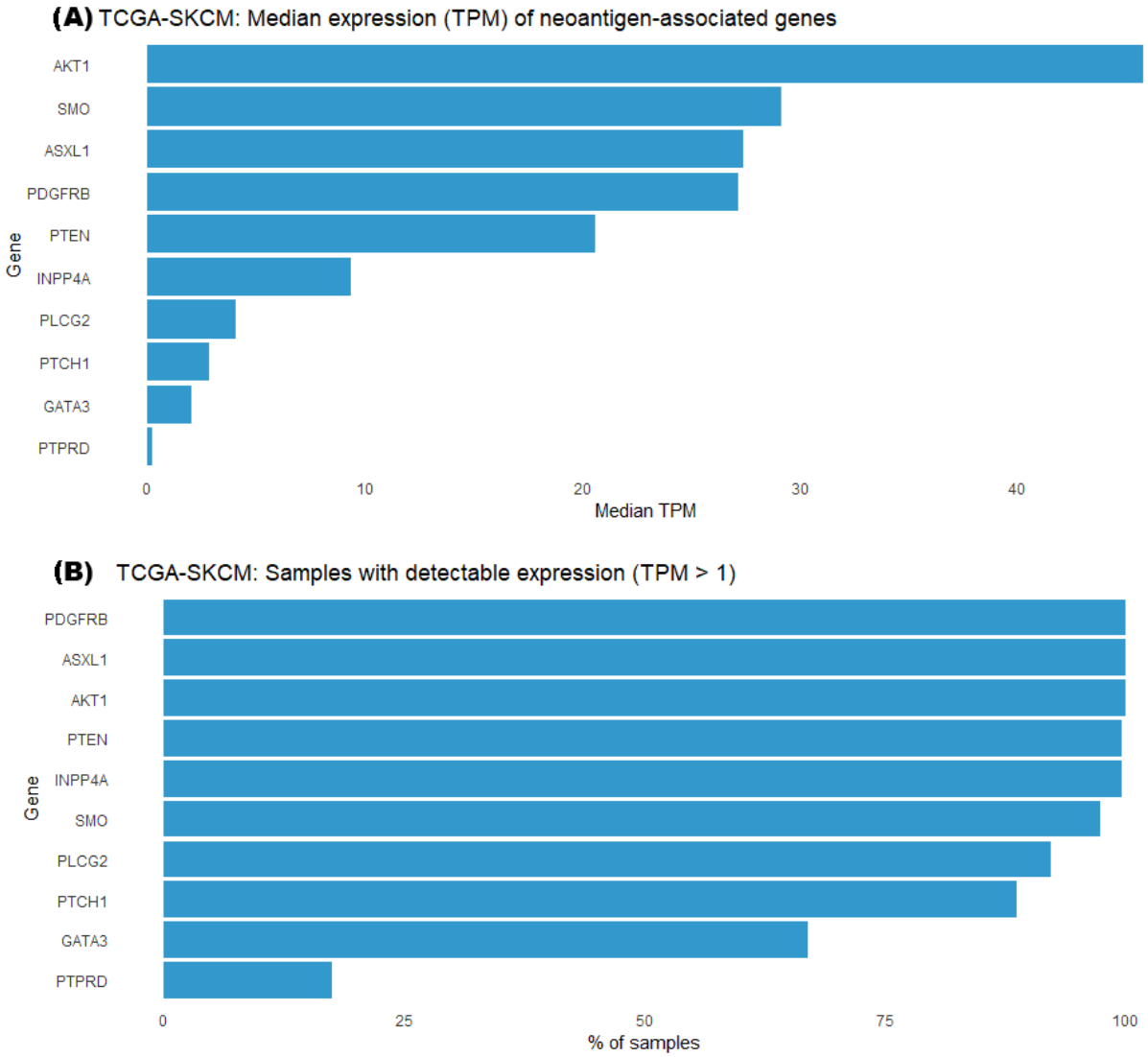

Figure 2. Transcriptomic context of genes associated with prioritized neoantigen candidates in the TCGA-SKCM cohort. (**A**) Median gene expression levels expressed as transcripts per million (TPM). (**B**) Percentage of samples with detectable expression (TPM > 1).

4.9. Overall Summary of Results

The computational workflow implemented in this study enabled the systematic identification and prioritization of candidate neoantigens derived from non-synonymous somatic mutations in melanoma. From an initial set of 16,311 coding variants, a total of 50,480 mutant peptide sequences were generated using a sliding-window approach.

Peptide–HLA binding predictions performed using MHCflurry 2.0 [15] identified a subset of 12,209 peptide–HLA combinations that met the predefined prioritization criterion based on percentile rank (≤2%), representing candidate peptides with predicted HLA binding potential.

Subsequent integration of complementary computational features, including proteasomal cleavage prediction and T-cell immunogenicity estimation, enabled further refinement of candidate neoantigens. Comparative analysis between mutant and wild-type peptides provided additional support for the predicted impact of somatic mutations on peptide–HLA binding properties.

The prioritized candidate peptides were derived from mutations affecting genes with diverse biological functions, including genes previously reported in melanoma genomic datasets.

The prioritized candidate peptides were derived from mutations affecting genes with diverse biological functions, including genes previously reported in melanoma genomic studies [6,7]. In addition, transcriptomic analysis using TCGA-SKCM data [25] provided cohort-level expression context for genes associated with prioritized neoantigen candidates, supporting their potential biological relevance.

Overall, these results describe a structured computational framework for the identification and prioritization of mutation-derived candidate neoantigens in melanoma. The findings should be interpreted within the context of in silico prediction and provide a basis for subsequent experimental validation and hypothesis-driven investigation in personalized cancer immunotherapy.

5. Discussion

The present study implemented a computational workflow for the identification and prioritization of candidate neoantigens derived from non-synonymous somatic mutations in melanoma. The analytical strategy integrated mutation filtering, mutant peptide generation, peptide–HLA class I binding prediction, mutant-versus-wild-type comparison, proteasomal cleavage prediction, sequence-based immunogenicity estimation, and transcriptomic contextualization using an independent melanoma cohort. Multi-step computational pipelines integrating these components are widely used in immunogenomic studies aimed at identifying mutation-derived peptide candidates in cancer [11–14,24].

Within this context, the study was explicitly framed as a computational prioritization framework intended to identify candidate neoantigens on the basis of in silico predictions. Accordingly, the results should be interpreted within the context of hypothesis generation and do not represent experimental validation of antigen processing, peptide presentation, or immunogenicity.

Predicted peptide–HLA interactions were observed across alleles from the HLA-A, HLA-B, and HLA-C loci, indicating that candidate peptides were distributed across the evaluated HLA repertoire. Within the analyzed allele panel, HLA-B*35:01 accounted for the largest proportion of peptide–HLA combinations within the selected percentile threshold.

Structural and computational studies of peptide–MHC complexes have shown that differences in peptide-binding repertoires among HLA molecules are associated with allele-specific binding motifs and the structural configuration of the peptide-binding groove [1,2]. However, the distribution observed in this study should be interpreted with caution, as the analysis was restricted to a representative panel of six HLA class I alleles.

The selection of HLA-A*02:01, HLA-A*24:02, HLA-B*35:01, HLA-B*39:05, HLA-C*04:01, and HLA-C*07:02 was intended to represent the three classical HLA class I loci while maintaining computational feasibility given the large number of mutant peptides generated. In particular, the inclusion of HLA-A*02:01 and HLA-A*24:02 facilitates comparison with prior neoantigen prediction frameworks and immunogenomic studies in which these alleles are frequently evaluated [12–14]. Nevertheless, restricting the analysis to a limited allele panel may influence the observed distribution of peptide–HLA interactions and limit the generalizability of the results.

Candidate prioritization relied primarily on peptide–HLA binding affinity and percentile rank predicted using MHCflurry 2.0 [15], a machine learning-based predictor trained on large experimental peptide–MHC binding datasets. Because percentile rank is commonly used in neoantigen discovery pipelines as a normalized metric that facilitates comparisons across different HLA alleles [11–13,15], it served as a central filtering criterion in the present analysis. In total, 12,209 peptide–HLA combinations met the predefined threshold of rank $\leq 2\%$. Even so, this relatively large candidate set reflects the permissive nature of this threshold and represents an initial prioritization step rather than a final selection of

therapeutic targets. More broadly, recent computational frameworks have emphasized that neoantigen prioritization benefits from the integration of multiple filtering and ranking criteria rather than reliance on peptide–HLA binding predictions alone [19,26–28].

An additional analytical component of the pipeline involved direct comparison between mutant peptides and their corresponding wild-type counterparts, enabling estimation of mutation-associated changes in predicted HLA binding using Δaffinity and Δrank metrics. In this respect, several mutant peptides exhibited lower predicted binding affinity values and improved percentile ranks relative to their corresponding wild-type sequences, suggesting that specific amino acid substitutions may influence predicted peptide–HLA interactions. Although these computational differences do not establish immunogenicity, they provide a useful criterion for identifying mutations associated with altered predicted binding properties [11–13,15,19,26].

Because peptide–HLA binding affinity alone does not fully capture antigen presentation probability, the pipeline incorporated proteasomal cleavage prediction using NetChop 3.1 [23]. Proteasomal processing represents a key step in the generation of peptides presented by HLA class I molecules, and computational prediction of cleavage sites is widely used to provide additional context in neoantigen discovery pipelines [14,23,27,28]. In line with this, the predicted cleavage patterns observed in this analysis indicate that several candidate peptides contain sequence features compatible with proteasomal processing.

The workflow also incorporated predicted T-cell immunogenicity estimation using the IEDB epitope immunogenicity tool [24]. As a complementary parameter to peptide-HLA binding affinity, immunogenicity prediction estimates sequence features associated with T-cell recognition. Previous studies have shown that peptide–HLA affinity alone does not determine immunogenic potential and that multiple structural, functional, and biochemical factors influence T-cell recognition [1,2,14,16,29,30]. Consistent with this, some peptides with strong predicted HLA binding in this study exhibited moderate immunogenicity scores, highlighting the multifactorial nature of T-cell epitope recognition.

To provide additional biological context, gene expression analysis was performed using RNA-seq data from the TCGA-SKCM cohort [25]. In this regard, the incorporation of mRNA abundance as an additional analytical layer represents an important refinement of the computational prioritization framework, as it enables evaluation of whether genes associated with candidate neoantigens are transcriptionally active in melanoma [25].

Because the MSK-IMPACT dataset used for mutation analysis does not include transcriptomic measurements, direct mutation–expression integration at the sample level was not possible. Instead, transcriptomic data from TCGA-SKCM [25], were used to assess expression patterns of genes associated with prioritized candidate peptides. Therefore, the expression results should be interpreted as population-level transcriptomic context rather than direct evidence of expression of specific mutated alleles.

Within this transcriptomic context, several genes linked to prioritized candidates, including AKT1, PDGFRB, ASXL1, PTEN, and INPP4A, were expressed across a substantial proportion of melanoma samples. Thus, the integration of expression data provides additional biological context for candidate prioritization and supports the relevance of these genes within the tumor transcriptional landscape.

Candidate peptides identified in this study were derived from mutations affecting genes previously reported in melanoma genomic analyses, including PTEN, AKT1, PLCG2, SMO, and PTCH1. As these genes are involved in signaling pathways frequently altered in melanoma, their presence in the prioritized candidate set reflects the mutational composition of the analyzed dataset, consistent with previous reports [6,7,25,31]. Similarly, comparison between the MSK-IMPACT and TCGA-SKCM datasets confirmed the presence of recurrent melanoma-associated genes such as BRAF, NRAS, NF1, TP53, and TERT.

Several limitations should be considered when interpreting these results. Mutation and expression data were derived from independent cohorts and were not matched at the individual sample level. Although proteasomal cleavage prediction and immunogenicity estimation were incorporated, additional components of antigen processing and presentation, including TAP-mediated transport, peptide trimming, and peptide–MHC complex stability, were not explicitly modeled [14,23]. Likewise, the study did not evaluate tumor clonal architecture, HLA loss of heterozygosity, or immune escape mechanisms, which could also affect neoantigen presentation [4,5]. Finally, the candidate peptides identified in this study represent computationally prioritized predictions and were not experimentally validated using immunopeptidomic or functional T-cell assays.

Taken together, these findings describe a reproducible computational prioritization framework for the identification of mutation-derived candidate neoantigens in melanoma. By integrating peptide–HLA binding prediction, mutant-versus-wild-type comparison, proteasomal cleavage analysis, immunogenicity estimation, and transcriptomic contextualization, this in silico framework may serve as a useful basis for future studies aimed at the experimental evaluation of candidate neoantigens in the context of personalized cancer immunotherapy [11–14].

6. Conclusions

This study presents a reproducible in silico bioinformatics workflow for the identification and prioritization of candidate neoantigens derived from non-synonymous somatic mutations in melanoma. The analytical framework integrates somatic mutation filtering, mutant peptide generation, peptide–HLA class I binding prediction using MHCflurry 2.0, comparison between mutant and wild-type peptides, proteasomal cleavage prediction, sequence-based immunogenicity estimation, and transcriptomic contextualization using an independent melanoma cohort.

Analysis of the MSK-IMPACT melanoma dataset (mel-mskimpact-2020) identified 16,311 non-synonymous somatic mutations, from which 50,480 mutant peptides were generated using a sliding-window approach. Among the resulting peptide–HLA combinations, 12,209 interactions met the predefined prioritization threshold (percentile rank $\leq 2\%$), indicating that a subset of the mutation-derived peptide repertoire satisfied the computational criteria applied in this study.

Within the evaluated allele panel, candidate peptide–HLA interactions were identified across alleles from the HLA-A, HLA-B, and HLA-C loci. A higher proportion of prioritized combinations was associated with HLA-B*35:01; however, this observation should be interpreted within the context of the selected allele panel and not as a definitive representation of HLA distribution in melanoma populations.

Several prioritized candidate peptides were derived from genes recurrently altered in melanoma genomic datasets, including PLCG2, GATA3, AKT1, PTEN, PTCH1, and SMO. In addition, comparison between mutant peptides and their corresponding wild-type counterparts identified sequence changes associated with differences in predicted peptide–HLA binding metrics, suggesting that specific amino acid substitutions may influence peptide–HLA interactions.

To provide additional biological context, transcriptomic data from the TCGA-SKCM cohort were analyzed to evaluate the expression of genes associated with prioritized candidate peptides. Although mutation and expression data were obtained from independent cohorts, the analysis indicated that many genes associated with prioritized peptides exhibited detectable expression in melanoma, thereby providing complementary transcriptomic context for the computationally prioritized candidates.

Taken together, these results describe a structured computational prioritization framework for the identification of mutation-derived candidate neoantigens in melanoma. It should be emphasized that the present study represents an in silico prioritization strategy based on publicly available genomic datasets and computational prediction tools. Therefore, the peptides identified in this study should be interpreted as computationally prioritized candidates, and their biological and clinical relevance remains to be established through further experimental and translational validation.

Limitations and Future Work

Several limitations should be considered when interpreting the results of the present study.

The analysis was restricted to a representative panel of six HLA class I alleles (HLA-A*02:01, HLA-A*24:02, HLA-B*35:01, HLA-B*39:05, HLA-C*04:01, and HLA-C*07:02), selected to represent the three classical HLA class I loci while maintaining computational feasibility given the large number of mutant peptides generated during the prediction pipeline. Although this design enabled the evaluation of peptide–HLA interactions across multiple loci, restricting the analysis to a limited allele panel may influence the observed distribution of prioritized peptide–HLA combinations and limit the generalizability of the results. Therefore, the allele-specific patterns observed in this study should be interpreted as exploratory rather than definitive. Future studies could expand the pipeline to include broader HLA allele panels in order to evaluate candidate repertoires across more diverse populations.

Somatic mutation data and transcriptomic data were derived from independent cohorts. Mutation analysis was performed using the MSK-IMPACT melanoma dataset, whereas gene expression analysis was conducted using RNA-seq data from the TCGA-SKCM cohort. As a result, mutation and expression measurements were not paired at the individual tumor sample level. Although the TCGA-based analysis provided additional context indicating that many genes associated with prioritized peptides are transcriptionally active in melanoma, the integration of matched genomic and transcriptomic datasets would enable a more precise evaluation of mutation–expression relationships in future studies [31].

Although the workflow incorporated proteasomal cleavage prediction (NetChop 3.1) and sequence-based immunogenicity estimation, the present pipeline does not explicitly model other key components of antigen processing and presentation, including TAP-mediated peptide transport, peptide trimming in the endoplasmic reticulum, and peptide–MHC complex stability. These factors are known to influence antigen presentation beyond peptide–HLA binding affinity and are increasingly considered in integrated neoantigen prediction frameworks [17,18]. Therefore, the current results should be interpreted within the scope of the analytical features included in this pipeline. Incorporating these additional layers would provide a more comprehensive representation of antigen presentation in future computational frameworks.

The present study does not include analyses of tumor clonal architecture or immune escape mechanisms, such as variant allele frequency (VAF), clonal versus subclonal mutation structure, HLA loss of heterozygosity, or alterations affecting beta-2 microglobulin (B2M). These factors may influence the generation, persistence, and immunological visibility of tumor-derived neoantigens during tumor evolution and have been associated with differential responses to immunotherapy [5,6]. Consequently, the absence of these parameters represents an additional limitation that may affect the biological interpretation of the prioritized candidates.

The candidate peptides identified in this study were prioritized exclusively through computational analyses and were not experimentally validated. Consequently, the present results do not demonstrate peptide processing, surface presentation, or T-cell recognition. Experimental validation using mass spectrometry–based immunopeptidomics, peptide–MHC stability assays, and functional T-cell activation or recognition experiments will be necessary to evaluate the biological relevance of the prioritized candidates [11,12].

Future work may also benefit from systematic comparisons among multiple neoantigen prediction frameworks, including MHCflurry, NetMHCpan, and complementary antigen-processing models, in order to assess inter-model robustness and improve prioritization strategies [3,14]. Such comparisons were not included in the present study because these predictors require additional parameterization and benchmarking across alleles. In addition, applying the workflow to additional melanoma cohorts and to other tumors characterized by high mutational burden may allow further evaluation of its reproducibility and transferability across different genomic contexts.

Overall, the computational prioritization framework presented here provides a structured basis for subsequent studies integrating broader HLA representation, matched transcriptomic datasets, additional antigen-processing models, clonal architecture analyses, and experimental validation approaches.

Author Contributions: Conceptualization, K.T.-V. and J.I.G.-T.; data acquisition K.T.-V., J.I.G.-T. and C.H.E.-S.; data curation, K.T.-V. and C.H.E.-S.; formal analysis, K.T.-V., J.I.G.-T. and C.H.E.-S.; funding acquisition, C.E.G.-T. and J.I.G.-T.; methodology, K.T.-V., J.I.G.-T. and C.H.E.-S.; software data analysis, K.T.-V.; supervision, K.E.V.-B., C.E.G.-T. and J.I.G.-T.; writing—original draft, K.T.-V.; writing—review and editing, V.M.-G., H.C.C.-A., C.E.G.-T., G.V.C.-R. and M.A.S.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: You can access the source code (in R-Studio Version 4.3.0 language implemented in RStudio 2022.02.3+492 "Prairie Trillium" Release (1db809b8323ba0a87c148d16eb84efe39a8e7785, 2022-05-20) for Windows Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.8 Chrome/69.0.3497.128 Safari/537.36) and the process data via the following link: https://github.com/Dantecore2/NeoantigenosData, accessed on 26 March 2026.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sim, M.J.; Sun, P.D. T Cell Recognition of Tumor Neoantigens and Insights Into T Cell Immunotherapy. *Front. Immunol.* **2022**, *13*, 833017. [CrossRef]
2. Mariuzza, R.A.; Wu, D.; Pierce, B.G. Structural basis for T cell recognition of cancer neoantigens and implications for predicting neoepitope immunogenicity. *Front. Immunol.* **2023**, *14*, 1303304. [CrossRef] [PubMed]
3. Gupta, R.G.; Li, F.; Roszik, J.; Lizée, G. Exploiting tumor neoantigens to target cancer evolution: Current challenges and promising therapeutic approaches. *Cancer Discov.* **2021**, *11*, 1024–1039. [CrossRef]
4. McGranahan, N.; Furness, A.J.; Rosenthal, R.; Ramskov, S.; Lyngaa, R.; Saini, S.K.; Jamal-Hanjani, M.; Wilson, G.A.; Birkbak, N.J.; Hiley, C.T.; et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **2016**, *351*, 1463–1469. [CrossRef]
5. Riaz, N.; Havel, J.J.; Makarov, V.; Desrichard, A.; Urba, W.J.; Sims, J.S.; Hodi, F.S.; Martín-Algarra, S.; Mandal, R.; Sharfman, W.H.; et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. *Cell* **2017**, *171*, 934–949.e16. [CrossRef]
6. Kim, H.J.; Kim, Y.H. Molecular Frontiers in Melanoma: Pathogenesis, Diagnosis, and Therapeutic Advances. *Int. J. Mol. Sci.* **2024**, *25*, 2984. [CrossRef]
7. Hayward, N.K.; Wilmott, J.S.; Waddell, N.; Johansson, P.A.; Field, M.A.; Nones, K.; Patch, A.M.; Kakavand, H.; Alexandrov, L.B.; Burke, H.; et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **2017**, *545*, 175–180. [CrossRef]

8. Oliveira, G.; Stromhaug, K.; Klaeger, S.; Kula, T.; Frederick, D.T.; Le, P.M.; Forman, J.; Huang, T.; Li, S.; Zhang, W.; et al. Phenotype, specificity and avidity of antitumour CD8+ T cells in melanoma. *Nature* **2021**, *596*, 119–125. [CrossRef]
9. Hu, Z.; Leet, D.E.; Allesøe, R.L.; Oliveira, G.; Li, S.; Luoma, A.M.; Liu, J.; Forman, J.; Huang, T.; Iorgulescu, J.B.; et al. Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat. Med.* **2021**, *27*, 515–525. [CrossRef] [PubMed]
10. Weber, J.S.; Carlino, M.S.; Khattak, A.; Meniawy, T.; Ansstas, G.; Taylor, M.H.; Kim, K.B.; McKean, M.; Long, G.V.; Sullivan, R.J.; et al. Individualised neoantigen therapy mRNA-4157 (V940) plus pembrolizumab versus pembrolizumab monotherapy in resected melanoma (KEYNOTE-942): A randomised, phase 2b study. *Lancet* **2024**, *403*, 632–644. [CrossRef] [PubMed]
11. Mattos-Arruda, L.D.; Vazquez, M.; Finotello, F.; Lepore, R.; Porta, E.; Hundal, J.; Amengual-Rigo, P.; Ng, C.K.; Valencia, A.; Carrillo, J.; et al. Neoantigen prediction and computational perspectives towards clinical benefit: Recommendations from the ESMO Precision Medicine Working Group. *Ann. Oncol.* **2020**, *31*, 978–990. [CrossRef]
12. He, J.; Xiong, X.; Yang, H.; Li, D.; Liu, X.; Li, S.; Liao, S.; Chen, S.; Wen, X.; Yu, K.; et al. Defined tumor antigen-specific T cells potentiate personalized TCR-T cell therapy and prediction of immunotherapy response. *Cell Res.* **2022**, *32*, 530–542. [CrossRef]
13. Battaglia, S. Neoantigen prediction from genomic and transcriptomic data. *Methods Enzymol.* **2020**, *635*, 267–281. [CrossRef]
14. Müller, M.; Huber, F.; Arnaud, M.; Kraemer, A.I.; Altimiras, E.R.; Michaux, J.; Taillandier-Coindard, M.; Chiffelle, J.; Murgues, B.; Gehret, T.; et al. Machine learning methods and harmonized datasets improve immunogenic neoantigen prediction. *Immunity* **2023**, *56*, 2650–2663.e6. [CrossRef]
15. Fotakis, G.; Trajanoski, Z.; Rieder, D. Computational cancer neoantigen prediction: Current status and recent advances. *Immuno-Oncol. Technol.* **2021**, *12*, 100052. [CrossRef]
16. Yossef, R.; Tran, E.; Deniger, D.C.; Gros, A.; Pasetto, A.; Parkhurst, M.R.; Gartner, J.J.; Prickett, T.D.; Cafri, G.; Robbins, P.F.; et al. Enhanced detection of neoantigen-reactive T cells targeting unique and shared oncogenes for personalized cancer immunotherapy. *JCI Insight* **2018**, *3*, e122467. [CrossRef]
17. Zhang, Y.; Chen, T.T.; Li, X.; Lan, A.L.; Ji, P.F.; Zhu, Y.J.; Ma, X.Y. Advances and challenges in neoantigen prediction for cancer immunotherapy. *Front. Immunol.* **2025**, *16*, 1617654. [CrossRef] [PubMed]
18. O'Donnell, T.J.; Rubinsteyn, A.; Laserson, U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Syst.* **2020**, *11*, 42–48.e7. [CrossRef] [PubMed]
19. Kim, S.; Kim, H.S.; Kim, E.; Lee, M.G.; Shin, E.C.; Paik, S.; Kim, S. Neopepsee: Accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Ann. Oncol.* **2018**, *29*, 1030–1036. [CrossRef]
20. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [CrossRef]
21. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, pl1. [CrossRef]
22. Cheng, D.T.; Mitchell, T.N.; Zehir, A.; Shah, R.H.; Benayed, R.; Syed, A.; Chandramohan, R.; Liu, Z.Y.; Won, H.H.; Scott, S.N.; et al. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **2015**, *17*, 251–264. [CrossRef]
23. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Sander, C.; Stuart, J.M.; Chang, K.; Creighton, C.J.; et al. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef]
24. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **2016**, *44*, e71. [CrossRef]
25. Bjerregaard, A.M.; Nielsen, M.; Jurtz, V.; Barra, C.M.; Hadrup, S.R.; Szallasi, Z.; Eklund, A.C. An analysis of natural T cell responses to predicted tumor neoepitopes. *Front. Immunol.* **2017**, *8*, 1566. [CrossRef] [PubMed]
26. Bai, P.; Li, Y.; Zhou, Q.; Xia, J.; Wei, P.C.; Deng, H.; Wu, M.; Chan, S.K.; Kappler, J.W.; Zhou, Y.; et al. Immune-based mutation classification enables neoantigen prioritization and immune feature discovery in cancer immunotherapy. *OncoImmunology* **2021**, *10*, 1868130. [CrossRef]
27. Tenzer, S.; Peters, B.; Bulik, S.; Schoor, O.; Lemmel, C.; Schatz, M.M.; Kloetzel, P.M.; Rammensee, H.G.; Schild, H.; Holzhütter, H.G. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell. Mol. Life Sci.* **2005**, *62*, 1025–1037. [CrossRef] [PubMed]
28. Benning, L.; Peintner, A.; Peintner, L. Advances in and the Applicability of Machine Learning-Based Screening and Early Detection Approaches for Cancer: A Primer. *Cancers* **2022**, *14*, 623. [CrossRef]
29. Oliveira, G.; Stromhaug, K.; Cieri, N.; Iorgulescu, J.B.; Klaeger, S.; Wolff, J.O.; Rachimi, S.; Chea, V.; Krause, K.; Freeman, S.S.; et al. Landscape of helper and regulatory antitumour CD4+ T cells in melanoma. *Nature* **2022**, *605*, 532–538. [CrossRef]

30. Li, X.; Zhu, Y.J.; Xue, Y.; Chen, T.T.; Sun, X.K.; Shi, H.Y. Neoantigen-Based Immunotherapy in Lung Cancer: Advances, Challenges and Prospects. *Cancers* **2025**, *17*, 1953. [CrossRef]
31. Yarchoan, M.; Hopkins, A.; Jaffee, E.M. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N. Engl. J. Med.* **2017**, *377*, 2500–2501. [CrossRef]