

Article



Molecular Mimicry in the Dark Immunopeptidome: Viral-Homologous Non-Canonical Tumor Peptides as Targets for Off-the-Shelf Cancer Vaccines

Angela Mauriello, Beatrice Cavalluzzo, Concetta Ragone, Biancamaria Cembrola, Simona Mangano, Luigi Buonaguro and Maria Tagliamonte *

Innovative Immunological Models Unit, National Cancer Institute IRCCS—“Fond G. Pascale”, Via Mariano Semmola, 52, 80131 Naples, Italy

* Correspondence: m.tagliamonte@istitutotumori.na.it; Tel.: +39-081-1777-0600

How To Cite: Mauriello, A.; Cavalluzzo, B.; Ragone, C.; et al. Molecular Mimicry in the Dark Immunopeptidome: Viral-Homologous Non-Canonical Tumor Peptides as Targets for Off-the-Shelf Cancer Vaccines. *Translational Insights* 2026, 1(1), 14. <https://doi.org/10.53941/ti.2026.100014>

Received: 25 May 2026

Revised: 29 June 2026

Accepted: 8 July 2026

Published: 21 July 2026

Abstract: Background: The identification of broadly shared, highly immunogenic tumor antigens remains a central challenge in cancer vaccine development. Non-canonical tumor peptides (nonC-TPs) arising from the cancer “dark genome” have recently emerged as a highly immunogenic class of tumor antigens with strong potential for off-the-shelf cancer vaccine development. These peptides originate from aberrant transcriptional and translational events involving transposable elements, endogenous retroviruses, untranslated regions, and alternative open reading frames, represent a promising antigen class, as they are absent from normal tissues, largely excluded from thymic selection, and shared across tumor histologies and patients. Molecular mimicry between tumor antigens and viral epitopes has been previously demonstrated for canonical tumor-associated antigens (TAAs), providing a mechanistic basis for cross-reactive anti-tumor immunity mediated by virus-specific memory T cells. This has never been investigated in the nonC-TP landscape of the dark genome. Methods: Starting from the proteogenomic dataset generated by Lozano-Rabella, comprising 65,504 canonical and non-canonical peptides identified in several primary cancer cell lines, we prioritized 244 tumor-exclusive non-canonical nonamers predicted as strong HLA binders using NetMHCpan4.1. Then we focused our analysis on restricted NonC peptides binding four HLA-A allotypes. Homology searches against viral proteomes were performed using BLAST, followed by HLA binding and stability prediction with NetMHCstabpan1.0. Structural modelling and molecular docking analyses were conducted to evaluate conservation of peptide-HLA molecular conformations between nonC peptides and viral homologs. Results: A significant subset of dark genome-derived nonC TP displayed significant sequence homology with viral epitopes across multiple HLA-A alleles, particularly HLA-A*02:01, HLA-A*03:01, and HLA-A*11:01. Homologous viral sequences were identified from clinically relevant pathogens. The homologous peptides frequently conserved amino acid residues predicted to mediate TCR interaction, supporting the possibility of cross-reactive recognition by pre-existing anti-viral memory CD8⁺ T cells. Structural modelling further demonstrated highly similar peptide-HLA docking conformations between nonC peptides and their viral counterparts, supporting the potential role of molecular mimicry within the dark immunopeptidome. Conclusions: This study provides the first systematic evidence that dark genome-derived nonC-TPs exhibit molecular mimicry with viral antigens,



extending a well-established immunological paradigm to a previously unexplored antigen class. The identified viral-homologous nonC-TPs combine tumor specificity, broad sharing across patients and histologies, and potential recognition by pre-existing virus-specific memory T cells, positioning them as uniquely attractive candidates for next-generation, off-the-shelf cancer vaccines. These findings open a new dimension of tumor immunosurveillance and provide a rational framework for the development of immunotherapeutic strategies that exploit the convergence of dark genome biology, molecular mimicry, and viral mimicry.

Keywords: non-canonical tumor peptide; molecular mimicry; cancer vaccine, dark genome

1. Introduction

The development of effective cancer vaccines relies on the identification of tumor antigens capable of eliciting potent and durable cytotoxic T lymphocyte (CTL) responses. For decades, two main antigen categories have dominated cancer immunotherapy: tumor-associated antigens (TAAs), self-proteins aberrantly overexpressed in tumor cells, and mutation-derived neoantigens (TSA) arising from tumor-specific somatic mutations [1]. While neoantigens are strictly tumor-specific and not affected by central tolerance, their clinical impact has been limited to now. Large-scale proteogenomic analyses consistently show that the majority of computationally predicted neoepitopes are not detectable on the HLA class I surface of tumor cells, and multiple clinical trials of neoantigen-based vaccines have produced modest results [2–5]. Moreover, the patient-specific nature of these antigens makes manufacturing complex and costly, precluding their use as broadly applicable, off-the-shelf therapeutic products. TAAs, on the other hand, while sharable across patients and tumors, suffer from partial central tolerance due to their expression in normal tissues, limiting the amplitude and quality of T cell responses they can elicit. To overcome these limitations, cancer vaccines should ideally be based on highly immunogenic peptides that are recognized as non-self, broadly shared across patients, and suitable for off-the-shelf formulations—a combination of properties that canonical tumor antigens have failed to deliver.

A paradigm shift has emerged from the recognition that up to 75% of the human genome can be transcribed and potentially translated, far beyond the 2% accounted for by protein-coding exons [6]. Epigenetic dysregulation leads to aberrant reactivation of normally silenced genomic regions, including transposable elements (TEs), endogenous retroviruses (ERVs), long non-coding RNAs (lncRNAs), transcribed ultra-conserved regions (T-UCRs), and non-canonical open reading frames (ncORFs) [7]. The peptides arising from these regions, collectively referred to as non-canonical tumor peptides (nonC-TPs) or “cancer dark matter”, are not derived from patient-specific mutations but from aberrant transcriptional and translational events shared broadly across tumor types and patients [8–12]. Because they are absent from normal somatic tissues and largely excluded from thymic selection, they are recognized as non-self and can elicit potent CTL responses without central tolerance constraints, positioning them as ideal candidates for shared, off-the-shelf cancer vaccines.

A key mechanism linking the dark genome to nonC-TPs generation is the production of Defective Ribosomal Products (DRiPs), aberrant translation products generated by specialized “immunoribosomes” activated in cancer cells. DRiPs arise from non-canonical translation events including initiation from alternative start codons, translation of 5' and 3' UTRs, and nuclear ribosome-mediated translation of introns and pre-mRNA exons [13–15]. Because DRiPs are degraded co-translationally and loaded directly onto HLA class I molecules with high efficiency, they constitute a privileged and shared source of immunogenic nonC-TPs, bypassing conventional antigen processing routes [16–18]. A landmark proteogenomic study by Lozano-Rabella and colleagues provided direct evidence of the breadth and immunological relevance of this antigen class, identifying numerous nonC-TPs presented on HLA-I molecules in patient-derived tumor cell lines from melanoma, gynecological, and head and neck cancers far more abundant than classical neoantigens, shared across histologies, and immunogenic upon in vitro sensitization [12].

A conceptually distinct but mechanistically complementary dimension of tumor antigen immunogenicity is provided by antigenic molecular mimicry, the structural and sequential homology between tumor antigens and pathogen-derived peptides [19–21]. The theoretical basis rests on the well-established degeneracy of TCR recognition: each T cell clone can potentially cross-react with multiple structurally similar peptide-HLA complexes, enabling virus-specific memory T cells to recognize homologous tumor epitopes. Our group has systematically investigated this phenomenon over the past decade, demonstrating that canonical TAAs can exhibit high sequence similarity to viral peptides at TCR-contacting residues and that cross-reactive CD8⁺ T cell responses can be elicited against paired tumor and viral epitopes from HIV, HBV, HPV, SARS-CoV-2, and

herpesviruses [22–25]. These findings establish that the vast pool of anti-viral memory T cells accumulated over a lifetime of infections represents a pre-existing immunological resource that can be redirected against structurally homologous tumor antigens.

While molecular mimicry between canonical TAAs and viral antigens has now been firmly established, an entirely unexplored dimension of this phenomenon concerns the nonC-TP landscape of the dark genome. This gap is critical for several reasons. First, nonC-TPs derived from TEs, ERVs, and ncORFs share evolutionary origins with viral sequences, ERVs are by definition remnants of ancient retroviral integrations [26–28]. Second, the aberrant translational events generating nonC-TPs in cancer cells parallel those occurring in virus-infected cells, where alternative reading frame usage is a well-documented source of immunogenic epitopes [29]. Third, the same epigenetic derepression producing nonC-TPs simultaneously generates double-stranded nucleic acids that activate innate immune sensors, a phenomenon termed “viral mimicry”, reinforcing the biological analogy between cancer and virus-infected cells and creating a pro-inflammatory microenvironment that supports adaptive immune activation [30–32].

Consequently, it is plausible that a subset of dark genome-derived nonC-TPs share sequence and/or structural homology with viral antigens, and may thus be recognized by pre-existing virus-specific memory CD8+ T cell populations. This would represent a previously uncharacterized and potentially powerful layer of tumor immunosurveillance, in which the anti-viral immune memory accumulated through a lifetime of infections and vaccinations contributes to cross-reactive anti-tumor responses directed against nonC-TPs presented by cancer cells.

Building on the proteogenomic dataset of canonical and nonC-TPs identified by Lozano-Rabella and colleagues and on our group’s established expertise in molecular mimicry between tumor and microbial antigens, the present study systematically investigates whether dark genome-derived nonC-TPs exhibit sequence and structural homology to viral epitopes. Using bioinformatic homology analysis against viral proteome databases, we identify a subset of nonC-TPs with significant similarity to known viral peptides and place our findings within the broader framework of DRiPs biology and molecular mimicry. We propose that the convergence of these mechanisms creates a uniquely immunogenic antigen class with strong implications for the rational design of next-generation, off-the-shelf cancer vaccines targeting the dark immunopeptidome.

2. Materials and Methods

2.1. Selection of the NonC Antigens

We starting from the list of 65,504 canonical and nonC-TPs, published by Lozano-Rabella, identified in 9 short-term cultured tumor cell line derived from 4 patients with gynecologic cancer (Gyn), 3 patients with melanoma (Mel), and 2 patients with head and neck (H&N) cancer, by LC/MS-MS using state-of-the-art procedures [12]. In particular, the authors identified Amino acid (Aa) sequences through a previously described pipeline, Peptide-PRISM [8]. In this study we prioritized 244 peptides (nonamers) predicted to be nonC-TPs expressed only by tumor samples (NonC-TL), highly shared across patients and the different tumor types and predicted to have high binding affinity (strong binders). The predicted binding affinity (in nM) and rank (%) of each peptide for the relevant HLA allotype were obtained from the NetMHCpan4.1 algorithm (<http://www.cbs.dtu.dk/services/NetMHCpan/> accessed on 08 April 2026) (the lower the value, the higher the affinity). The percentile rank threshold for predicted strong binders is inferior or equal to 0.5%. In total, four HLA allotypes were included in the present study: HLA-A*02:01, -A*03:01, -A*11:01, -A*24:02. Their prevalence in four populations (Caucasian, African, Hispanic, and Asian/Pacific Islanders) was calculated using published allele frequency data.

2.2. BLAST Homology Search

The nonC-TPs, identified as SB according to the NetMHCpan 4.1 prediction tool have been submitted to BLAST for a protein homology search against viral sequences (Viruses—taxid:10239) within the non-redundant protein sequences database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 15 April 2026). Homologous viral protein sequences have been extracted from the protein database of the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>, accessed on 22 April 2026) and epitope prediction has been performed with the NetMHCstabpan 1.0 tool, selecting the four HLA-A haplotypes.

2.3. Epitope Modelling and Molecular Docking.

The 3D structure of interaction between peptides and HLA-A*02:01 was generated using PyMol (PyMol ICM (<http://www.molsoft.com/> (accessed on 4 May 2026), version 3.8-7d) software. The PDB format of complex between HLA molecules and reference peptides were downloaded from RCS Protein Data Bank (PDB) website

(<https://www.rcsb.org>, accessed on 5 May 2026). In particular, HLA-A*02:01 (1AO7), HLA-A*03:01 (3RL1), HLA-A*24:02 (7JYV), HLA-A*11:01 (8RYO). The original peptide sequences in the downloaded structures were replaced with individual predicted epitopes using PyMol visualization system and Molsoft ICM software (version 3.9-5) was used to visualize the molecular docking.

2.4. Statistical Analysis

Absolute counts and percentage distributions were calculated for all nonC peptide categories, including their genomic origin, HLA allele binding preference, and tumor histology of origin. The proportion of strong binders (SB) per HLA allele and per tumor type was expressed as percentage of the total nonC-TL peptide pool. Similarly, the proportion of nonC peptides exhibiting homology to viral epitopes was calculated as percentage of the total SB pool, stratified by HLA allele and tumor histology. Population coverage for each HLA-A haplotype was estimated using published allele frequency data across four ethnic groups (Caucasian, African, Hispanic, and Asian/Pacific Islanders) applying the standard heterozygosity-corrected formula: prevalence = $[F^2 + 2(F(1 - F))]$ $\times 100$, where F corresponds to allele frequency. Cumulative predicted binding affinity values for viral homolog candidates were summarized per HLA allele. All graphical representations were generated using GraphPad Prism and/or Excel software. No inferential statistical testing was performed, as the study design was exploratory and hypothesis-generating.

3. Results

3.1. Selection of Non-Canonical Antigen Candidates

Starting from the comprehensive proteogenomic dataset published by Lozano-Rabella and colleagues, comprising 65,504 canonical and nonC-TPs identified by LC/MS-MS in 9 short-term cultured tumor cell lines derived from patients with gynecologic cancer (Gyn, n = 4), melanoma (Mel, n = 3), and head and neck squamous cell carcinoma (HNSCC, n = 2), we applied a multi-step selection pipeline to identify nonC peptide candidates for downstream homology and structural analysis (Figure 1A). Peptide sequences had been originally identified through the Peptide-PRISM pipeline [8]. From this dataset, we prioritized 244 nonamers predicted as non-canonical peptides expressed exclusively in tumor samples (nonC-TL), highly shared across patients and tumor types, and predicted to exhibit high HLA binding affinity (Supplementary Table S1). Binding affinity prediction was performed using the NetMHCpan 4.1 algorithm, retaining only peptides classified as strong binders (SB), defined by a percentile rank $\leq 0.5\%$.

Our analysis was focused on four HLA-A allotypes, A*02:01, A*03:01, A*11:01, and A*24:02, selected based on their high prevalence in Western populations. Their cumulative population coverage was estimated using published HLA allele frequency data across four major ethnic groups [33]. In particular, the A*02:01 and 24:02 are the most prevalent across the Continents; the A*11:01 is highly prevalent in South and South-East Asia as well as in Oceania. This selection strategy yielded a focused set of 93 immunologically relevant nonC peptide candidates for each of the four HLA alleles examined. In particular, for these HLA I selected there are a significant prevalence number of nonC-TPs for the HLA-A*11:01 (Figure 1B; (Supplementary Table S2)).

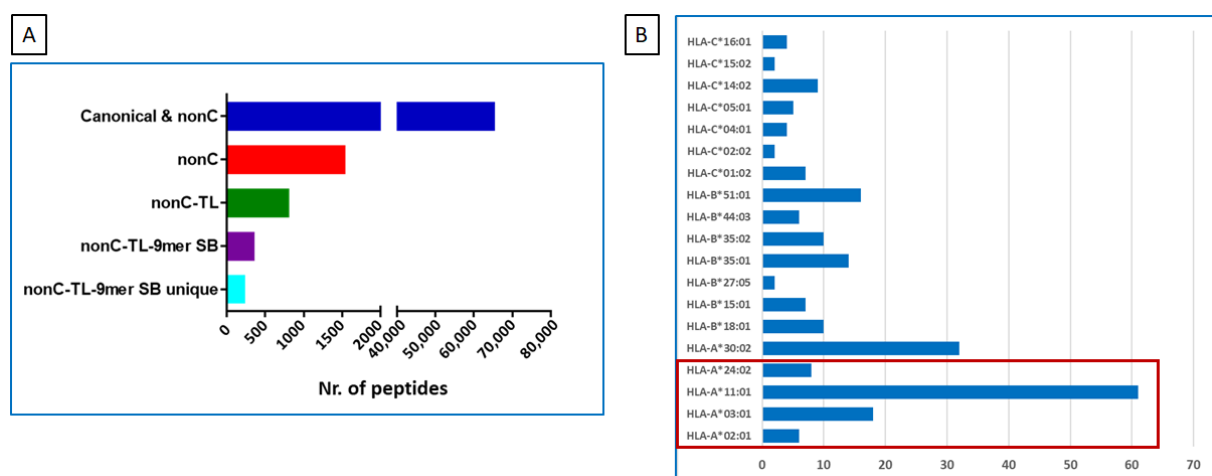


Figure 1. (A) Pipeline used for select nonC peptide candidates for downstream analysis. (B) Number of nonC-TPs from tumor cells identified for each HLA I. For each peptide, only the SB predicted by NetMHCpan4.1 was considered. The red box shows the HLAs selected for the analyses.

3.2. Characterization of Non-Canonical Antigen Candidates

The selected 93 nonC-TPs candidates, were analyzed for their genomic origin, distribution across the three tumor histologies and across the four HLA-A alleles selected in this study. NonC peptides were classified according to their dark genome source, including 5' UTR-derived, out-of-frame translated, lncRNA-derived, and other non-canonical genomic regions, revealing a diverse landscape of epigenetically reactivated genomic elements contributing to the nonC-TP pool (Figure 2A). The proportion of SB nonC peptides associated with each HLA-A allele was evaluated within each tumor type, highlighting both allele-specific and histology-specific patterns of nonC peptide presentation.

The highest percentage of nonC-TPs SBs (67%) was predicted in melanoma (Mel), followed by Head & Neck cancer (H&N) (20%) and by gynecologic cancer (Gyn) (16%) (Figure 2B). Then, we evaluated the percentage of SBs nonC peptides associated to each HLA-A selected in the different cancer types. Our analyses revealed a high variable prevalence of SBs specific to each HLA allele. In particular, the highest percentage was predicted for HLA-A*11:01 (69.59%), and was associated exclusively to melanoma. The second highest percentage were predicted for HLA-A*03:01 (19.35%) associated for 13.05% to H&N and for 4.30% to Gyn cancers (Figure 2C). As previously reported in other immunopeptidomic studies, non-canonical peptides display a preferential association with HLA-A*11:01 and HLA-A*03:01 alleles [12,34,35]. This bias likely reflects the highly similar peptide-binding motifs shared by these two HLA allotypes, which preferentially accommodate peptides containing basic residues at position 9 (P9), a major anchor position for nonamer binding within the HLA class I peptide-binding groove (Figure 3). To further evaluate the robustness of the selected non-canonical peptide repertoire, we compared our dataset with two recently published large-scale immunopeptidomic studies [36,37]. The comparison revealed fifteen nonC-TPs peptides shared with “ncHLAp” detected in PDAC organoid and thirteen detected in bulk tumors by Ely et al. [36]. Five of these shared peptides are identified in both data set described by Ely et al. [36]. Three of these are among the nonC-TPs from our analysis showing a viral homology. Moreover, we identify only two peptides shared with the dataset reported in breast tumor samples by Kina et al. [37] (Supplementary Table S3).

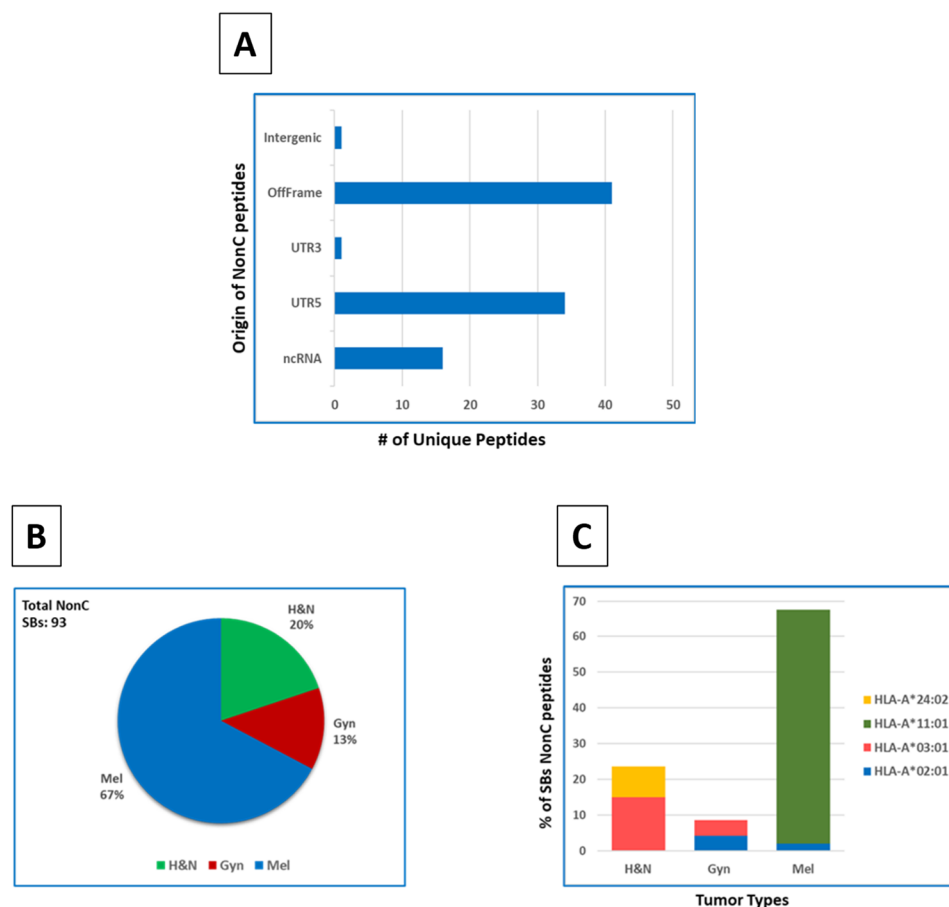


Figure 2. (A) Number of nonC-TL originated from each of the ORF categories noted. (B) Percentage of SBs nonC peptides identified in each tumor for HLA-A selected. (C) Percentage of SBs nonC peptides associated to each HLA-A identified in each tumor, listed by cancer.

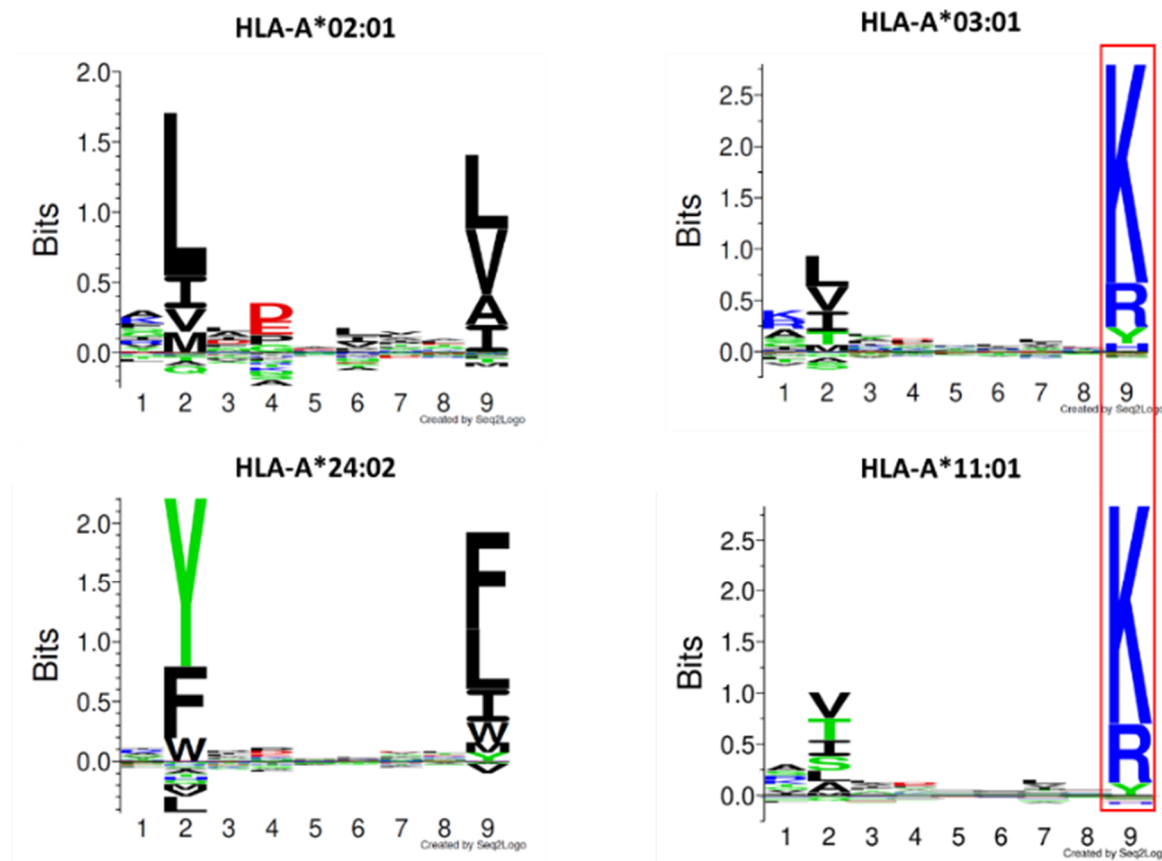


Figure 3. Consensus peptide binding motif of the two HLAs predicted to present the majority of the nonC peptides identified. Image downloaded from NetMHCpan4.1 motif viewer.

3.3. Identification of Homologous Viral Epitopes

To investigate the molecular mimicry between dark genome-derived nonC peptides and viral antigens, we performed a systematic homology analysis using BLAST against publicly available viral proteome databases within the NCBI non-redundant protein database. Homologous viral protein sequences were extracted from the NCBI protein database, and epitope prediction for the identified viral homologs was performed using the NetMHCstabpan 1.0 tool, which integrates both binding affinity and peptide-HLA complex stability predictions.

The analysis revealed that a reproducible subset of nonC peptides from each tumor histology exhibited significant sequence homology to viral epitopes across three out of four HLA alleles examined (HLA-A*02:01; HLA-A*03:01; HLA-A*11:01) and not for HLA-A*24:02. In particular, approximately 50% of nonC peptides associated with HLA-A*03:01 were homologous to viral peptides, while for HLA-A*02:01 and HLA-A*11:01 accounted for approximately 33% and 30%, respectively (Figure 4A). This allele-specific distribution suggests that the structural constraints and peptide-binding motifs of individual HLA allotypes may influence the likelihood of identifying viral-like non-canonical epitopes within the dark immunopeptidome.

In total, 20 viral peptides displaying sequence homology to the selected nonC-TPs were identified and predicted to bind the corresponding HLA alleles with affinities ranging from 4.69 nM, consistent with strong binders (SB), to 1000 nM, corresponding to weak binders (WB). Notably, the largest number of homologous viral peptides was identified in melanoma-derived nonC-TPs ($n = 15$), whereas only three and two homologous viral peptides were detected for head & neck and gynecologic cancers, respectively (Figure 4B).

The proportion of homologous viral antigens and their cumulative predicted binding affinity were evaluated for HLA allele. Only viral peptides homologous to melanoma specific NonC peptides showed strong binding to both HLA-A*02:01 and HLA-A*11:01 alleles. on the contrary, in H&N and Gyn Cancers we identified only homologous viral peptide with High affinity to HLA-A*03:01 (Figure 4C).

The cumulative predicted binding affinity analysis further demonstrated that viral homologs associated with HLA-A*11:01 displayed the strongest overall binding profile, with several peptides predicted within the high-affinity SB range (<50 nM), whereas homologous peptides associated with HLA-A*03:01 generally showed lower predicted affinity values (Figure 4D). These findings suggest that the degree of molecular mimicry between viral

epitopes and nonC-TPs may differ not only across tumor histologies but also according to the peptide-binding properties of individual HLA alleles.

Homologous viral peptides were derived from multiple clinically relevant human pathogens, including Dengue virus, Human Respiratory Syncytial Virus A, Human alphaherpesvirus 2, Enterovirus F4, Human adenovirus 4, Hepatitis B virus, Hepatitis C virus, Human papillomavirus, Herpesviridae family members, and Influenza A virus. Notably, several homologous peptide pairs conserved amino acid residues predicted to be exposed for TCR recognition, supporting the possibility of cross-reactive CD8⁺ T cell responses between viral antigens and dark genome-derived tumor peptides.

Among the most representative examples, the Dengue Virus 2-derived peptide TLARTKPTL shared 7 out of 9 identical amino acids with the melanoma-derived nonC peptide TLARTTPSL, including conservation of residues predicted to contribute to HLA anchoring and TCR interaction. Similar patterns of sequence conservation were observed across all viral/nonC peptide pairs, supporting the hypothesis that dark genome-derived tumor antigens may structurally mimic pathogen-associated epitopes recognized by pre-existing antiviral memory T cells.

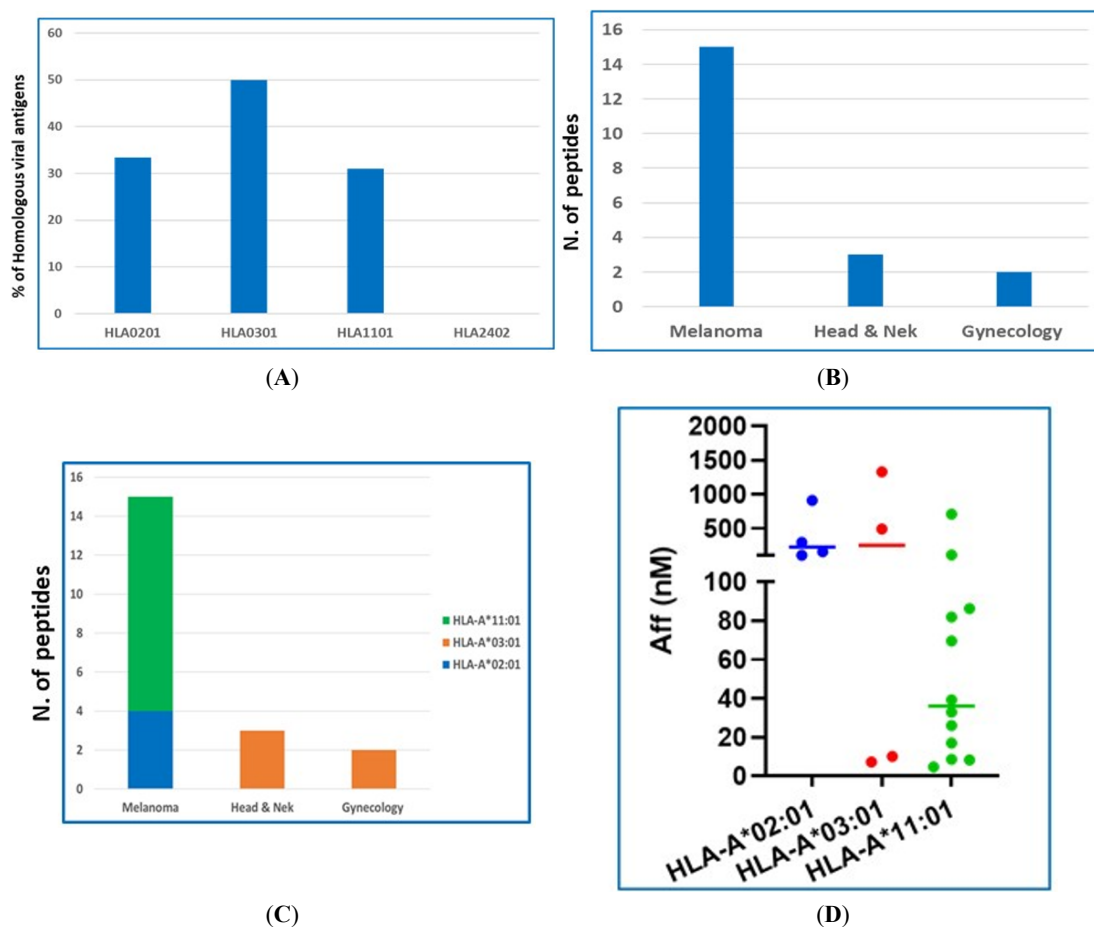


Figure 4. (A) Percentage of Viral antigens homologous to NonC Peptides for each HLA-A allele selected. (B) Cumulative number of Viral predicted antigens homologous to nonC peptides identified in each tumor. (C) Number of viral predicted antigens associated to each HLA allele selected homologous to nonC pep identified in each tumor. (D) Cumulative predicted affinity of Viral predicted antigens listed by HLA allele. The blue, red and green dots represent the viral predicted antigens for HLA*02:01, HLA*03:01 and HLA*11:01 respectively.

3.4. Peptide Modelling and Molecular Docking

To validate the predicted interactions between the identified nonC peptides, their viral homologs, and the relevant HLA molecules, three-dimensional modelling and molecular docking were performed using PyMol and Molsoft ICM software.

Reference HLA-peptide complex structures were downloaded from the RCSB Protein Data Bank for each allele: HLA-A*01:01 (PDB: 3BO8), HLA-A*02:01 (PDB: 1AO7), HLA-A*03:01 (PDB: 3RL1). The original peptide sequences within each downloaded structure were replaced with the individual predicted nonC and viral

epitopes using the PyMol visualization system, and molecular docking was performed and visualized using Molsoft ICM.

In detail, all viral/nonC peptide pairs analyzed shared identical contact patterns with the HLA molecule and extremely similar contact patterns with the alpha and beta chains of the TCR. In Figure 5 are shown two examples for the HLA-A*02:01 (A) and HLA-A* 11:01 (B) of the impact of residue differences on the molecular conformation. In the two example, the amino acid substitutions do not induce significant conformational changes in the structure of paired epitopes. Overall, the analysis confirmed that regardless of the position of the amino acid substitution along the peptide sequence, viral antigens may have a similar, if not identical, conformation and pattern of contact with the α and β chains of the TCR, as shown by the footprints of the paired peptides (Figure 5).

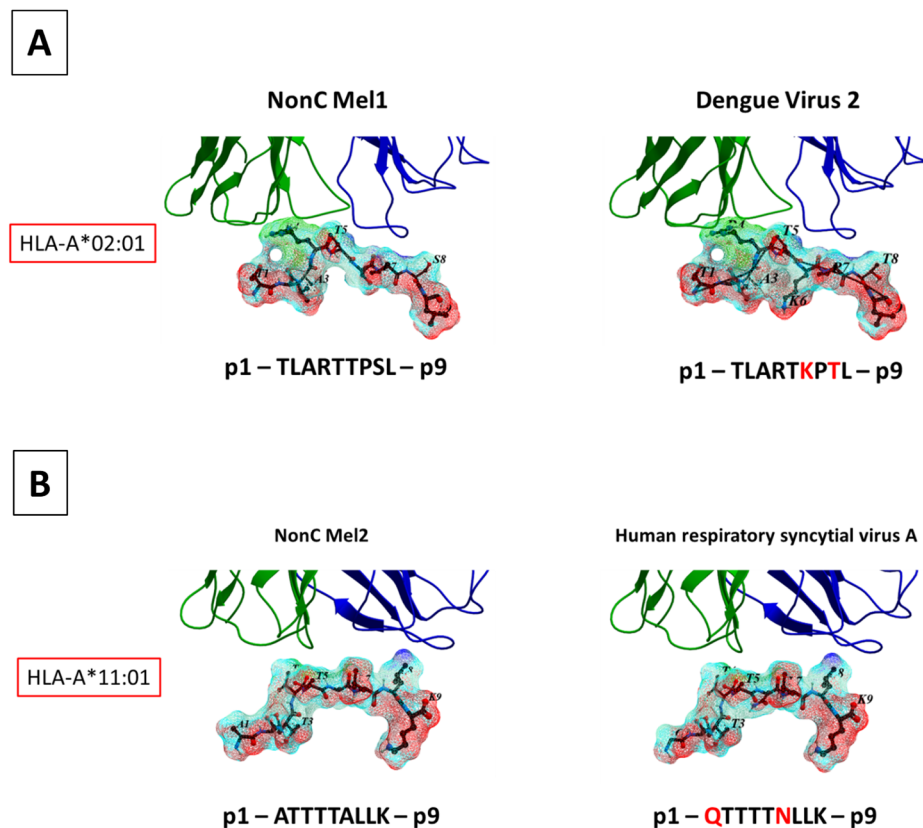


Figure 5. Impact of residue differences on the affinity to HLA alleles. Examples of different changes in the epitope sequences and their impact on the affinity to HLA-A*02:01 (A) and HLA-A* 11:01 (B) alleles as well as on TCR alpha and beta chains. Residues in the viral epitopes that differs from the nonC epitopes sequences are indicated in red color. Red areas = contact points with HLA-A molecule; blue areas = contact points with TCR α chain; green areas = contact points with TCR β chain.

4. Discussion

The present study explores, for the first time, the intersection between the dark immunopeptidome and antigenic molecular mimicry, providing evidence that non-canonical tumor peptides (nonC-TPs) generated through aberrant translational mechanisms can share significant sequence and structural similarity with viral epitopes. Building on previous proteogenomic evidence provided by Lozano-Rabella and integrating our previous findings on tumor viral molecular mimicry, we propose that dark genome-derived antigens may represent a previously unrecognized bridge between cancer immunogenicity and anti-viral immune memory.

The present work extends this paradigm to the nonC-TP landscape, representing a conceptual and methodological advancement of considerable significance. The demonstration that such peptides can additionally share homology with viral antigens opens a new dimension of potential immunogenicity that had not previously been considered.

Starting from a large proteogenomic dataset of 65,504 canonical and nonC-TPs identified across three tumor histologies, we applied a multi-step bioinformatic pipeline to select 244 HLA-A-restricted strong binders expressed exclusively in tumor cells and shared across patients and cancer types. A particularly important finding of the present study is that viral homologs were identified for nonC-TPs restricted by three of the four HLA alleles

analyzed: HLA-A*02:01, -A*03:01, and -A*11:01, while no homology was detected for HLA-A*24:02. This allele-specific pattern is consistent with the well-established structural differences in the peptide-binding grooves of HLA allotypes, which impose distinct constraints on the physicochemical properties of bound peptides and therefore on the probability of sequence overlap with viral proteomes [31]. Notably, a number of nonC-TPs identified in the present study are common to non-canonical peptides previously described [36,37] and three of them are among those showing viral homology. Such findings suggest that the mimicry could have a general relevance and be exploited in several tumor settings.

The viruses for which homologous nonC-TPs were identified (i.e., Dengue Virus 2, Human Respiratory Syncytial Virus A and hepatic viruses), are among the most globally prevalent human pathogens, associated with widespread and long-lasting immune memory in human populations. This is immunologically significant, indeed, the breadth and durability of memory CD8⁺ T cell responses to these pathogens implies that a substantial fraction of individuals in affected populations carry circulating virus-specific T cell clones that could, in principle, cross-recognize the homologous nonC-TPs presented by tumor cells. This concept aligns with the broader evidence that anti-microbial immune memory shapes the anti-tumor immune repertoire in ways that are only beginning to be understood.

The structural modelling and molecular docking analyses performed in the present study provide an important layer of mechanistic support for the functional plausibility of TCR cross-reactivity. By demonstrating that both nonC peptides and their viral homologs adopt compatible conformations within the relevant HLA binding grooves, we show that the sequence homology identified by BLAST translates into structural equivalence at the peptide-HLA interface, a prerequisite for shared TCR recognition. However, structural modelling demonstrated that homologous viral and non-canonical peptides can be accommodated within highly similar HLA-binding conformations, the present computational approach was not designed to quantitatively predict TCR recognition. Future studies integrating molecular dynamics simulations, surface electrostatic analyses, and TCR-peptide-HLA modelling will be valuable to further characterize the structural determinants underlying potential cross-reactive immune responses. Moreover, computational modelling cannot substitute for experimental validation of T cell cross-reactivity, but provides a rational and evidence-based framework for prioritizing candidate peptide pairs for downstream immunological testing.

The present findings should be interpreted within the broader biological context of dark genome reactivation in cancer. The same epigenetic dysregulation that generates nonC-TPs through aberrant translation of TEs, ERVs, lncRNAs, and ncORFs simultaneously produces double-stranded nucleic acids that activate innate immune sensors, RIG-I, MDA5, and the cGAS-STING pathway, triggering a state of viral mimicry in cancer cells that promotes type I interferon signaling, pro-inflammatory chemokine production, and immune cell recruitment to the tumor microenvironment [29]. This creates a self-amplifying immunological circuit in which the dark genome simultaneously provides both the antigenic substrate for adaptive immune recognition, through nonC-TPs, including those with viral homology and the innate immune activation signal necessary to sustain and amplify T cell responses. The convergence of molecular mimicry and viral mimicry within the same biological process represents a uniquely powerful immunological mechanism that may explain why immune-infiltrated tumors with high dark genome activity are more likely to respond to immunotherapy.

From a translational perspective, the shared, non-mutated nature of dark genome-derived nonC-TPs and their demonstrated homology to viral antigens position this antigen class as a highly attractive foundation for the development of next-generation, off-the-shelf cancer vaccines. Unlike patient-specific mutation-derived neoantigens, these peptides are broadly expressed across tumor histologies and patients, and their potential recognition by pre-existing virus-specific T cell memory may confer an inherent immunological advantage in terms of response magnitude and durability. Future studies should focus on the functional validation of cross-reactive T cell responses against the identified nonC peptide-viral epitope pairs, using ex vivo assays with PBMCs from cancer patients and healthy donors, as well as on the integration of DRiPs modulation strategies to enhance the in vivo presentation of these antigens on tumor cells.

Collectively, these results contribute to a growing body of evidence that the dark genome, far from being immunologically inert, represents a rich and clinically exploitable source of tumor antigens whose full potential is only beginning to be revealed.

Supplementary Materials

The additional data and information can be downloaded at: <https://media.sciltp.com/articles/others/2607201035014401/TI-26050190-SI.zip>.

Author Contributions

A.M. contributed to methods, validation; B.C (Beatrice Cavalluzzo) and C.R: contributed to software validation, writing—review & editing. S.M. and B.C (Biancamaria Cembrola): contributed to data analyses. M.T. and L.B.: conceived, conceptualized, supervised the project, contributed in method, writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding

The study was funded by the Italian Ministry of Health through Institutional “Ricerca Corrente” (Projects L2/30_25 to L.B.; L2/35_25 to M.T.); the PNRR Ministero Salute PNRR-POC-2022-12375769 “Molecular mimicry to improve liver cancer immunotherapy” (2023–2025) (to L.B.).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data are contained within the article and Supplementary Materials.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. Coulie, P.G.; van den Eynde, B.J.; van der Bruggen, P.; et al. Tumor antigens recognized by T lymphocytes: At the core of cancer immunotherapy. *Nat. Rev. Cancer* **2014**, *14*, 135–146.
2. Bassani-Sternberg, M.; Bräunlein, E.; Klar, R.; et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **2016**, *7*, 13404. <https://doi.org/10.1038/ncomms13404>.
3. Wu, D.W.; Jia, S.P.; Xing, S.J.; et al. Personalized neoantigen cancer vaccines: Current progression, challenges and a bright future. *Clin. Exp. Med.* **2024**, *24*, 229. <https://doi.org/10.1007/s10238-024-01436-7>.
4. Buonaguro, L.; Tagliamonte, M. Selecting target antigens for cancer vaccine development. *Vaccines* **2020**, *8*, 615. <https://doi.org/10.3390/vaccines8040615>.
5. Tagliamonte, M.; Petrizzo, A.; Tornesello, M.L.; et al. Antigen-specific vaccines for cancer treatment. *Hum. Vaccin. Immunother.* **2014**, *10*, 3332–3346.
6. Djebali, S.; Davis, C.A.; Merkel, A.; et al. Landscape of transcription in human cells. *Nature* **2012**, *489*, 101–108.
7. Camarena, M.E.; Theunissen, P.; Ruiz, M.; et al. Microproteins encoded by noncanonical ORFs are a major source of tumor-specific antigens in a liver cancer patient meta-cohort. *Sci. Adv.* **2024**, *10*, eadn3628. <https://doi.org/10.1126/sciadv.adn3628>.
8. Erhard, F.; Dölken, L.; Schilling, B.; et al. Identification of the cryptic HLA-I immunopeptidome. *Cancer Immunol. Res.* **2020**, *8*, 1018–1026. <https://doi.org/10.1158/2326-6066.CIR-19-0886>.
9. Saffern, M.; Samstein, R. MHCing the tumour’s dark genome. *Nat. Rev. Immunol.* **2023**, *23*, 140. <https://doi.org/10.1038/s41577-023-00839-z>.
10. Fox, B.A.; Urba, W.J.; Jensen, S.M.; et al. Cancer’s Dark Matter: Lighting the Abyss Unveils Universe of New Therapies. *Clin. Cancer Res.* **2023**, *29*, 2173–2175. <https://doi.org/10.1158/1078-0432.CCR-23-0422>.
11. Laumont, C.M.; Daouda, T.; Laverdure, J.P.; et al. Global proteogenomic analysis of human MHC class I-associated peptides derived from noncanonical reading frames. *Nat. Commun.* **2016**, *7*, 10238. <https://doi.org/10.1038/ncomms10238>.
12. Lozano-Rabella, M.; Garcia-Garijo, A.; Palomero, J.; et al. Exploring the immunogenicity of noncanonical HLA-I tumor ligands identified through proteogenomics. *Clin. Cancer Res.* **2023**, *29*, 2250–2265. <https://doi.org/10.1158/1078-0432.CCR-22-3298>.

13. Wei, J.; Yewdell, J.W. Immunoribosomes: Where's there's fire, there's fire. *Mol. Immunol.* **2019**, *113*, 38–42. <https://doi.org/10.1016/j.molimm.2017.12.026>.
14. Yewdell, J.W.; Holly, J. DRiPs get molecular. *Curr. Opin. Immunol.* **2020**, *64*, 130–136. <https://doi.org/10.1016/j.coi.2020.05.009>.
15. Yewdell, J.W.; Antón, L.C.; Bennink, J.R. Defective ribosomal products (DRiPs): A major source of antigenic peptides for MHC class I molecules? *J. Immunol.* **1996**, *157*, 1823–1826.
16. Yewdell, J.W. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. *Trends Immunol.* **2011**, *32*, 548–558. <https://doi.org/10.1016/j.it.2011.08.001>.
17. Komov, L.; Melamed Kadosh, D.; Barnea, E.; et al. The effect of interferons on presentation of defective ribosomal products as HLA peptides. *Mol. Cell. Proteom.* **2021**, *20*, 100105. <https://doi.org/10.1016/j.mcpro.2021.100105>.
18. Dersh, D.; Holly, J.; Yewdell, J.W. A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nat. Rev. Immunol.* **2021**, *21*, 116–128. <https://doi.org/10.1038/s41577-020-0372-7>.
19. Guven-Maiorov, E.; Tsai, C.J.; Nussinov, R. Pathogen mimicry of host protein-protein interfaces modulates immunity. *Semin. Cell Dev. Biol.* **2016**, *58*, 136–145.
20. Zitvogel, L.; Kroemer, G. Cross-reactivity between microbial and tumor antigens. *Curr. Opin. Immunol.* **2022**, *75*, 102171.
21. Buonaguro, L.; Cerullo, V. Pathogens: Our allies against cancer? *Mol. Ther.* **2020**, *29*, 10–12.
22. Cavalluzzo, B.; Ragone, C.; Mauriello, A.; et al. Identification and validation of viral antigens sharing sequence and structural homology with tumor-associated antigens (TAAs). *J. Immunother. Cancer* **2021**, *9*, e002694. <https://doi.org/10.1136/jitc-2021-002694>.
23. Buonaguro, L.; Mauriello, A.; Cavalluzzo, B.; et al. Microorganisms-derived antigens for preventive anti-cancer vaccines. *Mol. Asp. Med.* **2023**, *92*, 101192. <https://doi.org/10.1016/j.mam.2023.101192>.
24. Tagliamonte, M.; Cavalluzzo, B.; Mauriello, A.; et al. Molecular mimicry and cancer vaccine development. *Mol. Cancer* **2023**, *22*, 75. <https://doi.org/10.1186/s12943-023-01776-0>.
25. Mauriello, A.; Cavalluzzo, B.; Ragone, C.; et al. Molecular mimicry of SARS-CoV-2 antigens as a possible natural anti-cancer preventive immunization. *Front. Immunol.* **2024**, *15*, 1398002. <https://doi.org/10.3389/fimmu.2024.1398002>.
26. Cherkasova, E.; Scrivani, C.; Doh, S.; et al. Detection of an Immunogenic HERV-E Envelope with Selective Expression in Clear Cell Kidney Cancer. *Cancer Res.* **2016**, *76*, 2177–2185.
27. Rycaj, K.; Plummer, J.B.; Yin, B.; et al. Cytotoxicity of human endogenous retrovirus K-specific T cells toward autologous ovarian cancer cells. *Clin. Cancer Res.* **2015**, *21*, 471–483.
28. Garde, C.; Pavlidis, M.A.; Garces, P.; et al. Endogenous viral elements constitute a complementary source of antigens for personalized cancer vaccines. *npj Vaccines* **2025**, *10*, 54.
29. Wang, B.D.; Lee, N.H. Aberrant RNA Splicing in Cancer and Drug Resistance. *Cancers* **2018**, *10*, 458.
30. Nesci, S.; Marchi, S.; Hu, J.; et al. Inflammatory mitochondrial signalling and viral mimicry in cancer. *J. Transl. Med.* **2025**, *23*, 982. <https://doi.org/10.1186/s12967-025-06931-3>.
31. Chen, R.; Ishak, C.A.; De Carvalho, D.D. Endogenous Retroelements and the Viral Mimicry Response in Cancer Therapy and Cellular Homeostasis. *Cancer Discov.* **2021**, *11*, 2707–2725.
32. Hu, J.; Marincola, F.M.; Clay, E.; et al. The perfect storm: Viral mimicry meets cancer dark matter. *Transl. Insights* **2026**, *1*, 3.
33. Sidney, J.; Peters, B.; Frahm, N.; et al. HLA class I supertypes: A revised and updated classification. *BMC Immunol.* **2008**, *9*, 1. <https://doi.org/10.1186/1471-2172-9-1>.
34. Cruz-Tapias, P.; Castiblanco, J.; Anaya, J.M. Major histocompatibility complex: Antigen processing and presentation. In *Autoimmunity: From Bench to Bedside [Internet]*; Anaya, J.M., Shoenfeld, Y., Rojas-Villarraga, A., et al., Eds.; El Rosario University Press: Bogota, Colombia, 2013; Chapter 10. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK459467> (accessed on 22 May 2026).
35. Cresswell, P.; Ackerman, A.L.; Giodini, A.; et al. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. *Immunol. Rev.* **2005**, *207*, 145–157. <https://doi.org/10.1111/j.0105-2896.2005.00316.x>.
36. Ely, Z.A.; Kulstad, Z.J.; Gunaydin, G.; et al. Pancreatic cancer-restricted cryptic antigens are targets for T cell recognition. *Science* **2025**, *388*, eadk3487. <https://doi.org/10.1126/science.adk3487>.
37. Kina, E.; Laverdure, J.P.; Durette, C.; et al. Breast cancer immunopeptidomes contain numerous shared tumor antigens. *J. Clin. Investig.* **2024**, *134*, e166740. <https://doi.org/10.1172/JCI166740>.