HPV T-cell epitope landscape: systematic mapping of distribution, conservation, and HLA promiscuity towards rational vaccine design

Authors

Syandrez Prima Putra[1,2], Selin Cankat[1], Leo Swadling[1*]

Affiliations

[1]Division of Infection and Immunity, University College London, London, United Kingdom

[2]Department of Microbiology, Faculty of Medicine, Universitas Andalas, Padang, Indonesia

*Correspondence: l.swadling@ucl.ac.uk

Abstract

Background: Human papillomavirus (HPV) drives both malignant and benign tumors, yet the current prophylactic vaccines remain type-restricted and are not optimized for T-cell induction. Evidence shows that T-cells can both block and clear viral infections, supporting T-cell-based vaccines. We compile a comprehensive list of T-cell epitopes, analyze specificity and immunodominance, and perform *in silico* analysis to investigate conserved and HLA promiscuous T-cell epitopes to support immune monitoring and to guide prophylactic and therapeutic pan-HPV vaccine design.

Methods: Functionally validated HPV T-cell epitopes were curated from the Immune Epitope Database (IEDB), filtered by length, HLA restriction, and immunogenicity. Epitopes were mapped across viral proteins, genotypes and HLA. Additionally, we assessed conservation of each epitope across 454 complete HPV reference genomes and predicted their HLA-promiscuity.

Results: We identified 485 unique epitopes (133 studies; 1,494 functional assays). E6 and E7 proteins exhibited the highest epitope density and immunogenicity (CD4+/CD8+ response rate of 0.25/0.20 and 0.21/0.29, respectively). L2, despite high conservation, yielded only one validated epitope. High-risk HPV, especially HPV16 and HPV18, dominated the epitope landscape ($p<.001$), and favored $CD8^+$ epitopes (OR 6.18; 95% CI 3.40–11.22; $p<.001$). Epitopes were restricted by 27 Class I and 21 Class II HLA alleles, with broader Class II allele

restriction (OR 8.03; 95% CI 4.08-15.78; p<.001). *In silico* analysis showed epitopes in the capsid proteins L1/L2 and in high-risk HPV types were more conserved (p<.001). Conserved, immunodominant, and HLA-promiscuous epitopes were identified from existing data.

Conclusion: HPV epitope data are dominated by E6/E7, while conserved regions are underexplored, with L2 the most neglected. Epitope mapping and validation of conserved and HLA-promiscuous epitopes across broader HPV genomic datasets is critical for accurate assessment of cellular immunity to HPV and for rational design of next-generation vaccines.

Keywords: Human papillomavirus, T-cell epitope, vaccine design, cross-reactivity, conservation, immunodominance, HLA promiscuity, functional, prediction

Introduction

Human papillomavirus (HPV) is a circular double-stranded DNA virus with an ~8 kb genome, responsible for causing 5% of all human cancers, especially cervical, vulvar, vaginal, penile, anal, and head and neck (Baba et al., 2025). HPV-related cervical cancer is the second most common cause of death in women of reproductive age (Sung et al., 2021). HPV consists of more than 450 identified genotypes, of which 12 (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) are categorized as carcinogenic or high-risk types by the International Agency for Research on Cancer (IARC) (Wei et al., 2024). The development and roll-out of prophylactic bivalent, quadrivalent, and nonavalent L1-based virus-like particle vaccines has prevented ~70%-90% of HPV-related cancer incidence worldwide (Wang et al., 2020); however, these vaccines are type-restricted, non-universal, and have no therapeutic efficacy. In low- and middle-income countries where HPV-related cancers are common, such vaccinations are still not effectively implemented due to cost, alongside the insufficient screening and treatment modalities, meaning even vaccine-covered genotypes remain a significant burden (Boldeanu et al., 2025; Tsu et al., 2021). Thus, next-generation vaccines that are cross-protective, affordable, and effective in both prophylactic and therapeutic measures are required.

One promising approach lies in leveraging T-cell immunity (Dull et al., 2024; Eberhardt et al., 2021; Peng et al., 2023). Beyond neutralizing antibodies, T-cells play a crucial role in controlling viral infections by eliminating infected cells (Huber et al., 2014). Several lines of evidence point to a key role for T-cell immunity in controlling HPV infection and associated neoplasia (Van Bockel and Kelleher, 2023). HPV-specific T-cells can be found in the blood and

at higher numbers in the tumor site and draining lymph node in HPV infected individuals (Hilders et al., 1994; Peng et al., 2023). Those who progress to cervical intraepithelial neoplasia have a lower magnitude T-cell response than those who don't (Evans et al., 1997; Hilders et al., 1994). Expansion of HPV-specific T-cells has also been temporally correlated with regression of warts (Coleman et al., 1994). Genetic association studies have repeatedly found associations between outcome and human leukocyte antigen (HLA), in particular for class II alleles (Apple et al., 1994; Bontkes et al., 1998; Bowden et al., 2021) and HIV positivity and associated defects in cellular immunity are associated with higher rates of HPV acquisition, persistence and cancer progression (Clifford et al., 2017; Stelzle et al., 2021).

We, and others, have recently shown that T-cells can mediate protection from detectable infection even in the absence of antibodies (Fumagalli et al., 2024; Kalimuddin et al., 2025), in particular memory T-cell responses that recognize conserved viral regions, as observed in abortive SARS-CoV-2 infection (Swadling et al., 2022). T-cells can recognize all viral proteins - including highly-conserved non-structural regions that are not accessible to antibodies, are inherently cross-reactive, and tend to be more long-lived than humoral immunity (Dan et al., 2021; Eberhardt et al., 2021; Manisty et al., 2021). This highlights the potential of T-cells to provide broad and more durable anti-viral protection. Translating this principle to HPV requires consideration of its viral proteome, which comprises early proteins (E1–E7)—involved in replication and oncogenesis, and late proteins (L1, L2)—forming the viral capsid (Nelson and Mirabello, 2023). Within this landscape, T-cell epitopes that are conserved across genotypes may confer cross-type immunity (Eickhoff et al., 2019; Swadling et al., 2022). HLA molecules present viral peptides to T-cells, and their polymorphism determines population-level immunity as demonstrated in previous HPV studies (Nakagawa et al., 2007). Comprehensive mapping of HPV T-cell epitopes is therefore critical to identify the most promising candidates for broad-spectrum vaccine design.

To address this gap, we established a systematic framework to analyze HPV T-cell epitopes curated from published studies and public databases. The Immune Epitope Database (IEDB) was used as it provides the most comprehensive and standardized repository of experimentally validated T-cell epitopes, enabling reproducible large-scale analyses across pathogens. Our focus was threefold: to chart epitope distribution across proteins and genotypes, to evaluate conservation against a reference panel of complete genomes, and to

assess HLA- promiscuity using *in silico* predictions. This approach enables an assessment of how extensively HPV T-cell epitopes have been explored to date, and where key gaps remain. It also acts as a resource for individuals designing reagents (peptide pools, MHC multimers) to investigate cellular immunity to HPV in natural history or vaccine studies. Ultimately, the framework aims to guide the prioritization of conserved and promiscuous epitopes as candidates for next-generation HPV vaccines, while also laying the groundwork for broader application to other oncogenic viruses.

Methods

Study design, data collection and initial curation strategy

We conducted a meta-analysis integrated with *in silico* exploration by systematically curating a dataset of functionally validated HPV T-cell epitopes from IEDB (https://www.iedb.org; **Fig.1**) (Vita et al., 2025). For the inclusion criteria, we searched the database using the following filters: (1) filter options = T-cell (excluding B-cell and MHC binding-only assays), (2) epitope = any, (3) epitope source, organism = Human papillomavirus (all types), (4) T-cell assay outcome = positive, (5) MHC restriction = any, (6) host = *Homo sapiens* (human), (7) disease = any, and (8) reference = any. We exported the results from the 'Assays' query as full data columns. To maintain consistency, we merged epitopes with identical peptide sequences across studies and counted each sequence once as a unique epitope for distribution analyses, while preserving all associated assay information and references. We then screened the peptides by length, retaining only 8–11 amino acids for Major Histocompatibility Complex (MHC) Class I restricted $CD8^+$ T-cell epitopes and 12–23 amino acids for MHC Class II restricted $CD4^+$ T-cell epitopes, and excluded those outside these ranges (Grifoni et al., 2019). We used this curated dataset as a core for analysis (485 epitopes, 'proteome analysis' epitope list). To minimize bias, screening and extraction was repeated three times.

Proteome analysis

We downloaded the HPV16 reference sequence FASTA file from NCBI GenBank (NC_001526.4) and extracted the protein annotations to construct a representative HPV proteome map, showing the relative positions and amino acid lengths of each protein. Using

the IEDB core dataset, we then mapped all unique epitopes onto this proteome. Protein annotations were curated directly from the core dataset. Epitope density was defined for each HPV protein (E6, E7, E1, E2, E4, E5, L2, and L1) as the number of epitopes per 100 amino acids.

Genotype analysis

From the 'proteome analysis' dataset, we filtered unique T-cell epitopes based on HPV genotype annotation and excluded epitopes with unspecified type from genotype analysis to give the 'genotype analysis' epitope list (n=484). HPV risk groups were classified according to the IARC, comprising 12 high-risk types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59), with the remaining types designated as low risk. Epitopes with a single genotype annotation were classified as single-type, whereas those associated with more than two genotypes were classified as cross-type.

Immunogenicity analysis

Immunogenicity was defined by the response rate, calculated as the percentage of positive responses across all subjects tested in functional assays. For each unique epitope, the total number of subjects tested and the number of positive responders from different studies were pooled, and the response rate was recalculated (range 0–1). Where the number of subjects tested and with positive responses were not specified in IEDB the epitopes were excluded, leaving 316 epitopes in the 'immunogenicity' epitope list. Subject-level response frequencies were compared between $CD4^+$ and $CD8^+$ T-cell epitopes and across HPV proteins. The proportional contribution of each HPV protein to the overall T-cell epitope pool was illustrated using stacked horizontal bar plots, showing the relative percentage and number (n) of epitopes per protein. Epitope response profiles were visualized using immunome browser–style plots (IEDB), where response rates were mapped against amino acid positions within each protein sequence.

HLA analysis

Epitopes from the 'proteome analysis' dataset were filtered based on HLA (MHC class) annotation. All epitopes without defined HLA were excluded from the analysis, leaving 289 epitopes in the 'HLA analysis' epitope list. Epitope subsets were stratified into known MHC class I ($CD8^+$) and MHC class II ($CD4^+$) annotations according to IEDB data. The distribution of

epitopes with known HLA-annotated and restricted to allele-level was presented in absolute numbers. To characterize HLA promiscuity, epitopes annotated with a single HLA allele were classified as mono-allelic, whereas those with more than one allele were classified as multi-allelic, or HLA promiscuous epitope. Candidate epitopes for vaccine development were prioritized if they demonstrated a response rate >0.5 and were restricted by more than one HLA allele.

Conservation analysis

To evaluate conservation, each unique T-cell epitope from the 'proteome analysis' dataset (n=485) was mapped *in silico* against 454 complete HPV reference genomes retrieved from the Papillomavirus Episteme (PaVE, https://pave.niaid.nih.gov/explore/reference_genomes/human_genomes) (Dommer et al., 2025). Mapping was performed using six-frame translation to capture epitopes encoded on both forward and reverse strands. Conservation was defined as a 100% amino acid match. For downstream analyses, conservation was summarized at the HPV type level: epitope conservation was expressed as absolute counts (N) and percentages (%) of HPV types in which the epitope sequence was found identically across all 454 reference genomes. Conservation was further stratified by protein, peptide length, and T-cell type ($CD4^+$ or $CD8^+$). Statistical comparisons of conservation distributions were performed for proteins, T-cell subsets and risk groups. The top 30 HPV types with the highest number of mapped epitopes were identified, with annotation of their risk classification (high-risk vs low-risk). A proteome map was constructed to visualize the distribution of fully conserved (100%) epitopes across HPV proteins. The map was stratified by HPV risk group (high-risk vs low-risk) and further annotated according to T-cell subset ($CD8^+$ or $CD4^+$) and protein functional category (oncoprotein, replication, or structural).

HLA prediction analysis

To evaluate HLA promiscuity, each unique epitope from the 'proteome analysis' dataset (n=485) was assessed for predicted HLA binding *in silico* using NetMHCpan (v4.1) (Reynisson et al., 2020a) and NetMHCIIpan (v4.1) (Reynisson et al., 2020b) following the default recommended thresholds. Binding affinity to MHC class I alleles ($CD8^+$) was predicted across a panel of 20 representative HLA class I molecules, covering the major supertypes: HLA-

C*02:02, HLA-A*02:01, HLA-C*04:01, HLA-A*03:01, HLA-A*01:01, HLA-A*11:01, HLA-C*06:02, HLA-B*08:01, HLA-A*26:01, HLA-C*03:03, HLA-C*03:04, HLA-C*01:02, HLA-B*15:01, HLA-A*24:02, HLA-B*58:01, HLA-B*39:01, HLA-B*07:02, HLA-B*40:01, HLA-A*33:03, and HLA-B*27:05. For MHC class II epitopes ($CD4^+$), binding was predicted across the IEDB reference panel of HLA class II molecules (including HLA-DRB1*07:01, HLA-DQA1*01:02/DQB1*06:04, HLA-DQA1*02:01/DQB1*02:01, HLA-DRB1*09:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DQA1*01:02/DQB1*05:01, HLA-DPA1*01:03/DPB1*04:01, HLA-DRB1*13:01, HLA-DRB1*0101, HLA-DQA1*01:03/DQB1*05:01, HLA-DRB1*03:01, HLA-DPA1*01:03/DPB1*04:02, HLA-DRB1*11:01, HLA-DRB1*15:01, HLA-DRB1*04:01, HLA-DQA1*04:01/DQB1*03:01, HLA-DQA1*05:01/DQB1*03:01, HLA-DQA1*05:05/DQB1*03:01, HLA-DRB1*04:03, and HLA-DQA1*01:02/DQB1*06:02) (Sanchez-Mazas et al., 2024). Predicted binding affinity was expressed as percentile rank scores, with lower values indicating stronger binding. For MHC class I, epitopes were classified as strong binders (SB) when rank ≤0.5%, weak binders (WB) when 0.5–2%, and non-binders when >2%. For MHC class II, epitopes were classified as SB when rank ≤1%, WB when 1–5%, and non-binders when >5%. For each epitope, the number of predicted HLA binders (SB or WB) was calculated to quantify promiscuity. Epitope distributions were differentiated systematically for $CD8^+$/MHC I and $CD4^+$/MHC II subsets by: (1) non-binders versus binders, (2) mono- versus multi-allelic, (3) comparison of total HLA binders per epitope, (4) HPV proteins, and (5) the number of WB and SB in each HLA allele. Heatmaps were generated to depict the binding profiles of the top 25 promiscuous peptides within each HPV protein across the representative HLA allele panel.

Statistical analysis and reproducibility

To ensure rigorous evaluation of the data, we applied statistical methods tailored to the type and distribution of each variable. Categorical outcomes were analyzed using Pearson's chi-square test or Fisher's exact test, and the strength of association was expressed as odds ratios (OR) with 95% confidence intervals (CI). For continuous variables that were not normally distributed, comparisons were performed with the Mann–Whitney U test or Kruskal–Wallis test. Correlations between variables were assessed using Spearman's rank correlation (ρ). A p-value <0.05 was considered statistically significant. For immunogenicity, 95% CIs were estimated using the Wilson score method. All analyses were carried out in Python v3.12.11, employing established scientific libraries including pandas (v2.2.2), NumPy

(v2.0.2), SciPy (v1.16.1), statsmodels (v0.14.5), matplotlib (v3.10.0), seaborn (v0.13.2), and Biopython (v1.85). Figures were produced in vector format to ensure publication quality.

Data and Code Availability

All data analysed during this study are included in this published article and its Supplementary Information. Correspondence and requests for materials should be addressed to L.S. Custom scripts are available on request.

Results

Defining a core dataset of HPV T-cell epitopes

To define the antigenic landscape of HPV and inform rational vaccine strategies, we mapped T-cell epitopes by compiling 610 peptide sequences reported across 133 human studies/1,494 different functional assays from the IEDB (**Fig.1**). To ensure consistency in size, we excluded 125 peptides that did not meet peptide length criteria, resulting in a core dataset ('proteome analysis' epitope list) of 485 unique epitopes, consisting of 293 $CD4^+$ and 192 $CD8^+$ epitopes (**Supp Table. S1**). This curated dataset highlights both the diversity of epitope sources and the availability of high-quality, functionally validated sequences. Building on this curated dataset, we next explored six key aspects of HPV epitope biology: (1) proteome distribution, (2) genotype distribution, (3) immunogenicity, (4) HLA restriction, followed by *in silico* prediction of (5) epitope conservation and (6) HLA binding (**Fig.1**). For genotype analysis, we retained 484 epitopes after excluding one with an unspecified HPV type. For immunogenicity, we analyzed 316 epitopes with available response data. For HLA analysis, we focused on 272 epitopes with clearly defined HLA. Finally, *in silico* prediction was performed on all 485 epitopes. Together, these refinements provide a robust foundation for dissecting epitope conservation, breadth of recognition, and their implications for T-cell–mediated protection.

[**Fig.1** Data curation workflow for HPV T-cell epitope meta- and *in silico* analyses]

Proteome-wide mapping highlights research bias toward E6 and E7

First, we explored the antigen distribution patterns of the 485 unique functionally validated T-cell epitopes we identified across the entire HPV proteome (**Fig.2a-d**). We depicted the genomic structure of HPV proteins composed of oncoproteins (E6, E7, E5), replication

proteins (E1, E2, E4), as well as structural proteins (L1, L2) using a reference sequence of HPV16 (**Fig.2a**). $CD8^+$ epitopes were dominated by 9-mers, whilst $CD4^+$ epitopes were largely 15-mers (**Supp Fig.1a**). The number of studies mapping epitopes in each viral protein differed remarkably: E6/E7 were the most assayed proteins, with L2 being represented in just two studies (**Fig.2b**). E6 contributed the most epitopes for both $CD8^+$ and $CD4^+$ T-cells (70 and 90, respectively), followed by E7 (60 and 48) and L1 (16 and 77). In contrast, only a single epitope has been described in L2 (0 and 1) (**Fig.2c**), despite L2 being identified as a promising candidate antigen for vaccines due to its conserved regions (Putra et al., 2022). Many $CD4^+$ T-cell epitopes have been identified in the main target of currently licensed vaccines, L1, likely due to an interest in $CD4^+$ help for these antibody-based vaccines. On length normalization, E6 and E7 held the highest epitope density with a higher number than predicted with their length (110.2 and 106.0 per 100 amino acids, respectively) (**Fig.2d**). Epitopes in the smaller-sized oncoproteins E6/E7 were over-represented in the dataset relative to other proteins, with larger-sized structural proteins like L2 being underrepresented (**Supp Fig.1b**). There was no correlation between protein length and epitope number (ρ=0.10, p=0.823). To benchmark this observation, we compared the similar approaches to well-studied SARS-CoV-2 T-cell epitopes (Grifoni et al., 2021). SARS-CoV-2 showed a strong positive correlation between protein size and epitope number (ρ=0.89, p<.001), likely due to the more comprehensive mapping performed for SARS-CoV-2. Overall, this highlights significant knowledge gaps across a broader HPV antigenic universe, indicating a concerted effort to map epitopes and understand the T-cell response to the oncoproteins E6/E7, but limited exploration of other proteins.

[**Fig.2** Proteome-wide mapping and genotype-level distribution of HPV T-cell epitopes.]

[**Supp Fig.1** HPV T-cell epitopes length and immunodominance.]

High-risk types of HPV are overrepresented

To explore how genotype and clinical risk categories influence distribution among reported T-cell epitopes, we compared 484 unique HPV epitopes with known genotype information (**Fig.2e-h**). Most epitopes were identified in studies of high-risk types of HPV, with HPV16 and HPV18 possessing a substantially higher number of epitopes than low-risk types (**Fig.2e**). Epitopes selected from high-risk types were approximately six times as likely to be $CD8^+$ as $CD4^+$ (OR 6.18; 95% CI 3.40–11.22; p<.001) (**Fig.2f**). When stratified by protein, high-

risk types were represented with increased epitope representation in E5, E2, E1, E6, and E7 but reduced epitope representation in L1, L2, and E4, compared to low-risk ($p<.001$) (**Fig.2g**). This bias is expected as there is much interest in the therapeutic potential of cytotoxic CD8s targeting the oncoproteins of the high-risk genotypes. While most epitopes were reported to be genotype-specific, approximately 17% were found to be present among multiple types of HPV (**Fig.2h**). Very few studies assessed the conservation of epitopes or the cross-reactivity of T-cell responses across genotypes and so this is a significant underestimation of the potential for pan-genotype epitopes. Below we compile a reference sequence dataset and perform *in silico* analysis to better assess epitope conservation. These data overall reflect a high scientific emphasis upon high-risk types of HPV, particularly $CD8^+$-restricted epitopes, but reveal shared epitopes as promising candidates to generate broad-spectrum vaccines.

Immunodominance of E6 and E7 across and within HPV protein landscape

To better understand the immunogenic regions within HPV proteins, we analyzed 316 reported epitope T-cell responses, examined the frequency of positive responses among tested subjects according to T-cell types ($CD4^+$ and $CD8^+$) and proteins, and then mapped them to the viral proteome (**Fig.3a-b; Supp Fig.1c-d**). $CD8^+$ responses were more frequently detected than $CD4^+$ responses (OR=1.26, 95%CI=1.13-1.40, $p<.001$) (**Fig.3a**). Oncoproteins E5, E6 and E7 elicited higher positive responses, whereas early proteins like E1 and E2 were rarely immunogenic ($p<.001$) (**Fig.3b**). Many epitopes were derived from E6 (n=122) and E7 (n=73), together making up more than half of all reported epitopes (**Supp Fig.1c**). E6 and E7 showed the highest density of reported responses along their amino acid sequences, while other proteins—such as L1, E5, E2, and E1—were less frequently studied, and data for E4 and L2 were extremely limited (**Supp Fig.1d**). These findings suggest a strong research focus on E6 and E7 in HPV-specific T-cell immunity and highlighting the lack of data on proteins like L2, pointing to important gaps for future investigation.

[**Fig.3** HLA restriction landscape and immunogenicity patterns of HPV T-cell epitopes.]

Asymmetric HLA restriction reveals data coverage and research biases

To assess the potential breadth of T-cell cross-reactivity in genetically heterogeneous populations, we compared known HLA restriction profiles for 272 epitopes of HPV with available allelic data (**Fig.3c-f; Supp Fig.2**). A larger number of epitopes with HLA annotation

and restricting alleles have been identified for $CD8^+$ than $CD4^+$ responses (**Fig.3c**; **Supp Table. S2**). $CD4^+$ epitopes were eight times more likely to be restricted by multiple alleles (promiscuous) than $CD8^+$ epitopes (OR=8.03; 95% CI=4.08-15.78; p<0.001) (**Fig.3d**). Overall, HLA restriction profiles were highly biased, with a small number of alleles accounting for most epitopes described for MHC class I and MHC class II (**Fig.3e**). A total of 27 Class I and 21 Class II HLA alleles restricted the epitopes, with HLA-A*02:01 (70 epitopes) and HLA-DRB1*16:01 (31 epitopes) being the most frequent globally. The number of epitopes identified per HLA allele did not differ significantly between Class I and Class II molecules (p=0.536), suggesting a comparable breadth of HLA coverage across both T-cell subsets (**Supp Fig.2a**). The number of HLAs restricting epitopes within a protein were significantly correlated with the number of epitopes described for that protein, with E6 and E7 also being outliers for epitope density alongside HLA-engagement breadth (**Supp Fig2.b-c**).

To explore patterns of HLA allele restriction and epitope breadth, we examined 48 $CD4^+$ and 116 $CD8^+$ epitopes with available HLA data (**Fig.3f**). The most broadly restricted $CD4^+$ epitope binds 5 different HLA alleles, compared to a maximum of 4 for $CD8^+$ epitopes. $CD4^+$ epitopes showed a wider protein distribution, suggesting broader potential population coverage, while $CD8^+$ epitopes were more clustered. The most promiscuous epitopes were LLSVSTYTSLILLVL ($E5_{33\text{-}47}$, response rate of 0.63) and GLYYVHEGIRTYFVQ ($E2_{156\text{-}170,}$ response rate of 0.57) for $CD4^+$. For $CD8^+$, KLPDLCTEL ($E6_{13\text{-}21}$, response rate of 0.33) and EYRHYCYSL ($E6_{75–83}$, response rate of 0.32) were the most promiscuous but they had relatively low recorded response rates. These findings suggest that even though some HPV epitopes can bind to several HLA alleles, the database was highly biased toward a restricted set of common alleles and was biased by greater mapping of epitopes for certain HPV proteins.

[**Fig.3** HLA restriction landscape and immunogenicity patterns of HPV T-cell epitopes.]

In silico **prediction suggests limited cross-genotype conservation**

It is crucial to understand how conserved the HPV epitopes are across genotypes to facilitate identification of candidates with broad-spectrum T-cell immunity potential. To pursue this further, we performed an *in silico* screen against a database of 454 HPV reference genomes for the 485 functionally validated epitopes (**Fig.4**; **Supp Table. S3**). $CD8^+$ peptides generally covered more HPV types per peptide length than $CD4^+$ peptides. The maximum

coverage of HPV T-cell epitopes was 34 (7.5%) genotypes for $CD8^+$ epitope LQFIFQLCK (9-mer) and 21 (4.6%) for $CD4^+$ epitope FNKPYWLQRAQGHNN (15-mer) (**Fig.4a**). Structural protein L1 had higher conservation (median [IQR], 0.44% [0.22–0.66%]; p<.001) than the non-structural proteins (**Fig.4b**). There was no statistical distinction evident between overall $CD4^+$ versus $CD8^+$ epitope conservation (**Supp Fig.3a**). At the genotype level, epitope matches were contributed mostly by HPV16 (n=288, 59.4%), followed by HPV6 (n=42, 8.7%), then by HPV1 (n=37, 8.1%) and HPV18 (n=36, 7.4%) (**Fig.4c**). Epitopes of high-risk HPV were significantly more conserved than low-risk (median 0.2% vs. 0.0%, p<.001) (**Fig.4d**). Conserved epitopes were unevenly distributed across the HPV proteome, with dense clusters observed in E6/E7 and the structural protein L1, and replication proteins harbored relatively few epitopes (**Fig.4e**). High-risk HPV types exhibited a greater concentration of conserved $CD8^+$ and $CD4^+$ epitopes compared to low-risk types, highlighting immunodominant regions with potential relevance for broad vaccine design. These results suggest known HPV epitopes have low cross-genotype conservation.

[**Fig.4** Epitope conservation across HPV genomes following *in silico* analysis]

[**Supp Fig.3** *In silico* analysis of HPV T-cell epitope type coverage and donor response rate by T-cell subset and binding status]

Distinct patterns of predicted promiscuity in $CD4^+$ and $CD8^+$ epitopes across HLA alleles

For the epitopes described in IEDB, few studies have systematically considered the full breadth of HLA-allele promiscuity of each peptide, with most designating a single HLA allele only. To get a better appreciation of the promiscuity of HPV epitopes, we systematically predicted binding of the validated epitopes to a curated set of HLAs covering major super-types using NetMHCpan and NetMHCIIpan (**Fig.5, Supp Table. S4**). We investigated the association between binding capacity and epitope distribution by assessing the proportion of $CD8^+$ and $CD4^+$ T-cell epitopes among binders and non-binders. The analysis showed a significantly higher proportion of $CD8^+$ epitopes among binders compared to non-binders (OR=4.89; 95%CI=3.30-7.25; $p < 0.001$; **Fig.5a**). Following this, to evaluate the potential for false positives, we assessed the response rates across 316 epitopes from the immunogenicity dataset. $CD8^+$ non-binders displayed a significantly higher response rate than $CD8^+$ binders (p=0.048), suggesting an enrichment of artifacts in the non-binder category. Conversely, no

significant difference was detected between $CD4^+$ binders and non-binders (p=0.127) (**Supp Fig.3b**). Frequency analysis further showed that $CD4^+$ epitopes were generally predicted to be restricted to 10 HLA alleles, whereas $CD8^+$ epitopes, in some cases, reached 13 (**Fig.5b**). These findings describe the similar potential of both Class I and Class II epitopes to provide broader population coverage.

Protein-level stratification highlighted the differential distribution of promiscuous epitopes across the HPV proteome. Analysis of allelic restriction revealed that $CD8^+$ epitopes were predominantly multi-allelic across all HPV proteins (**Fig.5c**), whereas $CD4^+$ epitopes exhibited protein-specific patterns, with L1 showing a significant enrichment of multi-allelic binders (OR=2.80 versus all proteins, p=0.031) (**Fig.5d**). At the allele level, both $CD8^+$ and CD4+ binders were dominated by weak binders across all HLA alleles. The most promiscuous HLA for MHC class I were HLA-C*02:02 and HLA-A*02:01, while MHC class II were HLA-DRB1*07:01 and HLA-DQA*02:01/DQB1*02:01 (**Fig.5e-f**). Heatmap analysis revealed that the most promiscuous epitopes for both $CD8^+$ and $CD4^+$ were derived from major capsid protein L1, while early proteins were suggested to be more promiscuous within MHC Class I (**Fig.5g–h**). Nonetheless, the overall distribution remained skewed, with only a minority of epitopes achieving high-level promiscuity, and the majority restricted to one or a few alleles.

[**Fig.5** Epitope promiscuity across HLA alleles following *in silico* prediction]

Discussion

HPV is a major global health threat, causing large numbers of anogenital and oropharyngeal cancers globally despite effective prophylactic vaccines. Current vaccines, though remarkably effective against vaccine-included types, offer limited cross-type coverage without therapeutic effect against pre-existing infections, leaving a critical void for disease attenuation (Mo et al., 2022). There is increasing evidence to highlight the importance of T-cell immunity in infection blocking (Kalimuddin et al., 2025; Swadling and Maini, 2023) as well viral clearance and long-term viral control, making T-cell epitopes central targets for next-generation prophylactic vaccines as well as immunotherapeutics. Prophylactic approaches should benefit by moving beyond a narrow focus on neutralizing antibodies, to additionally exploiting the durability and cross-reactive potential of T-cell immunity. There are also many settings in which accurate quantification and assessment of the T-cell response to HPV may

aid in our understanding of pathogenesis and correlates of protection. Systematic epitope mapping throughout the entire HPV proteome is required to aid the design of reagents for the measurement of HPV-specific T-cell immunity and to inform design of vaccine immunogens.

T-cells are not all equally effective at anti-viral control. Specificity (at the level of TCR, epitope, protein and virus) and immunodominance (both in terms of magnitude of the response and proportion of individuals recognizing an epitope) influences how effective a T cells is and when a T-cell can act within an infection. An in-depth understanding of the locations of epitopes and response rates is essential. Our study brings together all currently available data on anti-HPV T cell specificity and immunodominance, highlighting key knowledge gaps, and using *in silico* tools to start to fill these gaps. This data offers a framework for a rationale-based vaccine design by describing conserved cross-reactive, HLA-promiscuous targets.

To the best of our knowledge, this study offers the first proteome-wide database of HPV T-cell epitopes that has been meticulously curated and thoroughly annotated by protein, genotype, and HLA restriction. Extending prior reports of HPV epitopes (Gallagher and Man, 2007; McInnis et al., 2023; Mercier-Letondal et al., 2018; Speetjens et al., 2022). Comparable efforts in alternate viral systems, i.e., influenza virus and SARS-CoV-2 (Chaisawangwong et al., 2022; Gaevert et al., 2021), have demonstrated the value of systematic epitope compilation and underscore the importance of such an enterprise for HPV.

HPV oncogenes E6 and E7 were highly represented among the epitope datasets, with peak epitope density reported despite their small size. These patterns support long-standing evidence that E6 and E7 (and to a lesser extent E1) (Peng et al., 2023) are common host T-cell targets, uniformly preserved throughout tumor cells in expression profiles, and major causes of HPV-containing cancers (Inturi and Jemth, 2021; Janiszewska et al., 2025; Letafati et al., 2024; Münger et al., 1989; Ranasinghe and McMillan, 2025). Epitopes within E6/E7 are biased to CD8+ responses and high-risk genotypes. Epitope density of E6/E7 within the literature has served to support development for some prospective therapeutic vaccine candidates (Fert et al., 2024; Lee et al., 2023; Zhang et al., 2025). In contrast, proteins such as L2, have been investigated in fewer studies, translating to fewer known epitopes, despite proven cross-type protective immunity possibilities (Tsukamoto et al., 2024; Wang and Roden, 2013; Zhao et al., 2020). Our previous work, which revealed minimal variation in HPV16 L2 sequences from

Indonesian cervical cancer samples, supported the notion that L2 contains conserved regions worthy of further study (Putra et al., 2022). We also highlighted a bias towards CD4 helper epitope discovery within L1, which aids design of prophylactic antibody-based vaccines. These conclusions present a clear bias in the epitope landscape.

It is evident that T-cell epitope coverage in literature is comprised mostly of high-risk HPV types with a special focus on HPV16, consistent with its established ranking among the most common genotypes associated with cancers (Nelson and Mirabello, 2023; Sadeghi et al., 2025). The importance of identifying and exploiting CD8 epitopes in oncogenes of high-risk genotypes is supported by previous work showing diminished $CD8^+$ function is accompanied by impaired viral clearance and higher risk for cancers (Jemon et al., 2016; Maskey et al., 2019; Mondatore et al., 2022). Protein targets that vary among high- compared to low-risk types also indicate a focus in previous research on early proteins central to viral replication as well as malignant conversion (Bhattacharjee et al., 2022). Of particular interest is the identification of epitopes that span multiple types of HPV to signify their importance as cross-protective antigens for a vaccine, which has already been implemented among several viruses, such as influenza (Eickhoff et al., 2019) and SARS-CoV-2 (Wang et al., 2023).

Importantly, analysis of HLA restriction revealed that HPV epitopes had broad coverage of a large repertoire of Class I and Class II molecules, however, they were concentrated disproportionately within a restricted subset of alleles, most notably HLA-A*02:01 and HLA-DR4. Such distribution mirrors an extensive experimental focus on more common alleles as has been described for other viruses (Medhasi and Chantratita, 2022; Sanchez-Mazas et al., 2024). The closed groove of MHC class I enforces stricter binding specificity, while the open-ended groove of MHC class II permits longer and more variable peptides (Rock et al., 2016). This is reflected in a preponderance towards mono-allelic restriction amongst $CD8^+$ epitopes suggests narrower coverage of individual CD8+ epitopes than $CD4^+$ epitopes, with a concern for population coverage within highly disparate genetic backgrounds (Sanchez-Mazas et al., 2024).

Our *in silico* analyses demonstrates that only a small percentage of known HPV T-cell epitopes exhibit broad conservation across types, with the majority being highly genotype-specific. In accordance with evidence that capsid-derived regions contain conserved epitopes that can induce cross-protection across diverse HPV types, structural proteins like L1 and L2

emerged as relatively richer sources of cross-type candidates (Ahmels et al., 2022; Tsukamoto et al., 2024). Although conserved fragments within E6/E7 have been reported and are being investigated as multigenotype therapeutic targets, epitopes from early proteins tended to be type-restricted (Hancock et al., 2019; Riemer et al., 2010).

Parallel HLA binding prediction showed that while most epitopes were allele-restricted, a small subset of highly promiscuous binders emerged. These epitopes are especially useful because HLA promiscuity extends the range of individuals who can present and target the epitope with T cell immunity (Grabowska et al., 2015; Rock et al., 2016). These findings underscore the dual challenge of limited conservation and restricted allele coverage for most HPV epitopes. Although no pan-genotype epitopes were identified here, we have highlighted a small number of epitopes with high response rates (immunodominant) and wide HLA restriction (promiscuous) for further investigation. Overall, these data show proof-of-concept for identifying HLA-promiscuous epitopes but call for systematic validation of HLA promiscuity and epitope conservation.

There are various limitations to this study. First, although one aim of this analysis was to identify knowledge gaps and bias in epitope mapping, these inherent biases limit our ability to accurately determine immunodominance at the epitope and protein level and to fully characterize HLA promiscuity and cross-genotype conservation. But through highlighting these gaps we hope to inform future efforts. Second, although informative, *in silico* predictions of HLA binding and conservation cannot completely replace experimental confirmation of T-cell responses in various human populations. We only considered an epitope to be pan-genotypic if sequence identity was 100%, however, T-cells have flexibility in recognition and cross-reactivity can be retained to epitope variants. Third, the analyses were limited to reference genomes, which might not accurately represent intra-type variation found in clinical isolates. Lastly, the evidence that is currently available may limit generalizability across the entire HPV spectrum by overemphasizing high-risk genotypes, especially HPV16. These drawbacks illustrate how crucial it is to combine methodical *in silico* techniques with more extensive experimental research to improve epitope selection and direct the creation of vaccines that are applicable to both populations and other types.

Conclusion

In conclusion, this study provides the first proteome-wide map of HPV T-cell epitopes, integrating information on distribution, conservation, HLA restriction, and immunogenicity. Our findings reveal the focus on E6 and E7, the underrepresentation of structural proteins such as L2, and the limited conservation and HLA coverage that constrain most epitopes. At the same time, the identification of conserved and HLA-promiscuous sequences highlights promising candidates for cross-type and broadly protective vaccines. By bridging prophylactic and therapeutic perspectives, these results offer a rational framework to expand HPV vaccine design beyond neutralizing antibodies, emphasizing the potential of T-cell–based immunity to address both prevention of new infections and clearance of established disease.

References

Ahmels, M., Mariz, F.C., Braspenning-Wesch, I., Stephan, S., Huber, B., Schmidt, G., Cao, R., Müller, M., Kirnbauer, R., Rösl, F., Hasche, D., 2022. Next generation L2-based HPV vaccines cross-protect against cutaneous papillomavirus infection and tumor development. Front. Immunol. 13, 1010790. https://doi.org/10.3389/fimmu.2022.1010790

Apple, R.J., Erlich, H.A., Klitz, W., Manos, M.M., Becker, T.M., Wheeler, C.M., 1994. HLA DR–DQ associations with cervical carcinoma show papillomavirus–type specificity. Nat. Genet. 6, 157–162. https://doi.org/10.1038/ng0294-157

Baba, S.K., Alblooshi, S.S.E., Yaqoob, R., Behl, S., Al Saleem, M., Rakha, E.A., Malik, F., Singh, M., Macha, M.A., Akhtar, M.K., Houry, W.A., Bhat, A.A., Al Menhali, A., Zheng, Z.-M., Mirza, S., 2025. Human papilloma virus (HPV) mediated cancers: an insightful update. J. Transl. Med. 23, 483. https://doi.org/10.1186/s12967-025-06470-x

Bhattacharjee, R., Das, S.S., Biswal, S.S., Nath, A., Das, D., Basu, A., Malik, S., Kumar, L., Kar, S., Singh, S.K., Upadhye, V.J., Iqbal, D., Almojam, S., Roychoudhury, S., Ojha, S., Ruokolainen, J., Jha, N.K., Kesari, K.K., 2022. Mechanistic role of HPV-associated early proteins in cervical cancer: Molecular pathways and targeted therapeutic strategies. Crit. Rev. Oncol. Hematol. 174, 103675. https://doi.org/10.1016/j.critrevonc.2022.103675

Boldeanu, L., Assani, M.-Z., Boldeanu, M.V., Siloși, I., Manolea, M.-M., Văduva, C.-C., Assani, A.-D., Dijmărescu, A.L., 2025. Cervical Cancer in the Era of HPV: Translating Molecular Mechanisms into Preventive Public Health Action. Int. J. Mol. Sci. 26, 8463. https://doi.org/10.3390/ijms26178463

Bontkes, H.J., Van Duin, M., De Gruijl, T.D., Duggan- Keen, M.F., Walboomers, J.M.M., Stukart, M.J., Verheijen, R.H.M., Helmerhorst, T.J.M., Meijer, C.J.L.M., Scheper, R.J., Stevens, F.R.A., Dyer, P.A., Sinnott, P., Stern, P.L., 1998. HPV 16 infection and progression of cervical intra-epithelial neoplasia: Analysis of HLA polymorphism and HPV 16 E6 sequence variants. Int. J. Cancer 78, 166–171. https://doi.org/10.1002/(SICI)1097-0215(19981005)78:2%253C166::AID-IJC8%253E3.0.CO;2-X

Bowden, S.J., Bodinier, B., Kalliala, I., Zuber, V., Vuckovic, D., Doulgeraki, T., Whitaker, M.D., Wielscher, M., Cartwright, R., Tsilidis, K.K., Bennett, P., Jarvelin, M.-R., Flanagan, J.M., Chadeau-Hyam, M., Kyrgiou, M., 2021. Genetic variation in cervical preinvasive and invasive disease: a genome-wide association study. Lancet Oncol. 22, 548–557. https://doi.org/10.1016/S1470-2045(21)00028-0

Chaisawangwong, W., Wang, H., Kouo, T., Salathe, S.F., Isser, A., Bieler, J.G., Zhang, M.L., Livingston, N.K., Li, S., Horowitz, J.J., Samet, R.E., Zyskind, I., Rosenberg, A.Z., Schneck, J.P., 2022. Cross-reactivity of SARS-CoV-2– and influenza A–specific T cells in individuals exposed to SARS-CoV-2. JCI Insight 7, e158308. https://doi.org/10.1172/jci.insight.158308

Clifford, G.M., Tully, S., Franceschi, S., 2017. Carcinogenicity of Human Papillomavirus (HPV) Types in HIV-Positive Women: A Meta-Analysis From HPV Infection to Cervical Cancer. Clin. Infect. Dis. 64, 1228–1235. https://doi.org/10.1093/cid/cix135

Coleman, N., Birley, H.D.L., Renton, A.M., Hanna, N.F., Ryait, B.K., Byrne, M., Taylor-Robinson, D., Stanley, M.A., 1994. Immunological Events in Regressing Genital Warts. Am. J. Clin. Pathol. 102, 768–774. https://doi.org/10.1093/ajcp/102.6.768

Dan, J.M., Mateus, J., Kato, Y., Hastie, K.M., Yu, E.D., Faliti, C.E., Grifoni, A., Ramirez, S.I., Haupt, S., Frazier, A., Nakao, C., Rayaprolu, V., Rawlings, S.A., Peters, B., Krammer, F., Simon, V., Saphire, E.O., Smith, D.M., Weiskopf, D., Sette, A., Crotty, S., 2021. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science 371, eabf4063. https://doi.org/10.1126/science.abf4063

Dommer, J., Van Doorslaer, K., Afrasiabi, C., Browne, K., Ezeji, S., Kim, L., Dolan, M., McBride, A.A., 2025. PaVE 2.0: Behind the Scenes of the Papillomavirus Episteme. J. Mol. Biol. 437, 168925. https://doi.org/10.1016/j.jmb.2024.168925

Dull, P.M., Achilles, S.L., Ahmed, R., Barnabas, R.V., Campos, N.G., Chirgwin, K., Cohen, J.A., De Sanjosé, S., Doorbar, J., Einstein, M.H., Emerson, C.I., Gottlieb, S.L., Hildesheim, A., Qiao, Y., Ruff, P., Sampson, J.N., Sasieni, P., Schiffman, M., Shin, H., Stanley, M.A., Trimble, C.L., Wentzensen, N., Riemer, A.B., Schiller, J.T., Kreimer, A.R., 2024. Meeting report: Considerations for trial design and endpoints in licensing therapeutic HPV16/18 vaccines to prevent cervical cancer. Vaccine 42, 126100. https://doi.org/10.1016/j.vaccine.2024.07.001

Eberhardt, C.S., Kissick, H.T., Patel, M.R., Cardenas, M.A., Prokhnevska, N., Obeng, R.C., Nasti, T.H., Griffith, C.C., Im, S.J., Wang, X., Shin, D.M., Carrington, M., Chen, Z.G., Sidney, J., Sette, A., Saba, N.F., Wieland, A., Ahmed, R., 2021. Functional HPV-specific PD-1+ stem-like CD8 T cells in head and neck cancer. Nature 597, 279–284. https://doi.org/10.1038/s41586-021-03862-z

Eickhoff, C.S., Terry, F.E., Peng, L., Meza, K.A., Sakala, I.G., Van Aartsen, D., Moise, L., Martin, W.D., Schriewer, J., Buller, R.M., De Groot, A.S., Hoft, D.F., 2019. Highly conserved influenza T cell epitopes induce broadly protective immunity. Vaccine 37, 5371–5381. https://doi.org/10.1016/j.vaccine.2019.07.033

Evans, E.M., Man, S., Evans, A.S., Borysiewicz, L., 1997. Infiltration of cervical cancer tissue with human papillomavirus-specific cytotoxic T-lymphocytes. Cancer Res. 57, 2943–50.

Fert, I., Douguet, L., Vesin, B., Moncoq, F., Noirat, A., Authié, P., Ciret, S., Le Chevalier, F., Blanc, C., Vitrenko, Y., Charneau, P., Majlessi, L., Anna, F., 2024. T-cell immunity induced and reshaped by an anti-HPV immuno-oncotherapeutic lentiviral vector. Npj Vaccines 9, 102. https://doi.org/10.1038/s41541-024-00894-0

Fumagalli, V., Ravà, M., Marotta, D., Di Lucia, P., Bono, E.B., Giustini, L., De Leo, F., Casalgrandi, M., Monteleone, E., Mouro, V., Malpighi, C., Perucchini, C., Grillo, M., De Palma, S., Donnici, L., Marchese, S., Conti, M., Muramatsu, H., Perlman, S., Pardi, N., Kuka, M., De Francesco, R., Bianchi, M.E., Guidotti, L.G., Iannacone, M., 2024. Antibody-independent protection against heterologous SARS-CoV-2 challenge conferred by prior infection or vaccination. Nat. Immunol. 25, 633–643. https://doi.org/10.1038/s41590-024-01787-z

Gaevert, J.A., Luque Duque, D., Lythe, G., Molina-París, C., Thomas, P.G., 2021. Quantifying T Cell Cross-Reactivity: Influenza and Coronaviruses. Viruses 13, 1786. https://doi.org/10.3390/v13091786

Gallagher, K.M.E., Man, S., 2007. Identification of HLA-DR1- and HLA-DR15-restricted human papillomavirus type 16 (HPV16) and HPV18 E6 epitopes recognized by CD4+ T cells from healthy young women. J. Gen. Virol. 88, 1470–1478. https://doi.org/10.1099/vir.0.82558-0

Grabowska, A.K., Kaufmann, A.M., Riemer, A.B., 2015. Identification of promiscuous HPV16-derived T helper cell epitopes for therapeutic HPV vaccine design: Promiscuous HPV T helper epitopes. Int. J. Cancer 136, 212–224. https://doi.org/10.1002/ijc.28968

Grifoni, A., Mahajan, S., Sidney, J., Martini, S., Scheuermann, R.H., Peters, B., Sette, A., 2019. A survey of known immune epitopes in the enteroviruses strains associated with acute flaccid myelitis. Hum. Immunol. 80, 923–929. https://doi.org/10.1016/j.humimm.2019.08.004

Grifoni, A., Sidney, J., Vita, R., Peters, B., Crotty, S., Weiskopf, D., Sette, A., 2021. SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19. Cell Host Microbe 29, 1076–1092. https://doi.org/10.1016/j.chom.2021.05.010

Hancock, G., Blight, J., Lopez-Camacho, C., Kopycinski, J., Pocock, M., Byrne, W., Price, M.J., Kemlo, P., Evans, R.I., Bloss, A., Saunders, K., Kirton, R., Andersson, M., Hellner, K., Reyes-Sandoval, A., Dorrell, L., 2019. A multi-genotype therapeutic human papillomavirus vaccine elicits potent T cell responses to conserved regions of early proteins. Sci. Rep. 9, 18713. https://doi.org/10.1038/s41598-019-55014-z

Hilders, C.G.J.M., Ras, L., Van Eendenburg, J.D.H., Nooyen, Y., Fleuren, G.J., 1994. Isolation and characterization of tumor-infiltrating lymphocytes from cervical carcinoma. Int. J. Cancer 57, 805–813. https://doi.org/10.1002/ijc.2910570608

Huber, S.R., van Beek, J., de Jonge, J., Luytjes, W., van Baarle, D., 2014. T cell responses to viral infections – opportunities for peptide vaccination. Front. Immunol. 5, 1–12.

Inturi, R., Jemth, P., 2021. CRISPR/Cas9-based inactivation of human papillomavirus oncogenes E6 or E7 induces senescence in cervical cancer cells. Virology 562, 92–102. https://doi.org/10.1016/j.virol.2021.07.005

Janiszewska, J., Kostrzewska-Poczekaj, M., Wierzbicka, M., Brenner, J.C., Giefing, M., 2025. HPV-driven oncogenesis—much more than the E6 and E7 oncoproteins. J. Appl. Genet. 66, 63–71. https://doi.org/10.1007/s13353-024-00883-y

Jemon, K., Leong, C.-M., Ly, K., Young, S.L., McLellan, A.D., Hibma, M.H., 2016. Suppression of the CD8 T cell response by human papillomavirus type 16 E7 occurs in Langerhans cell-depleted mice. Sci. Rep. 6, 34789. https://doi.org/10.1038/srep34789

Kalimuddin, S., Tham, C.Y.L., Chan, Y.F.Z., Hang, S.K., Kunasegaran, K., Chia, A., Chan, C.Y.Y., Ng, D.H.L., Sim, J.X.Y., Tan, H.-C., Syenina, A., Ngoh, A.Q., Hamis, N.Z., Chew, V., Leong, Y.S., Yee, J.X., Low, J.G., Chan, K.R., Ong, E.Z., Bertoletti, A., Ooi, E.E., 2025. Vaccine-induced T cell responses control Orthoflavivirus challenge infection without

neutralizing antibodies in humans. Nat. Microbiol. 10, 374–387. https://doi.org/10.1038/s41564-024-01903-7

Lee, Seonghyun, Yoon, H., Hong, S.H., Kwon, S.P., Hong, J.J., Kwak, H.W., Park, Hyeong-Jun, Yoo, S., Bae, S., Park, Hyo-Jung, Lee, J., Bang, Y., Lee, Y., Kim, J., Yoon, S., Roh, G., Cho, Y., Kim, Y., Kim, Daegeun, Park, S., Kim, Do-Hyung, Lee, Sowon, Oh, A., Ha, D., Lee, Soo-Yeon, Park, M., Hwang, E., Bae, G., Jeon, E., Park, S.H., Choi, W.S., Oh, H.R., Kim, I.W., Youn, H., Keum, G., Bang, E., Rhee, J.H., Lee, S.E., Nam, J., 2023. mRNA-HPV vaccine encoding E6 and E7 improves therapeutic potential for HPV-mediated cancers via subcutaneous immunization. J. Med. Virol. 95, e29309. https://doi.org/10.1002/jmv.29309

Letafati, A., Taghiabadi, Z., Zafarian, N., Tajdini, R., Mondeali, M., Aboofazeli, A., Chichiarelli, S., Saso, L., Jazayeri, S.M., 2024. Emerging paradigms: unmasking the role of oxidative stress in HPV-induced carcinogenesis. Infect. Agent. Cancer 19, 30. https://doi.org/10.1186/s13027-024-00581-8

Manisty, C., Treibel, T.A., Jensen, M., Semper, A., Joy, G., Gupta, R.K., Cutino-Moguel, T., Andiapen, M., Jones, J., Taylor, S., Otter, A., Pade, C., Gibbons, J., Lee, J., Bacon, J., Thomas, S., Moon, C., Jones, M., Williams, D., Lambourne, J., Fontana, M., Altmann, D.M., Boyton, R., Maini, M., McKnight, A., Chain, B., Noursadeghi, M., Moon, J.C., 2021. Time series analysis and mechanistic modelling of heterogeneity and sero-reversion in antibody responses to mild SARS-CoV-2 infection. EBioMedicine 65, 103259. https://doi.org/10.1016/j.ebiom.2021.103259

Maskey, N., Thapa, N., Maharjan, M., Shrestha, G., Maharjan, N., Cai, H., Liu, S., 2019. Infiltrating CD4 and CD8 lymphocytes in HPV infected uterine cervical milieu. Cancer Manag. Res. Volume 11, 7647–7655. https://doi.org/10.2147/CMAR.S217264

McInnis, C., Bhatia, S., Vijaykumar, B., Tian, Q., Sun, Y., Leistritz-Edwards, D., Quinn, C.T., Uppaluri, R., Egloff, A.M., Srinivasan, L., Pregibon, D.C., Coyle, A.J., Hanna, G.J., 2023. Identification of HPV16 E1 and E2-specific T cells in the oropharyngeal cancer tumor microenvironment. J. Immunother. Cancer 11, e006721. https://doi.org/10.1136/jitc-2023-006721

Medhasi, S., Chantratita, N., 2022. Human Leukocyte Antigen (HLA) System: Genetics and Association with Bacterial and Viral Infections. J. Immunol. Res. 2022, 1–15. https://doi.org/10.1155/2022/9710376

Mercier-Letondal, P., Marton, C., Deschamps, M., Ferrand, C., Vauchy, C., Chenut, C., Baguet, A., Adotévi, O., Borg, C., Galaine, J., Godet, Y., 2018. Isolation and Characterization of an HLA-DRB1*04-Restricted HPV16-E7 T Cell Receptor for Cancer Immunotherapy. Hum. Gene Ther. 29, 1202–1212. https://doi.org/10.1089/hum.2018.091

Mo, Y., Ma, J., Zhang, H., Shen, J., Chen, J., Hong, J., Xu, Y., Qian, C., 2022. Prophylactic and Therapeutic HPV Vaccines: Current Scenario and Perspectives. Front. Cell. Infect. Microbiol. 12, 909223. https://doi.org/10.3389/fcimb.2022.909223

Mondatore, D., Bai, F., Augello, M., Giovenzana, M., Pisani Ceretti, A., Bono, V., Opocher, E., d'Arminio Monforte, A., Marchetti, G.C., Tincati, C., 2022. Persistence of High Percentage of Peripheral Activated CD8+ T Cells Predict Cytologic HPV-Related Dysplasia in cART-Treated, HIV-Positive Subjects. Open Forum Infect. Dis. 9, ofac046. https://doi.org/10.1093/ofid/ofac046

Münger, K., Phelps, W.C., Bubb, V., Howley, P.M., Schlegel, R., 1989. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for

transformation of primary human keratinocytes. J. Virol. 63, 4417–4421. https://doi.org/10.1128/jvi.63.10.4417-4421.1989
Nakagawa, M., Kim, K.H., Gillam, T.M., Moscicki, A.-B., 2007. HLA Class I Binding Promiscuity of the CD8 T-Cell Epitopes of Human Papillomavirus Type 16 E6 Protein. J. Virol. 81, 1412–1423. https://doi.org/10.1128/JVI.01768-06
Nelson, C.W., Mirabello, L., 2023. Human papillomavirus genomics: Understanding carcinogenicity. Tumour Virus Res. 15, 200258. https://doi.org/10.1016/j.tvr.2023.200258
Peng, X., Woodhouse, I., Hancock, G., Parker, R., Marx, K., Müller, J., Salatino, S., Partridge, T., Nicastri, A., Liao, H., Kruppa, G., Hellner, K., Dorrell, L., Ternette, N., 2023. Novel canonical and non-canonical viral antigens extend current targets for immunotherapy of HPV-driven cervical cancer. iScience 26, 106101. https://doi.org/10.1016/j.isci.2023.106101
Putra, S., Putra, A., Astuti, I., Asmara, W., 2022. Human Papillomavirus Type 16 L2 Gene Sequence Variation Analysis in Indonesian Cervical Cancer Specimens. Asian Pac. J. Cancer Prev. 23, 2009–2016. https://doi.org/10.31557/APJCP.2022.23.6.2009
Ranasinghe, V., McMillan, N., 2025. Novel therapeutic strategies for targeting E6 and E7 oncoproteins in cervical cancer. Crit. Rev. Oncol. Hematol. 211, 104721. https://doi.org/10.1016/j.critrevonc.2025.104721
Reynisson, B., Alvarez, B., Paul, S., Peters, B., Nielsen, M., 2020a. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 48, W449–W454. https://doi.org/10.1093/nar/gkaa379
Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W.H., Peters, B., Nielsen, M., 2020b. Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. J. Proteome Res. 19, 2304–2315. https://doi.org/10.1021/acs.jproteome.9b00874
Riemer, A.B., Keskin, D.B., Zhang, G., Handley, M., Anderson, K.S., Brusic, V., Reinhold, B., Reinherz, E.L., 2010. A Conserved E7-derived Cytotoxic T Lymphocyte Epitope Expressed on Human Papillomavirus 16-transformed HLA-A2+ Epithelial Cancers. J. Biol. Chem. 285, 29608–29622. https://doi.org/10.1074/jbc.M110.126722
Rock, K.L., Reits, E., Neefjes, J., 2016. Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol. 37, 724–737. https://doi.org/10.1016/j.it.2016.08.010
Sadeghi, Z., Aboofazeli, A., Sarrafzadeh, S., Tayebi, N., Tajdini, R., Yarandi, F., Jazayeri, S.M., 2025. Investigating the link between HPV genotypes and cervical abnormality incidence in women with HPV infections: insights from a leading referral centre. Virol. J. 22, 264. https://doi.org/10.1186/s12985-025-02858-9
Sanchez-Mazas, A., Nunes, J.M., Di, D., Dominguez, E.A., Gerbault, P., Faye, N.K., Almawi, W., Andreani, M., Arrieta-Bolanos, E., Augusto, D.G., Buhler, S., De Groot, N.G., Díaz-Peña, R., D'Silva, S., Dubois, V., ElGhazali, G., Gorodezky, C., Grubić, Z., Hajeer, A., Hollenbach, J., Issler, H.C., Ivanova, M., Jawdat, D., Kanga, U., Kirijas, M., Lemieux, W., Loginova, M., Mahfoudh, N., Mehra, N., Mouron, D., Petřek, M., Petzl-Erler, M.L., Richard, L., Roelen, D.L., Sacchi, N., Sharma, A., Singh, B., Singh, M., Singh, Y., Tarassi, K., Vidan-Jeras, B., Villard, J., 2024. The most frequent HLA alleles around the world: A fundamental synopsis. Best Pract. Res. Clin. Haematol. 37, 101559. https://doi.org/10.1016/j.beha.2024.101559

Speetjens, F.M., Welters, M.J.P., Slingerland, M., Van Poelgeest, M.I.E., De Vos Van Steenwijk, P.J., Roozen, I., Boekestijn, S., Loof, N.M., Zom, G.G., Valentijn, A.R.P.M., Krebber, W.-J., Meeuwenoord, N.J., Janssen, C.A.H., Melief, C.J.M., Van Der Marel, G.A., Filippov, D.V., Van Der Burg, S.H., Gelderblom, H., Ossendorp, F., 2022. Intradermal vaccination of HPV-16 E6 synthetic peptides conjugated to an optimized Toll-like receptor 2 ligand shows safety and potent T cell immunogenicity in patients with HPV-16 positive (pre-)malignant lesions. J. Immunother. Cancer 10, e005016. https://doi.org/10.1136/jitc-2022-005016

Stelzle, D., Tanaka, L.F., Lee, K.K., Ibrahim Khalil, A., Baussano, I., Shah, A.S.V., McAllister, D.A., Gottlieb, S.L., Klug, S.J., Winkler, A.S., Bray, F., Baggaley, R., Clifford, G.M., Broutet, N., Dalal, S., 2021. Estimates of the global burden of cervical cancer associated with HIV. Lancet Glob. Health 9, e161–e169. https://doi.org/10.1016/S2214-109X(20)30459-9

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. 71, 209–249. https://doi.org/10.3322/caac.21660

Swadling, L., Diniz, M.O., Schmidt, N.M., Amin, O.E., Chandran, A., Shaw, E., Pade, C., Gibbons, J.M., Le Bert, N., Tan, A.T., Jeffery-Smith, A., Tan, C.C.S., Tham, C.Y.L., Kucykowicz, S., Aidoo-Micah, G., Rosenheim, J., Davies, J., Johnson, M., Jensen, Melanie P., Joy, G., McCoy, L.E., Valdes, A.M., Chain, B.M., Goldblatt, D., Altmann, D.M., Boyton, R.J., Manisty, C., Treibel, T.A., Moon, J.C., COVIDsortium Investigators, Abbass, H., Abiodun, A., Alfarih, M., Alldis, Z., Andiapen, M., Artico, J., Augusto, J.B., Baca, G.L., Bailey, S.N.L., Bhuva, A.N., Boulter, A., Bowles, R., Boyton, R.J., Bracken, O.V., O'Brien, B., Brooks, T., Bullock, N., Butler, D.K., Captur, G., Champion, N., Chan, C., Collier, D., De Sousa, J.C., Couto-Parada, X., Cutino-Mogue, T., Davies, R.H., Douglas, B., Di Genova, C., Dieobi-Anene, K., Ellis, A., Feehan, K., Finlay, M., Fontana, M., Forooghi, N., Gaier, C., Gilroy, D., Hamblin, M., Harker, G., Hewson, J., Hickling, L.M., Hingorani, A.D., Howes, L., Hughes, A., Hughes, G., Hughes, R., Itua, I., Jardim, V., Lee, W.-Y.J., Jensen, Melanie Petra, Jones, J., Jones, M., Joy, G., Kapil, V., Kurdi, H., Lambourne, J., Lin, K.-M., Louth, S., Mandadapu, V., McKnight, Á., Menacho, K., Mfuko, C., Mitchelmore, O., Moon, C., Munoz-Sandoval, D., Murray, S.M., Noursadeghi, M., Otter, A., Palma, S., Parker, R., Patel, K., Pawarova, B., Petersen, S.E., Piniera, B., Pieper, F.P., Pope, D., Prossora, M., Rannigan, L., Rapala, A., Reynolds, C.J., Richards, A., Robathan, M., Sambile, G., Semper, A., Seraphim, A., Simion, M., Smit, A., Sugimoto, M., Taylor, S., Temperton, N., Thomas, S., Thornton, G.D., Tucker, A., Veerapen, J., Vijayakumar, M., Welch, S., Wodehouse, T., Wynne, L., Zahedi, D., Van Dorp, L., Balloux, F., McKnight, Á., Noursadeghi, M., Bertoletti, A., Maini, M.K., 2022. Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. Nature 601, 110–117. https://doi.org/10.1038/s41586-021-04186-8

Swadling, L., Maini, M.K., 2023. Can T Cells Abort SARS-CoV-2 and Other Viral Infections? Int. J. Mol. Sci. 24, 4371. https://doi.org/10.3390/ijms24054371

Tsu, V.D., LaMontagne, D.S., Atuhebwe, P., Bloem, P.N., Ndiaye, C., 2021. National implementation of HPV vaccination programs in low-resource countries: Lessons, challenges, and future prospects. Prev. Med. 144, 106335. https://doi.org/10.1016/j.ypmed.2020.106335

Tsukamoto, K., Yamashita, A., Maeki, M., Tokeshi, M., Imai, H., Fukao, A., Fujiwara, T., Okudera, K., Mizuki, N., Okuda, K., Shimada, M., 2024. Enhanced Broad-Spectrum Efficacy of an L2-Based mRNA Vaccine Targeting HPV Types 6, 11, 16, 18, with Cross-Protection Against Multiple Additional High-Risk Types. Vaccines 12, 1239. https://doi.org/10.3390/vaccines12111239

Van Bockel, D., Kelleher, A., 2023. The crossroads: divergent roles of virus-specific $CD4^+$ T lymphocytes in determining the outcome for human papillomavirus infection. Immunol. Cell Biol. 101, 525–534. https://doi.org/10.1111/imcb.12650

Vita, R., Blazeska, N., Marrama, D., IEDB Curation Team Members, Shackelford, D., Zalman, L., Foos, G., Zarebski, L., Chan, K., Reardon, B., Fitzpatrick, S., Busse, M., Coleman, S., Sedwick, C., Edwards, L., MacFarlane, C., Ennis, M., Duesing, S., Bennett, J., Greenbaum, J., De Almeida Mendes, M., Mahita, J., Wheeler, D.K., Cantrell, J.R., Overton, J.A., Natale, D.A., Sette, A., Peters, B., 2025. The Immune Epitope Database (IEDB): 2024 update. Nucleic Acids Res. 53, D436–D443. https://doi.org/10.1093/nar/gkae1092

Wang, C.Y., Peng, W.-J., Kuo, B.-S., Ho, Y.-H., Wang, M.-S., Yang, Y.-T., Chang, P.-Y., Shen, Y.-H., Hwang, K.-P., 2023. Toward a pan-SARS-CoV-2 vaccine targeting conserved epitopes on spike and non-spike proteins for potent, broad and durable immune responses. PLOS Pathog. 19, e1010870. https://doi.org/10.1371/journal.ppat.1010870

Wang, J.W., Roden, R.B.S., 2013. L2, the minor capsid protein of papillomavirus. Virology 445, 175–186. https://doi.org/10.1016/j.virol.2013.04.017

Wang, R., Pan, W., Jin, L., Huang, W., Li, Y., Wu, D., Gao, C., Ma, D., Liao, S., 2020. Human papillomavirus vaccine against cervical cancer: Opportunity and challenge. Cancer Lett. 471, 88–102. https://doi.org/10.1016/j.canlet.2019.11.039

Wei, F., Georges, D., Man, I., Baussano, I., Clifford, G.M., 2024. Causal attribution of human papillomavirus genotypes to invasive cervical cancer worldwide: a systematic analysis of the global literature. The Lancet 404, 435–444. https://doi.org/10.1016/S0140-6736(24)01097-3

Zhang, Y., Qiu, K., Ai, J., Xu, M., Wang, B., Alu, A., Ye, C., Huang, X., Zhang, Y., Zhou, Y., Song, Z., Shi, J., Lu, Y., Wei, Y., Ren, J., Zhao, Y., Cheng, P., Wei, X., 2025. Ad-E6/7-HR vaccine improves the prophylactic and therapeutic efficacy in HPV-associated cancers. Clin. Transl. Med. 15, e70305. https://doi.org/10.1002/ctm2.70305

Zhao, X., Yang, F., Mariz, F., Osen, W., Bolchi, A., Ottonello, S., Müller, M., 2020. Combined prophylactic and therapeutic immune responses against human papillomaviruses induced by a thioredoxin-based L2-E7 nanoparticle vaccine. PLOS Pathog. 16, e1008827. https://doi.org/10.1371/journal.ppat.1008827

Figure Legends

Fig.1: Data curation workflow for HPV T-cell epitope meta- and in-silico analyses. Data were obtained from Immune Epitope Database (IEDB, http://www.IEDB.org) based on queries on August 14th, 2025.

Fig.2: Proteome-wide mapping and genotype-level distribution of HPV T-cell epitopes. Proteome analysis of 485 HPV T-cell epitopes with defined peptide length and protein annotations **(a-d)**, showing: **(a)** genomic map of HPV proteins by reading frame (HPV-16;

Ref.Seq: NC_001526.4), colored by functional class (oncoprotein E5, E6, E7=vermillion, replication E1, E2, E4=navy-hatched, structural L1, L2=green-dotted) with amino acid lengths indicated and dashed lines connecting spliced segments; **(b)** number of studies reporting epitopes per protein; **(c)** number of epitopes per protein based on T-cell subsets; and **(d)** epitope density showing the number of epitopes studied every 100 amino acids of each protein. Genotype analysis of 484 HPV T-cell epitopes with defined genotype annotations **(e-h)**, illustrating: **(e)** the number of epitopes per HPV type stratified by T-cell subset ($CD8^+$ =yellow-hatched, $CD4^+$=light blue-dotted), IARC-defined high-risk types are highlighted in bold orange; **(f)** proportion of epitopes restricted to high-risk versus low-risk groups, with p-value and odds ratio (OR, 95% CI) shown (reference=low-risk); **(g)** distribution of epitopes across HPV proteins (E1–L2) stratified by risk group (high-risk=vermilion-hatched, low-risk=gray-dotted) with global p-value indicated; and **(h)** proportion of epitopes restricted to a single HPV type (gray) versus those conserved across multiple types (cross-type epitopes, pink).

Fig.3: HLA restriction landscape and immunogenicity patterns of HPV T-cell epitopes. Immunogenicity analysis of 316 HPV T-cell epitopes with defined response rates **(a-b),** showing: **(a)** proportion of subjects tested with $CD4^+$ (light blue-dotted) and $CD8^+$ (yellow-hatched) epitopes showing positive (colored) versus negative (gray) immune responses, with p value and OR [95% CI] indicated; **(b)** proportion of subjects tested with epitope form oncoprotein E5-E7 (vermillion), replication E1-E4 (navy blue-hatched) and structural L1-L2 (green-dotted) protein epitopes showing positive (colored) versus negative (gray) immune responses, with global p value and OR [95% CI] indicated. Analysis of 289 epitopes with defined HLA-annotation and 219 allele-level-restricted **(c-f)**, illustrating: **(c)** number of $CD4^+$ (light blue-dotted) and $CD8^+$ (yellow-hatched) epitopes across all HLA-annotated (total) and allele-level restricted datasets, with total n indicated below each group; **(d)** proportion of class I (yellow-hatched) and Class II (light blue-dotted) epitopes classified as mono-allelic or multi-allelic (HLA-promiscuous) binders, with OR [95% CI] and p value indicated; **(e)** number of class I ($CD8^+$, yellow-hatched) and class II ($CD4^+$, light blue-dotted) restricted epitopes mapped to individual HLA alleles; **(f)** correlation between response rate and HLA promiscuity among $CD8^+$ and $CD4^+$ epitopes. Each dot represents a unique epitope, colored by HPV protein and scaled by the number of subjects tested (dot size, quartiles shown in legend).

Fig.4: Epitope conservation across HPV genomes following *in silico* analysis. Conservation analysis of 485 unique T-cell epitopes towards 454 reference genomes, showing: **(a)** distribution of $CD8^+$ (yellow) and $CD4^+$ (light blue) epitopes by peptide length and number of HPV types covered; **(b)** cross-type coverage of unique epitopes across HPV proteins; **(c)** number of unique T-cell epitopes identified across HPV genotypes, grouped by high-risk (orange) and low-risk (gray) types; **(d)** proportion and number of HPV types covered by each epitope, compared between high-risk (orange) and low-risk (gray) groups; and **(e)** epitope distribution across the HPV proteome, showing the localization of $CD8^+$ (yellow) and $CD4^+$ (light blue) T-cell epitopes in high-risk and low-risk types, grouped by functional protein classes (oncoprotein, replication, and structural).

Fig.5: Epitope promiscuity across HLA alleles following *in silico* prediction. HLA prediction analysis of 485 T-cell epitopes across 20 HLA allele supertypes for both $CD4^+$ and $CD8^+$ T cells, illustrating: **(a)** distribution of epitopes classified as non-binders versus predicted binders, stratified by T-cell type ($CD8^+$ vs $CD4^+$); **(b)** frequency distribution of the number of HLA alleles bound per epitope; **(c–d)** distribution of predicted binder epitopes across HPV proteins for $CD8^+$ and $CD4^+$, shown by mono-allelic (grey) versus multi-allelic epitopes (colored); **(e–f)** predicted binding capacity of HPV epitopes across HLA Class I ($CD8^+$) and Class II (CD4) alleles, showing the number of weak (light-dotted) and strong (dark) binders for each allele; and **(g–h)** heatmaps showing the distribution of predicted binding strength across HPV proteins and HLA alleles for Class I ($CD8^+$, yellow) and Class II ($CD4^+$, light blue), with strong binders (colored) and weak binders (gray) mapped across structural (green), oncoprotein (orange), and replication-associated (light blue) antigens.

Supp Fig.1: HPV T-cell epitopes length and immunodominance. Proteome analysis of 485 HPV T-cell epitopes with defined peptide length and protein annotations, showing: **(a)** number of experimentally validated $CD8^+$ (Class I) and $CD4^+$ (Class II) T-cell epitopes identified per HPV protein; and **(b)** correlation between protein length and epitope count for HPV and SARS-CoV-2 proteins, with linear regression (solid) and 95% confidence interval (shaded). Immunogenicity analysis of 316 HPV T-cell epitopes with defined response rates, illustrating: **(c)** distribution of epitopes with available response rate data, showing proportional representation of HPV proteins; and **(d)** immunogenicity profiles (response rate) across amino acid positions of individual HPV proteins, highlighting distinct regions of epitope clustering.

Supp Fig.2: Relationship between HLA restriction breadth and epitope count. Analysis of 219 HLA allele-restricted illustrating: **(a)** comparison of the number of epitopes per HLA allele between Class I ($CD8^+$) and Class II ($CD4^+$) restrictions; and **(b–c)** correlations between the number of unique HLA alleles and total epitopes per protein for $CD8^+$ (b, orange) and $CD4^+$ (c, light blue) T-cell epitopes, with Spearman's ρ and p-values indicated.

Supp Fig. 3: In silico analysis of HPV T-cell epitope type coverage and donor response rate by T-cell subset and binding status. Analysis of 485 unique T-cell epitopes mapped to 454 reference genomes showing **(a)** comparison of HPV types coverage between $CD4^+$ and $CD8^+$ T-cell epitopes, where each dot represents a unique epitope, and horizontal lines show medians with interquartile ranges. Analysis of 316 epitopes across 20 HLA-restricted alleles for both MHC Class I ($CD8^+$) and Class II ($CD4^+$) T cells, showing **(b)** response rate (% positive donors) comparison between binders and non-binders.

Supplementary Table S1: Meta-analysis of experimentally validated HPV T-cell epitopes with immunogenicity, HLA restriction, and assay information based on IEDB database.

Supplementary Table S2: List of experimentally validated HPV T-cell epitopes restricted by specific HLA based on IEDB database.

Supplementary Table S3: *In silico* analysis of conservation and HLA-binding profiles of HPV T-cell epitopes.

Supplementary Table S4: *In silico* analysis of HPV epitope binders across HLA alleles.

Acknowledgements

We would like to thank all members of the Swadling laboratory for helpful discussions.

Ethical approval

This study did not involve living subjects and thus was exempted from institutional ethical review board approval.

Author Contributions

SPP and LS designed the study. SPP and SC collected data, analyzed, and drafted manuscripts. LS validated the methodology and verified the data. All authors read and approved the final manuscript.

Funding

This work was supported by: an Indonesia Endowment Fund for Education (LPDP) grant number LOG-20409/LPDP.3/2024a (to SPP), a Rosetrees Trust and Pears Foundation Advancement Fellowship and a Wellcome Career Development award [302473/Z/23/Z] (to LS). A BBSRC LIDo-DTP studentship (BB/T008709/1)(to SC).

Competing interest

The authors declare no competing interests.

Fig. 1

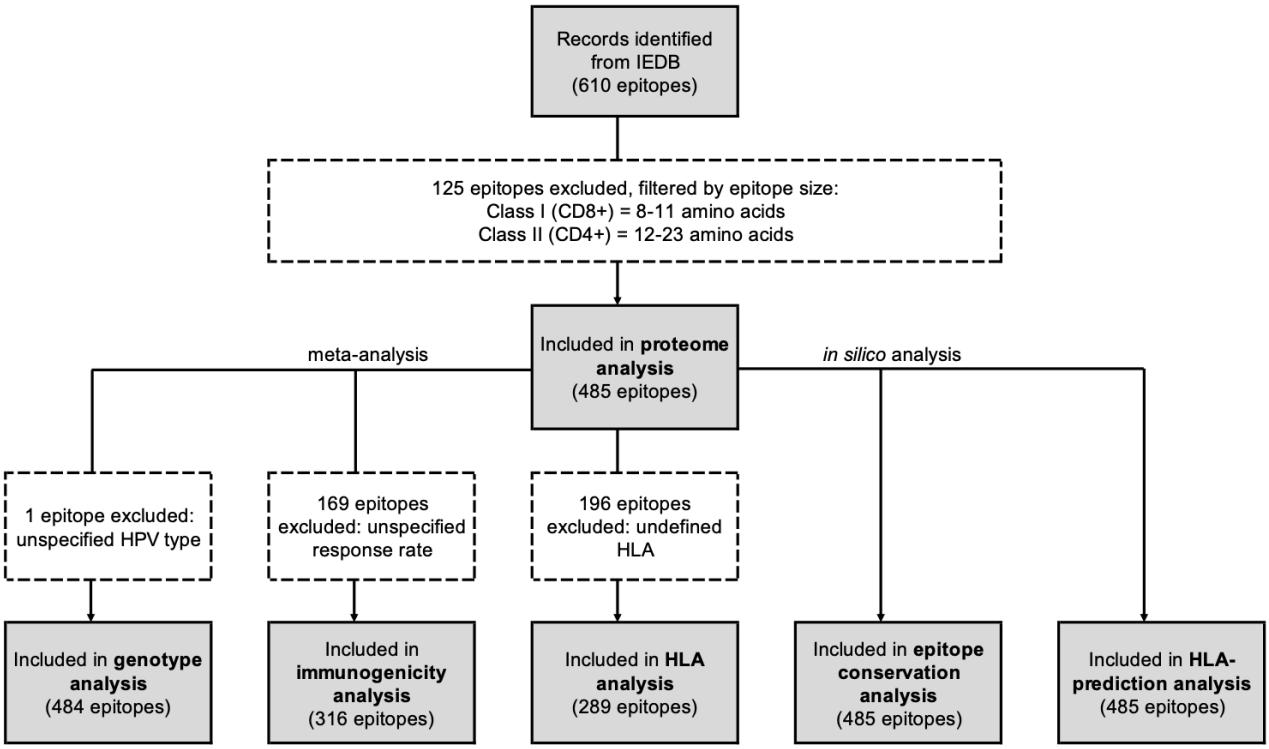
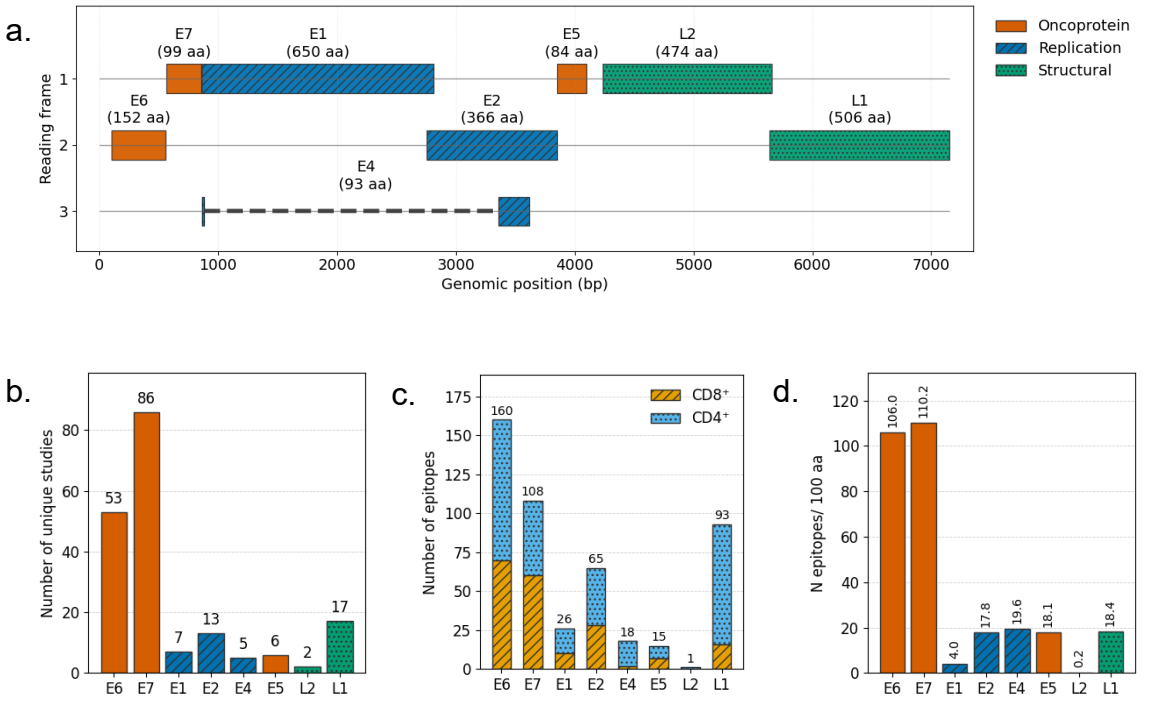

Records identified from IEDB (610 epitopes)
125 epitopes excluded, filtered by epitope size:
Class I (CD8+) = 8-11 amino acids
Class II (CD4+) = 12-23 amino acids
Included in proteome analysis (485 epitopes)
meta-analysis
in silico analysis
1 epitope excluded: unspecified HPV type
169 epitopes excluded: unspecified response rate
196 epitopes excluded: undefined HLA
Included in genotype analysis (484 epitopes)
Included in immunogenicity analysis (316 epitopes)
Included in HLA analysis (289 epitopes)
Included in epitope conservation analysis (485 epitopes)
Included in HLA-prediction analysis (485 epitopes)

Fig. 2

Proteome analysis (n=485)



Genotype analysis (n=484)

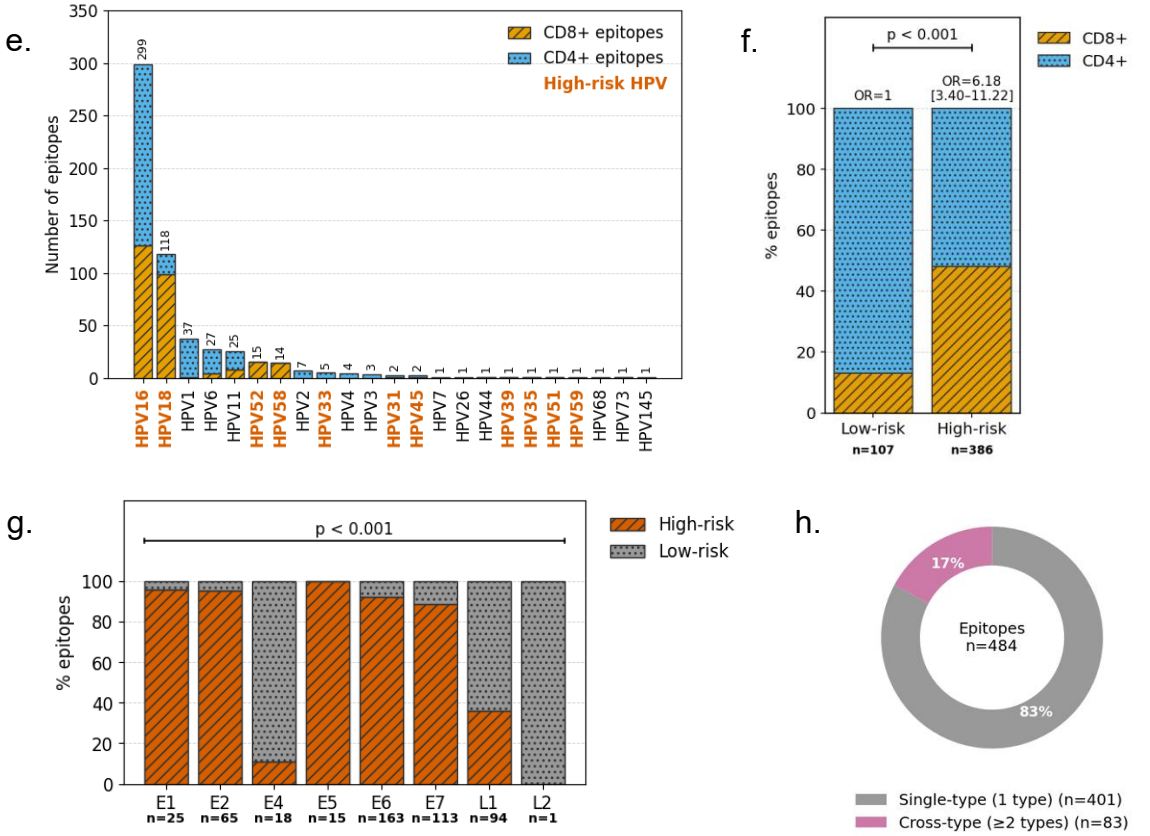
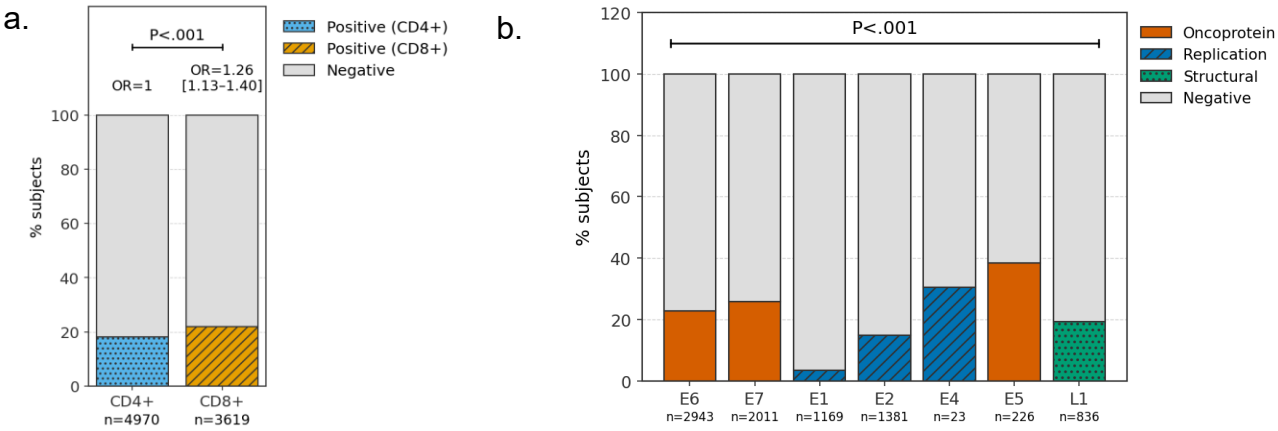


Fig. 3

Immunogenicity analysis (n=316)



HLA analysis (n=289)

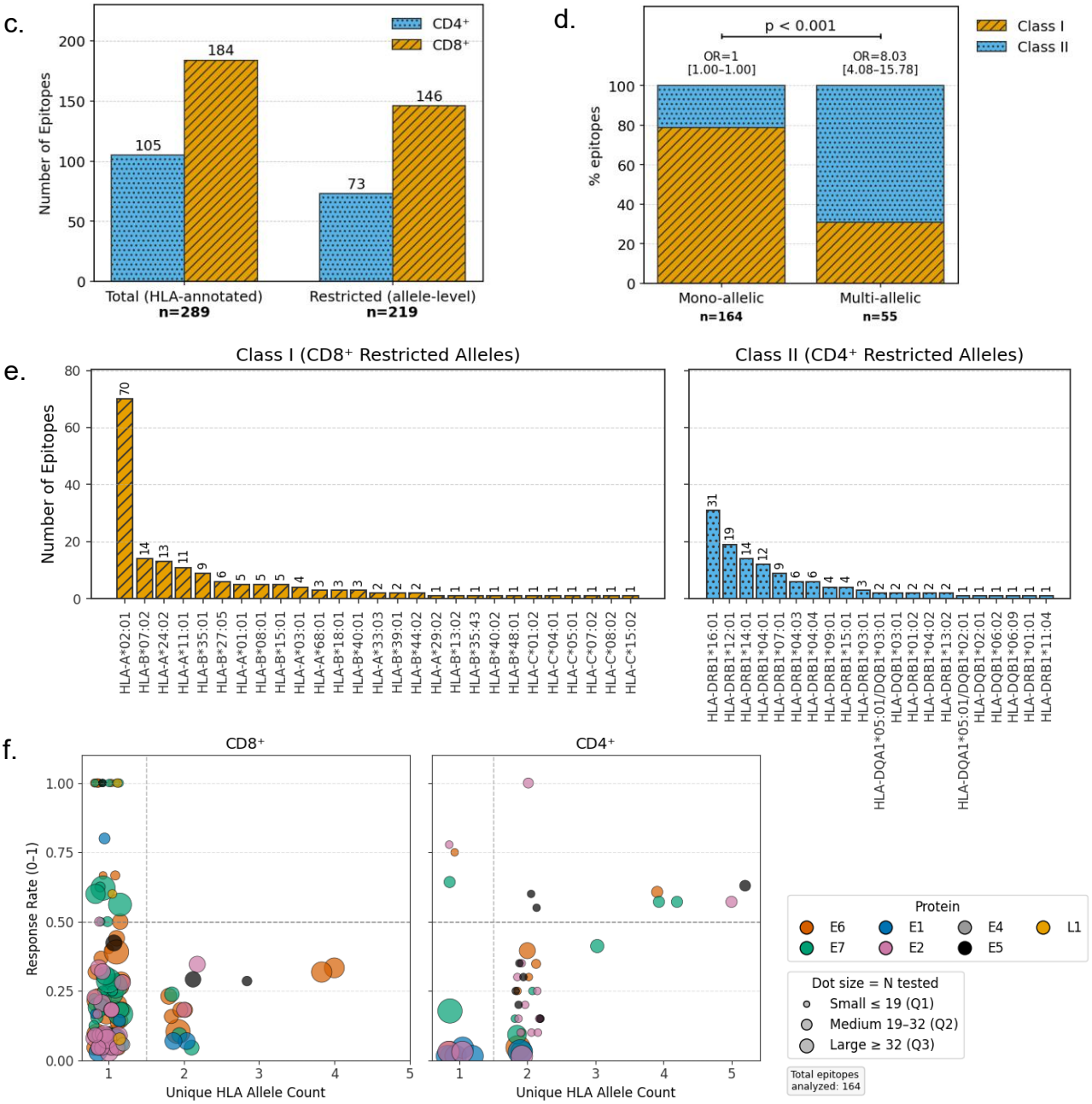


Fig. 4

Conservation analysis (n=485)

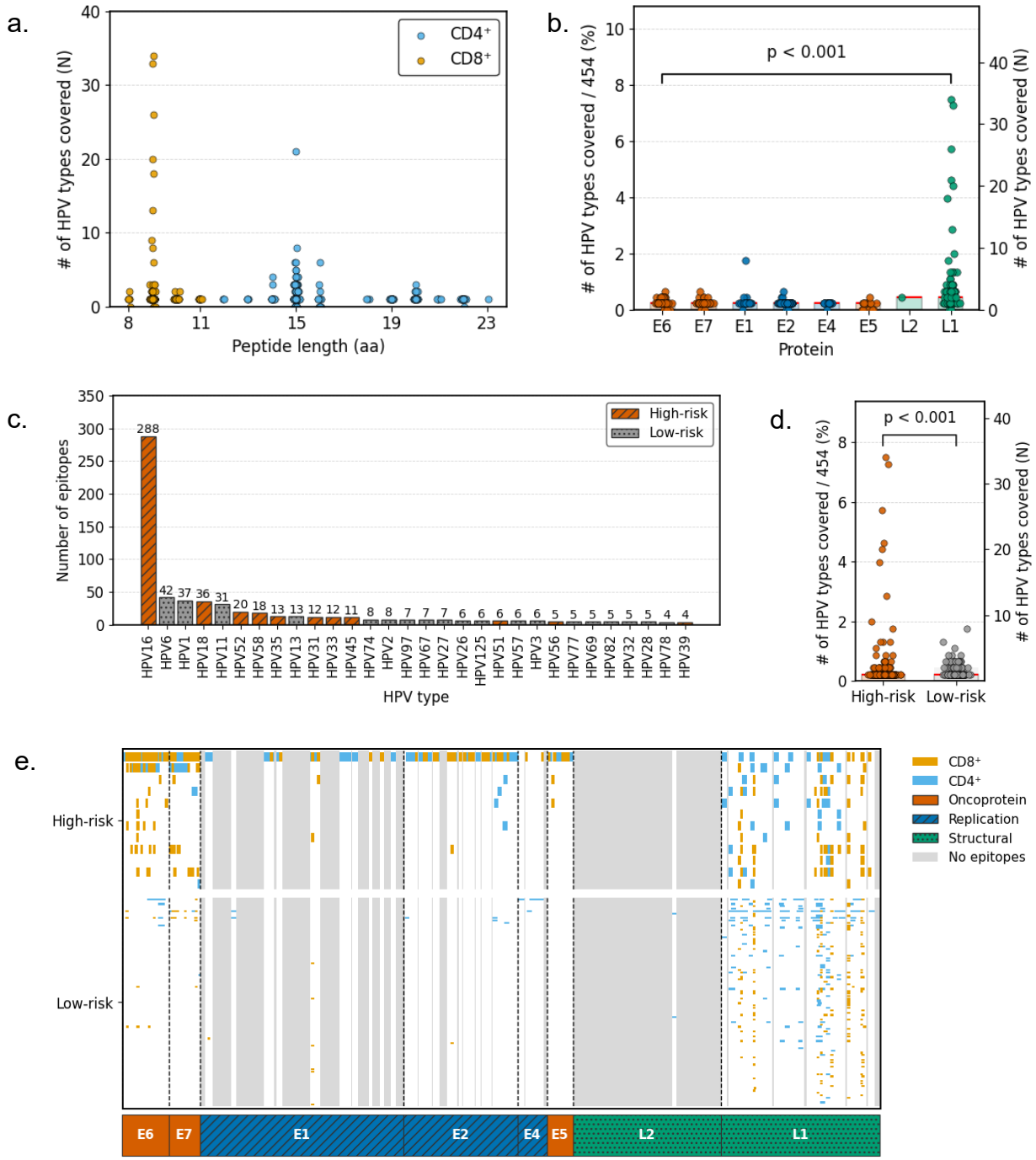
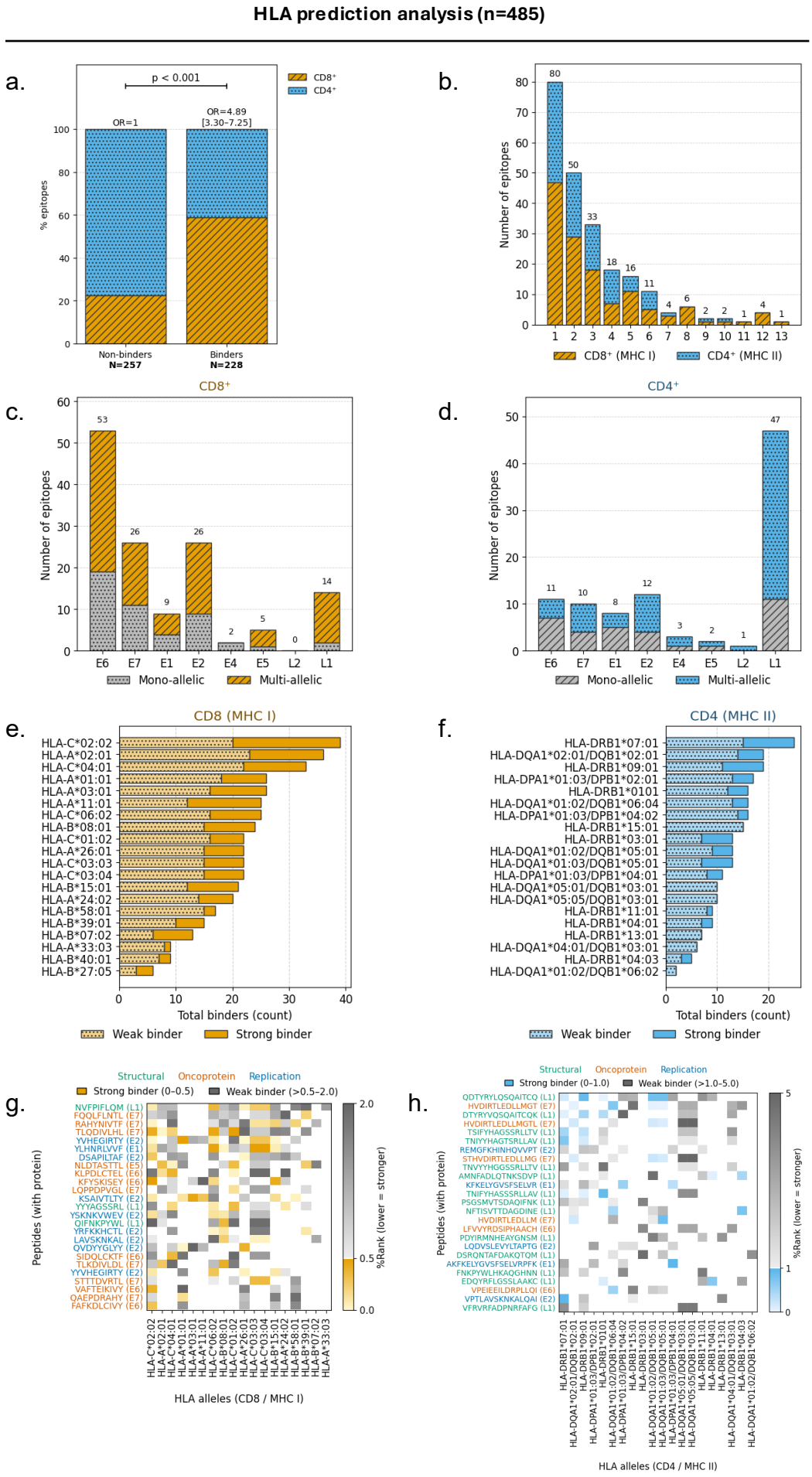
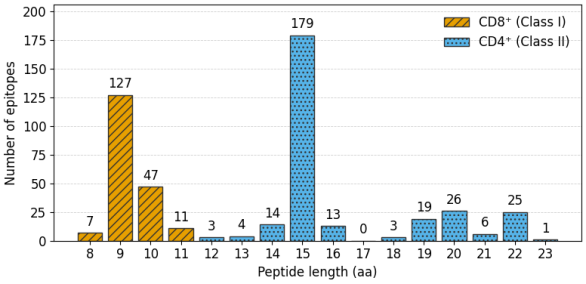


Fig. 5

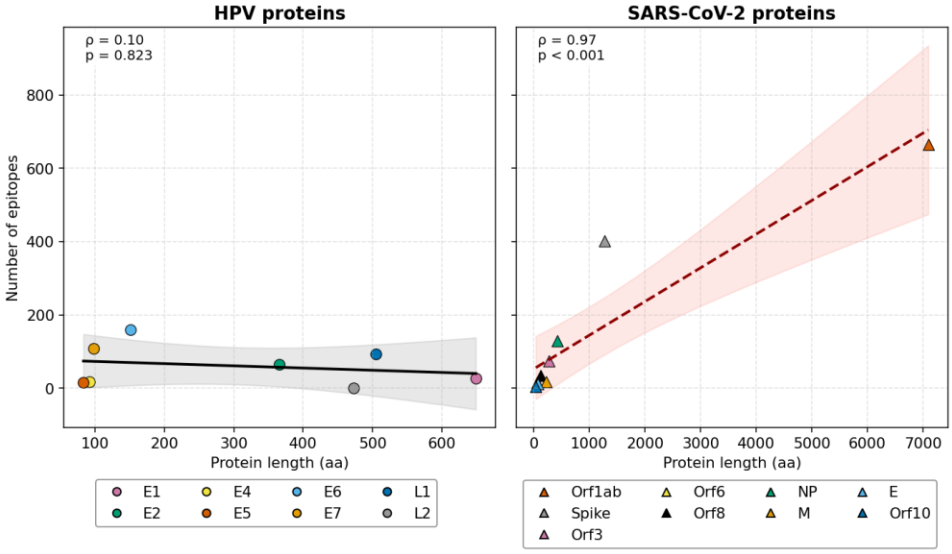

HLA prediction analysis (n=485)
a.
p < 0.001
CD8+
CD4+
OR=1
OR=4.89
[3.30–7.25]
% epitopes
Non-binders
N=257
Binders
N=228
b.
Number of epitopes
CD8+ (MHC I)
CD4+ (MHC II)
c.
CD8+
Number of epitopes
E6
E7
E1
E2
E4
E5
L2
L1
Mono-allelic
Multi-allelic
d.
CD4+
Number of epitopes
Mono-allelic
Multi-allelic
e.
CD8 (MHC I)
HLA-C*02:02
HLA-A*02:01
HLA-C*04:01
HLA-A*01:01
HLA-A*03:01
HLA-A*11:01
HLA-C*06:02
HLA-B*08:01
HLA-C*01:02
HLA-A*26:01
HLA-C*03:03
HLA-C*03:04
HLA-B*15:01
HLA-A*24:02
HLA-B*58:01
HLA-B*39:01
HLA-B*07:02
HLA-A*33:03
HLA-B*40:01
HLA-B*27:05
Total binders (count)
Weak binder
Strong binder
f.
CD4 (MHC II)
HLA-DRB1*07:01
HLA-DQA1*02:01/DQB1*02:01
HLA-DRB1*09:01
HLA-DPA1*01:03/DPB1*02:01
HLA-DRB1*0101
HLA-DQA1*01:02/DQB1*06:04
HLA-DPA1*01:03/DPB1*04:02
HLA-DRB1*15:01
HLA-DRB1*03:01
HLA-DQA1*01:02/DQB1*05:01
HLA-DQA1*01:03/DQB1*05:01
HLA-DPA1*01:03/DPB1*04:01
HLA-DQA1*05:01/DQB1*03:01
HLA-DQA1*05:05/DQB1*03:01
HLA-DRB1*11:01
HLA-DRB1*04:01
HLA-DRB1*13:01
HLA-DQA1*04:01/DQB1*03:01
HLA-DRB1*04:03
HLA-DQA1*01:02/DQB1*06:02
Total binders (count)
Weak binder
Strong binder
g.
Structural
Oncoprotein
Replication
Strong binder (0–0.5)
Weak binder (>0.5–2.0)
Peptides (with protein)
%Rank (lower = stronger)
HLA alleles (CD8 / MHC I)
h.
Structural
Oncoprotein
Replication
Strong binder (0–1.0)
Weak binder (>1.0–5.0)
Peptides (with protein)
%Rank (lower = stronger)
HLA alleles (CD4 / MHC II)

Supp Fig. 1

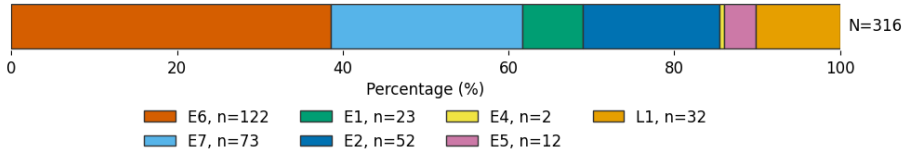
a.



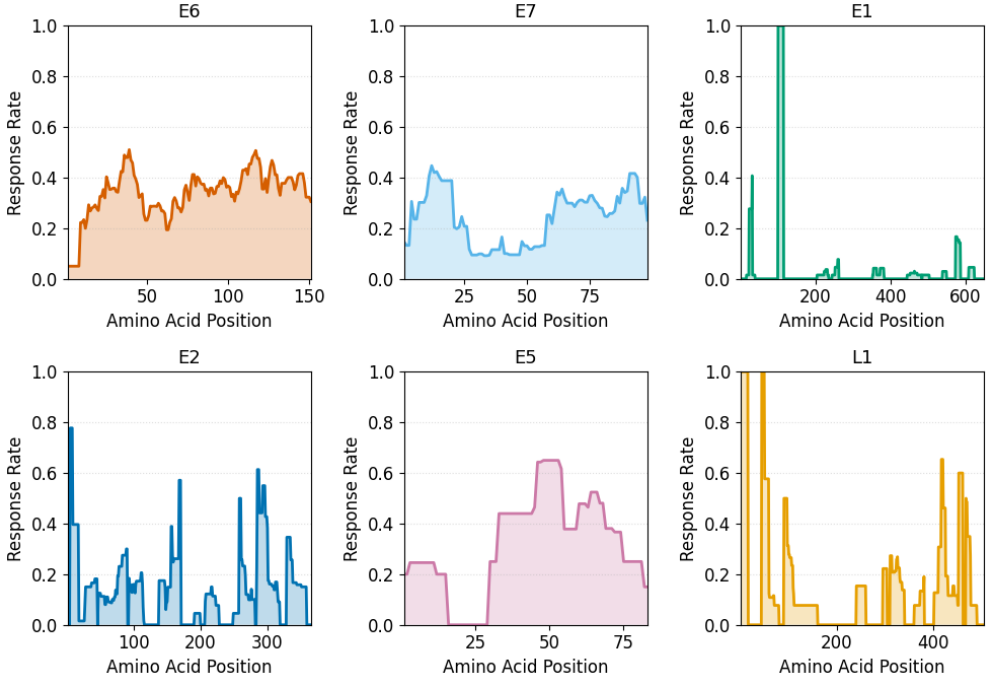
b.



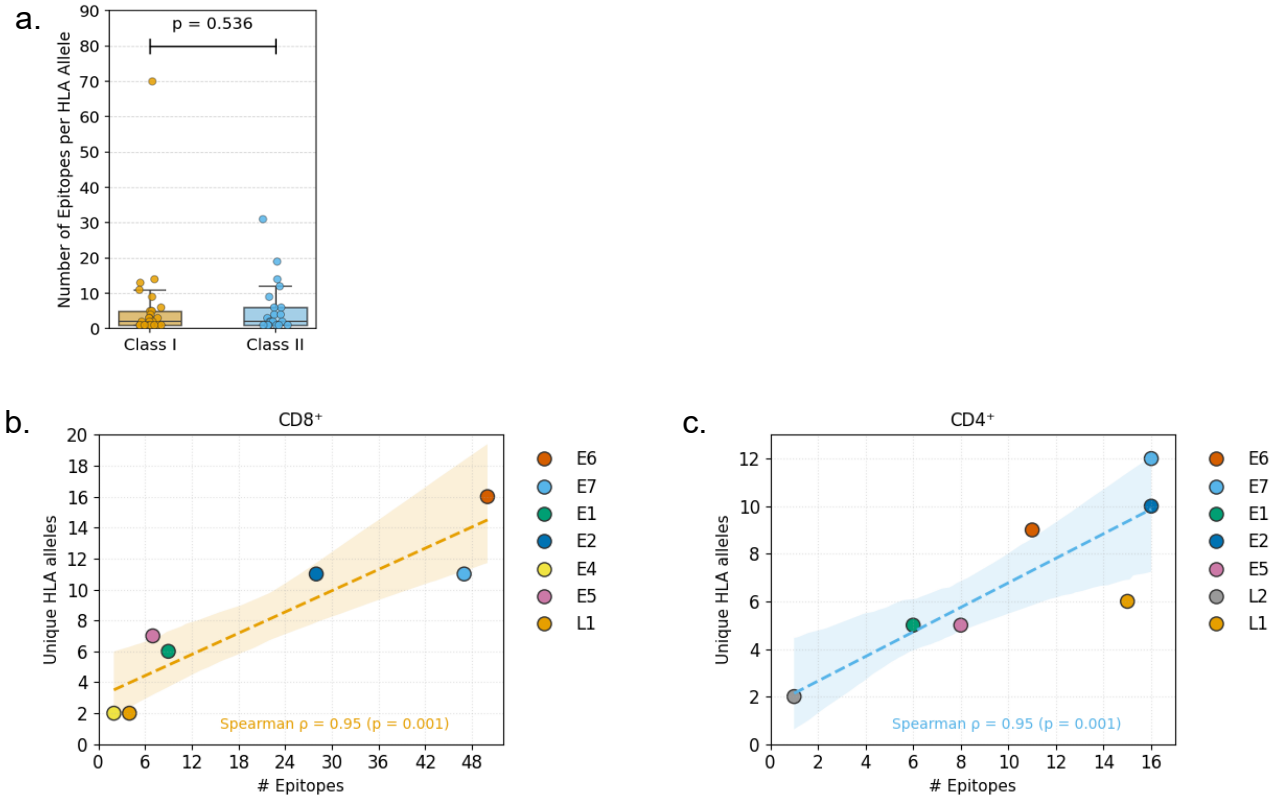
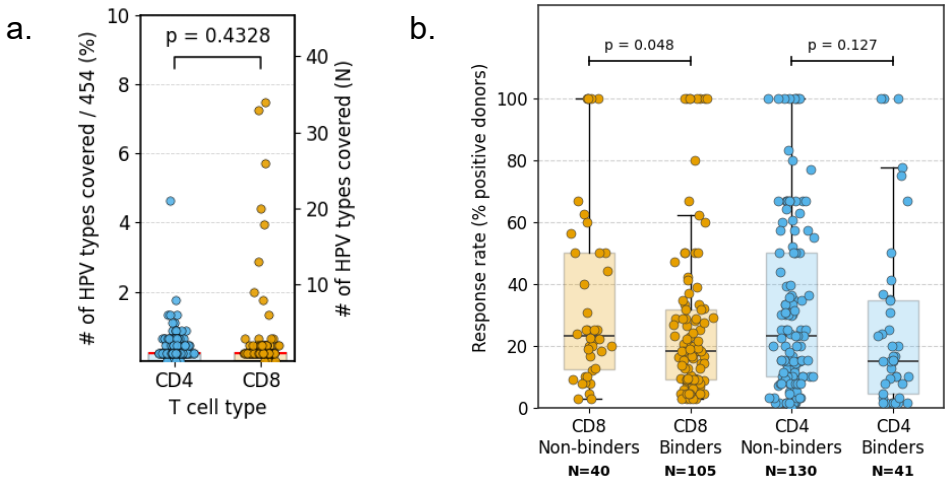
c.



d.



Supp Fig. 2



Supp Fig. 3