



Identification of HPV16 E1 and E2-specific T cells in the oropharyngeal cancer tumor microenvironment

Christine McInnis ¹, Shilpa Bhatia,¹ Brinda Vijaykumar,¹ Qiaomu Tian,¹ Yanbo Sun,¹ Del Leistritz-Edwards,¹ Charles T Quinn,² Ravi Uppaluri ³, Ann Marie Egloff,³ Lakshmi Srinivasan,¹ Daniel C Pregibon,¹ Anthony J Coyle,¹ Glenn J Hanna²

To cite: McInnis C, Bhatia S, Vijaykumar B, *et al.* Identification of HPV16 E1 and E2-specific T cells in the oropharyngeal cancer tumor microenvironment. *Journal for ImmunoTherapy of Cancer* 2023;**11**:e006721. doi:10.1136/jitc-2023-006721

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jitc-2023-006721>).

AJC and GJH are joint senior authors.

Accepted 09 March 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Repertoire Immune Medicines, Cambridge, Massachusetts, USA

²Center for Head and Neck Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

³Division of Otolaryngology-Head and Neck Surgery, Brigham and Women's Hospital, Boston, Massachusetts, USA

Correspondence to

Dr Christine McInnis;
cmcinnis@repertoire.com

ABSTRACT

Background High-risk human papillomavirus (HPV) is a primary cause of an increasing number of oropharyngeal squamous cell carcinomas (OPSCCs). The viral etiology of these cancers provides the opportunity for antigen-directed therapies that are restricted in scope compared with cancers without viral components. However, specific virally-encoded epitopes and their corresponding immune responses are not fully defined.

Methods To understand the OPSCC immune landscape, we conducted a comprehensive single-cell analysis of HPV16+ and HPV33+ primary tumors and metastatic lymph nodes. We used single-cell analysis with encoded peptide-human leukocyte antigen (HLA) tetramers to analyze HPV16+ and HPV33+ OPSCC tumors, characterizing the ex vivo cellular responses to HPV-derived antigens presented in major Class I and Class II HLA alleles.

Results We identified robust cytotoxic T-cell responses to HPV16 proteins E1 and E2 that were shared across multiple patients, particularly in HLA-A*01:01 and HLA-B*08:01. Responses to E2 were associated with loss of E2 expression in at least one tumor, indicating the functional capacity of these E2-recognizing T cells and many of these interactions validated in a functional assay. Conversely, cellular responses to E6 and E7 were limited in quantity and cytotoxic capacity, and tumor E6 and E7 expression persisted.

Conclusions These data highlight antigenicity beyond HPV16 E6 and E7 and nominate candidates for antigen-directed therapies.

BACKGROUND

Human papillomavirus (HPV) is a major risk factor for head and neck cancers, particularly within the oropharynx.¹ HPV-associated oropharyngeal squamous cell carcinoma (OPSCC) is characterized by distinct biology and heterogenous mutational and immune landscapes compared with HPV-negative tumors.^{2–5} In general, HPV+ OPSCC is associated with a better prognosis compared with HPV-negative cancers, but treatment-related morbidities are substantial following standard

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Most research and immunotherapies for human papillomavirus (HPV)-associated cancers focus on proteins E6 and E7 from HPV16 but these therapies have had limited clinical success, leaving an unmet clinical need for additional targets for immunotherapies. Additionally, all HPV+ cancers are treated with the same standard of care regardless of HPV genotype which could influence outcomes.

WHAT THIS STUDY ADDS

⇒ We find here robust T-cell recognition of HPV16 proteins E1 and E2 which are expressed in many HPV16+ oropharyngeal squamous cell carcinoma (OPSCC) and T cells recognizing these epitopes are shared across patients.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These data indicate that proteins HPV16 E1 and E2 are good candidates for immunotherapy in HPV16+ OPSCC, and that HPV subtyping at diagnosis might provide benefit to stratify some patients for therapy deintensification.

of care surgery, radiation, and chemotherapy.⁶ Furthermore, for patients who experience locoregional recurrence or distant metastatic spread, subsequent therapy options are limited.⁶ Immune checkpoint inhibitors have now been approved for patients with recurrent/metastatic OPSCC,^{7–9} but response rates are generally low and overall survival remains poor.¹⁰ While preventative vaccines provide protection against prevalent high-risk HPV genotypes, they do not appear to offer therapeutic benefit and are not projected to make a significant impact on HPV-driven malignancies for decades to come.¹¹

The HPV genome is composed of early (E) and late genes involved in different aspects of the viral life cycle.^{12 13} Human leukocyte antigens (HLAs) are genes encoding proteins

that are critical components of antigen recognition and are influential in the outcome of viral infections.^{14 15} Historically, numerous studies have focused on HPV16 E proteins E6 and E7 as targets of T-cell reactivity¹⁶ and immunotherapy due to the established role of E6 and E7 in neoplastic transformation.^{17–23} Despite limited early clinical successes targeting these antigens, durable complete tumor regression has been an elusive goal. Other HPV genes have been relatively understudied and may offer promise for therapeutic intervention. E1 and E2 play a role in the viral life cycle and can be involved in carcinogenesis.^{24 25} It has been suggested that E1 and E2 expression is downregulated in more advanced tumors,²⁶ but other studies indicate that these antigens could in fact be relevant tumor antigens.²⁷ In this regard, a recent study by Eberhardt *et al*, identified E2-reactive T cells in HPV-related OPSCC,²⁸ suggesting that further evaluation of the key antigenic drivers of HPV-reactive T cells may be merited.

Though HPV16 is associated with most HPV-related cancers, there are other genotypes responsible for a subset of HPV+ cancers. Beyond HPV16, there is limited information on the differential prognosis linked to HPV genotype differences that could impact disease outcomes. Despite these differences, OPSCC driven by different HPV genotypes are similarly treated at all stages of disease.²⁹

Here, we employed high-throughput, single-cell immune synapse profiling using multiplexed barcoded tetramers and single-cell RNA sequencing of HPV16+ and HPV33+ OPSCC tumors to further characterize disease-relevant T cells to identify new candidates for immunotherapeutic intervention. We observed T cell reactivity against both HPV16 E1 and E2 antigens that were highly expanded, cytotoxic and exhibited clonotype sharing across patients in HPV16+ patients. We compared these findings to HPV33+ tumors. Most notably, we found HLA differences that appear to influence disease susceptibility and T cell recognition of viral proteins. Overall, our findings provide deep insights into the immune repertoire of HPV-associated OPSCC combined with heightened understanding of disease-relevant targets and their interplay with the immune machinery that will lay the foundation for future immunotherapies.

RESULTS

We used dissociated fresh surgical resections of 13 untreated and 1 previously treated HPV+ malignancies to profile HPV+ OPSCC tumors at single-cell resolution. Patient characteristics are shown in [table 1](#). The cohort comprised both male and female patients with a median age of 63. A majority of these patients were white non-Hispanic in origin. Approximately 50% of these patients had a history of alcohol consumption and/or smoking. The tumors analyzed in this study primarily originated either in the oropharynx (base of tongue or tonsil) or were collected from involved metastatic neck lymph nodes. For each sample, we performed *ex vivo* characterization

of cell phenotype via single-cell transcriptomics, expression of HPV genes, and paired TCRA/TCRB sequences of tumor infiltrating T cells (Schema in [figure 1](#)). For those samples stained with tetramer (n=9), we also quantified peptide-major histocompatibility complex (MHC) binding for subsequent determination of T-cell receptor (TCR) specificity. This provided a comprehensive characterization of the immune cell landscape in OPSCC and a deep characterization of T cell reactivity.

Tumor microenvironment characterization

Consistent with the prognosis of HPV+ OPSCC, we observed that 4/19 (21%) patients within our cohort developed recurrence during a median follow-up time of 13.6 months (range: 4.2–29.1+) ([table 1](#)).^{30 31} Using cell type inference from single-cell RNA sequencing (RNA-seq) of dissociated tumors, we found that the tumor microenvironment (TME) was composed of a complex mix of cell types aside from tumor, T cells, and B cells (online supplemental figure 1A,B). B cells from 13 different tumors where B cells were detected (DFCI17 had no B cells) were further re-clustered (online supplemental figure 1C) to discern different B cell subpopulations. One patient who recurred had an extremely high proportion of macrophages, but this was not representative of the other two patients who recurred (online supplemental figure 1B).

Heterogeneity in HPV gene expression is correlated with CD8+ T-cell infiltration

HPV gene expression patterns in tumors can vary and this can occasionally be associated with loss of E1, E2 and E5 and lower overall HPV expression.³² It has also been reported that patients with higher expression of E2, E4 and E5 have better outcomes compared with those where E6 and E7 expression predominates.³³ We report HPV gene expression pattern and tumor size, nodal status (number of involved nodes) or smoking status ([figure 2A](#)) in a subgroup of 13 patients. The patient with high macrophages had no detectable tumor cells and therefore was excluded from this analysis. Among the other two patients who recurred, one of them showed E6/E7 expression only, but the other had expression of E1, E2, E4 and E5 with particularly high expression of E2.

We next sought to understand the relationship between HPV gene expression and tumor immune cell infiltrates. Differential composition of T cells across all biopsies is displayed following re-clustering analysis from ([figure 2B](#)). T cell subtype distributions were heterogeneous among patients. We calculated an adjusted Pearson correlation with an adjusted correlation coefficient for multiple comparisons. We found a correlation between E1 gene expression and CD8 effector T cells ([figure 2C](#); $p=0.026$, $R^2=0.394$) and a trend between E1 gene expression and CD8 memory T cells ([figure 2D](#) [figure 2C](#); $p=0.101$, $R^2=0.242$). We detected CD4 Tregs as defined by *FOXP3* expression, and there was a trend towards negative association with E1 gene expression ([figure 2E](#) [figure 2C](#);

Table 1 Patient characteristics

Sample ID	Ethnicity	Tumor site	Stage	Extranodal extension (ENE)	Smoking history	Alcohol use	Prior treatment	Later recurrence in patient	Time to recurrence (months)	Last follow-up	HPV type	HPV reads (fragments/mL)	A02:01+	Biopsy GEX	Tumor HPV GEX
DFC11	W-N	BOT	pT3N2-2	>2mm	Y	N	ChemoRT for Bca, Chemo for CLL	Y	5	19.5	16	12,971	N	Y	Y
DFC12	W-N	Oropharynx	pT1N1-1	N	Y	Y	No	N	19.5	19.5	16	344	Y	Y	Y
DFC13	W-N	BOT	cT1N1	N	N	Y	Basal cell carcinoma	N	17.7	17.7	33	237	Y	Y	Y
DFC14	W-N	Tonsil	pT1N1-1	N	N	Y	No	Y	13.6	19.4	16	10,292	N	Y	Y
DFC15	W-N	Tonsil	pT1N1-1	>2mm	N	N	No	N	16.6	16.6	16	10	N	Y	Y
DFC16	B-N	Tonsil	pT1N1-1	N	Y	Y	No	N	16.7	16.7	33	142	Y	Y	Y
DFC18	W-N	BOT	cT1N1M0	Y	Y	Y	ChemoRT for ThCa	N	5.8	5.8	16	935	Y	N	N
DFC19	W-N	Larynx	pT1N1-1	N	Y	Rare	Prior RT for TNBC	N	10.1	10.1	33	12	Y	N	N
DFC111	W-N	Tonsil	pT1N1-1	N	Y	Y	No	N	14.9	14.9	16	43	N	Y	Y
DFC112	W-N	Tonsil	cT1N1-3M0	Y	Y	Unknown	ChemoRT (TestCA)	N	17.7	17.7	16	3,895.4	N	N	N
DFC114	W-N	Met lymph node	pT1N2-2	>2mm	Y	Unknown	No	N	10.8	10.8	33	1,702	Y	Y	Y
DFC115	W-N	Tonsil	pT2N1-1	N	N	Y	No	N	12.2	12.2	16	47	Y	Y	Y
DFC116	W-N	Tonsil	pT2N1-1	>2mm	N	N	No	N	13.6	13.6	16	Indeterminate	N	Y	Y
DFC117	W-N	Met lymph node	cT1N1M0	N	N	Rare	ChemoRT	Y	4.1	29.1	16	Tissue positive	N	Y	N
DFC118	W-N	Tonsil	pT2N1-1	<2mm	Y	Y	No	N	10.2	10.2	16	21,310	N	Y	Y
DFC119	W-H	Tonsil	1-N1N1d	N	N	Y	No	Y	4.6	10.7	16	67	N	N	N
DFC120	W-N	Pharyngeal wall	pT2N0-1	N	N	N	No	N	7.7	7.7	16	117	Y	Y	Y
DFC126	W-N	Tonsil	pT1N1-1	N	N	Unknown	No	N	7.4	7.4	33	12	Y	N	N
DFC135	W-N	Lymph Nodes	pT0N1-1	N	Y	Y (rarely)	Mohs surgery for basal cell carcinoma	N	6.6	6.6	33	778	N	Y	Y

Bca, breast cancer; B-N, black non-Hispanic; BOT, base of tongue; CLL, chronic lymphocytic leukemia; disease stage c, clinical staging; disease stage p, pathologic staging; GEX, gene expression; Met lymph node, metastatic lymph node; N, no; RT, radiation therapy; TestCA, testicular cancer; ThCa, thoracic cancer; TNBC, triple negative breast cancer; W-H, white Hispanic; W-N, white non-Hispanic; Y, yes.

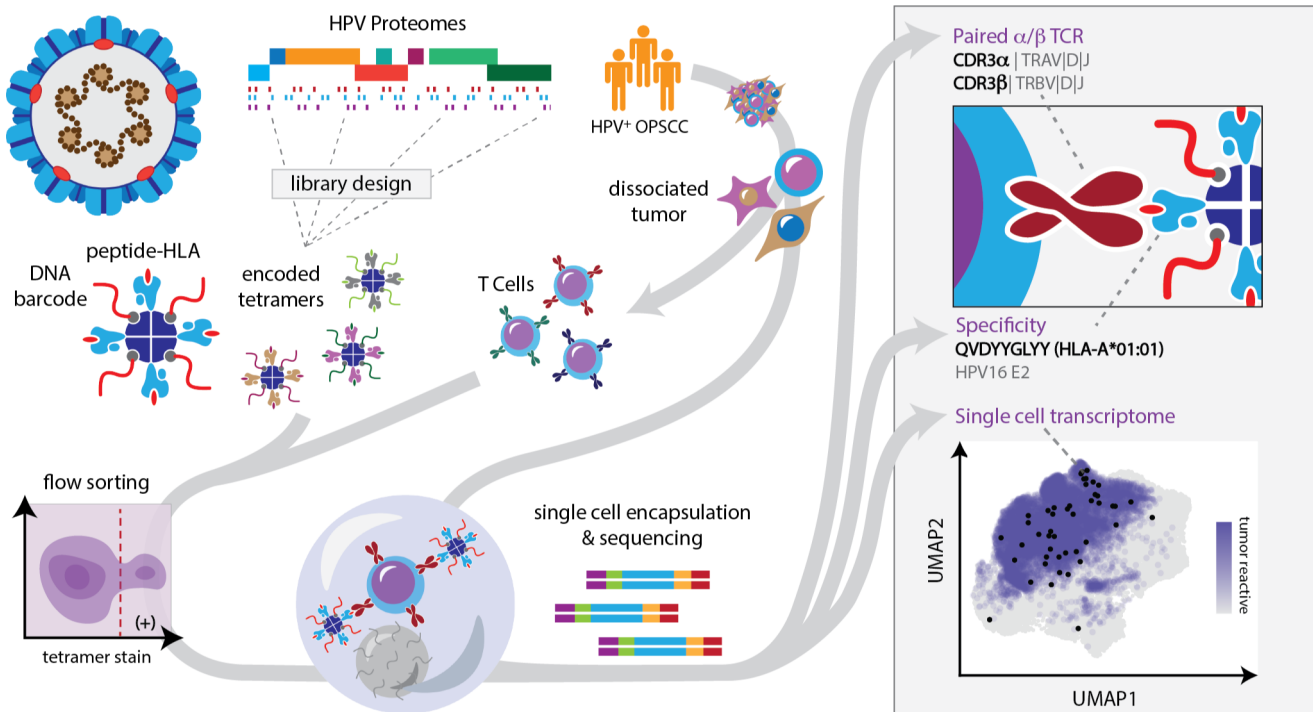


Figure 1 Experimental workflow. Biopsies were dissociated and subjected to scRNA-seq. Separately, tumor-infiltrating lymphocytes were isolated and probed with DNA-encoded peptide-MHC tetramers, sorted using flow cytometry and subjected to scRNA-seq. Readout of TCR, GEX, and tetramer sequencing data allowed for the characterization of TCRs, cellular phenotype, and tetramer binding. GEX, gene expression; HLA, human leukocyte antigen; HPV, human papillomavirus; MHC, major histocompatibility complex; OPSCC, oropharyngeal squamous cell carcinoma; scRNA-seq, single-cell RNA sequencing; TCR, T-cell receptor; UMAP, uniform manifold approximation and projection.

$p=0.108$ $R^2 = 0.233$). We also analyzed correlations between HPV E6 gene expression and T-cell subsets and found that E6 expression negatively correlated with CD4 Tfh cells ($p=0.021$, $R^2=0.412$). Together, these data indicate that higher HPV E1 gene expression results in a more robust proinflammatory TME associated with greater CD8 infiltration, activation and differentiation in virally-driven OPSCC.

HPV16 E1 and E2 drive robust public cytotoxic T cell responses

We observed broadly that expression of some HPV genes was positively correlated with CD8 infiltration, so we sought to analyze if tumor infiltrating lymphocytes (TIL) were recognizing HPV antigens. We probed T cells from four HPV16+ tumors with tetramers from HPV16 E1, E2, E4, E5, E6 and E7 and three with tetramers from E6 and E7 alone. We probed two HPV33+ tumors with tetramers from HPV16 E1, E2, E4, E5, E6 and E7. Peptide sequences in the tetramer libraries can be found in online supplemental table 1 and a list of alleles tested per patient can be found in online supplemental table 2. HPV16-reactive T cells were abundant in tumors, particularly those reacting to E2, E1 and E6 (figure 3A). We found no reactivity to any HPV33 proteins in the two alleles probed, HLA-A*02:01 and HLA-A*24:01 despite other viral controls in the experiment working well. Similarly, we found limited or no reactivity to HPV16 proteins in HLA-A*02:01 and HLA-A*24:02. We found limited reactivity to proteins

E5 and E7 (figure 3A,B, filled circles). In contrast we found abundant reactivity to proteins HPV16 E1 and E2 in two patients, particularly in A*01:01 and B*08:01 (figure 3A,B, filled circles). In patient DFCI20, we saw robust T-cell recognition of E2 and a near complete loss of E2 gene expression. This may indicate the functionality of these T cells in lysing E2+ tumor cells. T-cell reactivity to the epitope QVDYYGLYY-A*01:01 from HPV16 E2 was previously reported by Eberhardt *et al* from an HPV+ OPSCC patient.²⁸ We also identified many T cells specific for this epitope which made up approximately 1.6% of TIL in patient DFCI18. HPV-reactive CD8s recognizing E1 and E2 and showed moderate expression of TOX high mobility group box family member 2 (TOX2) indicative of exhaustion and showed high expression of genes associated with tumor reactivity³⁴ (figure 3C). Most HPV-reactive CD4s recognized E5, E6 and E7 and showed moderate-to-low expression of interferon (IFN)- γ , and moderate-to-high expression of TOX2. The combination of low IFN- γ and high TOX2 may indicate that these clonotypes may have a limited ability to carry out cytotoxic gene programs (figure 3D).^{35 36}

It has been reported that paired CDR3 TCRA and TCRB amino acid sequences can be shared across individuals, and similar TCR sequences may recognize the same epitope.^{37 38} We sought to look for TCR homology across patients with matched HLAs to see if they may have similar or identical TCRs recognizing the same antigen

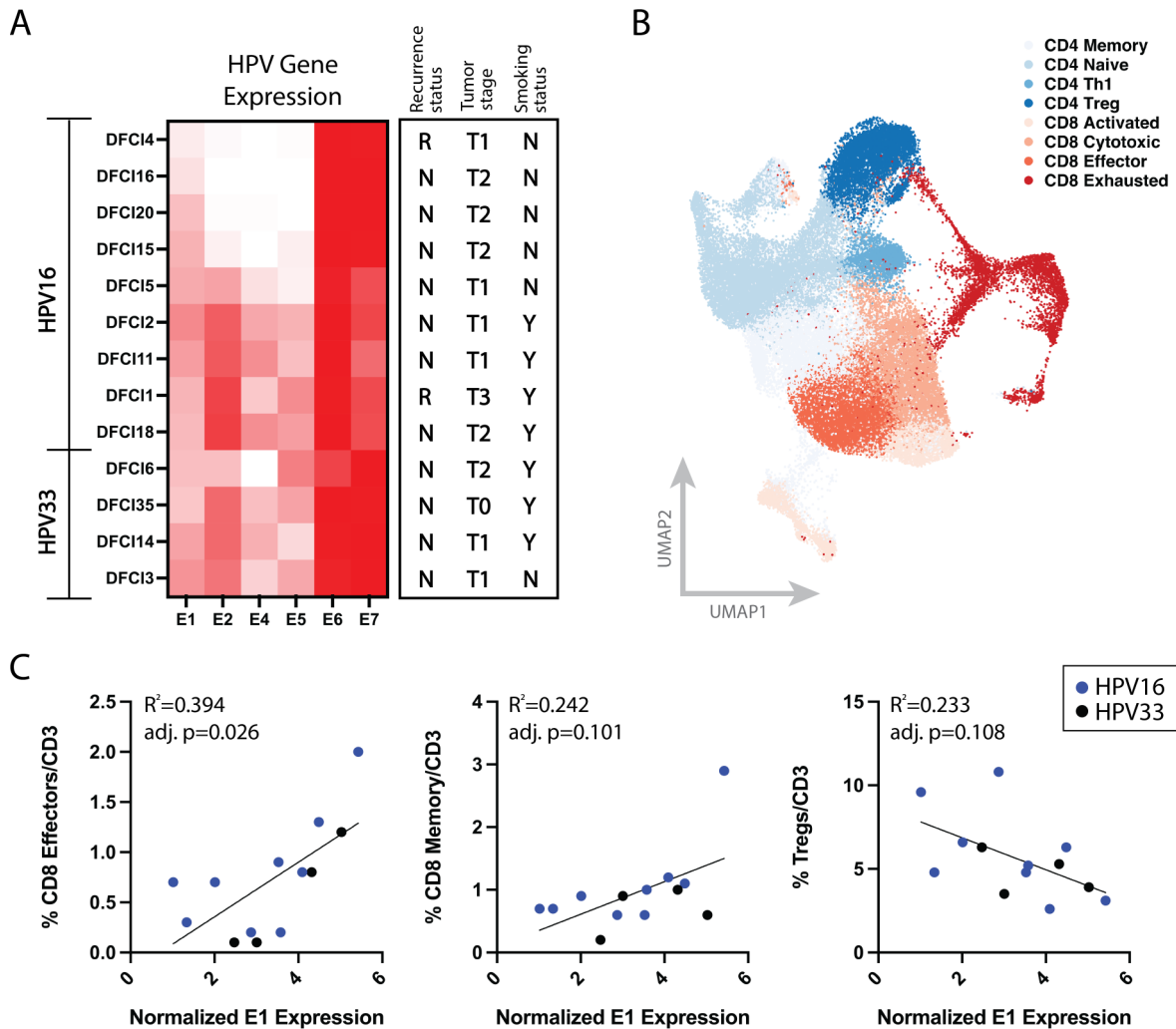


Figure 2 Heterogeneity in HPV gene expression influences T cell infiltrates. (A) The percentage of HPV+ tumor cells expressing each HPV gene assessed is heterogenous. (B) UMAP showing cell classifications of T cell subtypes in TME. (C). HPV E1 expression is positively correlated with CD8 memory infiltration. HPV E1 expression is positively associated with CD8 effector T cell infiltration. HPV E1 expression is negatively associated with CD4 Treg infiltration. adj., adjusted; HPV, human papillomavirus; Th1, type 1 T helper cells; TME, tumor microenvironment; Treg, regulatory T cells; UMAP, uniform manifold approximation and projection.

allowing us to infer T cell reactivity. We were unable to probe dissociated tumor cells from all patients with tetramer due to resection size limitations or tetramer availability at the time of collection. We aggregated TCR sequences across single-cell sequencing data of patient tumor resections ($n=14$) and assessed homology of TCR alphas and TCR betas using TCRDist3. We then cross-mapped our tetramer-identified TCRs to infer specificity onto TCR clusters.

We found a striking level of alpha and beta sharing at the amino acid and gene level across these patients particularly in T cells reacting to viral antigens, with identical matches or encoded sequences differing by only 1-2aa. In our data we found that clonotype sharing existed among a number of previously characterized virally-associated epitopes.^{39 40} Examining HPV-relevant antigens, T cells recognizing QVDYYGLYY-A*01:01 from HPV16 E2 displayed the most clonotype sharing across patients primarily from patients DFCI1, DFCI11, DFCI20,

DFCI18, with a common alpha motif (figure 3E) recognizing QVDYYGLYY-A*01:01 (HPV16 E2). By tetramer, we identified numerous TCRs binding to the QVDYYGLYY-A*01:01 (HPV16 E2) tetramer sharing the alpha CAVDTGGFKTIF or CAVNTGGFKTIF between patients DFCI18 and DFCI20. We also saw a robust clonotype network among T cells reacting to YSKNKVWEV-B*08:01 from HPV16 E2 that were united by a common beta motif (figure 3F).

To test the validity of these TCR-epitope interactions, recombinant TCRs were purchased and expressed in J76-CD8-NFAT-GFP cells, incubated overnight with cognate peptide and HLA-matched antigen presenting cells. Activation was assessed by nuclear factor of activated T cells (NFAT) and CD69 signal by flow cytometry (Schema in figure 4A). We chose a subset of TCR-epitope pairs identified by tetramer to validate in this functional assay.⁴¹ Results of this validation confirmed TCRs reactive to several epitopes and are presented in figure 4B. We also

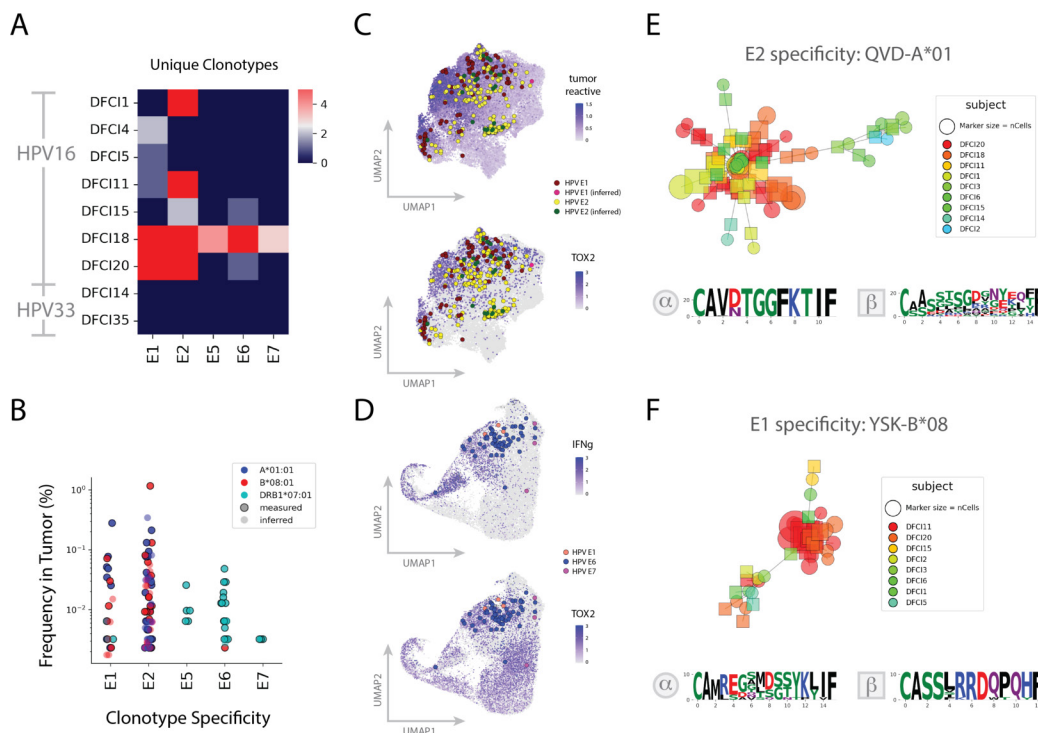


Figure 3 Cellular response to HPV antigens in OPSCC. (A) Number of clonotypes by subject measured and inferred with high confidence to be specific to epitopes derived from HPV antigens. (B) Frequency of HPV-reactive clonotypes in dissociated tumor samples. Measured clonotypes are displayed with filled dots and inferred clonotypes are displayed with open dots. (C) UMAP of CD8+ TIL showing HPV-specific clonotypes overlaid on tumor reactive and TOX2 gene scores. (D) UMAP of CD4+ TIL showing HPV-specific clonotypes overlaid on IFN- γ and TOX2 gene scores. (E) HPV16 E2 QVDYYGLYY-A*01:01-specific T cells show strong clonotype sharing within the TCR-alpha chain across patients. We observed T cells from several different patients appearing in this cluster, with heaviest contributions from DFCI1, DFCI11, DFCI20 and DFCI18, all patients that were HLA-A*01:01+ and HPV16+. (F) HPV16 E2 YSKNKVWEV-B*08:01-specific T cells show strong clonotype sharing within the TCR-beta chain across patients. We observed T cells from several different patients appearing in this cluster, with heaviest contributions from DFCI1, DFCI11, DFCI20 and DFCI15, all patients that were HLA-B*08:01+ and HPV16+. HLA, human leukocyte antigen; HPV, human papillomavirus; IFN, interferon; OPSCC, oropharyngeal squamous cell carcinoma; TCR, T-cell receptor; TIL, tumor-infiltrating lymphocytes; TOX2, TOX high mobility group box family member 2; UMAP, uniform manifold approximation and projection.

tested a subset of the T cells inferred to be specific to QVDYYGLYY-A*01:01 and YSKNKVWEV-B*08:01 (HPV16 E2). From the YSKNKVWEV-B*08:01 network we selected three clonotypes to test in our TCR-epitope validation assay from patient DFCI11, whose TIL we were unable to probe directly with tetramer. All three of these clonotypes validated, including those constituting 1.20%, 0.13% and 0.025% of that patient's TIL (figure 4C). For QVDYYGLYY-A*01:01, we tested T cells with homology both in the alpha and beta chain and found that T cells with alpha chains homologous to those measured by tetramer validated, while those united by beta homology did not.

(figure 4D). Overall, we see that in line with prior reports showing homology and clonotype sharing in virally-reactive T cells, HPV16 E1 and E2-reactive T cells are observed to induce clonotype sharing among patients.

In summary, we observed that HPV genes E1 and E2 elicited robust T cell responses including expression of genes reported to be associated with tumor-reactive T cells.³⁴ We also observed that HPV proteins E1 and E2 elicited robust public or pseudo-public T cell responses. These T cells expressed genes reported to be associated

with tumor-reactive T cells.³⁴ Together, these data indicate that HPV16 E1 and E2 gene expression results in a robust proinflammatory TME associated with CD8+ T cell infiltration, activation and differentiation in virally-driven OPSCC.

HPV33 and HPV16-driven tumors are immunologically distinct and display differences in lymphocyte infiltration and HLA-A allele usage

We performed an exploratory analysis of the tumor microenvironment to assess the proportion of dendritic cells, fibroblasts, monocytes, natural killer cells, neutrophils or overall T-cell number between HPV16 (n=10) and HPV33+ (n=4) patients (online supplemental figure 2A). Consistent with Chatfield-Reed *et al.*⁴² we found that there was a trend for fewer cytotoxic T cells accounting for 1.45% of dissociated tumor cells in HPV33+ patients compared with 7.33% of HPV16+ patients (p=0.02398) (online supplemental figure 2B). There were no differences in any other CD4 or CD8 subsets between genotypes (online supplemental figure 2B).

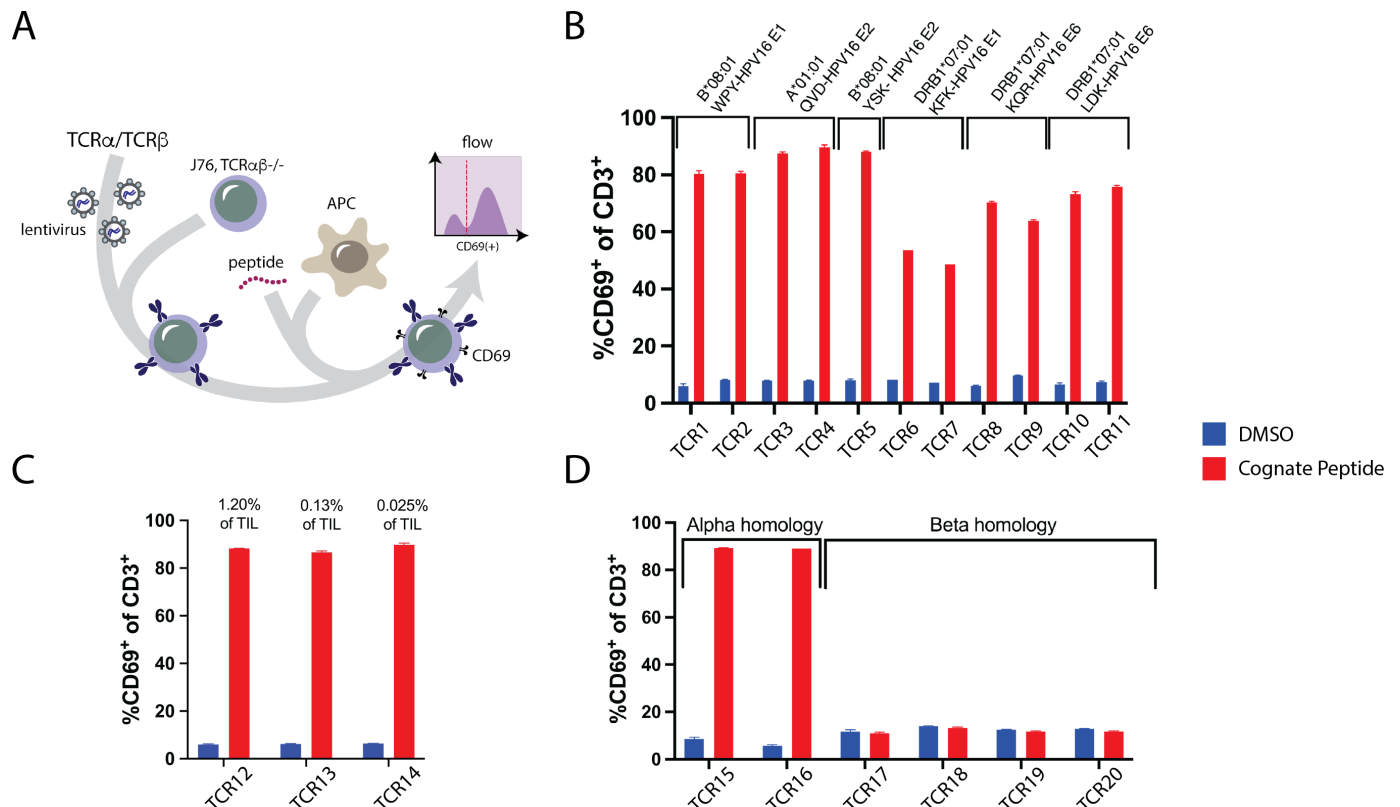


Figure 4 Functional validation of TCR-epitope pairs. (A) TCR-epitope interactions across HPV16 genes were tested using this workflow. (B) The TCRs derived from patient TIL tested signaled when presented with cognate peptide as measured by CD69 induction. Peptides tested included WPYLHNRLV-B*08:01, QVDYYGLYY-A*01:01, YSKNKWVEV-B*08:01, KFKELYGVSFSELVR-DRB1*07:01, QQRHLDDKKQRFHNIQRWTGR-DRB1*07:01 and LDKKQRFHNIQRWTGR-CMSC-DRB1*07:01. (C) Three TCRs derived from tumor of patient DFC11 tested with beta chain homology to YSKNKWVEV-B*08:01-reactive T cells signal when presented with cognate peptide and make up 1.2%, 0.13% and 0.025% of TIL. (D) TCRs with beta chains homologous to QVDYYGLYY-reactive T cells fail to signal but both TCRs with similar alphas signal when presented with cognate peptide. APC, allophycocyanin; DMSO, dimethylsulfoxide; TCR, T-cell receptor; TIL, tumor-infiltrating lymphocytes.

Next, we compared the HLA-A*02 allele distribution across our patient cohort (total $n=13$ (HPV16+); $n=6$ (HPV33+)) with the US Caucasian population. Intriguingly, we noted that HLA-A*02 allele frequency was higher among HPV33+ patients in our cohort compared with the US Caucasian population (HLA-A*02 allele frequency: 0.5 (DFCI data set) compared with 0.28 (US Caucasian)) (online supplemental figure 2C). The data in our patient cohort were concordant with an independent data set of patients with OPSCC (Tempus) where total $n=16$ HPV33+ and $n=202$ HPV16+ patients were analyzed (online supplemental figure 2C). Therefore, we found 1.77-fold more HPV33+ patients carrying the HLA-A*02 allele than in US Caucasians ($p=0.0247$) (online supplemental figure 2C). We next sought to define epitopes well presented in HPV16 but missing in HPV33 which could potentially explain susceptibility to HPV33 but not HPV16-driven OPSCC among HLA-A*02:01+ patients. We performed single-cell HPV gene expression in tumor cells from four patients with HPV33 and nine patients with HPV16. Two HPV33+ patients in our cohort had no tumor gene expression data. Comparing the ratios of E7/E6 expression across this cohort, HLA-A*02:01+ HPV33+ patients had higher E7 gene expression than E6 (online

supplemental figure 2D). In contrast, all patients with HPV16 had lower E7/E6 expression regardless of HLA type, and the single HLA-A*02:01- HPV33+ had E7/E6 gene expression resembling the HPV16+ patients (online supplemental figure 2D).

To test if there was a recognition deficit in A*02:01 of HPV33 E7 versus HPV16 we used NetMHCpan I to predict potentially presented epitopes from HPV16 versus HPV33 in HLA-A*02:01. While HPV16 and HPV33 are similar in sequence, they are not identical and differences in their amino acid sequence results in different predicted epitopes for many regions of the HPV genome (online supplemental figure 2E). There were several predicted epitopes across HPV E1, E2, E4, E5 and E6 with no clear deficiencies in HPV33 compared with HPV16. However, E7 epitope prediction revealed that each genotype had three (HPV33) or four (HPV16) predicted epitopes under 200nM affinity within E7 with sequence, with differences between these epitopes (online supplemental figure 2E). One epitope predicted to have high affinity in HPV16 E7- amino acids 11–19, was not predicted to be presented in HPV33 (online supplemental figure 2E). HPV16 E7_{11–19} has been reported to be a critical epitope in HPV16 recognition and has been the target of several

clinical trials seeking to treat HPV16-driven malignancies.^{21 43 44} Therefore, we infer that a deficiency in HPV E7 recognition by T cells in HPV33 HLA-A*02:01+ may pose an elevated risk of virally-driven malignancy for this HPV genotype.

DISCUSSION

Broadly, we demonstrate that the antigenic landscape of HPV-driven OPSCC is far more complex than the nearly unilateral focus on HPV16 E6 and E7 would imply. This focus has historically been both because HPV16 is the most common cancer-causing HPV genotype and because E6 and E7 are the only proteins required to maintain malignancy. However, HPV16 E6 and E7 targeted therapies have had only modest success and there is much room for improvement. We find that other HPV16 proteins, particularly E1 and E2 drive highly expanded and cytotoxic T cell responses and these antigens could be promising targets for immunotherapy.

In at least one patient, DFCI20, these E2-reactive T cells were associated with E2 loss, potentially indicating the functional relevance of these T cells. One patient we examined with high E2 expression who failed to mount an E2 response did poorly and relapsed quickly on standard treatment. This patient might be a good candidate for an E2-targeted immunotherapy as expression of the antigen remains. We also observed robust clonotype sharing among E1 and E2-reactive T cells. This phenomenon has been previously reported particularly among T cells reacting to viral antigens such as influenza, human cytomegalovirus and Epstein-Barr virus.⁴⁵ To our knowledge this is the first report to show this occurring in HPV-reactive T cells. This is important because it highlights the strength of these antigens in inducing strong public T cell responses and the universal availability of reactive precursors.

Work done to understand clearance of precancerous HPV-driven lesions may also be applicable to OPSCC. Two studies have reported that the clearance of these lesions is associated with the emergence of peripheral E2, but not E6 or E7, T cell responses.^{46 47} An HPV16 E2 vaccination has been used as a therapy with striking efficacy, though the study was closed due to low enrollment issues.⁴⁸ Additionally, a separate study reported that greater E1 T cell response was associated with improved prognosis in cervical cancer.⁴⁹ Many clinical trials have targeted proteins E6 and E7 in HPV-driven OPSCC, but none have been particularly successful. Although these are viral targets usually associated with robust immune responses, both proteins are small (E7 98aa and E6 158aa) with a limited number of presentable epitopes. Both E1 and E2 are substantially larger at 649aa and 365aa, respectively, and may be a better therapeutic target partially for this reason. This literature around precancerous lesion clearance and E1 reactivity in relation to cervical cancer outcome paired with our findings of robust cytotoxic tumor T cell reactivity to E1 and E2 make HPV16

E1 and E2 promising candidates for immunotherapy development.

Additionally, we find that HPV33-driven OPSCC may be associated with HLA-A*02:01 as a risk allele and is characterized by lower cytotoxic T cell infiltration that may partially explain reports of worse outcomes. We were unable to find tumor T cells recognizing any HPV33 antigens in the two alleles that we probed. Clinically, p16 immunohistochemistry is commonly used as a surrogate to detect high-risk HPV infection as p16 is typically upregulated by HPV proteins. This approach fails to delineate the HPV genotype driving malignancy and therefore, all patients regardless of their HPV genotype are currently subjected to the same standard of care treatment which may not be an optimal strategy. HPV33-driven malignancies have worse outcomes compared with HPV16 genotype in patients with both OPSCC and cervical cancer^{42 50} and the disparity in clinical outcomes can be partly attributed to the unique genomic and immunologic characteristics.^{51–53} Despite the very limited sample size, our observation of fewer cytotoxic T cells in HPV33-driven tumors replicated previously published reports.⁴² Plasma tumor tissue-modified viral (TTMV)-HPV DNA testing can help clarify HPV genotype if integrated as a part of the diagnostic evaluation to potentially provide a more tailored approach.

The HPV16 E7_{11–19} epitope is important for T cell mediated recognition of HPV16 E7 and has been the focus of multiple clinical trials.⁵⁴ Others have noted that HPV33+ OPSCCs do not show the same affinity for immune-protected tonsillar crypts in the way that HPV16+ OPSCCs do, potentially confirming that these tumors are less immunogenic than HPV16+ cancers.⁵⁵ Regardless, we find an unmet need that presents a unique opportunity for novel precision-based immunotherapeutic approaches for non-HPV16 malignancies and to improve outcomes for this growing patient population.

There are several limitations to our work. First, the number of patients we were able to study was small. Though we were able to do a comprehensive analysis on each patient our ability to extrapolate our findings is limited by sample size and it is likely that with a larger sample size, more variables significantly impacting recurrence and tumor milieu would be uncovered. Additionally, most of the patients that we examined were newly diagnosed and therefore the translatability to patient tumors who have failed treatment is unknown.

In conclusion, we find that the biology of HPV+ OPSCC is exquisitely complex. We find evidence that differences in the underlying OPSCC HPV genotype results in heterogeneity of TME and patient outcomes. This indicates that HPV typing at diagnosis could provide significant benefit in terms of clinical management and selection of patients for de-intensified or more aggressive therapy. Additionally, we find that reactivity to proteins HPV16 E1 and E2 contribute meaningfully to T cell recognition of HPV+ tumors and could be valuable targets for immunotherapies.

MATERIALS AND METHODS

Patient accrual and clinical annotation

Patients with newly diagnosed HPV-driven OPSCC (base of tongue, tonsil, and unknown primary) were identified and prospectively enrolled over a 12-month study period if they were undergoing confirmatory tissue biopsy or definitive oncologic transoral robotic-assisted resection with neck management of the primary tumor and/or involved cervical neck nodes. Patients with known distant metastatic disease were excluded, otherwise all clinical stages of curable disease were eligible (stages I, II, and III by American Joint Commission on Cancer 2017 eighth edition). Patients were consented to an existing, institutional review board-approved head and neck cancer tissue collection protocol (DF/HCC#09-472) prior to enrollment and sample acquisition. Clinical information and demographics were recorded for each participant along with initial treatment and response data. Patients were followed post-treatment to evaluate for biopsy-confirmed recurrence, survival status, and therapy-related toxicity. Recurrence was defined as locoregional tumor regrowth or distant metastatic spread (either or both).

Tissue acquisition and collection

Following informed consent, at the time of intra-office or interventional radiology facilitated core needle biopsy for diagnosis and/or transoral operative resection tissue specimens were prospectively collected. On the day of the procedure, freshly acquired tissue samples were submerged fully in pre-filled 1.5 mL cryovials or 5 mL Eppendorf tubes with magnetic-activated cell sorting (MACS) tissue storage solution (Miltenyi Biotec) and immediately labeled and stored at 4°C. Samples were then shipped directly to Repertoire Immune Medicines within 24 hours (using Biocair life science shipping), avoiding late in the week collections to avoid weekend-related processing delays.

Blood and buccal swab collection

On the same day as the planned biopsy or operative procedure for tissue procurement, patients underwent collection of 5×6 mL or 3×10 mL green top heparin (or ethylenediaminetetraacetic acid (EDTA) purple top tubes based on availability) stored at ambient temperature and similarly shipped directly to Repertoire within 24 hours of venipuncture. Additionally, two buccal cheek swabs (RSC Buccal Swab DNA kit, Maxwell, Madison, Wisconsin, USA) were obtained from each participant pretreatment with collection supervised by a study team researcher and stored at ambient temperature; shipped directly to Repertoire with matching patient blood samples.

HPV plasma and tissue diagnostic testing

Fine needle aspiration cytology, core needle, or surgical biopsy specimens were obtained from all patients to ensure clinical diagnostic accuracy. Immunostaining for p16 was performed in all cases and required 70% or great expression to be deemed positive, as is convention.

Confirmatory tissue-based in situ hybridization (ISH) or PCR testing was attempted in all cases to confirm HPV causality. Our in-house workflow is to perform a platform test for HPV E6/7 transcripts of subtypes 16/18 by RNA ISH, followed by reflex PCR to identify other high-risk subtypes if ISH returned negative. Results are interpreted by two independent anatomic pathologists with head and neck subspecialty training. All patient biopsies included here tested positive both by p16 and RNA ISH.

We performed baseline peripheral blood testing pretreatment on all participants for TTMV-HPV DNA (NavDx, Naveris, Waltham, Massachusetts, USA)^{56 57} at the time of enrollment on blood and/or tumor. This ultrasensitive digital droplet PCR assay detects the five most common high-risk subtypes of HPV (16, 18, 31, 33, and 35) and reports TTMV-HPV DNA fragments/mL of plasma (normalized across samples to generate a score) categorized as positive (>7 HPV16 and >12 non-HPV16), negative (<5 fragments/mL), or indeterminate (5–7 HPV16 and 5–12 non-HPV16) result for reporting purposes.

Tumor dissociation

Fresh tumor biopsy samples obtained from patients with HPV+ OPSCC were dissociated using a human tumor dissociation kit (Miltenyi Biotec, Bergisch Gladbach, Germany) per manufacturer guidelines. Briefly, tumors were weighed, and cut into small pieces (2–4mm) in 1–2 mL Roswell Park Memorial Institute (RPMI)-1640 media containing tumor dissociation enzymes. Tumor chunks were further minced on the gentleMACS Octo Dissociator (Miltenyi Biotec) at 37°C for 15 min. Dissociated tumor cells were pelleted and resuspended in 1–2 mL RPMI-1640 containing 10% human serum albumin. The cell suspension was filtered through a MACS Smart-Strainer (70 µm), pelleted, and resuspended in 1 mL of RPMI-1640 with 10% serum. The dissociated tumor cells were counted, and viability was determined using the ViaStain viability dye (Nexcelom Bioscience, Lawrence, Massachusetts, USA) using an automated cellometer (Nexcelom Bioscience).

HLA typing

The genomic DNA (gDNA) was extracted from the peripheral blood of patients with HPV+ OPSCC by using GeneCatcher gDNA Blood Kit (Invitrogen). The quantity and quality of gDNA was evaluated in NanoDrop 8000 (Thermo Fisher). The gDNA was then amplified using LinkS_{eq} HLA-ABCD_RD_QDP+384 Kit, and the dissociation profile was subjected to SureType software (One Lambda) to determine the patient's HLA type. HLA typing was validated by buccal swab through a commercial service (Cisco Genetics).

Antigen library design

The amino acid sequences of HPV16 and HPV33 E1, E2, E4, E5, E6 and E7 were run through NetMHCpan⁵⁸ in order to predict all possible 9-mers (class I), 15-mers or

21-mer (class II) binding to their respective MHCs with an affinity of 500 nM or stronger. We applied this strategy to predict HPV peptides for A*01:01, A*02:01, A*24:02, B*07:02, B*08:01 and DRB1*07:01. Finally, we added into libraries a set of well-defined viral epitopes from cytomegalovirus, Epstein-Barr virus, influenza viruses (CEF peptide pool) and SARS-CoV-2 that elicit T-cell responses in the population at large. Antigenic peptides with 500 nM affinity or lower were then selected for inclusion.

Production of peptide MHC library pools

Class I MHC extracellular domains were expressed in *Escherichia coli* and refolded along with beta-2-microglobulin and ultraviolet (UV)-labile place-holder peptides.^{59–62} A C-terminal sortase recognition sequence on the HLA was modified by sortase transpeptidation^{63 64} with a synthetic alkynylated linker peptide, featuring an N-terminal triglycine connected to propargylglycine via a PEG linker (Genscript, Piscataway, New Jersey, USA). The modified HLA monomer was then purified by size-exclusion chromatography (SEC). Full-length streptavidin with an N-terminal Flag tag and a C-terminal sortase recognition sequence and 6xHisTag was prepared by expression and purification from *E. coli* using immobilized metal affinity chromatography and SEC. Streptavidin was modified by sortase transpeptidation with a synthetic azidylated linker peptide, featuring an N-terminal triglycine connected to picolyl azide via a PEG linker (Click Chemistry Tools, Scottsdale, Arizona, USA). HLA tetramers were produced by mixing alkynylated HLA monomers and azidylated streptavidin in 0.5 mM copper sulfate, 2.5 mM BTAA (2-(4-((Bis((1-(tert-butyl)-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-1,2,3-triazol-1-yl)acetic acid) and 5 mM ascorbic acid for up to 4 hours on ice, followed by purification of highly multimeric fractions by SEC. Individual peptide exchange reactions containing 500 nM HLA tetramer and 60 mM peptide were exposed to long-wave UV (366 nm) at a distance of 2–5 cm for 30 min at 4°C, followed by 30 min incubation at 30°C. A biotinylated oligonucleotide barcode (Integrated DNA Technologies) was added to each individual reaction followed by 30 min incubation at 4°C. Individual tetramer reactions were then pooled and concentrated using 30 kDa molecular weight cut-off centrifugal filter units (Amicon). Tetramer production was quality controlled using SEC, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and UV-mediated peptide exchange by assessing binding to peptide-expanded cell lines (data not shown).

The MHC Class II α -chain and β -chain extracellular domains were recombinantly expressed with C-terminal Myc and His tag sequences, respectively. For DRB1*15:01 the Myc tag was replaced with a V5 tag. The N-terminus of the β -chain was fused to class II-associated invariant chain (CLIP) peptide followed by a flexible Factor Xa-cleavable linker. Both α -chain and β -chain were co-expressed in Chinese hamster ovary (CHO) cells and secreted into the expression medium as a stable CLIP-loaded heterodimer. Heterodimerization of the α -chain and β -chain of

DRB1*07:01 was forced using a fusion of an engineered human IgG1-Fc protein to each chain. Following CHO expression, the heterodimer was purified by immobilized metal ion affinity chromatography and SEC and was subjected to an overnight biotinylation reaction at 4°C using a commercial BirA biotin-protein ligase reaction kit (Avidity). Biotinylated proteins were further purified via SEC and digested with Factor Xa protease according to the manufacturer's instructions (New England BioLabs). The proteolytic reaction was stopped by the addition of Factor Xa inhibitor, 1,5-Dansyl-Glu-Gly-Arg Chloromethyl Ketone, Dihydrochloride (Sigma-Aldrich). Class II proteins were concentrated to ~30 μ M and stored at -80°C prior to preparation of the corresponding barcoded peptide libraries. Individual peptide exchange reactions containing 7 μ M Class II heterodimer and 350 μ M peptide in a sodium citrate buffer (100 mM sodium citrate, 100 mM sodium chloride, 0.1% octyl glucoside, SIGMAFAST Protease Inhibitor Cocktail Tablets) were incubated at pH 5.5 and 30°C overnight. The individual reactions were then neutralized using 1 M Tris-HCL at pH 10 to a final concentration of 0.14 M Tris-HCL. In a separate container, Klickmers (Immudex, Denmark) were individually barcoded using biotinylated oligonucleotide barcodes (Integrated DNA Technologies) and incubated for 30 min at 4°C. Individually exchanged Class II heterodimers were then added to the individually barcoded Klickmers and allowed to incubate for 30 min at 4°C. The 2 mM D-Biotin (Invitrogen) was then added to each individual reaction to quench the streptavidin, and allowed to incubate for 30 min at 4°C. These individual reactions were then pooled and concentrated using 30 kDa molecular weight cut-off centrifugal filter units (Amicon), and subsequently washed using cold PBS (Gibco) until a desired number of washes was achieved. The pool was then concentrated to the desired final concentration for staining.

Cell staining

Dissociated tumor cells obtained from the patients with OPSCC were subjected to MACS to get an enriched CD3+, CD8+ or CD4+ T cells using CD3T cell or CD8 T cell isolation kit, respectively, (Miltenyi Biotec) following the manufacturer's protocol. The enriched fraction of T cells was then stained with 1 nM final concentration tetramer library in the presence of 2 mg/mL *salmon sperm* DNA in phosphate-buffered saline (PBS) with 0.5% bovine serum albumin solution for 20 min. Cells were then labeled with anti-TCR antibody-derived tag (ADT, IP26, BioLegend, San Diego, California, USA) for 15 min followed by washing. Tetramer bound cells were then labeled with phycoerythrin (PE) conjugated anti-DKDDDDK-Flag antibody (BioLegend) followed by dead cell discrimination using 7-amino-actinomycin D. The live, tetramer positive cells were sorted using a Sony MA900 Sorter (Sony Biotechnology, San Jose, California, USA). Final counts were determined and viability was assessed using

an automated cell counter (Nexcelom Bioscience) for use in single-cell encapsulations.

Sample multiplexing

To ensure sufficient complementary DNA production in single-cell sequencing, we used sample multiplexing for several experiments. When applied, samples were stained and sorted separately according to the protocol described above using anti-TCR ADTs with unique barcodes. Labeled samples were combined prior to encapsulation and single-cell sequencing.

Single-cell sequencing

Single-cell encapsulations were generated using 5' v1 Gem beads from 10x Genomics (Pleasanton, California, USA) on a 10x Chromium controller and downstream TCR, and Surface marker libraries were made following manufacturer recommended conditions. All libraries were quantified on a BioRad CFX 384 (Hercules, California, USA) using Kapa Biosystems (Wilmington, Massachusetts, USA) library quantified kits and pooled at an equimolar ratio. TCRs, surface markers, and tetramer-generated libraries were sequenced on Illumina (San Diego, California, USA) NextSeq 550 instruments. Sequencing data were processed using the Cell Ranger Software Suite (V.3). Samples were demultiplexed and unique molecular identifier (UMI) counts were quantified for TCRs, tetramers, and gene expression.

Single-cell transcriptomic analysis

Hydrogel-based RNA-seq data were analyzed using the Cell Ranger package from 10x Genomics (V.3.1.0) with the GRCh38 human expression reference (V.3.0.0). Except where noted, the Seurat R package (V.1.6.0) was used to perform the subsequent single-cell analyses. Any exogenous control cells identified by TCR clonotype were removed before further gene expression processing. Hydrogels that contain UMIs for less than 300 genes were excluded. Genes that were detected in less than three cells were also excluded from further analysis. Several additional quality control thresholds were also enforced. To remove data generated from cells likely to be damaged, upper thresholds were set for per cent UMIs arising from mitochondrial genes (10%). To exclude data likely arising from multiple cells captured in a single drop, upper thresholds were set for total UMI counts based on individual distributions from each encapsulation (from 1500 to 3000 UMIs). Any hydrogel outside of any of the thresholds was omitted from further analysis. A total of 133,182 hydrogels were carried forward. Gene expression data were normalized to counts per 10,000 UMIs per cell (CP10K) followed by log_{1p} transformation: ln(CP10K+1).

Highly variable genes were identified (1,567) and scaled to have a mean of zero and unit variance. They were then provided to Seurat (V.3.0),⁶⁵ to perform batch integration and dimension reduction. The data were then used to generate shared principal components which were in turn used to generate a UMAP representation.

Shared neighbor graphs and Leiden clustering were subsequently performed. The hydrogel data (not scaled to mean zero, unit variance, and before extraction of highly variable genes) were assigned labels using SingleR (V.1.4.0,⁶⁶) using the Human Primary Cell Atlas Data reference from CellDex (V.1.0.0,⁶⁷).^{66,67} SingleR was used to annotate the clusters with their best-fit match from the cell types in the references. Clusters that yielded cell types other than types of the T-cell lineage were removed from consideration and the process was repeated starting from the batch integration step. The best-fit annotations from SingleR after the second round of clustering and the annotation was assigned as putative labels for each Leiden cluster. Further clustering of transcriptomic data were performed using KMeans in sklearn (V.0.24) with n_clusters set to 8. As the method has a preference to assign like-sized clusters, further consolidation of two central memory clusters was performed.

In order to provide corroboration for the SingleR best-fit annotations and further evidence as to the phenotype of the clusters, gene panels representing functional T-cell categories (Naïve, Effector, Memory, Exhaustion, Proliferation) were used to score each hydrogel's expression profiles using scanpy's 'score_genes' function which compares the mean expression values of the target gene set against a larger set of randomly chosen genes that represent background expression levels. The gene panels for each class were: Naïve - TCF7, LEF1, CCR7; Effector - GZMB, PRF1, GNLY; Memory - AQP3, CD69, GZMK; Exhaustion - PDCD1, TIGIT, LAG3; Proliferation - MKI67, TYMS. The gene expression matrix for all hydrogels was first imputed using the MAGIC algorithm (V.2.0.4,⁶⁸). These functional scores were the only data generated from imputed expression values.

B cell subtyping was performed using previously reported gene sets. The gene panels for each class were: Naïve - IGD, CD23, low CD80, low CD86, low CD95; Memory - low IGD, low CD23, CD21, CD40, CD80, CD86, CD95; Germinal center - PRDM1, IRF4, XBPI; CXCR4+ Centroblasts are subsets of GC B cells that also show an upregulation of BCL6, PAX5, and BACH2; Plasma - CD19-, CD138+.⁶⁹⁻⁷²

HPV calling

A transcriptome reference was built including different genotypes from HPV (hvp16, hvp18, hvp31, hvp33, hvp45, hvp52 and hvp 58). Fastq files from 10x single cell data were pseudo-aligned to this reference using Kallisto V.0.46.0. BUSTools V.0.39.4 was used to error correct barcodes, collapse UMIs, and produce gene count matrices to use for further analysis. HPV counts were log transformed and normalized to tumor content across patients to compare relative expression levels.

Scoring peptide-HLA-TCR interactions

Tetramer data analysis was performed using Python (V.3.7.3). For each single-cell encapsulation, tetramer UMI counts (columns) were matrixed by cell (rows)

and log-transformed. Duplicates of this matrix were independently Z-score transformed by row or column, and subsequently median-centered by the opposite axis (column or row), respectively. For each peptide-HLA-cell interaction, this provided two scores—inter-tetramer (Z_{tet}) and inter-cell (Z_{cell}), which were used to calculate a classifier for unique CDR3 a/b clonotypes across N cells as $N \times \bar{Z}_{tet} \times \bar{Z}_{cell}$. Classifier thresholds for positive interactions were set at 16.

TCR network analysis

TCR motif analysis was performed in Python (V.3.8.5) using TCRdist3 (V.0.2.2) on both alpha and beta chains. Network graphs were assembled using NetworkX (V.2.7.1). Nodes representing alpha (circle) and beta chains (square) were connected by edges to each other within a clonotype and to corresponding chains from other cells when distances were less than or equal to 11 and 23, respectively. Graphs were drawn using a spring layout with spring weights inversely proportional to distances. Once clusters were identified, sequence alignment was performed using the pairwise2 module in Biopython (V.1.78) and visualized using logomaker (V.0.8).

Recombinant TCR validation

TCRs identified from patient samples were ordered from TWIST Biosciences in the pLVX-EF1a lentiviral backbone (Takara) as a bicistronic TCRb-T2A-TCRa vector. Viral supernatants from transfected HEK 293T cells were collected 48 and 72 hours after transfection (MIR6655 TransIT Lentivirus System) and added to the parental TCR $\alpha/\beta^{-/-}$ Jurkat J76 cell line expressing CD8 and a NFAT-green fluorescent protein (GFP) reporter, referred to as J76-CD8-NFAT-GFP. Recombinant TCR surface expression was confirmed throughflow cytometry by staining transduced J76-CD8-NFAT-GFP cells with anti-CD3-PE (Clone UCHT1) and anti-TCR α/β -allophycocyanin (APC) antibodies (Clone IP26).

To assess functional activity of recombinant TCRs, J76-CD8-NFAT-GFP expressing recombinant TCRs were incubated at a 1:1 ratio with the HLA-A*01:01⁺, HLA-B*08:01⁺ or DRB1*07:01 expressing B-LCL cell line, with a final concentration of 0.5% dimethylsulfoxide (vehicle) or 50 μ M of cognate peptide (Vivitide, >95% pure). Cell mixtures were incubated at 37°C, 5% CO₂. At 16 hours, cells were washed and blocked with staining buffer (BD 554656), stained with anti-CD3-PE-Cy7 (Clone UCHT1) and anti-CD69-APC (Clone FN50) antibodies, and analyzed using the Sartorius iQue Screener Plus and FlowJo V.10. CD69 activity was measured as percent positive of CD3⁺ cells.

External data set

Records of 250 patients with HPV+ OPSCC were licensed from Tempus. HLA class I typing for each sample was performed using Optitype to 4-digit resolution on DNA-seq data. HPV typing on each sample was performed using RNA-seq data using Kallisto-DESeq2 workflow. Kallisto

was used for quantifying abundances of HPV genotype specific transcripts followed by count normalization using DESeq2 R package.^{73 74} HPV genotypes include HPV16, HPV18, HPV31, HPV33, and HPV45. Genes include E1, E2, E4, E5, E6, E7, L1 and L2 specific to each genotype. HPV genotype was assigned based on detection of genotype specific HPV E* genes. Specifically, for a given sample at least one E* gene for that genotype must have expression greater than/equal to 0.75 units.

Statistical analysis

Detailed description on data processing, normalization, and handling is included in each of the methods' subsections. Statistical analysis was performed using GraphPad Prism V.9 software and R. Statistical analyses of differences were performed using Shapiro-Wilk tests for normality and Mann-Whitney U tests for gene expression differences between HPV types. Pearson's correlation analysis was run to determine association between HPV gene expression and different cell types.

Twitter Ravi Uppaluri @DrUppaluri

Acknowledgements Phil Bardwell, Shweta Chavan, Conor Dempsey, Jenna Dong, Timothy Drescher, Isabel Goldaracena, William Gordon, Kane Hadley, Kobe Haley, Olivia Huber, Karen Kaye, Gyuna Kim, Taylor Lynch, Justin Mastroianni, Joshua Nickerson, Catarina Nogueira, Tyler Nguyen, Pooja Patel, Amy Perea, Violeta Rayon, Ben Roscoe, Susana Ruiz, Veronica Sanchez, Prashant Shambharkar, Aswathy Sheena, Katherine Triebel and Marcos Wille.

Contributors SB, DL-E and QT performed most of the experiments. CM, SB, DCP and LS composed most of the manuscript. BV and YS designed, executed and drafted critical portions of the analyses. CM, DCP, GJH and AJC designed the overall study. GJH, CTQ, AME and RU contributed to patient accrual sample acquisition and guidance of clinical components. CM is the guarantor.

Funding AME reports support from NIH/NIDCR UO DE029188.

Competing interests CM, BV, QT, DL-E, DCP and AJC are employees and/or stockholders of Repertoire Immune Medicine. GJH reports grants and institutional support from ASCO CCF, Bicara, BMS, Gateway for Cancer Research, GSK, Kite, KSQ, Kura Oncology, ImmunityBio, Regeneron. Consulting/honoraria from Bicara, BMS, Coherus, Excure, Kura, Maverick, Merck, Naveris, Regeneron, and SIRPant. AME reports support from NIH/NIDCR UO DE029188. RU serves on a Merck advisory board.

Patient consent for publication Consent obtained directly from patient(s).

Ethics approval This study involves human participants and was approved by (DF/HCC#09-472), Dana Farber Cancer Institute. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Christine McInnis <http://orcid.org/0000-0002-8449-4954>Ravi Uppaluri <http://orcid.org/0000-0001-5988-6828>

REFERENCES

- Mork J, Lie AK, Glatte E, *et al.* Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N Engl J Med* 2001;344:1125–31.
- Mroz EA, Rocco JW, Math, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* 2013;49:211–5.
- Kürten CHL, Kulkarni A, Cillo AR, *et al.* Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat Commun* 2021;12:7338.
- Powell SF, Vu L, Spanos WC, *et al.* The key differences between human papillomavirus-positive and -negative head and neck cancers: biological and clinical implications. *Cancers (Basel)* 2021;13:5206.
- Gillison ML, Akagi K, Xiao W, *et al.* Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. *Genome Res* 2019;29:1–17.
- Vokes EE, Agrawal N, Seiwert TY. Hpv-associated head and neck cancer. *J Natl Cancer Inst* 2015;107:djv344.
- Ferris RL, Blumenschein G Jr, Fayette J, *et al.* Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med* 2016;375:1856–67.
- Cohen EEW, Soulières D, Le Tourneau C, *et al.* Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. *Lancet* 2019;393:156–67.
- Burtneß B, Harrington KJ, Greil R, *et al.* Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. *Lancet* 2019;394:1915–28.
- Baum J, Seiwert TY, Pfister DG, *et al.* Pembrolizumab for platinum- and cetuximab-refractory head and neck cancer: results from a single-arm, phase II study. *J Clin Oncol* 2017;35:1542–9.
- Ghani L, Chiocca S. High risk-human papillomavirus in HNSCC: present and future challenges for epigenetic therapies. *Int J Mol Sci* 2022;23:3483.
- Longworth MS, Laimins LA. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol Mol Biol Rev* 2004;68:362–72.
- Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond)* 2006;110:525–41.
- MacDonald KS, Fowke KR, Kimani J, *et al.* Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 2000;181:1581–9.
- Ochoa EE, Huda R, Scheibel SF, *et al.* Hla-associated protection of lymphocytes during influenza virus infection. *Virology* 2020;17:128.
- Santegoets SJ, Welters MJP, Schrikke DS, *et al.* The common HLA class I-restricted tumor-infiltrating T cell response in HPV16-induced cancer. *Cancer Immunol Immunother* 2022;16.
- Stevanović S, Helman SR, Wunderlich JR, *et al.* A phase II study of tumor-infiltrating lymphocyte therapy for human papillomavirus-associated epithelial cancers. *Clin Cancer Res* 2019;25:1486–93.
- Draper LM, Kwong MLM, Gros A, *et al.* Targeting of HPV-16+ epithelial cancer cells by TCR gene engineered T cells directed against E6. *Clin Cancer Res* 2015;21:4431–9.
- Norberg SM, Hinrichs CS. Advances in adoptive cell therapy for head and neck cancer. *Otolaryngol Clin North Am* 2021;54:761–8.
- Voskens CJ, Sewell D, Hertzano R, *et al.* Induction of MAGE-A3 and HPV-16 immunity by Trojan vaccines in patients with head and neck carcinoma. *Head Neck* 2012;34:1734–46.
- A phase 1 study in patients with HPV16+ recurrent/ metastatic head and neck squamous cell carcinoma. n.d. Available: <https://clinicaltrials.gov/ct2/show/NCT03978689>
- Chung CH, Douglas Adkins ADC, Gibson MK, *et al.* A phase 1 dose-escalation and expansion study of CUE-101, A novel HPV16 E7-phla-IL2-fc fusion protein, given alone and in combination with pembrolizumab in patients with recurrent/metastatic HPV16+ head and neck cancer. *J Clin Oncol* 2022;40:6045.
- Sousa LG de, Rajapakse K, Rodriguez Canales J, *et al.* ISA101 and nivolumab for HPV-16+ cancer: updated clinical efficacy and immune correlates of response. *J Immunother Cancer* 2022;10:e004232.
- Stevanović S, Pasetto A, Helman SR, *et al.* Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science* 2017;356:200–5.
- Münger K, Baldwin A, Edwards KM, *et al.* Mechanisms of human papillomavirus-induced oncogenesis. *J Virol* 2004;78:11451–60.
- Romanczuk H, Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci U S A* 1992;89:3159–63.
- Balaji H, Demers I, Wuerdemann N, *et al.* Causes and consequences of HPV integration in head and neck squamous cell carcinomas: state of the art. *Cancers (Basel)* 2021;13:4089.
- Eberhardt CS, Kissick HT, Patel MR, *et al.* Functional HPV-specific PD-1+ stem-like CD8 T cells in head and neck cancer. *Nature* 2021;597:279–84.
- Vermorken JB, Mesia R, Rivera F, *et al.* Platinum-Based chemotherapy plus cetuximab in head and neck cancer. *N Engl J Med* 2008;359:1116–27.
- Guo T, Kang SY, Cohen EEW. Current perspectives on recurrent HPV-mediated oropharyngeal cancer. *Front Oncol* 2022;12:966899.
- Guo T, Qualliotine JR, Ha PK, *et al.* Surgical salvage improves overall survival for patients with HPV-positive and HPV-negative recurrent locoregional and distant metastatic oropharyngeal cancer. *Cancer* 2015;121:1977–84.
- Parfenov M, Pedamallu CS, Gehlenborg N, *et al.* Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A* 2014;111:15544–9.
- Ren S, Gaykalova DA, Guo T, *et al.* Hpv E2, E4, E5 drive alternative carcinogenic pathways in HPV positive cancers. *Oncogene* 2020;39:6327–39.
- He J, Xiong X, Yang H, *et al.* Defined tumor antigen-specific T cells potentiate personalized TCR-T cell therapy and prediction of immunotherapy response. *Cell Res* 2022;32:530–42.
- Doering TA, Crawford A, Angelosanto JM, *et al.* Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory. *Immunity* 2012;37:1130–44.
- Khan O, Giles JR, McDonald S, *et al.* Tox transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* 2019;571:211–8.
- Meysman P, De Neuter N, Gielis S, *et al.* On the viability of unsupervised T-cell receptor sequence clustering for epitope preference. *Bioinformatics* 2019;35:1461–8.
- Glanville J, Huang H, Nau A, *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* 2017;547:94–8.
- DeWitt WS 3rd, Smith A, Schoch G, *et al.* Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *Life* 2018;7:e38358.
- Chiou S-H, Tseng D, Reuben A, *et al.* Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery. *Immunity* 2021;54:586–602.
- Jutz S, Leitner J, Schmetterer K, *et al.* Assessment of costimulation and coinhibition in a triple parameter T cell reporter line: simultaneous measurement of NF- κ B, NFAT and AP-1. *J Immunol Methods* 2016;430:10–20.
- Chatfield-Reed K, Gui S, O'Neill WQ, *et al.* HPV33+ HNSCC is associated with poor prognosis and has unique genomic and immunologic landscapes. *Oral Oncol* 2020;100:104488.
- Nagarsheth NB, Norberg SM, Sinkoe AL, *et al.* TCR-engineered T cells targeting E7 for patients with metastatic HPV-associated epithelial cancers. *Nat Med* 2021;27:419–25.
- Trial to test safety and efficacy of vaccination for incurable HPV 16-related oropharyngeal, cervical and anal cancer. n.d. Available: <https://clinicaltrials.gov/ct2/show/NCT02865135>
- Venturi V, Price DA, Douek DC, *et al.* The molecular basis for public T-cell responses? *Nat Rev Immunol* 2008;8:231–8.
- Jacobelli S, Sanaa F, Moyal-Barracco M, *et al.* Anti-HPV16 E2 protein T-cell responses and viral control in women with usual vulvar intraepithelial neoplasia and their healthy partners. *PLoS One* 2012;7:e36651.
- Woo YL, van den Hende M, Sterling JC, *et al.* A prospective study on the natural course of low-grade squamous intraepithelial lesions and the presence of HPV16 E2-, E6- and E7-specific T-cell responses. *Int J Cancer* 2010;126:133–41.
- Rosales R, López-Contreras M, Rosales C, *et al.* Regression of human papillomavirus intraepithelial lesions is induced by MVA E2 therapeutic vaccine. *Hum Gene Ther* 2014;25:1035–49.
- Ma M, Feng Y, Fan P, *et al.* Human papilloma virus E1-specific T cell immune response is associated with the prognosis of cervical cancer patients with squamous cell carcinoma. *Infect Agent Cancer* 2018;13:35.
- Ferdousi J, Nagai Y, Asato T, *et al.* Impact of human papillomavirus genotype on response to treatment and survival in patients receiving

- radiotherapy for squamous cell carcinoma of the cervix. *Exp Ther Med* 2010;1:525–30.
- 51 Zhang S, Wang B, Ma F, *et al.* Characteristics of B lymphocyte infiltration in HPV+ head and neck squamous cell carcinoma. *Cancer Sci* 2021;112:1402–16.
- 52 Ruffin AT, Cillo AR, Tabib T, *et al.* B cell signatures and tertiary lymphoid structures contribute to outcome in head and neck squamous cell carcinoma. *Nat Commun* 2021;12:3349.
- 53 Zhou X, Su Y-X, Lao X-M, *et al.* Cd19 (+) IL-10 (+) regulatory B cells affect survival of tongue squamous cell carcinoma patients and induce resting CD4 (+) T cells to CD4 (+) FOXP3 (+) regulatory T cells. *Oral Oncol* 2016;53:27–35.
- 54 Kräutler NJ, Suan D, Butt D, *et al.* Differentiation of germinal center B cells into plasma cells is initiated by high-affinity antigen and completed by Tfh cells. *J Exp Med* 2017;214:1259–67.
- 55 Morgan D, Tergaonkar V. Unraveling B cell trajectories at single cell resolution. *Trends Immunol* 2022;43:210–29.
- 56 Chera BS, Kumar S, Beaty BT, *et al.* Rapid clearance profile of plasma circulating tumor HPV type 16 DNA during chemoradiotherapy correlates with disease control in HPV-associated oropharyngeal cancer. *Clin Cancer Res* 2019;25:4682–90.
- 57 Chera BS, Kumar S, Shen C, *et al.* Plasma circulating tumor HPV DNA for the surveillance of cancer recurrence in HPV-associated oropharyngeal cancer. *J Clin Oncol* 2020;38:1050–8.
- 58 Reynisson B, Alvarez B, Paul S, *et al.* 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* 2020;48:W449–54.
- 59 Altman JD, Davis MM. MHC-peptide tetramers to visualize antigen-specific T cells. *Curr Protoc Immunol* 2016;115:17.
- 60 Rodenko B, Toebe M, Hadrup SR, *et al.* Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat Protoc* 2006;1:1120–32.
- 61 Bakker AH, Hoppes R, Linnemann C, *et al.* Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -a3, -A11, and -B7. *Proc Natl Acad Sci U S A* 2008;105:3825–30.
- 62 Frøsig TM, Yap J, Seremet T, *et al.* Design and validation of conditional ligands for HLA-B*08:01, HLA-B*15:01, HLA-B*35:01, and HLA-B*44:05. *Cytometry A* 2015;87:967–75.
- 63 Rabuka D. Chemoenzymatic methods for site-specific protein modification. *Curr Opin Chem Biol* 2010;14:790–6.
- 64 Rashidian M, Dozier JK, Distefano MD. Enzymatic labeling of proteins: techniques and approaches. *Bioconjug Chem* 2013;24:1277–94.
- 65 Hao Y, Hao S, Andersen-Nissen E, *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–87.
- 66 Aran D, Looney AP, Liu L, *et al.* Reference-Based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019;20:163–72.
- 67 Mabbott NA, Baillie JK, Brown H, *et al.* An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics* 2013;14:632.
- 68 van Dijk D, Sharma R, Nainys J, *et al.* Recovering gene interactions from single-cell data using data diffusion. *Cell* 2018;174:716–729.
- 69 Caron G, Le Gallou S, Lamy T, *et al.* Cxcr4 expression functionally discriminates centroblasts versus centrocytes within human germinal center B cells. *J Immunol* 2009;182:7595–602.
- 70 Liu YJ, Barthélémy C, de Bouteiller O, *et al.* Memory B cells from human tonsils colonize mucosal epithelium and directly present antigen to T cells by rapid up-regulation of B7-1 and B7-2. *Immunity* 1995;2:239–48.
- 71 Tangye SG, Liu YJ, Aversa G, *et al.* Identification of functional human splenic memory B cells by expression of CD148 and CD27. *J Exp Med* 1998;188:1691–703.
- 72 Ellyard JI, Avery DT, Phan TG, *et al.* Antigen-selected, immunoglobulin-secreting cells persist in human spleen and bone marrow. *Blood* 2004;103:3805–12.
- 73 Bray NL, Pimentel H, Melsted P, *et al.* Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;34:525–7.
- 74 github. Available: <https://github.com/mikelove/DESeq2>

