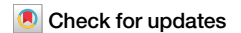


<https://doi.org/10.1038/s43856-026-01401-3>

A highly resolved integrated single-cell atlas of HPV-negative head and neck cancer



Lina Kroehling ^{1,2}, Andrew Chen ^{1,2}, Anthony Spinella ³, Eric Reed⁴, Maria Kukuruzinska ⁵,
Xaralabos Varelas ³ & Stefano Monti ^{1,2,6}

Abstract

Background Head and Neck Squamous Cell Carcinomas (HNSCC) are the seventh most prevalent form of cancer and are associated with human papilloma virus infection (HPV-positive) or with tobacco and alcohol use (HPV-negative). HPV-negative HNSCCs have a high recurrence rate, and individual patients' responses to treatment vary greatly due to the high level of cellular heterogeneity of the tumor and its microenvironment.

Methods Here, we describe a HPV-negative HNSCC single cell atlas, which we created by integrating six publicly available datasets encompassing over 230,000 cells across 54 patients. We classify cell types, subpopulations, and their expression programs in the immune, mesenchymal, endothelial and epithelial compartments. We interrogate the relationship between cell types through hierarchical clustering, cell-cell communication analysis and correlating populations changing together across patients.

Results We resolve the myeloid and fibroblast compartments, revealing an *IL1B*⁺ myeloid population previously unexplored in HNSCC and clarifying two immune cancer associated fibroblast populations that are frequently conflated, identify sex-associated changes in cell type proportions, and a unique interaction between CXCL8-positive fibroblasts and vascular endothelial cells.

Conclusions We utilize the atlas to contextualize the relationships between existing signatures and cell populations, harmonize nomenclature across studies, and show the power of this large-scale resource to robustly identify associations between transcriptional signatures and clinical phenotypes that would not be possible to discover using fewer patients. Beyond our findings, the atlas serves as a public resource for the high-resolution characterization of tumor heterogeneity of HPV-negative HNSCC.

Plain Language Summary

HPV-negative head and neck squamous cell carcinomas (HNSCC) are a type of cancer that develops in and around the mouth and is often linked to smoking and drinking alcohol. They have a high recurrence rate and respond differently to treatment due to the high level of variation between patients in both their tumor cells and surrounding cells. In this study, we combined cell-level data from many people with HPV-negative HNSCC to create a large, unified dataset of the cell types found in these tumors. Using computational analyses, we identified and characterized the different cell types and their features. We then related those cellular features to how the tumors behaved in different patients. These results advance our understanding of how particular combinations of tumor and surrounding cells can influence cancer growth and treatment response. This knowledge could help guide more precise, personalised approaches to treating HNSCC.

Head and Neck Squamous Cell Carcinomas (HNSCCs) arise from the mucosal epithelium lining of the oral cavity, larynx, and pharynx^{1,2}. The two major subtypes are HPV-positive and HPV-negative HNSCC, the former more frequently occurring in the pharynx and associated with human papilloma virus infection, especially the HPV-16 strain^{3,4}, and the latter more frequently occurring in the oral cavity and larynx and associated with carcinogens in alcohol, tobacco and betel quid, among others.

Amongst all cancers, HNSCCs have the seventh highest incident rate, accounting for almost 7% of cancers worldwide, with 75% being HPV-negative^{3,5,6}. According to the GLOBOCAN database, cancers of the lip and oral cavity are the leading cause of cancer death for men in India⁶. HPV-negative cancers tend to have higher mutational burdens than HPV-positive cancers and lower levels of immune infiltration in the tumor micro-environment (TME). Together, these factors result in clinicopathological

¹Bioinformatics Program, Faculty of Computing and Data Science, Boston University, Boston, MA, USA. ²Section of Computational Biomedicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA. ³Department of Biochemistry and Cell Biology, Boston University Medical Center, Boston, MA, USA. ⁴Department of Medicine, Albert Einstein College of Medicine, New York, NY, USA. ⁵Department of Molecular and Cell Biology, Department of Translational Dental Medicine, Boston University Medical Center, Boston, MA, USA. ⁶Department of Biostatistics, School of Public Health, Boston University, Boston, MA, USA. ✉e-mail: smonti@bu.edu

differences and a significant difference in 5-year survival rates between the two subtypes, with a 75–80% rate for HPV-positive HNSCCs and a 55% rate for HPV-negative HNSCCs^{4,7–9}. Consequently, the two subtypes have now been separated in the American Joint Committee of Cancer (AJCC) staging system and are commonly considered distinct diseases^{7,10}.

Current treatments for the disease include radiation, chemotherapy, combination therapy, and immune checkpoint inhibitors (ICIs). However, response to treatment as well as mode of effective treatment varies greatly across individuals due to the high degree of heterogeneity both in the genetic profiles and plasticity of the cancer cells, as well as in the cell type composition of the TME, even within HPV-negative HNSCCs^{11,12}. HNSCC patients have a poor overall response to ICI therapy, with less than 20% presenting a durable response^{11,13}.

Significant progress has been made in characterizing the HNSCC TME. HPV-negative tumor epithelial cells are commonly characterized by tumor suppressor inactivating mutations in genes *TP53*, *CDKN2a*, and *PTEN*, and amplifications in *CCND1*, which together enable the progression of the disease^{4,14}. Previously identified transcriptional signatures in tumor cells include response to hypoxia, stress, epithelial-to-mesenchymal transition (EMT), and partial EMT (pEMT), a molecular phenotype in which tumor cells become more plastic to acquire a mesenchymal-like state and enhanced migratory capabilities¹⁵. Additionally, immune and stromal cells present in the TME play a major role in affecting tumor growth through TME interactions. For example, cancer-associated fibroblasts (CAFs) have been shown to affect tumor cell migration by remodeling the extracellular matrix¹⁵. Natural killer (NK) and CD8 + T cells have been found to take on effector phenotypes and kill tumor cells, whereas myeloid-derived suppressor cells (MDSCs) and Tregs secrete cytokines that suppress the immune system⁴. Further delineation of these complex dynamics in the non-HPV TME is critical for discovering new potential avenues of therapeutic intervention that may drive heterogeneous patient responses to ICI.

Single-cell RNA-sequencing (scRNA-seq) has emerged as the technology of choice to profile the individual transcriptomes of tumor cells and the surrounding TME. scRNA-seq enables the characterization of tumor heterogeneity within and across patients through the identification of *cell types* – here defined as categories of cells that perform a specific function (e.g., T cell, Fibroblast), and specific *cell states* – here defined as transient molecular changes in gene expression in response to endogenous and exogenous cues that multiple cell types can share (e.g., dysfunction, proliferation)^{15–18}. While individual single-cell HNSCC datasets have produced invaluable findings, given the relatively small sample size of single studies, and the relatively small number of publicly available single-cell HNSCC datasets, we hypothesized that their integration and combined analysis would facilitate contextualization of findings from individual studies, and provide increased statistical power to identify novel associations between molecular and clinical phenotypes¹⁹. Other HNSCC atlases have been previously developed, however the focus has been on a single cell type^{20,21}, contrasting the HPV-positive and HPV-negative subtypes¹⁶, or on the stepwise progression of the disease^{22,23}, leaving a need for the comprehensive characterization of the HPV-negative subtype.

To this end, we have integrated six publicly available datasets (Puram et al.¹⁵, Cillo et al.²⁴, Kurten et al.²⁵, Peng et al.²⁶, Choi et al.²³, Quah et al.²⁷) for the high-resolution characterization of the cells in the TME of HPV-negative HNSCC. Creation of the atlas allowed us to perform four types of analyses. *Contextualization*, in which we taxonomically organize different

cell types and identify how closely related previously identified populations are to each other, and identify cell populations whose proportional changes are associated across patients. *Harmonization*, in which previously described populations with the same characteristics but differently labeled in the individual studies are assigned a consistent nomenclature. *Identification and characterization of subpopulations and interactions*, in which we leverage the increased number of cell profiles in the atlas to identify and annotate cell types and predict interactions. *Association with clinical phenotypes*, in which the increased sample size of the atlas increases the statistical power to detect significant associations between cell type compositional changes or cell state signatures, and clinical phenotypes such as stage and sex. In summary, we leverage published data to create and annotate an HPV-negative HNSCC transcriptomic atlas comprised of over 50 patients and 230,000 cells that may serve as a public resource. The atlas integrates six existing datasets to produce a comprehensive landscape of HNSCC tumors and their micro-environment, supporting novel findings, providing a reference for generating and contextualizing new hypotheses, and connecting molecular phenotypes to clinical end points.

Methods

Data acquisition

Six datasets were integrated in this study, totaling 54 patients and 232,015 cells. For five datasets, the processed matrix files containing raw counts and gene and barcode tsv files were downloaded from GEO: Peng²⁶ (GSE172577), Choi²³ (GSE181919), Cillo²⁴ (GSE139324), Quah²⁷ (GSE225331), and Kurten²⁵ (GSE164690). Due to availability, processed normalized data were downloaded for the last dataset, Puram¹⁵ (GSE103322). HPV-positive samples were removed to retain only HPV-negative samples for use in the atlas. The studies were approved by the following Institutional Review Boards. Puram: Patients at the Massachusetts Eye and Ear Infirmary (MEEI) were consented preoperatively to take part in the study following Institutional Review Board approval (Protocol #11-024H). Quah: The study of patient tumor samples was approved by SingHealth Centralized Institutional Review Board (CIRB: 2014/2093, 2018/2512 and 2016/2757) with each patient’s written consent. Choi: All the experiments with patient samples were performed under the approval of the Ajou University Institutional Review Board, using the approved protocol AJIRB-BMR-SMP-18-150 with the written informed consent of all patients. Cillo: All patients provided informed written consent, and the study was approved by the Institutional Review Board (University of Pittsburgh Cancer Institute, Protocol 99-069). Kurten: For all human patient samples, informed written consent was obtained prior to donation. The University of Pittsburgh Cancer Institute Review Board (Protocol 99-069) approved the study. Peng: The study was approved by the Institutional Review Board of Sun Yat-Sen Memorial Hospital, China (No. SYSEC-KY-KS-2021-040). Informed written consent was obtained from each participant.

Dataset pre-processing

Before integration, each dataset was pre-processed individually using *Seurat* v4²⁸ (RRID:SCR_007322) to eliminate low-quality cells and annotate the cells by cell type. If not pre-filtered in the originating study, cells with fewer than a minimum number of genes detected per cell (nFeatures ≤ 200) or more than a maximum percent of reads coming from mitochondrial genes (≥ 20%) were filtered out, as summarized in Table 1. These cutoffs were

Table 1 | Processing Steps for Pre-Integration

Dataset	Puram	Kurten	Cillo	Peng	Quah	Choi
Individual Dataset Pre-processing						
Filtering	Published data prefiltered	Filtered	Filtered	Filtered	Published data prefiltered	Published data prefiltered
nGenes/Cell	2000, or average housekeeping expression level <2.5	200	200	200	200	200
Percent mito	NA	20%	20%	20%	20%	10%

largely consistent with those chosen by the individual studies' authors. We retained all cells for Puram, Quah, and Choi, which were already prefiltered.

To annotate cells, each dataset was normalized by library size and log transformed using Seurat's *NormalizeData* function, except for Puram, which was already normalized. Data was clustered using the Louvain algorithm with 2000 variable features and 30 PCs at a resolution of 0.3. Cells were annotated at a single-cell level using *SingleR* v2.6.0²⁹ (RRID:SCR_023120) with both the *Blueprint*^{30–32} (RRID:SCR_003844) and *Human Primary Cell Atlas (HPCAD)*^{32,33} databases, accessed through the *celldex* v1.13.3 package³². Each cluster was assigned a broad cell label (Epithelial, Endothelial, Fibroblast, T/NK, Immune, B Cell, or Mast), based on the consensus majority label amongst Blueprint and HPCAD, and when available, the paper-provided annotations.

Integration

Integration was performed via Seurat's reciprocal principal component analysis (rPCA) method³⁴. For each pre-processed dataset, *SCTransform* was used to normalize the raw counts using *method* = "glmGamPoi", except for Puram, in which the already normalized counts were used. *SelectIntegrationFeatures* was used to identify 3000 integration features (genes highly variable amongst all datasets). The *FindIntegrationAnchors* function was used to identify 20 anchors in 30 dimensions. The integration was then performed using *IntegrateData*, producing a single Seurat object containing two slots; an integrated slot which contains integrated-corrected counts for the 3000 integration features, used only for clustering, and an RNA slot containing the raw counts for all genes. The raw data in the RNA slot was normalized by library size and log transformed using *NormalizeData*, as well as saved to a new SCT slot in which it was normalized using *SCTransform*, with *vars.to.regress* = "dataset" (again, the pre-normalized Puram data was copied into the SCT slot and not renormalized). After integration, dimensionality reduction and clustering were performed on the integrated slot using 30 PCA dimensions, using the Louvain algorithm, at a resolution of 0.4. *scDblFinder* v1.21.2 (RRID:SCR_022700) was used to identify doublets using the random based method and *db.r.ds* = 1. No cluster contained more than 20% doublets.

Signature scoring

Transcriptional signatures were collected from multiple sources (Table 2 Supplementary Data 2). Select genesets from the KEGG³⁵ (RRID:SCR_012773), Hallmarks³⁶, and Reactome³⁷ (RRID:SCR_003485) compendia were also used. Signatures of HNSCC tumor core versus leading edge were compiled from Arora et al.³⁸.

For all gene expression analyses, the normalized data slot of the RNA assay was used. In each compartment, cells were scored for a given signature using Seurat's *AddModuleScore*, whereby the average expression level of the genes in the signature is computed at the single cell level, subtracted by the aggregated expression of randomly selected control feature³⁹. Thus, taking as input the gene-by-cell normalized expression matrix and a list of signatures, signature scoring outputs a signature-by-cell matrix *S*, with entry *s_{ij}* representing the signature score for a signature *i* in cell *j*.

Analysis of variance (ANOVA) was performed on the signature scores across T cell clusters to test for significant differences between clusters. Post-hoc pairwise comparisons were conducted using Tukey's Honest Significant Difference (TukeyHSD) test using the *stats* v4.4.0 (RRID:SCR_025678) package to identify differences between specific clusters.

Creation of effector cell score and correlation with stage

All T cells were scored using the dysfunctional and cytotoxic signatures from Luen et al.⁴⁰ using the same signature scoring procedure described above. The dysfunctional state is characterized both by increased expression of immune inhibitory receptors such as *TIGIT* and *LAG3* and loss of cytotoxic genes such as *IFNG* and *IL2*⁴¹. It's been shown that T cells can express genes that are characteristic of both the cytotoxic and dysfunctional states simultaneously, suggesting a continuous transition between these states as opposed to a binary switch. Because of this, we created an additional score to

Table 2 | Signatures used in atlas

Publication	Derivation	Signature Type	nSignatures
Szabo et al.	Pan-cancer	T cell	7
Chu et al.	Pan-cancer	T cell	9
Luen et al.	Pan-cancer	T cell	4
Rose et al.	Human blood	Memory T cell	2
Mulder et al.	Cross-tissue	Myeloid	7
Coulton et al.	Pan-cancer	Myeloid	1
Luoma et al.	HNSCC	Myeloid	1
Wei et al.	Pan-cancer	Myeloid	2
Coulton et al.	Pan-cancer	Myeloid	1
Cheng et al.	Pan-cancer	Myeloid	15
Ma et al.	Pan-cancer	Fibroblast	1
Cho et al.	Pan-cancer	Fibroblast	11
Cords et al. Review	Pan-cancer	Fibroblast	5
Cords et al.	BrCa	Fibroblast	2
Puram et al.	HPV + , HPV- HNSCCs	Epithelial states, TME, Cell Types	9,43,22
Arora et al.	Spatial HNSCC	Epithelial	3
Kegg		General	
Hallmarks		Tumor	
Reactome		General	

infer the dominant state of the T cells that we termed 'effector capacity score' by subtracting the dysfunctional score from the cytotoxic score for each cell, so that the higher the score, the more purely cytotoxic the cell is assumed to be. CD8+T cells were selected by taking clusters 0, 1, 8, 15, and 23 from the immune compartment clustered at resolution 0.8. The relationship between the effector capacity score and stage was determined using a mixed effect model using *lme4*⁴² (RRID:SCR_015654), with dataset and patient as nested random effects across the CD8+T cells.

$$\text{Effector Capacity Score} \sim \text{Stage} + (1|\text{dataset}) + (1|\text{dataset} : \text{patient})$$

The association of effector capacity score with stage was also performed separately within each dataset using the same model, but with just patient as a random effect.

$$\text{Effector Capacity Score} \sim \text{Stage} + (1|\text{patient})$$

Identification of marker genes

Identification of marker genes per cluster was performed using Seurat's *FindAllMarkers* with the parameters *test.use* = "MAST" (RRID:SCR_016340), and *latent.vars* = "dataset" to control for dataset as a fixed effect⁴³.

Gene set enrichment analyses

Gene set enrichment analyses were performed by hypergeometric testing with the KEGG³⁵ (RRID:SCR_012773), Hallmarks³⁶, WikiPathway⁴⁴ (RRID:SCR_002134), and Biocarta⁴⁵ (RRID:SCR_006917) compendia using *hyper* v2.4.0⁴⁶, with all genes in the atlas used as the background.

Taxonomic organization of cell type clusters

K2Taxonomer v1.0.7⁴⁷ was used to elucidate the relationship between different clusters and define transcriptional programs enriched at data-driven cluster subgroups. K2Taxonomer was applied separately to the immune and the epithelial compartments, using the scaled, integrated slot of the dataset. The preprocessing step was performed using *featMetric* = "F", *nBoots* = 400, and *clustFunc* = *cKmeansWrapperSubsample*. The K2 dendrogram was

Table 3 | Summary of datasets used in atlas

Dataset	GEO	# Patients HPV(-)	# Cells	Sorting	Age Range	Sex (#F/#M)	LNM (primary/LNM)	T Stage (1/2/3/4)
Puram	GSE103322	10	4199		41–88	4/6	5/5	1/3/2/4
Kurten	GSE164690	12	80523		35–80–66	5/7	12/0	1/2/7/2
Cillo	GSE139324	6	20692	CD45+ only	43–66	1/5	6/0	1/1/1/3
Peng	GSE172577	6	57984		32–72	2/4	6/0	0/2/2/2
Quah	GSE225331	7	53459		18–70	2/5	0/7	0/0/4/3
Choi	GSE181919	13	15158		53–77	4/9	12/1	2/8/0/3
Total		54	232015		18–88	18/36	49/13	5/15/16/18

produced using the main *K2tax* function with default settings. The normalized RNA slot data was used to perform downstream differential expression and gene set enrichment analyses, with the dataset source included as a covariate.

Correlation of cell type proportions across patients

The correlations of cell type proportions were calculated on the immune (after removal of PBL cells) and non-immune compartments separately to identify clusters manifesting coordinated proportional changes across patients. Per patient cluster proportions were estimated as the number of cells in a cluster divided by the total number of cells in the given compartment. Spearman correlation was calculated among all clusters across all patients. FDR corrections were applied to the lower-triangle of the correlation matrix. Correlations with a FDR < 0.05 and absolute value of the coefficient greater than 0.3 were considered significant.

Inference of copy number variations

*InferCNV v1.20.0*⁴⁸ (RRID:SCR_021140) was used to detect copy number variations (CNVs) in the epithelial cells, and was run on each dataset separately. Epithelial cells were considered cancer cells and used as ‘observations’. A set of 1000 cells randomly sampled from T cells, B cells, and vascular endothelial cells were used as reference cells. As a negative control, a subset of 100 of the 1000 reference cells was removed and added to the observation set of epithelial cells in which CNVs should not be found. The gene position file hg38_gencode_v27.txt was used for gene locations. Raw counts were used as input, and *inferCNV* was run using *cutoff* = 0.1, *cluster_by_groups* = TRUE, *denoise* = TRUE and *HMM* = FALSE.

The output of *InferCNV* is a gene-by-cell matrix containing CNV scores centered at one, where a value less than one corresponds to a loss in copy number and a number above one corresponds to a copy number gain. A malignancy score for each cell was calculated by subtracting one from the copy number score to center at zero, taking the absolute value of the sum of scores per cell, and dividing by the number of genes.

$$\text{Malignancy Score} = \frac{\sum |\text{CNVscore} - 1|}{n\text{Genes}}$$

To identify if epithelial and salivary cells were cancer cells or not, we calculated the top 10% of the malignancy scores of the normal control cells for each dataset. All epithelial and salivary cells that lie within or above this top 10% were defined as cancer cells, and all those below were defined as normal cells.

Inference of cell potency

*CytoTrace2 v1.0.0*⁴⁹ was run separately on each dataset, and on the immune and non-immune compartments separately. The counts slot of the RNA assay was used as input. The final potency annotation was added to the atlas for visualization across datasets.

Detection of cell type proportion changes across sex

*Sccomp v1.99.18*⁵⁰ was used to detect changes in cell type proportions between males and females. *Sccomp* was applied to the immune and non-

immune compartments of the atlas separately to avoid artefacts arising from CD45+ and CD45- cell being sorted and sequenced separately which was the case in many of the individual datasets. *Sccomp* was first applied to each compartment using all datasets in the atlas. Composition and variability models were both included, where dataset was included as a random effect in both cases, and with *bimodal_mean_variability_association* = TRUE. *Sccomp* was then applied to the immune and non-immune compartments on each dataset separately using the same parameters, but without including dataset as a random effect.

Cell cell communication analyses

*CellChat v2.1.2*⁵¹ (RRID:SCR_021946) was used to perform cell cell communication analyses on the RNA slot between clusters as specified in Supplementary Fig 7a, b. The communication probability pathway was computed using the “truncatedMean” method with a cutoff of 0.15, requiring a minimum of 10 cells. A second *CellChat* analysis focused on signaling between the specific myeloid clusters was performed which used the clustering specified in Supplementary Fig 7a, b, but in which the macrophage and mac/mono/neutrophil clusters were resolved at a higher resolution as detailed as the six clusters using the “trimean” method. Lastly, a *CellChat* analysis focused on signaling occurring between the fibroblast populations was performed using the clustering shown in Supplementary Fig. 7a, b, but in which clusters Fib1, Fib2, and Fib3 were resolved at the higher resolution of 0.7, again using the “trimean” method.

Statistics and reproducibility

95% confidence intervals were used for all linear regression models. FDR-corrected hurdle p-values were used for all differential expression testing through *MAST* (RRID:SCR_016340).

Analysis of variance (ANOVA) was performed on the signature scores across T cell clusters to test for significant differences between clusters. Post-hoc pairwise comparisons were conducted using Tukey’s Honest Significant Difference (TukeyHSD) test to identify differences between specific clusters.

GLMs were used to model the effector score with respect to stage. FDR-corrected p-values < 0.05 were used to identify significant associations.

For cell type proportion correlation across patients, spearman correlations were used. P-values from the AS 89 algorithm were FDR-corrected, and significant correlations were defined as those with FDR < 0.05 and the absolute value of the coefficient > 0.3.

Cell-cell communication analyses used empirical p-values in dotplots showing specific ligand-receptor interactions between cell types. P-values are calculated through the *CellChat* package as the fraction of (cell type label) permuted interaction scores that are as large or larger than the original score.

FDR q-values are used for all differential cell type abundance analyses calculated with the *sccomp* package.

Results

Generation of an integrated atlas

Six publicly available datasets of human HPV-negative HNSCC tumors were integrated, encompassing 54 patients and a total of 232,015 cells (Table 3). Aggregated patient metadata is provided in Supplementary Data 1. All patients were treatment-naïve, and cells were derived from

primary tumors and, in some cases, from local lymph node metastases. The integrated UMAP demonstrated both effective dataset mixing and preservation of cell-type classification: cells assigned to a given cell type prior to integration clustered together after integration (Fig. 1a, b). Clusters comprised cells across sexes, anatomical sites, ages, disease stages, and datasets (Supplementary Fig 1a, b). Despite known biological differences between anatomical sites⁵², cells from the larynx and pharynx were well-mixed with those from the oral cavity, which contributed the majority of cells (Supplementary Fig 1a). Canonical marker genes for broad cell types showed specificity in their expression patterns (Fig. 1c). The datasets varied substantially in the number of cells recovered from each patient, reflecting both biological and technical factors. For example, the Puram dataset, generated with Smart-seq2, yielded the fewest cells while the other datasets employed 10X Genomic platforms. The distribution of recovered cell types also differed between patients, likely due to tumor microenvironment heterogeneity as well as dataset-specific sampling strategies (e.g., Cillo and Kurten also profiled peripheral blood lymphocytes). (Fig. 1d and Supplementary Fig 1b).

Separation of immune and non-immune compartments

Because some datasets were sorted prior to integration (e.g., Cillo et al. contained only CD45+ cells), we divided the atlas into the immune and non-immune compartments for downstream analyses. To this end, we first performed global clustering on the entire integrated dataset at a resolution of 0.4 and identified major cell populations within. The non-immune compartment was defined as the union of cells classified as adipocytes, endothelial cells, epithelial cells, fibroblasts, and muscle cells. The immune compartment was defined as the union of cells classified as T cells, NK cells, B cells, Myeloid (labeled broadly as “immune”), and Mast cells. The resulting immune and non-immune subsets were reclustered at resolutions of 0.8 and 0.2, respectively (Supplementary Fig 2).

Immune compartment

Within the immune compartment we identified 30 clusters which we classified into five broad cell types: myeloid, T/NK cells, B cells, Plasma cells, and Mast cells (Fig. 2a, b). This classification was based on a consensus annotation that integrated singleR classification with canonical marker gene expression. Expression of *CD14* and *CD16* (*FCGR3A*) separated monocytes from macrophages within the myeloid group, while *CD4*, *CD8A*, and *CD160* expression clearly separated CD4 + T cells, CD8 + T cells and NK cells within the lymphoid cell cluster (Fig. 2b). We further identified plasma cells (*IGHG1*-positive), B cells (*CD20/MS4A1*-positive), and Mast cells (*TSAB1*-positive). Enrichment analysis of marker genes and signatures further resolved clusters corresponding to cDCs (clusters 14, 25, 26), pDCs, (19), macrophages (5, 7), and monocytes (2, 17, 24) (Supplementary Fig 3a, b).

Annotation of the T cell compartment and association of cellular programs with clinical phenotypes

To resolve cell subtypes higher resolution within the broad T/NK cell group, we scored T cells using published signatures (Supplementary Data 2) and evaluated their average expression at the cluster level, as described in Methods (Supplementary Fig 3c). This analysis identified clusters corresponding to naïve T cells, T follicular helper cells (Tfh), regulatory T cells (Tregs), and central memory T cells (Tcm) within the CD4+ compartment, and cytotoxic, central memory, and dysfunctional cells within the CD8+ compartment, alongside populations of NK and proliferating T cells (Fig. 2a–c and Supplementary Fig 3d).

Naïve T cells were represented by clusters 4, 9, 10, 20 and 21; Tregs by cluster 6; and Tfh cells by cluster 12, each expressing *CD4* (Supplementary Fig 3c). Across published studies, the naïve population has been referred to by various terms – such as naïve (Choi²³), CD4+ (Kurten²⁵), naïve-like (Quah²⁷), conventional T-helper (Puram¹⁵, Cillo²⁴), and TCF7+ cells (Peng²⁶) – despite expressing a consistent set of genes (*IL7R*, *LTB*, *CCR7*, *SELL*, *TCF7*).

Natural Killer (NK) and CD8 + T cells fall on a trajectory from cytotoxicity to dysfunction

Next, we focused on CD8+ and NK cells. Clusters 13 and 16 showed significantly higher expression of the NK cell signature compared to all other clusters (Supplementary Fig 3c and Supplementary Data 3). Clusters 0, 1, 8, and 23 were identified as CD8 + T cell clusters. Each cell was further annotated by *dysfunctional* and *cytotoxic* scores as previously described⁵³. CD8 + T cells mediate cytotoxic killing of tumor cells but can transition to a dysfunctional state following prolonged stimulation by tumor antigens. This state is characterized by reduced cytokine production and effector killer function, alongside upregulation of immune checkpoint markers such as *LAG3*, *TIGIT*, *CTLA4*, and *PD-L1*⁵⁴. To model this transition, we calculated an overall effector capacity score by subtracting the dysfunctional score from the cytotoxic score, with higher values indicating greater cytotoxicity and reduced dysfunction. NK cells (clusters 13, 16) exhibited the highest cytotoxic score, consistent with their known cytotoxic role, followed by CD8 + T cell clusters 0, 8, and 23. Notably, CD8 + T cell cluster 0 displayed the highest dysfunctional score, reflecting a multi-state phenotype characterized by expression of both cytotoxicity and dysfunction programs.

The balance of T cell cytotoxicity and dysfunction is associated with tumor stage

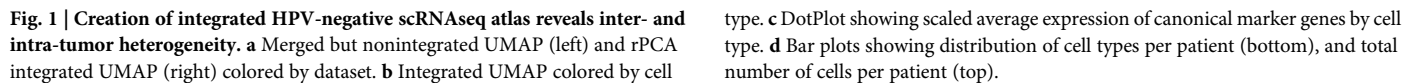
We investigated the relationship between the T cell effector capacity score (cytotoxic minus dysfunctional) and tumor stage. After isolating the CD8 + T cells, we applied a linear mixed effects model to regress the effector capacity score on stage. We found a significant positive association with stage T1 ($t = 0.318$, FDR = 0.018), suggesting a cytotoxic T cell response at the earliest disease stage. (Supplementary Fig 3e, Supplementary Data 4). When performing the same analysis within the individual datasets, a significant association with stage T1 is not achieved. Within the Peng dataset there was an almost significant ($t = 0.326$, FDR = 0.06) association between the effector capacity score and stage T2, which was not observed in other datasets. This highlights the utility of the large sample size in the atlas to either boost or temper confidence in identifying associations between cellular programs and clinical phenotypes.

Contextualization of immune cell populations

We next sought to elucidate the relationship between cell-type annotated clusters within the immune compartment (Fig. 2c).

CytoTRACE2 is an interpretable AI method for predicting cellular differentiation potential from scRNA-seq data originally developed in a developmental biology context. We applied CytoTRACE2 to the immune compartment and found that most immune cells were in a fully differentiated state (Fig. 2d). In contrast, naïve T cells and proliferating T cells were classified as unipotent or oligopotent, reflecting their ability to differentiate into various T cell subtypes, such as effector and memory cells, upon antigen stimulation⁴⁹. Interestingly, some Treg cells were also annotated as unipotent. This finding is consistent with previous studies showing that Tregs exhibit high plasticity and may switch between states in response to different cues^{55,56}.

Application of K2Taxonomer (K2T) to the immune cell compartment revealed the taxonomic relationships between the clusters therein⁴⁷ (Fig. 2e). The top-level partition broadly separated myeloid and lymphoid subsets, with the exception of mast cells, which clustered with lymphoid cells, likely due to their lack of expression of some myeloid markers, such as *LYZ* and *LST1*. The four monocyte clusters (2, 7, 17, 24, and 29) formed a hierarchical group. Cluster 5 (macrophages) was most closely related to cDC clusters 14, 22, and 26, due to elevated expression of complement genes shared by macrophages and cDCs relative to other immune cells (Supplementary Fig 4a,b). Within the lymphoid lineage, T cells and NK cells grouped hierarchically and further segregated by cell states: naïve (4, 9, 10, 20, 21), cytotoxic (8, 13, 16, 23), and dysfunctional (0).



type. **c** DotPlot showing scaled average expression of canonical marker genes by cell type. **d** Bar plots showing distribution of cell types per patient (bottom), and total number of cells per patient (top).

Interrogation of the top genes enriched in each cluster revealed two clusters marked by high expression of canonical macrophage markers *CIQB*, *CD4*, and *CD163* (Fig. 3b, c and Supplementary Data 6). One cluster was defined by *CXCL9* expression and enriched

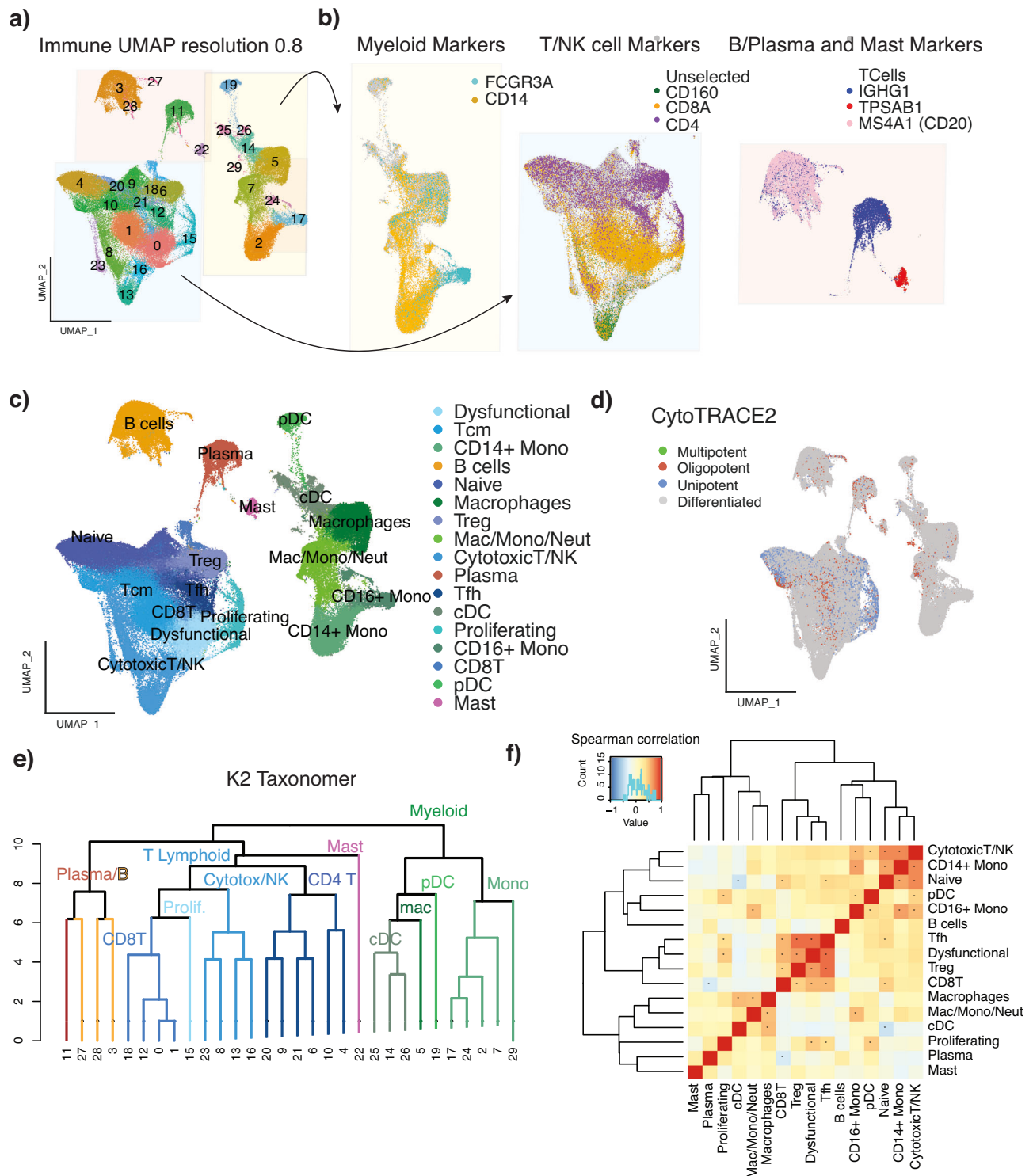


Fig. 2 | Characterization of cell types in the immune microenvironment.

a Immune compartment UMAP clustered at resolution 0.8, across 53 patients. **b** Identification of broad cell types by expression of marker genes. *FCGR3A* and *CD14* expression on myeloid cells (left), *CD160*, *CD8A* and *CD4* expression on T and NK cells (center), and *IGHG1*, *TPSAB1* and *MS4A1* expression on B cells, plasma cells, and mast cells (right). Cells highlighted are those expressing more than the average gene expression across the immune compartment for the given gene. **c** UMAP of immune compartment with cluster names based on signature enrichment. **d** UMAP of immune compartment colored by CytoTRACE2 potency

classifications. **e** K2 Taxonomer dendrogram of immune compartment clusters at resolution 0.8. **f** Heatmap of correlation of immune cell type proportions across patients. Spearman correlation was calculated on 16 sample-level immune sub-cluster frequencies. Positive co-occurrence patterns are shown in red colors and negative co-occurrence patterns are shown in blue colors, with color intensity proportional to the spearman correlation coefficient. Asterisks indicate statistical significance based on FDR-adjusted two-sided p values $P < 0.05$ from the AS 89 algorithm, and $|R| \geq 0.3$. n Patients used is 53.

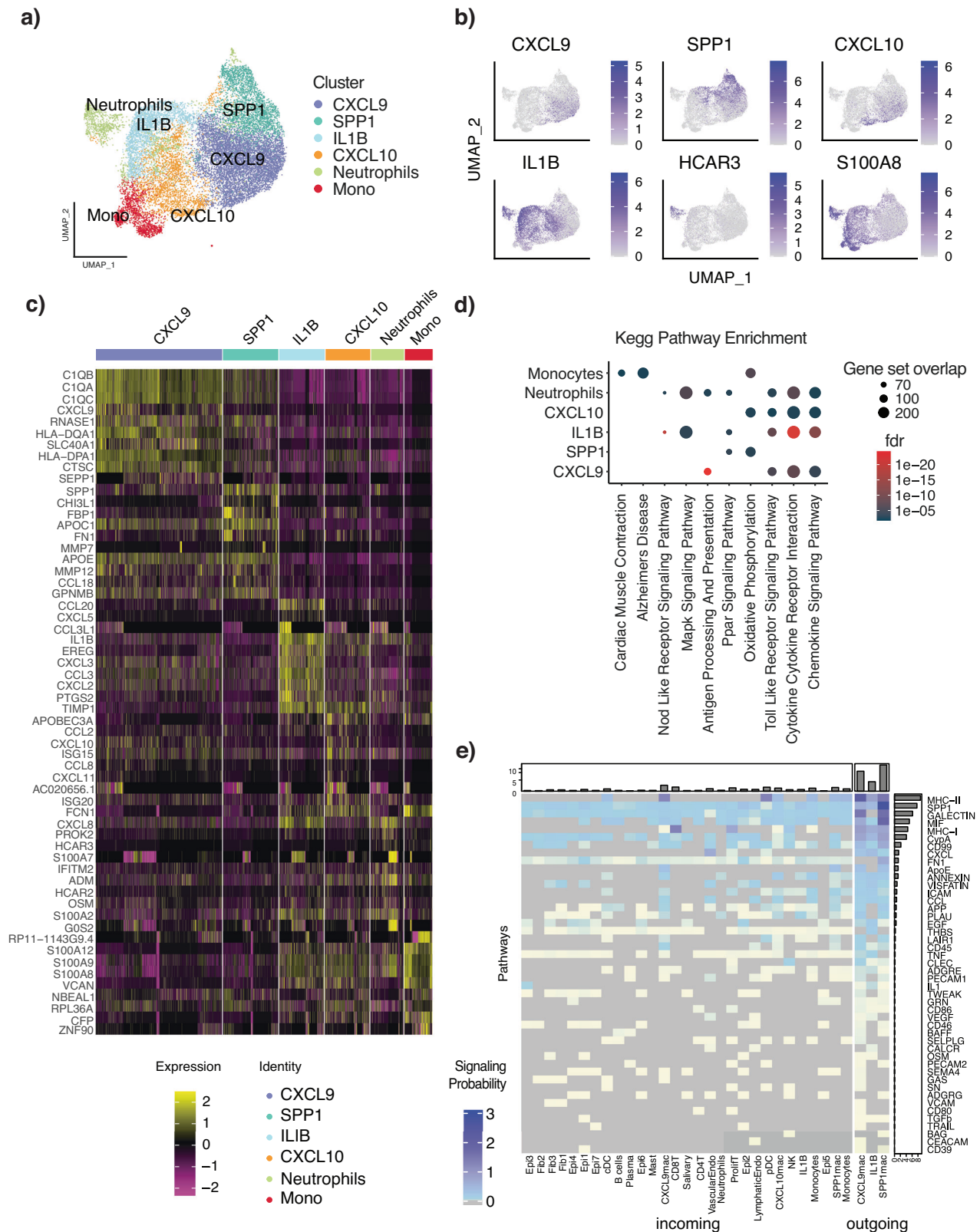


Fig. 3 | Identification of myeloid subclusters and signaling pathways. a UMAP of immune compartment clusters 5 and 7 (from Fig. 2a) reclustered at resolution 0.3. Six clusters were identified. *N* patients = 49. **b** UMAPs of marker gene expression for the six clusters, *CXCL9*, *SPP1*, *CXCL10*, *IL1B*, *HCAR3*, and *S100A8*. **c** Heatmap of top 10 marker genes per cluster. Expression is scaled row-wise. **d** KEGG signature enrichment for macrophage subclusters based on signature genes using

hypergeometric test with FDR-corrected significance values. **e** Cell-cell communication analysis by CellChat showing pooled signaling probability across pathways. Signals shown are sent from the *CXCL9*, *SPP1*, and *IL1B* clusters (outgoing) to the other clusters (incoming). Bar plots on the top and right of the heatmap show total signaling per cluster.

for antigen presentation-related genes (Fig. 3b–d). The other cluster expressed high levels of *SPP1* and metalloproteinases (MMPs), but showed a relative paucity of KEGG or Hallmark pathway enrichments (Fig. 3c, d). Notably, these macrophage clusters were distinguished more by *CXCL9* and *SPP1* than by canonical M1/M2 markers (Supplementary Fig 4c), supporting an alternative macrophage categorization by polarization states in HNSCC recently described by Bill et al.⁶¹.

The remaining four clusters lacked expression of canonical macrophage markers *CIQB*, *CD4*, *APOE* and *CD163* (Supplementary Fig 4a). Among these, one cluster was *CXCL10*-positive, also overexpressing interferon induced proteins with tetratricopeptide repeats (IFIT) genes and *CCL2*, and was enriched for interferon alpha (*IFN-α*) and gamma (*IFN-γ*) response pathways (Fig. 3c, d). This cluster is similar to a population described by Zhang et al. in inflammatory conditions, expressing genes such as *IDO1* and *GBP1*, and stimulated by *IFN-γ* and *TNF-α*⁶⁵.

Another distinct cluster exhibited features consisted with myeloid-derived suppressor cells (MDSCs)⁶⁶, characterized by high expression of inflammatory cytokines including *IL1B*, but lacking macrophage markers *SPP1* and *CXCL9*. This population expressed inflammatory genes such as the aforementioned *IL1B*, known to promote drug resistance through induction of *ICAM1* on tumor cells⁶⁷, as well as *IL6* and *IL10*, cytokines implicated in Treg-mediated immunosuppression⁶⁸. It was also enriched for the cytokine-cytokine receptor interaction KEGG pathway (Fig. 3d). This cluster shares partial transcriptional similarity with previously described HNSCC tumor-associated macrophages (TAMs) classified as M2-like macrophages⁶⁹. However, it lacks expression of several canonical M2 markers such as a *CD163* and complement genes (Supplementary Fig. 4a)⁷⁰.

To contextualize the *IL1B*+ population within the broader myeloid landscape, we compared it with signatures from pan-cancer macrophage datasets (Supplementary Data 2). This cluster showed transcriptional similarity to *VCAN*- and *REG*-expressing TAMs associated with angiogenesis identified in a pan-cancer macrophage atlas (AngioMac)⁶³, a pan-tissue *IL1B*-expressing monocyte population⁷¹, a pan-cancer *CXCL3* + TAM cluster⁶⁴, and a HNSCC *CXCL8* + TAM cluster⁷² (Supplementary Fig. 5d). While nomenclature referencing this cluster varies, sometimes referring to these cells as macrophages or monocytes, this population emerges consistently across datasets. While studies indicate comparable levels of macrophages (identified using *CD68* as a marker) between HPV-positive and -negative patients⁷³, elevated levels correlate with worse prognosis, suggesting this population may play important roles in both contexts.

Given the *IL1B*+ cluster's enrichment for cytokine and chemokine signaling pathways, we examined signaling pathway activity unique to this cluster relative to other myeloid subtypes (Fig. 3d, e). Thrombospondin (*THBS*) signaling was specific to the *IL1B* cluster, and tumor necrosis factor (*TNF*) and *IL-1β* signaling pathways were elevated compared to the *SPP1*+ and *CXCL9*+ clusters (Fig. 3e). Although *IL-1β* and *TNF-α* secretion by *SPP1*-positive macrophages has been linked to HNSCC progression⁷⁴, our atlas highlights the *IL1B*+/*SPP1*-population as a distinct subpopulation associated with unique signaling pathways.

Finally, we identified a cluster of Neutrophils characterized by low expression of *HLA-DRA*, *HLA-DRB1*, and *CD74* expression, and positive expression of *IFITM2*, *S100A8*, *PROK2*, *HCAR3*, *CMTM2*, and *CSF3R*, among others (Fig. 3c)^{75,76}. Gene set enrichment analysis confirmed enrichment for cytokine-cytokine receptor interactions in this cluster (Fig. 1d), and cell-type classification using HPCAD and Blueprint databases corroborated these findings (Supplementary Fig 4d, Supplementary Data 7). The myeloid subcluster annotations are shown in the context of the larger immune compartment in Supplementary Fig 4e, and cluster-specific gene signatures are provided in Supplementary Data 6.

Non-immune compartment

Clustering analysis of the non-immune compartment based on Seurat at resolution 0.2 identified 16 clusters. These included distinct clusters of vascular and lymphatic endothelial cells, three fibroblast clusters, seven tumor epithelial cell clusters, and four salivary cell clusters, each defined by canonical marker genes and cell type signatures (Fig. 4a, b).

Epithelial compartment characterization and tumor cell identification

We used InferCNV to detect copy number variations (CNVs) in epithelial and salivary cells, as described in Methods. An aggregated malignancy score was computed per cell, defined as the total CNV burden in a cell regardless of direction (amplification or loss), with higher scores indicating more alterations. Within each dataset, epithelial and salivary cells exceeding the 90th quantile of malignancy scores derived from “normal” cells were designated cancer cells; others were classified as normal (Supplementary Fig. 5a, b, and Supplementary Data 8). Malignancy scores varied significantly by cluster, with patient-specific clusters Epi2 and Epi3 displaying the highest and lowest scores, respectively, within the epithelial compartment (Supplementary Fig 5c).

Normal epithelial cells segregate on a differentiation continuum

Annotation of the epithelial compartment revealed a gradient of differentiation signatures, with cluster Epi1 representing the least differentiated state and increasing differentiation across Epi6, Epi5, and Epi4 (Fig. 4c). This trajectory reflects the maturation of non-tumor epithelial cells from basal cells to squamous cells, the latter being the tumor's cells of origin. Epi6 expressed high levels of *KRT5* and *KRT14*, markers of basal cells⁷⁷, indicating its position at the early stage of the differentiation trajectory. Epi5 showed high expression of suprabasal markers *KRT13*, *KRT4*, and *SPRR3*⁷⁸. Epi1 expressed high levels of *CD44*, a marker of cancer stem cells (CSCs)⁷⁹, and *KRT18*, associated with progenitor and malignant epithelial cells⁸⁰. The pEMT program was also enriched in Epi1, further suggesting its malignant and plastic phenotype.

To further characterize differentiation status, we applied CytoTRACE2 to the non-immune compartment, which predicts cellular developmental potential on a continuous scale from 0 (fully differentiated) to 1 (totipotent)⁴⁹. While originally developed in a developmental biology context, this tool can be applied to cancer datasets with the ‘potency’ score quantifying a tumor cell's degree of plasticity, or a normal cell's differentiation state. Most the tumor cells were predicted to be oligopotent (plastic capability), whereas other tumor microenvironment cell types such as endothelial cells were largely fully differentiated (Supplementary Fig 5d). Within epithelial clusters, Epi1 exhibited the highest CytoTRACE2 potency scores, followed by Epi6, whereas Epi4 and Epi5 had lower scores. This suggests an inverse correlation between potency and non-tumor epithelial cell differentiation, and positive correlation with tumor cell plasticity and pEMT. Interestingly, some cancer-associated fibroblasts (CAFs) were also predicted to be oligopotent (defined as a lineage-restricted potential to differentiate into 2–3 downstream cell types), which may reflect their capacity for reprogramming⁸¹.

Some epithelial clusters are patient-specific

Hierarchical organization of the non-immune compartment using K2T first partitioned cells primarily into stromal and epithelial groups, with epithelial cluster pairs Epi4 and Epi5 grouped together, consistent with their differentiation state relative to the tumor cell of origin (Fig. 4d). Interestingly, Epi7 grouped within the stromal partition. Examination of patient distribution across epithelial clusters revealed that clusters Epi2, Epi3, and Epi5 were largely patient-specific (Supplementary Fig 5e). These clusters showed unique pathway enrichments, such as the pentose phosphate pathway and xenobiotic drug metabolism in Epi2, and translation in Epi3, with Epi7 consistent with a cycling cell population (Fig. 4c). Notably, Epi2 uniquely expressed *KRT19*, a gene shown to reprogram CSCs in breast cancer toward a more drug-sensitive state⁸².

[illegible]

stromal cells and distinct spatial localization, we leveraged an external spatial transcriptomics dataset, Arora et al.³⁸, which defined tumor leading-edge and core signatures in HNSCC. Projection these signatures onto our epithelial clusters showed that Epi1 was enriched for the leading-edge signature, whereas Epi4, and to a lesser extent Epi5, were enriched for the tumor core signature (Supplementary Fig 5f). These findings align with

prior reports locating pEMT-expressing tumor cells at the invasive front, facilitating interaction with CAFs¹⁵. Additionally, Epi1 strongly expressed *TGFBI*, a secreted ECM protein linked to aggressive cancers^{83,84}.

Detailed characterization of epithelial cells

Reclustering epithelial cells at higher resolution (0.4) identified 18 distinct clusters (Supplementary Fig 5g). Marker genes were shared across multiple clusters suggesting that the initial lower resolution clustering was sufficient to capture the most salient cell differences (e.g., *ACKR1C3* in 0 and 4) (Supplementary Fig 5h). Gene set enrichment analysis using the Hallmarks compendium and signatures from Puram et al.¹⁶ revealed pathways co-enriched in multiple epithelial cell clusters (Supplementary Fig 5i). These included cholesterol homeostasis, known to be important for proper keratinization, and epithelial cell differentiation, both enriched in clusters 7, 8, 9 and 13⁸⁵. Clusters 3, 11, and 17 were enriched for translation, Notch signaling, apoptosis and the p53 pathway, suggesting Notch signaling is driving changes in these clusters, as notch signaling has been shown to interact with the p53 pathway to inhibit apoptosis and drive aberrant translation^{86,87}. Clusters 10, 12, 15, and 16 showed enrichment for hypoxia, *TGF-β* and *IL6* signaling, consistent with hypoxia-induced *TGF-β* and *IL6* signaling in cancer^{88,89}. Oxidative stress and the NRF2 pathway, which are induced by cigarette smoke⁹⁰, were enriched in clusters 0, 4, and 10, implicating these clusters as ones which may be more prevalent in HPV-negative patients.

Epithelial cell signaling

To examine the global patterns of interactions among clusters, we performed cell-cell communication analysis and looked at the total number of interactions between clusters in a pairwise manner (Fig. 4f). Epi1, Epi2, Fib1, Fib3 and vascular endothelial clusters exhibited the highest number of estimated outgoing signals, whereas epithelial cells generally received the most incoming signals. We next characterized signaling pathways exchanged between epithelial clusters and the TME (Fig. 4g, h). Epi1 and Epi2 were unique in sending extracellular matrix (ECM) and tissue physiology signals, including elevated laminin, fibronectin (FN1, primarily from Epi1), collagen (mainly from Epi1), and thrombospondin (THBS, exclusive to Epi1), compared to the other Epithelial clusters (Fig. 4g). Incoming signals to epithelial clusters featured collagen and laminin pathways predominantly originating from the fibroblasts, with increased strength to Epi1 and Epi2 (Fig. 4h). Unique incoming signals to Epi1 and Epi2 included angiopoietin (Angpt) signaling from endothelial cells, and visfatin. Notably, only Epi1 received *TGF-β* signals, consistent with the EMT pathway enrichment in Epi1 subclusters (1, 5, 15, 14) (Supplementary Fig. 5i), and aligning with *TGFβ*'s known role in EMT induction and poor prognosis in HNSCC^{84,91}.

Mesenchymal cell characterization and nomenclature harmonization

Next, we examined mesenchymal cells, a heterogeneous population that includes fibroblasts, pericytes and cancer associated fibroblasts (CAFs), which play critical roles in tumor progression^{15,92–94}.

Clustering of the non-immune compartment at a resolution of 0.7 identified seven mesenchymal populations (Fig. 5a). Expression of canonical CAF markers (*FAP*, *PDPN*, *CTGF*) was observed primarily in clusters 0, 15, 16, suggesting these are CAFs as opposed to resting fibroblasts, while myofibroblasts markers (*ACTA2* and *MYL9*) marked clusters 10 and 14. (Fig. 5b)¹⁵. We then assessed expression of CAF signatures gathered from pan-cancer and HNSCC-specific studies (Table 4 and Supplementary Data 2) to classify subpopulations (Fig. 5c and Supplementary Data 9). Clusters 0 and 15 expressed classic fibroblast¹⁶ and matrix-associated CAF (mCAF) signatures. Clusters 10 and 14 expressed myofibroblast and vessel-associated CAF (vCAF) signatures, with cluster 14 uniquely expressing vCAF markers *BCAM* and *MYH11* (Fig. 5c, Supplementary Fig 6a). Cluster 16 expressed the inflammatory CAF (iCAF) signature. Notably, while cluster 0 expressed

chemokines such as *CXCL1*, *CXCL8*, and *IL11*, cluster 16 uniquely expressed canonical iCAF genes *CXCL12* and *CXCL14*, suggesting that the iCAF signature^{95,96} may encompass at least two distinct subtypes in HNSCC. Some genes such as *CCL2*, *CXCL2* and *IL6* were shared between both clusters (Supplementary Fig. 6a). Interestingly, cluster 0 also expressed the metabolic CAF (meCAF) signature (Fig. 5c), first identified in pancreatic ductal adenocarcinoma⁹⁵. Although antigen-presenting CAFs (apCAFs) are common in several cancer types^{93,97}, there is a lack of a strong localized expression of the corresponding signature in the atlas (Supplementary Fig. 6a). Cluster 22 likely represents myocyte muscle cells due to their expression of myosin (*MYL9*) (Fig. 5d). Cluster 32 expressed both fibroblast markers (*COL1A2*, *PDGFRB*) and vascular endothelial marker *PECAM1*, suggesting this may be a cluster of doublets.

To contextualize atlas fibroblast clusters with CAF populations identified in the originating studies, we scored atlas clusters using the published CAF gene signatures (Supplementary Fig 6b). Cluster 15 was consistently classified as CAFs by most studies^{15,23,25,27}, while clusters 10, 14, 22 and 32 were classified as myofibroblasts, and were termed either myofibroblasts or pericytes in most studies^{15,23,25,27}. Cluster 0 and 16, the two immune-related fibroblast populations were variably termed 'elastic' or 'resting' fibroblasts^{15,25}. Recent studies identified *CXCL8*+ inflammatory CAFs¹⁶ and *IL11* + CAFs⁹⁸ in HNSCC, prompting our harmonization of nomenclature across studies to define equivalent populations in our atlas (Supplementary Fig 6b, Table 2). Cluster-specific gene signatures were defined using a log2 fold change cutoff > 1 and FDR < 0.05 (Supplementary Data 10).

Cell-cell communication

Due to the high expression of multiple chemokines in CAF clusters 0, 15, and 16, and the observed high correlation between fibroblast cluster 1 and the vascular endothelial cell cluster across patients ($R = 0.5$, $FDR = 0.008$) (Fig. 5e), we examined fibroblast-endothelial cell signaling. Fibroblasts sent collagen signals preferentially to lymphatic, but not vascular, endothelial cells (Fig. 5e). Cluster 0 showed increased signaling strength to vascular endothelial cells via the CXCL pathway via enhanced *CXCL1-ACKR1* and *CXCL8-ACKR1* interactions (Fig. 5f). Interestingly, *CXCR1* and *CXCR2*, the canonical receptors for *CXCL1* and *CXCL8*, were not expressed in any atlas cells, whereas *ACKR1*, an alternative receptor for these ligands⁹⁹, was exclusively expressed in vascular endothelial cells (Fig. 5g). Similarly, cluster 16 showed increased signaling strength to the vascular endothelial cells through *CXCL2-ACKR1* interactions (Fig. 5f).

Changes in cell type proportions with sex

The large sample size of the atlas enabled us to evaluate potential associations between TME heterogeneity and sex, recognizing that the extent and composition of immune cell infiltration in the TME can play an important role in tumor aggressiveness. Differential cell type composition analyses were performed within the immune and non-immune compartments, separately (Supplementary Data 11)⁵⁰. For simplicity, related clusters were merged: naïve, Treg, CD4+ Tcm and Tfh clusters were combined into a single CD4+ T cell cluster; the cytotoxic and dysfunctional CD8+ T cells were merged into one CD8+ T cell cluster; and the four salivary clusters were consolidated (Supplementary Fig 7a, b). Several significant sex-associated changes were observed within the immune compartment. Male patients exhibited increased proportions of macrophages, and CD8+ and proliferating T cells, whereas female patients had higher proportions of Plasma cells, Monocytes and NK cells (Fig. 6a). A pan-cancer study of changes in cell type proportions using deconvolution in TCGA datasets similarly reported increased CD8+ T cells in male patients, alongside multiple immune populations enriched in female patients¹⁰⁰. In the non-immune compartment, the proportions of Epi1 and Epi4 were greater in male patients (Fig. 6b). Equivalent analyses performed within the individual studies largely failed to identify significant sex-associated changes (Supplementary Fig 7c,d), underscoring the advantage of the integrated atlas in detecting such

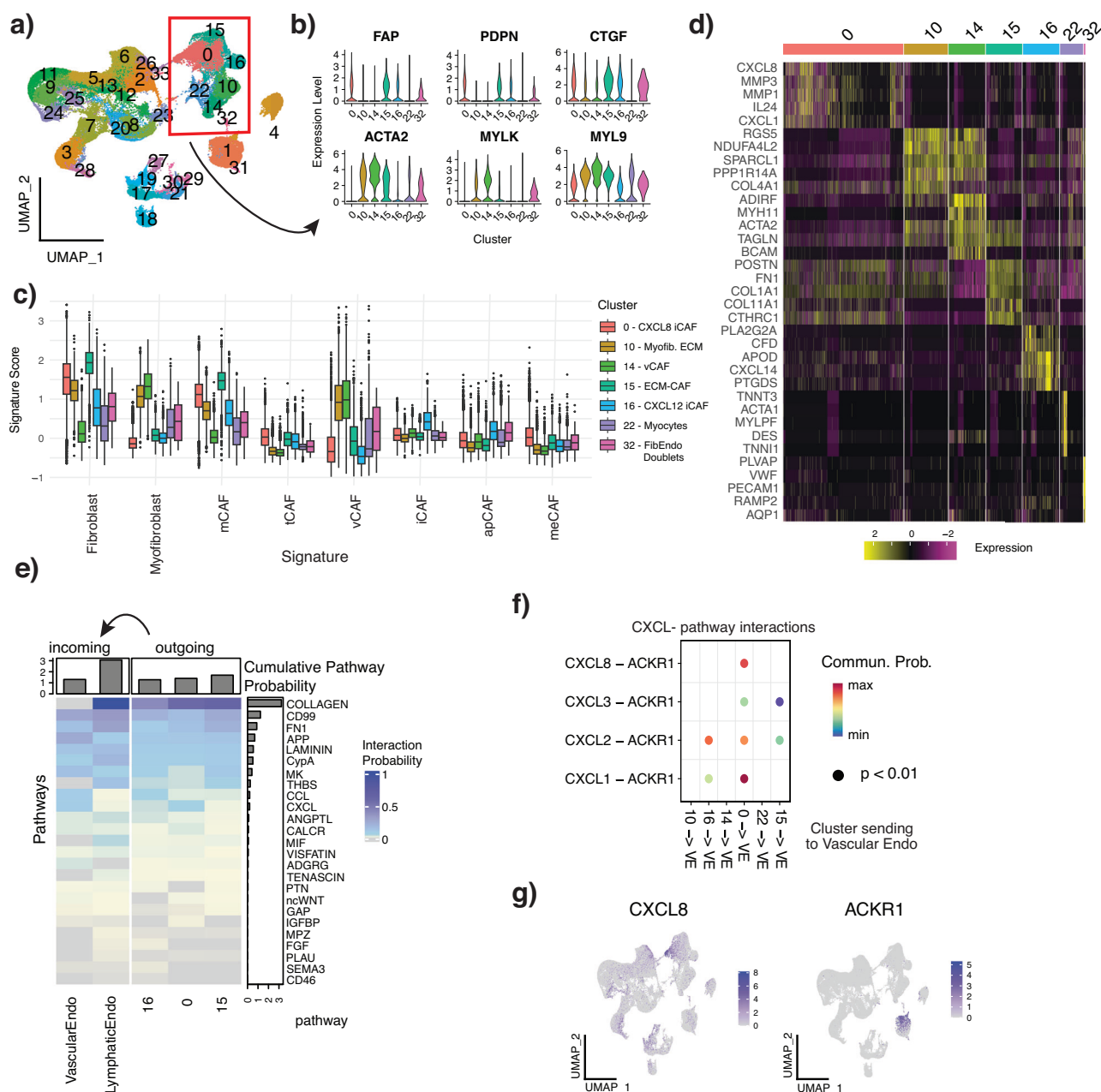


Fig. 5 | Fibroblast classification and signaling. a UMAP clustered at resolution 0.7 for a total of 34 clusters. n Patients = 47. **b** Violin plots showing normalized expressed values for *FAP*, *PDPN*, *CTGF*, *ACTA2*, *MYLK*, and *MYL9* across the fibroblast clusters. n Patients with cells in c0: $n = 45$; c15: $n = 36$; c16: $n = 40$; c10: $n = 38$; c22: $n = 34$; c14: $n = 38$; c32: $n = 20$. n cells per fibroblast cluster: c0: $n = 8514$; c15: $n = 2588$; c16: $n = 2575$; c10: $n = 3110$; c22: $n = 1621$; c14: $n = 2596$; c32: $n = 119$. **c** Boxplots showing the average signature scores per fibroblast cluster. Boxes represent median $\pm$ interquartile range; whiskers represent $1.5 \times$ interquartile range. **d** Heatmap of top five DE genes per fibroblast cluster. Values shown are

z-scaled normalized expression values. **e** Cell-cell communication analysis by CellChat showing pathways involved in sending signals from fibroblast clusters 0, 15, and 16 (outgoing) to the vascular and lymphatic endothelial cells (incoming). Color intensity represents the sum of the signaling strength of all interactions in each pathway. Bar plots represent total signaling strength per cluster (top), and per pathway (right). **f** Specific ligand and receptor pairs involved in CXCL pathway interactions occurring between fibroblast clusters 0, 15, and 16 and the vascular endothelial cells. P-values shown represent empirical p values. **g** Expression of *CXCL8* and *ACKR1* in non-immune compartment.

associations thanks to the increased statistical power afforded by a large number of patients.

Discussion

In this study, we integrated six publicly available HNSCC scRNA-seq datasets to establish a comprehensive HPV-negative HNSCC atlas. This resource is thoroughly annotated with patient metadata, cell types, and their corresponding genes signatures. The large cohort size attained enabled statistically powered evaluation of associations between

gene expression, cell type composition, and molecular and clinical phenotypes.

Recent work uncovered the *SPP1/CXCL9* polarization axis that characterizes distinct macrophage populations⁶¹. However, other myeloid populations remain less well defined in HNSCC. Our atlas recapitulates *SPP1*+ and *CXCL9*+ macrophage subsets and further identifies an *IL1B*+ population variably labeled in prior studies as *CXCL3* TAMs, monocytes, *CXCL8* TAMs, or AngioMacs, highlighting the ongoing challenge of inconsistent nomenclature across studies. We find that this *IL1B*+ cluster

Table 4 | Fibroblast nomenclature harmonization

Cluster	0	15	16	14	22	10	32
Dataset							
Atlas	Imm-CAF (CXCL8)	ECM-CAF	Imm-CAF (CXCL12)	vCAF	Myocytes	Myofibroblast-ECM	Fib-Endo Doublets
Puram	CAF1	CAF1	CAF2	Myofibroblasts	NA	CAF1	Myofibroblasts
Choi	cf4	cf4	cf2,cf4	cf3	Myocytes	cf3	cf3,cf4
Kurten	CAF	CAF	Elastic	Pericytes	Pericytes	Pericytes	Pericytes
Quah	CAF	CAF	CAF	Myofibroblast	Myofibroblast	Myofibroblast	CAF, Myofibroblast
Peng	No Fib analysis						
Cillo	No CD45- cells						

secretes immunosuppressive cytokines (e.g., *IL10*) and interacts with other TME cells via *IL-1β* and *TNF* signaling pathways, both implicated in HNSCC progression⁷¹, as well as through unique pathways such as thrombospondin signaling. The role of the thrombospondin pathway in HNSCC is complex; it has been reported both to promote invasion¹⁰¹ and inhibit angiogenesis¹⁰², leaving its net effect on tumor progression unclear. However, CD47, a major thrombospondin receptor, is a promising therapeutic target in HNSCC, with anti-CD47 treatment shown to delay tumor growth and stimulate effector T cells¹⁰³. Prior studies have linked macrophage populations to poor prognosis, specifically *SPP1*+ macrophages in HNSCC⁶¹ and M2 macrophages across cancers, which have also been effectively targeted by therapies⁷⁰. Furthermore, recent evidence showed macrophage-secreted *IL-1β* to promote docetaxel resistance⁶⁷. Given that these macrophage clusters are major sources of *IL1B* and immunosuppressive cytokines, their inhibition may improve both drug efficacy and immune response, although further validation is needed to demonstrate their precise role in the HNSCC TME.

The landscape of CAFs has also been increasingly refined, both in pan-cancer contexts and specifically in HNSCC^{97,104}, and we were able to extend this characterization further within the atlas. We distinguish two separate iCAF populations – *CXCL8*+ and *CXCL12*+ iCAFs—that are frequently conflated in the literature^{97,105,106}. We also identify vCAFs, which have been described in pan-cancer studies but are frequently termed myofibroblasts in HNSCC. Jenkins et. al. recently reported enrichment of both myCAFs and iCAFs (the iCAFs align with the *CXCL8*+ iCAFs) in HPV-negative HNSCC, while a population termed FRC-like CAFs (which align with the *CXCL12*+ iCAFs) was more abundant in HPV-positive tumors⁹⁸. Interestingly, their spatial transcriptomics data revealed *CXCL12*+ iCAFs correlate with CD4+ T cells and B cells and are associated with better survival, whereas *CXCL8*+ iCAFs correlate with *IL1B*+ inflammatory monocytes but show no survival association.

We identify a specific interaction between *CXCL8*+ CAFs and *ACKR1* on vascular endothelial cells that, based on current literature, may be associated with worse outcome through promoting angiogenesis, although experimental validation is needed. The *CXCL8*+ population was previously characterized by Quah et al.²⁷ who demonstrated its induction by galectin-7 (*LGALS7*) expression in malignant epithelial clusters. A recent spatial transcriptomics study of HNSCC also showed co-expression of *CXCL1*, *CXCL8* and *ACKR1*³⁸, corroborating our finding that *CXCL8*+ CAFs and vascular endothelial cell abundances are correlated across patients. Prior studies also showed that *CXCL8* and *CXCL1* promote the angiogenic potential of endothelial cells and affect their proliferation and survival *in-vitro*¹⁰⁷, and a similar myCAF population expressing *CXCL8* and *IL24* was recently linked to metastasis and angiogenesis in esophageal squamous cell carcinoma¹⁰⁸. Both *CXCL1* and *CXCL8* have been associated with poor patient survival in HNSCC¹⁰⁹. Collectively, these findings suggest that the *CXCL8*+ iCAF population may contribute to worse outcomes for HNSCC patients by stimulating angiogenesis and could represent an effective target for anti-angiogenic therapies. However, studies in other contexts have reported that *ACKR1* may function as a decoy receptor limiting

angiogenesis¹¹⁰. Thus, experimental validation is essential to ascertain *ACKR1*'s role in HNSCC.

A central challenge in HNSCC management is overcoming resistance to treatment. It has been suggested that many cancers, including HNSCCs, maintain a population of CSCs or tumor cells with stem-like properties capable of self-renewal and therapy resistance^{109,111–113}. Our integrated analysis of epithelial and tumor cells across datasets identified shared transcriptional programs across patients, exemplified by the Epi1 cluster, which exhibits stem-like properties, pEMT and EMT signatures, and preferential localization to the tumor edge where it interacts with stromal populations, recapitulating key findings from Puram et al.¹⁵ Cell-cell communication analyses revealed extensive bidirectional signaling involving extracellular matrix components (e.g. laminin, collagen) which influences tumor stiffness, cell adhesion and migration^{114,115}, with increased signaling directed to Epi1 compared to other epithelial clusters. Uniquely, Epi1 receives *TGF-β* signals, consistent with its EMT program enrichment. Since *TGF-β* has been implicated in treatment resistance¹¹⁶, co-targeting *TGF-β* alongside ECM-related pathways may enhance therapeutic response while suppressing tumor progression. As additional drug resistance signatures emerge¹¹⁷, they can be mapped onto the atlas to identify specific cell types and transcriptional programs underpinning resistance.

Sex-associated differences in cellular composition across cancers have been noted but remain understudied in HNSCC. Leveraging the atlas, we observed that male patients harbor higher proportions of *PDL1*-expressing CD8+ T cells, which correlates with improved response to immune checkpoint inhibitors (ICIs)¹¹⁸, suggesting potential sex-based differences in ICI efficacy. Within the non-immune compartment, the Epi1 cluster, enriched for both the pEMT signature and proliferating tumor cells, was more prevalent in male patients, consistent with a study demonstrating increased carcinogen-induced proliferation in male mice, leading to more aggressive and metastatic cSCC¹¹⁹.

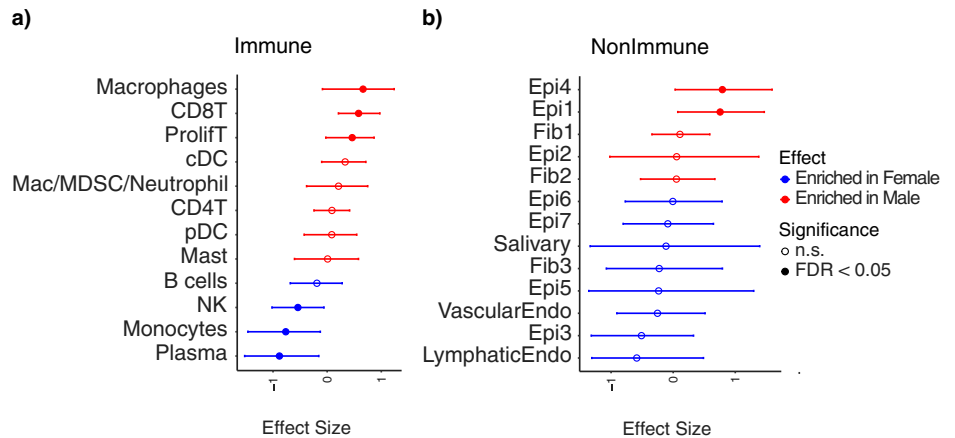
Limitations of this study include variability in cell types collected and sequenced per patient, as well as heterogeneity in tissue anatomical origin. Although the current cohort enables robust analyses, fully disentangling technical from biological variation will require continued expansion of the atlas with new datasets. Moreover, integration updates require manual reprocessing. Scalable integration strategies such as non-negative matrix factorization¹²⁰ or transfer learning approaches such as scArches¹²¹ may streamline future expansions. Beyond transcriptional profiling, future efforts may aim to integrate spatial transcriptomics and multi-modal datasets capturing paired transcriptomic and epigenetic information to deepen understanding of HNSCC heterogeneity.

Data availability

The following publicly available datasets (raw count.mtx, gene and cell barcode files) were accessed via GEO through the accession numbers: Peng²⁶ (GSE172577), Choi²³ (GSE181919), Cillo²⁴ (GSE139324), Quah²⁷ (GSE225331), and Kurten²⁵ (GSE164690). Due to availability, processed normalized data were downloaded for the last dataset, Puram¹⁵ (GSE103322). The atlas has been deposited to cellxgene for online

Fig. 6 | Sex-specific cell type proportion changes.

a Cell type proportion changes identified by sccomp within the immune compartment. Significance shown is based on FDR-corrected p -values. Female $n = 18$. Male $n = 35$. Error bars represent 95% credible intervals. **b** Cell type proportion changes identified by sccomp within the non-immune compartment. Significance shown is based on FDR-corrected n -values. Female $n = 16$. Male $n = 31$. Error bars represent 95% credible intervals.



visualization by the research community through the cellxgene repository, an interactive data explorer, at the following [link](#). The full atlas is also available as a Seurat.rds object for download on zenodo [<https://doi.org/10.5281/zenodo.17880437>]¹²² using the following [link](#). Source data is provided with this publication on Zenodo. Source data for Fig. 1a–d, Fig. 2a–f, Fig. 3a–e, Fig. 4a–h, Fig. 5a–e, g–f, Fig. 6a, b, Supplementary Figs. 1a, b, Supplementary Figs. 3b,c, Supplementary Figs. 4a–f, Supplementary Figs. 5a–f, Supplementary Fig. 6a, Supplementary Figs. 7c, d, are in SourceData.xlsx. Source data for Supplementary Fig. 5h is in SourceData Fig5h.csv. Source data for Supplementary Fig. 3e is in SuppData4_Teffstage_data_SFfig3e.xlsx.

Code availability

All code is available on [github](#) and [zenodo](#).¹²³

Received: 4 March 2025; Accepted: 14 January 2026;

Published online: 11 March 2026

References

- Chow, L. Q. M. Head and neck cancer. *N. Engl. J. Med.* **382**, 60–72 (2020).
- White, A. C. & Lowry, W. E. Defining the role for adult stem cells as cancer cells of origin. *Trends Cell Biol.* **25**, 11–20 (2015).
- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Johnson, D. E. et al. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Prim.* **6**, 92 (2020).
- Tumban, E. A current update on human papillomavirus-associated head and neck cancers. *Viruses* **11**, 922 (2019).
- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J. Clin.* **71**, 209–249 (2021).
- Sabatini, M. E. & Chiocca, S. Human papillomavirus as a driver of head and neck cancers. *Br. J. Cancer* **122**, 306–314 (2020).
- Wagner, S. et al. Human papillomavirus-related head and neck cancer. *Oncol. Res. Treat.* **40**, 334–340 (2017).
- Ghiani, L. & Chiocca, S. High risk-human papillomavirus in HNSCC: present and future challenges for epigenetic therapies. *Int J. Mol. Sci.* **23**, 3483 (2022).
- Amin, M. B. et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more ‘personalized’ approach to cancer staging. *CA Cancer J. Clin.* **67**, 93–99 (2017).
- Heath, B. R. et al. Head and neck cancer immunotherapy beyond the checkpoint blockade. *J. Dent. Res.* **98**, 1073–1080 (2019).
- Ferris, R. L. et al. Nivolumab vs investigator’s choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. *Oral. Oncol.* **81**, 45–51 (2018).
- Bila, M. et al. Exploring long-term responses to immune checkpoint inhibitors in recurrent and metastatic head and neck squamous cell carcinoma. *Oral. Oncol.* **149**, 106664 (2024).
- Califano, J. et al. Genetic progression model for head and neck cancer: implications for field cancerization1. *Cancer Res.* **56**, 2488–2492 (1996).
- Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624.e24 (2017).
- Puram, S. V. et al. Cellular states are coupled to genomic and viral heterogeneity in HPV-related oropharyngeal carcinoma. *Nat. Genet.* **55**, 640–650 (2023).
- Calbo, J. et al. A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* **19**, 244–256 (2011).
- Prazanowska, K. H. & Lim, S. B. An integrated single-cell transcriptomic dataset for non-small cell lung cancer. *Sci. Data* **10**, 167 (2023).
- Johnston, K. G., Grieco, S. F., Nie, Q., Theis, F. J. & Xu, X. Small data methods in omics: the power of one. *Nat. Methods* 1–6. <https://doi.org/10.1038/s41592-024-02390-8> (2024).
- Zhang, Q. et al. Integrated analysis of single-cell RNA-seq and bulk RNA-seq reveals distinct cancer-associated fibroblasts in head and neck squamous cell carcinoma. *Ann. Transl. Med.* **9**, 1017–1017 (2021).
- Li, K. et al. Single cell analysis unveils B cell-dominated immune subtypes in HNSCC for enhanced prognostic and therapeutic stratification. *Int J. Oral. Sci.* **16**, 1–12 (2024).
- Liu, Z. L. et al. Single cell deciphering of progression trajectories of the tumor ecosystem in head and neck cancer. *Nat. Commun.* **15**, 2595 (2024).
- Choi, J.-H. et al. Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat. Commun.* **14**, 1055 (2023).
- Cillo, A. R. et al. Immune landscape of viral- and carcinogen-driven head and neck cancer. *Immunity* **52**, 183–199.e9 (2020).
- Kürten, C. H. L. et al. Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nat. Commun.* **12**, 7338 (2021).
- Peng, Y. et al. Single-cell profiling of tumor-infiltrating TCF1/TCF7+ T cells reveals a T lymphocyte subset associated with tertiary lymphoid structures/organs and a superior prognosis in oral cancer. *Oral. Oncol.* **119**, 105348 (2021).

27. Quah, H. S. et al. Single cell analysis in head and neck cancer reveals potential immune evasion mechanisms during early metastasis. *Nat. Commun.* **14**, 1680 (2023).
28. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
29. Dvir Aran, Aaron Lun, Daniel Bunis, Jared Andrews, Friederike Dündar. SingleR. Bioconductor <https://doi.org/10.18129/B9.BIOC.SINGLER>.
30. Martens, J. H. A. & Stunnenberg, H. G. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* **98**, 1487–1489 (2013).
31. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
32. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
33. Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C. & Hume, D. A. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genom.* **14**, 632 (2013).
34. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
35. Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
36. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
37. Fabregat, A. et al. Reactome graph database: efficient access to complex pathway data. *PLoS Comput. Biol.* **14**, e1005968 (2018).
38. Arora, R. et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat. Commun.* **14**, 5029 (2023).
39. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
40. van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8+ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).
41. Ruffin, A. T. et al. Improving head and neck cancer therapies by immunomodulation of the tumour microenvironment. *Nat. Rev. Cancer* **23**, 173–188 (2023).
42. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
43. Nguyen, H. C. T., Baik, B., Yoon, S., Park, T. & Nam, D. Benchmarking integration of single-cell differential expression. *Nat. Commun.* **14**, 1570 (2023).
44. Agrawal, A. et al. WikiPathways 2024: next generation pathway database. *Nucleic Acids Res.* **52**, D679–D689 (2024).
45. Rouillard, A. D. et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).
46. Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307–1308 (2020).
47. Reed, E. R. & Monti, S. Multi-resolution characterization of molecular taxonomies in bulk and single-cell transcriptomics data. *Nucleic Acids Res.* **49**, e98–e98 (2021).
48. Tickle, T., Tirosh, I., Georgescu, C., Brown, M. & Haas, B. inferCNV of the trinity CTAT project. In *Klarman Cell Observatory* (Broad Institute of MIT and Harvard, 2019).
49. Kang, M. et al. Improved reconstruction of single-cell developmental potential with CytoTRACE 2. *Nat. Methods* **22**, 2258–2263 (2025).
50. Mangiola, S. et al. sccomp: Robust differential composition and variability analysis for single-cell data. *Proc. Natl. Acad. Sci.* **120**, e2203828120 (2023).
51. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
52. SHIGA, K. et al. Differences between oral cancer and cancers of the pharynx and larynx on a molecular level. *Oncol. Lett.* **3**, 238–243 (2012).
53. Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
54. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* **12**, 492–499 (2011).
55. Hatzioannou, A. et al. Regulatory T cells in autoimmunity and cancer: a duplicitous lifestyle. *Front. Immunol.* **12**, 731947 (2021).
56. Wu, C. et al. Galectin-9-CD44 interaction enhances stability and function of adaptive regulatory T cells. *Immunity* **41**, 270–282 (2014).
57. Wagner, S. et al. CD56-positive lymphocyte infiltration in relation to human papillomavirus association and prognostic significance in oropharyngeal squamous cell carcinoma. *Int. J. Cancer* **138**, 2263–2273 (2016).
58. Fialová, A., Koucký, V., Hajdušková, M., Hladíková, K. & Špišek, R. Immunological network in head and neck squamous cell carcinoma — a prognostic tool beyond HPV status. *Front. Oncol.* **10**, 1701 (2020).
59. Han, N. et al. Increased tumor-infiltrating plasmacytoid dendritic cells predicts poor prognosis in oral squamous cell carcinoma. *Arch. Oral. Biol.* **78**, 129–134 (2017).
60. Badoual, C. et al. Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and neck cancers. *Clin. Cancer Res.* **12**, 465–472 (2006).
61. Bill, R. et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* **381**, 515–524 (2023).
62. Cha, S. M. et al. SPP1+ macrophages in HR+ breast cancer are associated with tumor-infiltrating lymphocytes. *npj Breast Cancer* **10**, 83 (2024).
63. Coulton, A. et al. Using a pan-cancer atlas to investigate tumour associated macrophages as regulators of immunotherapy response. *Nat. Commun.* **15**, 5665 (2024).
64. Wei, C. et al. Tumor-associated macrophage clusters linked to immunotherapy in a pan-cancer census. *npj Precis. Onc.* **8**, 176 (2024).
65. Zhang, F. et al. IFN-γ and TNF-α drive a CXCL10+ CCL2+ macrophage phenotype expanded in severe COVID-19 lungs and inflammatory diseases with tissue inflammation. *Genome Med.* **13**, 64 (2021).
66. Veglia, F., Sanseviero, E. & Gabrilovich, D. I. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. *Nat. Rev. Immunol.* **21**, 485–498 (2021).
67. Hsieh, C.-Y. et al. Macrophage secretory IL-1β promotes docetaxel resistance in head and neck squamous carcinoma via SOD2/CAT-ICAM1 signaling. *JCI Insight* **7**, e157285 (2022).
68. Hao, Z. et al. Landscape of myeloid-derived suppressor cell in tumor immunotherapy. *Biomark. Res.* **9**, 77 (2021).
69. Mhaidly, N. et al. Macrophage profiling in head and neck cancer to improve patient prognosis and assessment of cancer cell-macrophage interactions using three-dimensional coculture models. *Int. J. Mol. Sci.* **24**, 12813 (2023).
70. Li, B., Ren, M., Zhou, X., Han, Q. & Cheng, L. Targeting tumor-associated macrophages in head and neck squamous cell carcinoma. *Oral. Oncol.* **106**, 104723 (2020).
71. Mulder, K. et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* **54**, 1883–1900.e5 (2021).
72. Luoma, A. M. et al. Tissue-resident memory and circulating T cells are early responders to pre-surgical cancer immunotherapy. *Cell* **185**, 2918–2935.e29 (2022).
73. Lechien, J. R. et al. HPV involvement in the tumor microenvironment and immune treatment in head and neck squamous cell carcinomas. *Cancers* **12**, 1060 (2020).
74. Liu, C. et al. SPP1+ macrophages promote head and neck squamous cell carcinoma progression by secreting TNF-α and IL-1β. *J. Exp. Clin. Cancer Res.* **43**, 332 (2024).

75. Shao, X. et al. A single-cell landscape of human liver transplantation reveals a pathogenic immune niche associated with early allograft dysfunction. *Engineering* <https://doi.org/10.1016/j.eng.2023.12.004> (2024).
76. Garrido-Trigo, A. et al. Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. *Nat. Commun.* **14**, 4506 (2023).
77. Abashev, T. M., Metzler, M. A., Wright, D. M. & Sandell, L. L. Retinoic Acid signaling regulates Krt5 and Krt14 independently of stem cell markers in submandibular salivary gland epithelium. *Dev. Dyn.* **246**, 135–147 (2017).
78. Soboleva, A. et al. Gene-expression patterns of tumor and peritumor tissues of smoking and non-smoking hpv-negative patients with head and neck squamous cell carcinoma. *Biomedicines* **12**, 696 (2024).
79. Prince, M. E. P. & Ailles, L. E. Cancer stem cells in head and neck squamous cell cancer. *JCO* **26**, 2871–2875 (2008).
80. Chen, B. et al. KRT18 modulates alternative splicing of genes involved in proliferation and apoptosis processes in both gastric cancer cells and clinical samples. *Front. Genet.* **12**, 635429 (2021).
81. Huang, J., Tsang, W.-Y., Li, Z.-H. & Guan, X.-Y. The origin, differentiation, and functions of cancer-associated fibroblasts in gastrointestinal cancer. *Cell Mol. Gastroenterol. Hepatol.* **16**, 503–511 (2023).
82. Saha, S. K., Kim, K., Yang, G.-M., Choi, H. Y. & Cho, S.-G. Cytokeratin 19 (KRT19) has a role in the reprogramming of cancer stem cell-like cells to less aggressive and more drug-sensitive cells. *Int. J. Mol. Sci.* **19**, 1423 (2018).
83. Sarubo, M. et al. Involvement of TGFBI-TAGLN axis in cancer stem cell property of head and neck squamous cell carcinoma. *Sci. Rep.* **14**, 6767 (2024).
84. Wang, B. et al. TGFBI promotes tumor growth and is associated with poor prognosis in oral squamous cell carcinoma. *J. Cancer* **10**, 4902 (2019).
85. Palmer, M. A., Blakeborough, L., Harries, M. & Haslam, I. S. Cholesterol homeostasis: links to hair follicle biology and hair disorders. *Exp. Dermatol.* **29**, 299–311 (2020).
86. Li, X. et al. The Notch signaling pathway: a potential target for cancer immunotherapy. *J. Hematol. Oncol.* **16**, 45 (2023).
87. Nair, P., Somasundaram, K. & Krishna, S. Activated Notch1 inhibits p53-induced apoptosis and sustains transformation by human papillomavirus Type 16 E6 and E7 oncogenes through a PI3K-PKB/Akt-dependent pathway. *J. Virol.* **77**, 7106–7112 (2003).
88. Mallikarjuna, P., Zhou, Y. & Landström, M. The synergistic cooperation between TGF- β and hypoxia in cancer and fibrosis. *Biomolecules* **12**, 635 (2022).
89. Castillo-Rodríguez, R. A., Trejo-Solís, C., Cabrera-Cano, A., Gómez-Manzo, S. & Dávila-Borja, V. M. Hypoxia as a modulator of inflammation and immune response in cancer. *Cancers* **14**, 2291 (2022).
90. Xue, J. et al. Cigarette smoke-induced oxidative stress activates NRF2 to mediate fibronectin disorganization in vascular formation. *Open Biol.* **12**, 210310 (2022).
91. Wang, X., Eichhorn, P. J. A. & Thiery, J. P. TGF- β , EMT, and resistance to anti-cancer treatment. *Semin. Cancer Biol.* **97**, 1–11 (2023).
92. Cho, S. J. et al. Intercellular cross-talk through lineage-specific gap junction of cancer-associated fibroblasts related to stromal fibrosis and prognosis. *Sci. Rep.* **13**, 14230 (2023).
93. Reid, S. E. et al. Cancer-associated fibroblasts rewire the estrogen receptor response in luminal breast cancer, enabling estrogen independence. *Oncogene* **43**, 1113–1126 (2024).
94. Wang, Y. et al. Cancer-associated fibroblasts in the invasive tumour front promote the metastasis of oral squamous cell carcinoma through MFAP5 upregulation. *Gene* **876**, 147504 (2023).
95. Ma, C. et al. Pan-cancer spatially resolved single-cell analysis reveals the crosstalk between cancer-associated fibroblasts and tumor microenvironment. *Mol. Cancer* **22**, 170 (2023).
96. Lavie, D., Ben-Shmuel, A., Erez, N. & Scherz-Shouval, R. Cancer-associated fibroblasts in the single-cell era. *Nat. Cancer* **3**, 793–807 (2022).
97. Cords, L., de Souza, N. & Bodenmiller, B. Classifying cancer-associated fibroblasts—the good, the bad, and the target. *Cancer Cell* **42**, 1480–1485 (2024).
98. Jenkins, B. H. et al. Single cell and spatial analysis of immune-hot and immune-cold tumours identifies fibroblast subtypes associated with distinct immunological niches and positive immunotherapy response. *Mol. Cancer* **24**, 3 (2025).
99. Cambier, S., Gouwy, M. & Proost, P. The chemokines CXCL8 and CXCL12: molecular and functional properties, role in disease and efforts towards pharmacological intervention. *Cell Mol. Immunol.* **20**, 217–251 (2023).
100. Han, J. et al. Pan-cancer analysis reveals sex-specific signatures in the tumor microenvironment. *Mol. Oncol.* **16**, 2153–2173 (2022).
101. Haughton, P. D. et al. Differential transcriptional invasion signatures from patient derived organoid models define a functional prognostic tool for head and neck cancer. *Oncogene* **43**, 2463–2474 (2024).
102. Bornstein, P. Thrombospondins function as regulators of angiogenesis. *J. Cell Commun. Signal.* **3**, 189–200 (2009).
103. Wu, L. et al. Anti-CD47 treatment enhances anti-tumor T-cell immunity and improves immunosuppressive environment in head and neck squamous cell carcinoma. *Oncoimmunology* **7**, e1397248 (2018).
104. Sahai, E. et al. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat. Rev. Cancer* **20**, 174 (2020).
105. Hu, C., Zhang, Y., Wu, C. & Huang, Q. Heterogeneity of cancer-associated fibroblasts in head and neck squamous cell carcinoma: opportunities and challenges. *Cell Death Discov.* **9**, 124 (2023).
106. Öhlund, D. et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J. Exp. Med.* **214**, 579–596 (2017).
107. Karl, E. et al. Unidirectional crosstalk between Bcl-xL and Bcl-2 enhances the angiogenic phenotype of endothelial cells. *Cell Death Differ.* **14**, 1657–1666 (2007).
108. Guo, W. et al. Single-cell RNA sequencing and spatial transcriptomics of esophageal squamous cell carcinoma with lymph node metastases. *Exp. Mol. Med.* **57**, 59–71 (2025).
109. Li, Y. et al. Analysis of the Prognosis and Therapeutic Value of the CXCL Chemokine Family in Head and Neck Squamous Cell Carcinoma. *Front Oncol.* **10**, 570736 (2021).
110. Fu, W. et al. Cellular features of localized microenvironments in human meniscal degeneration: a single-cell transcriptomic study. *eLife* **11**, e79585 (2022).
111. Adams, A. et al. ALDH/CD44 identifies uniquely tumorigenic cancer stem cells in salivary gland mucoepidermoid carcinomas. *Oncotarget* **6**, 26633–26650 (2015).
112. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
113. Clara, J. A., Monge, C., Yang, Y. & Takebe, N. Targeting signalling pathways and the immune microenvironment of cancer stem cells — a clinical update. *Nat. Rev. Clin. Oncol.* **17**, 204–232 (2020).
114. Meireles Da Costa, N. et al. Potential therapeutic significance of laminin in head and neck squamous carcinomas. *Cancers* **13**, 1890 (2021).
115. Huang, J. et al. Extracellular matrix and its therapeutic potential for cancer treatment. *Sig Transduct. Target Ther.* **6**, 153 (2021).
116. Gauss, C. et al. Overcoming resistance to standard-of-care therapies for head and neck squamous cell carcinomas. *Cells* **13**, 1018 (2024).

117. Keddar, M. R. et al. Pan-cancer analysis in the real-world setting uncovers immunogenomic drivers of acquired resistance post-immunotherapy. *CANCER-CELL-D-25-00908*, Available at: <https://ssrn.com/abstract=5371151>.
118. Ott, P. A. et al. T-cell-inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J. Clin. Oncol.* **37**, 318–327 (2019).
119. Budden, T. et al. Female immunity protects from cutaneous squamous cell carcinoma. *Clin. Cancer Res.* **27**, 3215–3223 (2021).
120. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
121. Lotfollahi, M. et al. Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* **40**, 121–130 (2022).
122. Kroehling, L. & Monti, S. A highly resolved integrated single-cell atlas of HPV-negative head and neck cancer. *Zenodo*. <https://doi.org/10.5281/zenodo.17880437> (2025).
123. Kroehling, L. montilab/kroehling_et_al_hpvneg_hnsc_atlas: kroehling_et_al_hpvneg_hnsc_atlas. *Zenodo* <https://doi.org/10.5281/zenodo.18198004> (2026).

Acknowledgements

This study has received funding from: System-Level Analyses of Multi-Omics Data to Reveal Mechanisms of Head & Neck Cancer; R01DE031831 (SM); National Institute of Dental and Craniofacial Research Defining immune-evasive mechanical signaling in head and neck cancer.; R01DE033519 (XV, SM, MAK); National Institute of Dental and Craniofacial Research Elucidating mechanisms of cellular communication critical for head and neck cancer progression and metastasis.; NIH 1F31DE033292-01 (LK); NIH/NIDCR Defining the β -catenin/CBP axis in head and neck cancer, R01DE030350 (MAK, SM, XV); National Institute of Dental and Craniofacial Research NIH/NCATS grant BU-CTSI 1UL1TR001430 (SM) Find the Cause Breast Cancer Foundation findthecausebcf.org (SM)

Author contributions

S.M. and L.K. designed the study. L.K. performed all analyses and wrote the manuscript. A.C. and E.R. provided computational guidance and feedback. A.S., M.K., and X.V. provided biological guidance. S.M. oversaw the project, provided feedback, and guided the research. The manuscript is edited by all authors.

Competing interests

The authors declare no competing interests.

Consent to publish

Not applicable. All datasets included are already published.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-026-01401-3>.

Correspondence and requests for materials should be addressed to Stefano Monti.

Peer review information *Communications Medicine* thanks Vanessa Porter, Noah Sasa and the other, anonymous, reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026