Divergent Genomic Trajectories Specify Distinct Tumor Ecotypes and Reveal Targetable Immune Evasion in Oral Carcinogenesis

Yuanfu Zhang[1,2,7], Rong-quan Xiao[1,2,7], Yingying Feng[3,7], Yi Ren[1], Hongtao Jin[4], Jiayi Ye[5,6], Tianxin Wang[3], Zheng Zhang[3], Jianming Tang[3], Hongliang Xie[3], Chenchen Ma[1,2], Jiaxuan Li[1,2], Ran Zhang[5], Guoquan Zhang[3*], Shimin Shuai[1,2,8*]

[1] Department of Human Cell Biology and Genetics, School of Medicine, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China

[2] SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China

[3] Medical Center of Stomatology, Shenzhen People's Hospital (The First Affiliated Hospital, Southern University of Science and Technology; The Second Clinical Medical College, Jinan University), Shenzhen, Guangdong 518020, China

[4] Department of Pathology, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University: The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, Guangdong 518020, China

[5] Department of Oral Pathology, Peking University School and Hospital of Stomatology, 22 South Zhongguancun Avenue, Haidian District, Beijing 100081, China

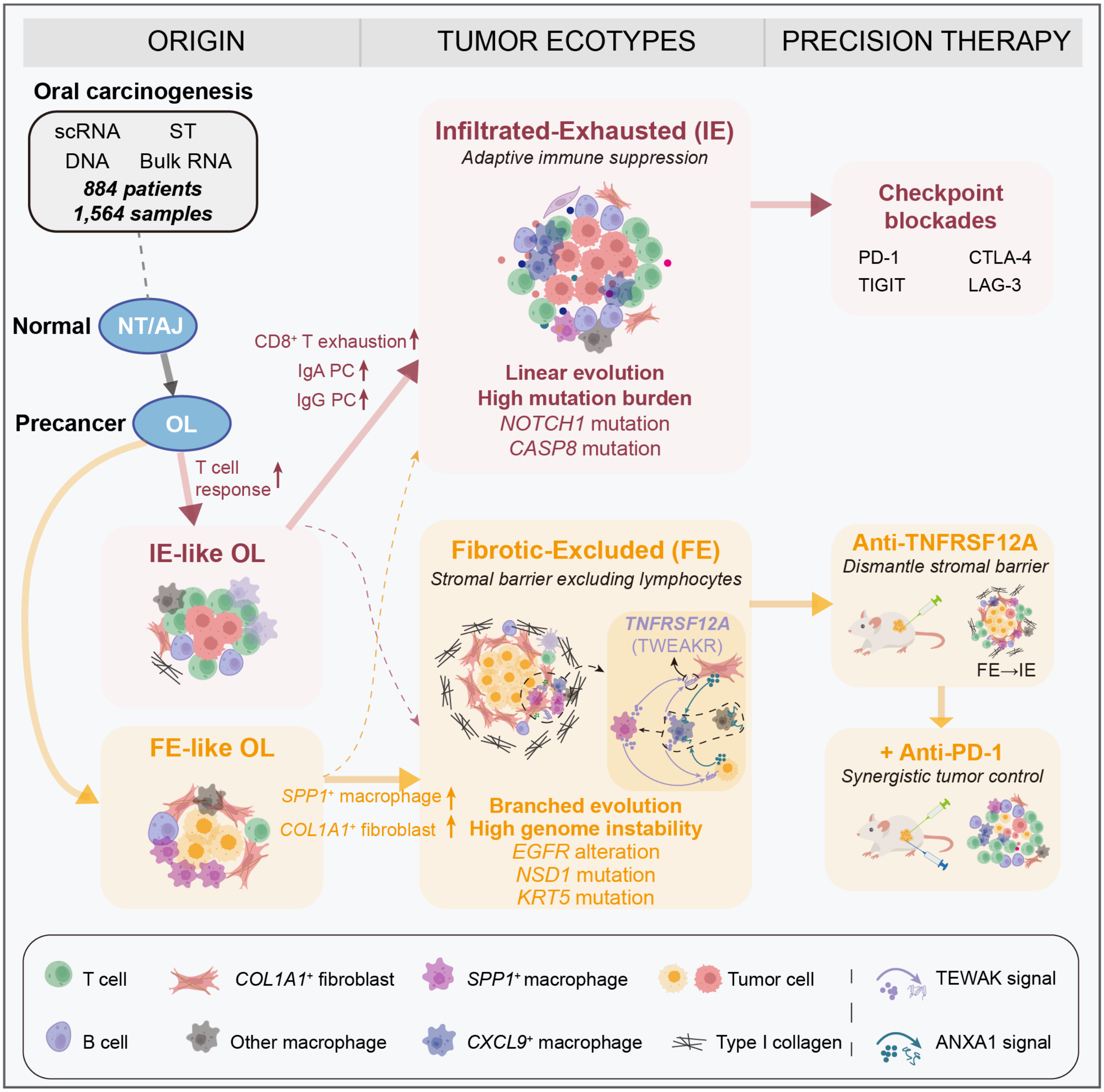
[6] The Affiliated Stomatological Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi 330006, China

[7] These authors contributed equally

[8] Lead contact

*Correspondence: zhang.guoquan@szhospital.com (GZ), shuaism@sustech.edu.cn (SS)

Graphical Abstract



Highlights

- Oral cancers organize into two fundamental ecotypes with distinct outcomes
- Each ecotype arises from a distinct genomic evolutionary trajectory
- Tumor microenvironment fate is imprinted at the precancerous stage
- Targeting the fibrotic barrier converts cold tumors to immunotherapy-responsive

Summary

Oral squamous cell carcinoma, the predominant subtype of head and neck cancer, exhibits limited immunotherapy efficacy, yet the biological basis remains unclear. Through integrative multi-omics analysis spanning the full carcinogenesis continuum, we identify two fundamental tumor ecotypes: Infiltrated-Exhausted (IE) with adaptive T cell suppression, and Fibrotic-Excluded (FE) with a stromal barrier physically excluding lymphocytes. IE tumors follow a linear, hypermutation-driven trajectory enriched for *NOTCH1/CASP8* mutations, while FE tumors emerge via branched, chromosomally unstable evolution with *EGFR* alterations. These divergent programs are already detectable at the precancer stage. We identify the TWEAK-TNFRSF12A signaling circuit involving *SPP1*$^+$ macrophages, partial-EMT tumor cells and *COL1A1*$^+$ fibroblasts as the architectural keystone of the FE niche. Targeting TNFRSF12A *in vivo* dismantles the exclusion barrier, converting cold tumors to an inflamed state and sensitizing them to anti-PD-1 therapy. This study links genomic evolution to microenvironmental fate and provides an ecotype-guided framework for precision immunotherapy.

Keywords

Oral squamous cell carcinoma, head and neck cancer, oral leukoplakia, tumor microenvironment, cancer evolution, cancer multi-omics, cancer-associated fibroblasts, immunotherapy resistance, cancer subtyping, TNFRSF12A

Introduction

Oral squamous cell carcinoma (OSCC), the most common type of head and neck cancer, remains a major health burden with 5-year survival rates near 50%[1-3]. Anti-PD-1/PD-L1 therapy benefits only a minority (<20%) of patients[3-5], reflecting profound heterogeneity in immune evasion and tumor ecosystem architecture.

Single-cell technologies now resolve the cellular and spatial complexity of tumors, revealing specialized immune-stromal niches that govern progression and therapy response[6-12]. In OSCC, recent single-cell transcriptomics studies have uncovered stromal-immune crosstalk, such as *TDO2*$^+$ myofibroblasts inducing CD8$^+$ T cell exhaustion via the kynurenine-AhR axis[6]. Yet existing work is limited by small cohorts, incomplete coverage of the histological continuum, and absence of ecosystem subtype (ecotype)-specific immune escape analysis.

Oral carcinogenesis often arises from oral leukoplakia (OL), a premalignant lesion with 10-20% malignant transformation risk[13,14]. OL offers a critical interception window, but its evolutionary routes and ecosystem remodeling remain poorly understood. Previous pseudo-temporal modeling of adjacent normal (AJ), OL, and OSCC tissues suggest stage-specific cellular dynamics and molecular signatures[6,10,12], yet face challenges including evolutionary discontinuity, limited sample size, and the resolution limits of bulk profiling. While OSCC driver mutations are known[15], genetic determinants of OL progression and their link to immune escape remain elusive.

Here, we integrate bulk, single-cell, and spatial transcriptomics with genomic profiles to construct a comprehensive atlas of oral carcinogenesis. We reveal that OSCC is not a stochastic collection of subtypes but organizes into two fundamental ecotypes: an Infiltrated-Exhausted (IE) state, defined by high mutational burden and adaptive immune resistance, and a Fibrotic-Excluded (FE) state, driven by chromosomal instability and a specific stromal barrier. We trace these divergent fates back to the precancerous stage and identify the *TNFRSF12A* signaling axis as the keystone of the excluded niche. Targeting this axis *in vivo* effectively reprograms the tumor ecosystem and shrink the tumor, offering a validated strategy to overcome immunotherapy resistance.

Results

A Multi-Modal Atlas Reveals the Cellular Landscape of Oral Carcinogenesis

To construct a definitive, multi-dimensional map of oral carcinogenesis, we generated a substantial new multi-omics dataset from OSCC patients and integrated it with extensive public resources. Our contribution includes single-cell transcriptomics (scRNA-seq) from 30 samples, whole-genome sequencing (WGS) from 79 samples, and 49 bulk RNA-seq profiles, with a unique focus on synchronously collected adjacent normal (AJ), oral leukoplakia (OL), and squamous cell carcinoma (CA) tissues from the same patients (**Table S1**). This dataset was combined with public cohorts to assemble a comprehensive atlas encompassing 227 scRNA-seq samples, 918 bulk transcriptomic samples, 407 genomic profiles, and 12 spatial transcriptomic samples (**Figure 1A**; **STAR Methods**; **Table S1-4**). This resource spans the full disease spectrum, from normal tissue (NT) to lymph node metastasis (LN) across three distinct anatomical sites (tongue, gingiva, and buccal mucosa). To capture the full clinical landscape, our malignant scRNA-seq cohort comprised primarily treatment-naive primary tumors (n = 82), with subsets of recurrent (n = 4) and treated (n = 20) samples; these were analyzed collectively to define universal disease features, unless otherwise specified.

Unsupervised clustering of the integrated 979,352-cell scRNA-seq dataset resolved 11 major cell lineages, which were further sub-clustered into 51 distinct immune and stromal subsets (**Figure 1B and S1-3; Table S5**). This atlas revealed a clear and coordinated remodeling of the tissue microenvironment along the progression from normal tissue to invasive carcinoma (**Figure S4; Table S6**). The most striking global change was a progressive shift from a stromal-dominant to an immune-dominant landscape, characterized by a marked increase in T and B cells and a relative decrease in fibroblasts (**Figure 1C and S4A-C**). While the baseline cellular composition of normal tissues was shaped by anatomical context, such as the enrichment of T cells in the tongue and mast cells in gingival tissues, this site-specific signature diminished in carcinoma, suggesting a convergent pressure imposed by the malignant program (**Figure 1D and S5A**). Similarly, behavioral factors such as alcohol consumption were associated with specific TME alterations, including fewer fibroblasts and more B cells (**Figure S5B-D**). Despite these context-dependent variations, the overarching trajectory of profound immune infiltration and stromal reorganization emerged as the dominant signature of malignant transformation.

We further dissected the epithelial compartment, classifying cells across all tissue types into six subsets (C1-C6), distinguishing squamous (C1-C5) from glandular (C6) lineages (**Figure S3A-D**). While normal and malignant epithelium shared broad lineage features, we resolved malignant cells into four distinct functional states (Tu_C1-Tu_C4) based on copy number variation (CNV) and spatial gene signatures (**Figure S3E-G**). Tu_C1 cells exhibited differentiation features characteristic of the tumor core (*CLDN4*, *SPRR1B*), while Tu_C2 represented a transitory state. Crucially, Tu_C3 and Tu_C4 were defined by markers associated with the invasive margin (*LAMC2*, *ITGA5*) and partial epithelial-mesenchymal transition (pEMT) programs. As detailed later, these invasive-front subpopulations play a pivotal role in orchestrating the tumor's spatial ecology.

Together, this comprehensive multi-modal atlas provides an unprecedented high-resolution framework of oral carcinogenesis. It establishes that tumor development is not a random process but is characterized by stereotyped shifts in the cellular ecosystem, setting the stage for a deeper investigation into the specific cell states that orchestrate this process and define a tumor's ultimate microenvironmental fate.

Unsupervised Analysis Reveals Four Core Tissue Ecotypes Across Oral Carcinogenesis

The coordinated cellular remodeling observed during OSCC progression suggested that tumors do not evolve randomly, but instead assemble into a limited number of stable, recurring microenvironmental architectures. To test this hypothesis, we performed unsupervised consensus clustering on the relative abundances of 50 immune and stromal cell subsets (excluding double-negative T cells) across 192 tissue samples with at least 500 cells (**STAR Methods**; **Table S7**). This analysis initially yielded eight clusters (**Figure S6A–C**), which were subsequently merged into four robust and distinct tissue ecotypes according to their dominant cell populations (**Figure 2A**). The fidelity of these ecotypes was further confirmed by the Jaccard similarity index, which quantified specific patterns of cell population co-occurrence (**Figure 2B**).

The non-malignant samples largely defined the first two states. The Immune-Quiescent (IQ) ecotype represented a spectrum of stroma-dominant states (**Figure 2A-C**). The Lymphoid-Enriched (LE) ecotype was defined by a massive influx of B and naïve T cells, and the majority

(26/30) were lymph node samples (**Figure 2A-C**). Strikingly, malignant OSCC samples predominantly segregated into two fundamental but functional opposing tumor ecotypes (**Figure 2C and S6D**): the Infiltrated-Exhausted (IE) ecotype and the Fibrotic-Excluded (FE) ecotype. Consistent with this dichotomy, correlation analysis of tumor-resident cell subsets revealed a negative correlation between lymphocytes and stromal cells (**Figure S6E**), further highlighting the contrasting immune and stromal landscapes that define the two ecotypes.

The IE ecotype represents an inflamed TME, characterized by the successful penetration of the tumor by high numbers of CD4$^+$ and CD8$^+$ T cells and adaptive immune activation (**Figure 2A-B and S6F**). This robust immune presence, however, exists along a clinically relevant spectrum of exhaustion. Our analysis stratified this ecotype into two major states: an early or immature immune escape state (IE.1), which retained a higher proportion of lymphatic endothelial cells (Endo_ly), stress-responsive CD4$^+$ T cells (CD4$^+$ Tstr) and higher expression of interferon gamma (IFNγ) and TCF-1 (*TCF7*); a terminal T cell exhausted state (IE.2) where the T-cell infiltrate was profoundly dysfunctional and characterized by high levels of terminal exhaustion markers like PD-1, TIM-3, and LAG-3 (**Figure 2D and S6G**). This heterogeneity within the IE ecotype suggests that IE.1 tumors are more likely to benefit from single immune checkpoint therapies such as anti-PD-1/PD-L1[16-18], while IE.2 tumors represent a burnt-out immune state requiring combinatorial reactivation strategies.

In stark contrast, the FE ecotype represents an immune-excluded environment, defined by the failure of lymphocytes to penetrate the tumor parenchyma. This ecotype is characterized by a dense, remodeled stroma dominated by the *COL1A1*$^+$ fibroblast population and pro-tumorigenic myeloid cells (**Figure 2A-B**), which together form a physical and signaling barrier against anti-tumor immunity. Functionally, FE tumors exhibited enriched signaling pathways related to EMT, angiogenesis, TGFβ, TNFα, and hypoxia, consistent with an aggressive, immunologically cold phenotype (**Figure S6F**).

Ecotype Classification Predicts Clinical Outcomes and Aligns with Molecular Subtypes

Building upon our identification of two functionally distinct tumor ecotypes (IE and FE) in malignant OSCC samples, we sought to validate their cellular characteristics and spatial distribution within tissue architecture using immunohistochemistry (IHC) and spatial transcriptomics (ST). To this end, we performed IHC staining for representative immunomarkers (CD3 and CD68) and fibrosis markers (type I collagen) in our OSCC cohort with matching scRNA-Seq. In line with our single-cell analysis, IE tumors showed increased $CD8^+$ T cell infiltration, while FE tumors showed a coordinated increase in macrophages and type I collagen accompanied by decreased lymphoid T cell infiltration (**Figure 3A-C**). Notably, these markers exhibited spatial co-occurrence or exclusion, suggesting the possible existence of ecotype-specific niches. Moreover, complementary spatial transcriptomics analysis further confirmed these ecotype-specific patterns (**Figure 3D-E**).

To validate the clinical relevance and generalizability of this ecotype framework, we developed a robust gene-pair-based classifier to distinguish IE from FE tumors using bulk transcriptomic data (**STAR Methods; Figure S7A-C; Table S8**). When applied to four independent OSCC cohorts, this classification was a powerful predictor of patient survival: patients with IE-classified tumors exhibited significantly better prognoses than those with FE tumors (**Figure 3F**). Notably, this prognostic power was highly specific to head and neck cancer, as it was not observed in other gastrointestinal tumors (**Figure S7D**). To explore the potential cellular determinants underlying these prognostic differences, we mapped the gene expression signatures of cell subsets identified from our single-cell data onto the bulk transcriptome and correlated their abundance with overall survival across patient cohorts. As expected, several B cell and T cell subsets enriched in the IE ecotype were consistently associated with favorable outcomes in at least two independent datasets (**Figure 3G**). In contrast, stromal components which was enriched in the FE ecotype, were strongly associated with poor clinical outcomes. These findings suggest that specific immune and stromal cell populations contribute to the distinct prognostic landscapes of each tumor ecotype. Furthermore, our predicted ecotypes correlated strongly with established classifications (**Table S9**); IE tumors were enriched in HPV-positive tumors and the atypical RNA subtype from TCGA, while FE tumors were predominant in the classical RNA subtype and associated with specific copy number subtypes, providing an early link to underlying genomic drivers (**Figure 3H and S7E-H**).

Taken together, these orthogonal validations establish that OSCC is not a homogeneous entity but is fundamentally stratified into two principal ecotypes with distinct cellular architectures, molecular programs, and clinical outcomes. This framework provides a new lens through which to investigate the specific mechanisms of immune evasion that define them.

Ecotype-Specific Cellular Remodeling Diverges at the Precancerous Stage

To determine whether the Infiltrated-Exhausted (IE) and Fibrotic-Excluded (FE) ecotypes arise through distinct developmental programs, we analyzed immune and stromal lineage dynamics across the AJ-OL-CA continuum. While comparative analyses without ecotype stratification identified broad shifts in 31 immune and 9 stromal cell subsets during disease progression (**Figure S8A**; **Table S6**), stratifying this trajectory by ecotype revealed that these changes are not uniform but follow two divergent paths that separate as early as the precancerous stage (**Figure 4A-B**).

Specifically, the IE trajectory was characterized by the early infiltration of B and T lymphocytes beginning in the AJ and OL stages. In contrast, the FE trajectory was marked by the premature accumulation of stromal populations, including fibroblasts and pericytes, at these same early timepoints. While these distinct microenvironmental signatures were detectable in precancerous lesions, the divergence intensified during the transition to malignancy. The OL-to-CA transition in the IE lineage was defined by the specific depletion of cytotoxic T cells and a compensatory increase in IgG-secreting plasma cells. Conversely, the FE lineage underwent a profound stromal expansion characterized by the surge of *SPP1*$^+$ macrophages and *COL1A1*$^+$ fibroblasts, establishing the hallmarks of the immunosuppressive, fibrotic niche.

To validate these divergent trajectories in larger cohorts, we projected our single-cell signatures onto bulk gene expression profiles of OL and CA samples. Strikingly, unsupervised clustering revealed that precancerous OL lesions did not cluster together as a single intermediate state, but instead segregated into two distinct groups resembling the canonical IE and FE tumor ecotypes (**Figure 4C and S8B**). This dichotomy was further supported by pseudotime trajectory analysis, which reconstructed two independent progression routes from OL to CA that aligned with the IE-like and FE-like states (**Figure 4D and S8C-D; Table S10**). Crucially, longitudinal analysis of paired samples revealed that the vast majority (10/11) of OL-CA pairs retained the same ecotype identity during malignant transformation (**Figure 4E**).

Collectively, these findings demonstrate that the remodeling of the tumor microenvironment is not a stochastic late-stage event. Instead, immune and stromal compositions follow ecotype-specific evolutionary tracks that are imprinted at the precancerous stage, suggesting that the immune microenvironment of the tumor is shaped prior to the onset of invasion (**Figure 4F**).

Distinct Genomic Evolutionary Patterns Underlie the Divergence of Tumor Ecotypes

To investigate the genetic mechanisms underlying these distinct microenvironmental states, we performed whole-genome sequencing (WGS) on 30 OSCC patients, analyzing matched AJ, OL and CA tissues. Genome-wide copy number analysis revealed that genomic instability is initiated early in the disease course. While OL lesions exhibited modest alterations compared to frank carcinoma, they already harbored detectable somatic CNVs (SCNVs) (**Figure 5A**). Crucially, the magnitude of this early instability stratified with the future tumor phenotype: lesions associated with the Fibrotic-Excluded (FE) ecotype displayed a trend toward higher SCNV burdens than those associated with Infiltrated-Exhausted (IE) tumors (**Figure 5B; Table S11**). This suggests that the propensity for chromosomal instability, a hallmark of the FE trajectory, is a feature emerging at the precancerous stage.

To reconstruct the clonal history of these lesions, we inferred copy number profiles at single-cell resolution. We mapped subclonal architectures across the AJ-OL-CA continuum, identifying dominant meta-clones (Clones A-D) and tracing their expansion. Two distinct phylogenetic patterns emerged. The IE lineage predominantly followed a linear evolutionary trajectory (3/4 patients), characterized by shared ancestral clones between OL and CA followed by the stepwise acquisition of private alterations (**Figures 5C and S9A**). In contrast, the FE lineage exhibited branched evolutionary dynamics (3/5 patients), marked by minimal clonal overlap and the parallel emergence of divergent lineages (**Figures 5D and S9B**).

Somatic mutation analysis further delineated these evolutionary strategies. IE tumors exhibited a trend toward elevated tumor mutation burden (TMB), evident even in OL lesions (**Figure 5E; Table S11**). To quantify these opposing selection forces, we developed a Chromosomal Instability (CIN) Index, which stratified lesions into two dominant trajectories: high genomic instability (CIN Index > 0, associated with FE) versus mutational accumulation (CIN Index < 0, associated with

IE) (**Figure 5F-G**). Notably, in 70.6% (12/17) of matched OL-CA pairs, this genomic polarity was consistent across stages, suggesting that ecotypic identity is often imprinted prior to malignant transformation, although a minority of cases demonstrated polarity reversal, indicating potential plasticity.

Finally, we identified specific ecotype-associated SCNVs and mutations (SNV/INDEL) that characterize each trajectory. The FE genome was defined by focal amplifications in oncogenes (e.g., *EGFR* at 7p11.2, *TMEM184B* at 22q13.1) and deletions in key immunoregulatory genes (e.g., *CD74*, *CCL21*, *CD274*, *JAK2*), which directly correlated with reduced gene expression (**Figure 5H-I**; **Table S12**). Integration with single-cell data confirmed that these alterations were present in specific malignant subpopulations, linking genotype to phenotype (**Figure 5J and S9C**). Notably, 60.5% (92/152) of FE tumors harbored at least one *EGFR* alteration (SNV/INDEL or SCNV), compared to only 37.0% (71/192) of IE tumors. Conversely, the IE genome was enriched for mutations in *NOTCH1*, *CASP8*, and *FLG*. The co-occurrence of *NOTCH1* and *CASP8* mutations, which was mutually exclusive with *TP53* alterations, emerged as a specific mutational signature of the IE trajectory (**Figure 5K and S9D**; **Table S12**).

Taken together, these findings demonstrate that OSCC evolution is governed by two distinct genomic programs: an adaptive, mutation-driven linear trajectory that shapes the IE ecotype, and a chromosomally unstable, branched trajectory that drives the formation of the FE ecotype.

Distinct Immune Evasion Strategies Define the Infiltrated-Exhausted and Fibrotic-Excluded Ecotypes

The profound differences between the IE and FE ecotypes suggested they employ fundamentally distinct strategies to subvert anti-tumor immunity. To dissect these mechanisms, we integrated their pseudo-bulk transcriptomes and selected 453 immunoregulatory genes within the cancer-immune cycle framework for differential gene expression analysis, ultimately identifying 56 ecotype-specific immunoregulatory genes (**Figure 6A-B; Table S13; STAR Methods**). This revealed two divergent modes of immune escape.

The IE ecotype was characterized by an immunogenic paradox. On one hand, it showed high expression of pro-inflammatory cytokines such as *IFNG*, *IL2*, and *IL21*, indicating a robust, ongoing immune response. On the other hand, this was countered by the significant upregulation of a broad suite of co-inhibitory checkpoint receptors, including *PDCD1* (PD-1), *CTLA4*, *LAG3*, *TIGIT*, *KIR3DL2* and *KLRG1* (**Figure 6B; Table S13**). This molecular signature confirms that the primary immune evasion strategy in the IE ecotype is the active, adaptive suppression of tumor-infiltrating lymphocytes via multiple, redundant exhaustion pathways.

In stark contrast, the Fibrotic-Excluded ecotype suppressed anti-tumor immunity at an earlier stage in the cancer-immunity cycle by erecting a complex stromal barrier. This was driven by the overexpression of genes involved in extracellular matrix (ECM) remodeling (e.g., *COL1A1* and *LOXL2*), chemotaxis of neutrophils and helper T cells (*CXCL1/5/8*, *CCL20*), and anti-inflammatory cytokines (*IL10*, *IL1A*, *IL1RN*). Critically, alongside this stromal program, FE tumors upregulated a distinct set of non-canonical immune modulatory genes. Among the most prominent were *CD276* (B7-H3) and, notably, *TNFRSF12A* (also known as Fn14 and TWEAKR), which were both highly expressed in the FE ecotype and strongly associated with poor overall survival across multiple independent OSCC cohorts (**Figure 6B-D**).

The expression dynamics of *TNFRSF12A* and *CD276* marked them as a key potential orchestrator of the FE phenotype. Their expression was not only a hallmark of the FE ecotype but also progressively increased during malignant transformation from OL to CA as shown in two independent RNA-Seq datasets (**Figure 6E-F**). Cellular-level analysis revealed that *TNFRSF12A* was predominantly expressed by the two key cellular architects of the FE niche: the *COL1A1*$^+$ fibroblast population and specific malignant epithelial cells (**Figure 6C-D and S10A-B**), which was also confirmed in another independent scRNA-Seq cohort[19] (**Figure S10C**). Consistent with its role in an immune-excluded TME, *TNFRSF12A* and *CD276* expression was negatively correlated with T/NK and B cell infiltration and positively correlated with the infiltration of fibroblasts and myeloid cells in tumors (**Figure 6G and S10D**). Moreover, immunohistochemistry and ST analyses revealed that *TNFRSF12A* is preferentially overexpressed in the FE ecotype and exhibits an inverse correlation with lymphocyte infiltration (**Figure 6H-J and 3A-E**). By manipulating macrophages to be classified based on distinct transcriptional markers[20], we found that *TNFRSF12A* and *CD276* expression levels were positively correlated with macrophage markers *IL1B* and *SPP1*, while negatively correlated with the *CXCL9:SPP1* ratio (**Figure 6K** and **S10E**). Previous studies have demonstrated that a higher *CXCL9:SPP1* ratio is positively correlated with increased lymphocyte infiltration and is associated with improved patient prognosis[19].

Unraveling the spatial choreography of tumor ecosystem that stratifies OSCC molecular subtypes remains critically understudied. Using a public spatial transcriptome dataset of OSCC[11], we assessed cell type abundance and identified colocalization of specific cell subsets in different tumor ecotypes (**STAR Methods**). Interestingly, we found that *COL1A1*$^+$ fibroblast was mainly enriched in the area adjacent to the tumor edge and was spatially closer to tumor epithelial cells than other stromal cells like *CXCL12*$^+$ fibroblast (**Figure S10F**), suggesting that this cell subset may be potentially associated with the stronger tumor cell progression in FE ecotype. Notably, *COL1A1*$^+$ fibroblast exhibited spatial co-localization with *SPP1*$^+$/*CXCL9*$^+$ macrophages and tumor epithelial cells in representative FE-like samples (**Figure 6L and S11A**), a pattern also observed in IE-like lesions although with notably diminished *COL1A1*$^+$ fibroblast abundance (**Figure S11B**).

Together, these results delineate the distinct immune evasion programs of the two OSCC ecotypes. While the IE ecotype relies on classic T-cell exhaustion, the FE ecotype builds a complex, fibrotic niche orchestrated by a distinct set of molecular players, prominently featuring the *TNFRSF12A* signaling axis as a potential therapeutic vulnerability.

A *TNFRSF12A*$^+$ Cellular Circuit Orchestrates the Fibrotic-Excluded Niche

Having established the Fibrotic-Excluded (FE) ecotype as a clinically aggressive, lymphocyte-poor state, we sought to identify the core cellular players and signaling pathways that construct and maintain this niche. The defining cellular architects of the FE ecotype were the *COL1A1*$^+$ fibroblast population and a specific subset of malignant cells (Tu_C3 and Tu_C4) enriched for invasive, mesenchymal-like features (**Figure 4A and 7A-B**). Critically, a key molecular link between these two dominant cell populations was their high and relatively specific co-expression of the immunoregulatory gene *TNFRSF12A*, which we had identified as a top prognostic marker for the FE ecotype (**Figure S2G and S5G**).

While a direct lineage relationship between these cell types remains to be definitively established, multiple lines of evidence pointed to a close functional and transcriptional relationship between the pEMT-like Tu_C3/Tu_C4 tumor cells and the *COL1A1*$^+$ fibroblasts (**Figure 7B-C and S12A**). Tu_C3 and Tu_C4 tumor cells shared many marker genes with *COL1A1*$^+$ fibroblast determined by Jaccard Index (**Figure 7C**). These shared genes are linked to pEMT (e.g., *TGFBI*, *FN1*, *TNC*) and ECM remodeling (e.g., *COL5A1*, *MMP1* and *LGALS1*). Transcriptional trajectory inference suggested a high degree of similarity (**Figure 7D**), and importantly, genetic analysis revealed that *COL1A1*$^+$ fibroblasts shared SCNV subclone with tumor cells (**Figure S12B**), and even more somatic mutations with the tumor clone than any other stromal cell type, indicating a uniquely intimate connection to the malignant compartment (**Figure 7E and S12C**).

Given this close association, we hypothesized that these *TNFRSF12A*$^+$ tumor cells and fibroblasts form a core interaction network that sustains the FE niche. We used CellChat to systematically map the intercellular communication patterns and found that IE tumors exhibited elevated intercellular communication (**Figure S12D**), primarily characterized by robust signal emission and reception by T cells, while FE ecotype displayed only increased signaling activity from epithelial cells and macrophages (**Figure S12E**). In specific signal pathways, the FE niche was characterized by the activation of pathways involved in stromal remodeling (TWEAK), fibrotic niche establishment (*SPP1*), and inflammation suppression (*IL10*, *ANXA1*), which mechanistically support lymphocyte exclusion **(Figure 7F; Table S14**). Notably, the immunosuppressive signal *IL10* is already present at the precancerous stage of FE tumors (**Figure S12F**).

Dissection of these pathways revealed a core, self-reinforcing signaling circuit in FE ecosystem. We found that *SPP1*$^+$ and *CXCL9*$^+$ macrophages were the primary source of TWEAK (*TNFSF12*), the ligand for TWEAKR (*TNFRSF12A*). This macrophage-derived TWEAK acts directly on both *TNFRSF12A*$^+$ tumor cells (Tu_C3/Tu_C4) and *TNFRSF12A*$^+$/*COL1A1*$^+$ fibroblasts to promote their proliferation and pro-fibrotic activity (**Figure 7G-H**). In turn, these tumor cells and fibroblasts secrete factors such as *ANXA1* that can recruit and polarize macrophages, closing the loop. Spatial analysis confirmed the physical basis of this circuit, showing the intimate co-localization of these three cell types and the enrichment of TWEAK and SPP1 signaling within the FE niche (**Figure S12G**).

Taken together, our analysis reveals that the Fibrotic-Excluded ecotype is actively maintained by a self-reinforcing signaling triad of *TNFRSF12A*$^+$ tumor cells, *TNFRSF12A*$^+$/*COL1A1*$^+$ fibroblasts, and *SPP1*$^+$/*CXCL9*$^+$ macrophages. This circuit, with the *TNFSF12-TNFRSF12A* (TWEAK pathway) axis at its core, represents a critical architectural hub and a prime therapeutic target for dismantling the immunosuppressive FE niche.

Targeting TNFRSF12A Reprograms the Fibrotic-Excluded Niche and Sensitizes Tumors to Immunotherapy

To validate *TNFRSF12A* as a therapeutic target for reprogramming the Fibrotic-Excluded (FE) microenvironment, we utilized the MOC2 murine OSCC model. This line was selected because it establishes rapidly progressing tumors with dense stroma and limited T-cell infiltration[21], closely recapitulating the human FE phenotype (**Figure 8A**). We generated *Tnfrsf12a* knockout (Fn14-KO) MOC2 cells using CRISPR/Cas9 and confirmed effective gene deletion and loss of protein expression (**Figure 8B** and **S13A**).

Disruption of *TNFRSF12A* significantly impaired tumor progression *in vivo*. In immunocompetent C57BL/6J mice, tumors derived from Fn14-KO cells exhibited significantly reduced volume and weight compared to wild-type controls without affecting overall body weight (n = 4 mice/group; $p < 0.001$, Wilcoxon rank-sum test; **Figure 8C-D**). Crucially, this growth suppression was markedly attenuated when the experiment was repeated in immunodeficient nude mice (**Figure S13B-E**), demonstrating that targeting *TNFRSF12A* is not merely cell-intrinsic but relies on the restoration of an adaptive anti-tumor immune response.

To dissect the microenvironmental changes underlying this reactivation, we performed single-cell RNA sequencing on representative Fn14-KO (n = 2) and wild-type murine tumors (n = 2). The results revealed a striking ecotype switch. Fn14 knockout effectively dismantled the FE architecture, leading to a significant reduction in collagen-secreting fibroblasts and a concurrent surge in T-cell infiltration (**Figure 8E-F and S13F-H**). This was corroborated by immunohistochemistry showing increased CD8$^+$ T-cell density in murine tumors (**Figure 8G**).

Mechanistically, this remodeling was driven by a fundamental rewiring of the tumor's signaling network. Fn14-KO tumor cells exhibited enhanced antigen presentation capacity and inflammatory signaling (e.g., IFN-α/γ), replacing the proliferative and DNA-repair programs that dominated wild-type cells (**Figure 8H**). Consistent with our human atlas, the macrophage compartment also shifted towards a pro-inflammatory phenotype, evidenced by a significant increase in the *Cxcl9*:*Spp1* ratio (**Figure 8I-J**). Furthermore, cell-cell interaction analysis confirmed a shift from exclusion to engagement: the knockout tumors displayed enhanced communication via T-cell recruitment (*CXCL*) and regulation (*PD-L1*, *CD80/86*) axes (**Figure S13I-J**). Interestingly, the infiltrating CD8$^+$ T cells in the knockout tumors displayed elevated expression of exhaustion markers like *Pdcd1* (PD-1) and *Lag3* (**Figure 8K**), indicating that the tumor had transitioned from a cold FE state to an inflamed but exhausted IE-like state.

This successful conversion to an IE-like phenotype suggested that *TNFRSF12A* inhibition renders the tumor susceptible to checkpoint blockade. To test this combinatorial strategy, we treated wild-type MOC2-bearing mice with an *Fn14* antagonist (L524-0366), anti-*PD-1*, or the combination. While both monotherapies showed moderate efficacy, the combination of *Fn14* antagonist and anti-*PD-1* yielded superior anti-tumor activity (**Figure 8L-N**). Immunohistochemical experiments also showed enhanced CD8$^+$ T cell infiltration in the treatment group (**Figure S13K-L**).

Together, these data provide *in vivo* proof-of-concept that targeting the *TNFRSF12A* axis can dismantle the immunosuppressive FE barrier, converting cold tumors into hot ones and unleashing the therapeutic potential of anti-PD-1 immunotherapy.

Discussion

Despite advances in molecular characterization, the clinical management of OSCC remains largely one-size-fits-all, contributing to stagnating survival rates[1,22]. Although molecular subtypes have been defined using bulk transcriptomics (e.g., TCGA)[23], these classifications have failed to translate into precision therapeutic strategies, largely because they obscure the complex cellular interactions within the tumor microenvironment and treat the tumor as a static entity without a history. Furthermore, current classification systems typically focus solely on established malignancy, overlooking the critical evolutionary window of the precancerous stage[15,23-25]. Here, by constructing a high-resolution multi-omics atlas spanning the full continuum from normal tissue to carcinoma, we resolve these limitations. We identify two fundamental, clinically actionable tumor ecotypes, Infiltrated-Exhausted (IE) and Fibrotic-Excluded (FE), and demonstrate that they are the endpoints of distinct evolutionary trajectories traceable back to precancerous oral leukoplakia.

A key conceptual advance of this work is linking specific modes of genomic evolution to distinct immune landscapes. While we cannot infer strict causality from our current data, we observe a parallelism between the tumor's genomic trajectory and its resulting microenvironment. The IE ecotype aligns with a linear, hypermutator trajectory characterized by elevated tumor mutation burden and frequent *NOTCH1* and *CASP8* mutations. This pattern is consistent with established models where high mutational load generates neoantigens that drive T-cell infiltration, subsequently exerting selective pressure for adaptive resistance mechanisms[26]. In contrast, the FE ecotype aligns with a branched, chromosomal unstable trajectory characterized by frequent *EGFR* alterations. This association resonates with emerging evidence in other malignancies where high levels of chromosomal instability actively suppress anti-tumor immunity[27,28]. Our observation that some of these genomic and microenvironmental features are detectable in precancerous leukoplakia suggests that the tumor's immune trajectory is often imprinted prior to malignancy.

We further elucidate the cellular mechanisms by which the FE tumors construct a physical barrier. We identify a mechanism of "niche engineering" where pEMT-phenotype tumor cells and *COL1A1*$^+$ fibroblasts cooperate to construct a fibrotic shield. Our transcriptomic analysis reveals a striking functional convergence between these two populations: pEMT-activated tumor cells acquire a fibroblast-like, matrix-producing phenotype, upregulating collagens and the *TNFRSF12A* receptor. While the definitive lineage relationship between these compartments remains to be fully resolved, the identification of shared somatic alterations between tumor cells and a subset of *COL1A1*$^+$ fibroblasts suggests a complex interplay, potentially involving stromal mimicry by malignant cells. This aligns with previous studies in head and neck cancer identifying pEMT programs at the invasive front that actively remodel the extracellular matrix[9]. Regardless of whether these cells arise via direct trans-differentiation or represent a distinct state of epithelial plasticity, the functional outcome is clear: the tumor autonomously supplements the stromal pool, accelerating the construction of the immune-excluded FE niche.

Our atlas also reframes the cancer-associated fibroblast compartment. We resolve three transcriptionally distinct fibroblast subpopulations, of which *CXCL12*$^{+}$ and *F13A1*$^{+}$ fibroblasts predominate in adjacent normal and precancerous tissues, whereas *COL1A1*$^{+}$ fibroblasts specifically expand during malignant transformation and are enriched in the FE ecotype. This population exhibits a distinct immunomodulatory profile characterized by high expression of *MMP1* and *LRRC15*, both identified as key mediators of fibroblast-driven immune escape in a recent pan-cancer study[30]. Prior single-cell studies in OSCC have identified specific CAF subsets related to immunosuppression and immune exclusion, including *TDO2*$^{+}$, *POSTN*$^{+}$, and *CXCL8*$^{+}$ CAFs[6,10,29]. Consistent with their cancer-associated identity, our data confirm that these are functional states within the *COL1A1*$^{+}$ population. However, a previously unrecognized consequence is that all these CAF subsets are, by extension, enriched in the FE ecotype. Our study identifies the TWEAK-TNFRSF12A axis as an additional route through which *COL1A1*$^{+}$ fibroblasts actively maintain the FE niche, linking them to a self-reinforcing signaling circuit that physically bars lymphocyte infiltration.

Our findings further support a paradigm shift toward ecotype-based precision immunotherapy, where therapeutic selection is guided by the specific cellular architecture of the tumor ecosystem. Within this framework, we identify distinct targetable vulnerabilities: adaptive checkpoints (e.g., PD-1, CTLA-4) for the IE ecotype, and structural exclusion drivers (e.g., TNFRSF12A, CD276) for the FE ecotype. TNFRSF12A (Fn14/TWEAKR) is a well-characterized driver of pathological tissue remodeling in chronic inflammatory diseases like rheumatoid arthritis and inflammatory bowel disease, where it promotes fibrosis and angiogenesis[31-33]. Our data suggest that FE tumors co-opt this conserved mechanism, establishing a self-reinforcing signaling circuit with *SPP1*$^{+}$ macrophages to physically exclude adaptive immunity. The progressive upregulation of *TNFRSF12A* during the OL-to-CA transition further underscores its role as a dynamic mediator of niche evolution rather than merely a static marker of the established tumor.

This mechanistic understanding resolves the paradox of why previous therapies have failed and validates a new combinatorial approach. While PD-1 blockade fails in excluded environments, and

TWEAKR-targeted agents (e.g., enavatuzumab) have shown limited efficacy as monotherapies in early clinical trials[34], our *in vivo* data demonstrate that these strategies are synergistic. Targeting TNFRSF12A effectively dismantles the fibrotic barrier, converting a cold FE tumor into a hot phenotype. Crucially, this induced infiltration renders the tumor susceptible to checkpoint blockade, as the newly recruited T cells require protection from adaptive immune suppression. This validates a dismantle and reactivate strategy[35,36]: using anti-TNFRSF12A to breach the physical barrier and anti-PD-1 to sustain the anti-tumor response.

In summary, we advance the understanding of OSCC by establishing a comprehensive ecotype framework that redefines the disease landscape from a static classification of molecular subtypes to a dynamic model of genetically constrained evolutionary tracks. By delineating the Infiltrated-Exhausted and Fibrotic-Excluded ecotypes as endpoints of these tracks, we provide a unified biological rationale for patient stratification that integrates genomic history, stromal architecture, and immune phenotype. The validation of TNFRSF12A as a druggable node for converting cold tumors to immunotherapy-responsive states offers a concrete translational path forward. Future efforts should focus on developing biomarkers to detect these ecotypic signatures in the precancerous stage, potentially enabling molecular interception strategies that target the fibrotic niche before malignancy is fully established. Ultimately, this framework bridges the gap between genomic evolution and precision oncology, offering an ecotype-guided path to improve outcomes in head and neck cancer.

Limitations of this study

Our study has limitations. First, while the concordance of ecotypes in paired precancerous and malignant tissues implies early imprinting, confirming the determinism of these trajectories requires larger longitudinal cohorts and mechanism studies. Second, our link between genomic architecture and immune ecotypes is associative; definitive proof that specific alterations play a causal role requires functional perturbation models. Third, while shared somatic mutations strongly suggest a tumor-to-fibroblast transition, establishing the precise rate of this plasticity requires lineage-tracing models. Finally, prospective trials are necessary to confirm that our ecotype classifier can accurately stratify patients for ecotype-directed therapy.

Data availability

The scRNA-seq, RNA-seq, and WGS data generated in this study can be obtained from the NGDC GSA database (https://ngdc.cncb.ac.cn/) under project number PRJCA039175 with accession number HRA011232, HRA011241, HRA011251, and CRA039278. Public scRNA-seq data were obtained from multiple databases; details are shown in **Table S2**. Public transcriptome data can be obtained from the NCBI GEO database (accession numbers GSE41613, GSE42743, GSE30784, GSE85195, GSE131568, GSE246050). The GEO accession number for public spatial transcriptomics data is GSE208253. Detailed information on all public datasets is shown in **Tables S2** and **S4**.

Acknowledgments

We thank Prof. Songlin Wang (SUSTech) for providing help on data collection. This work was supported by Shenzhen Medical Research Fund (A2303044), National Natural Science Foundation of China (82270239), Shenzhen Science and Technology Program (JCYJ20250604144238050, JCYJ20220530115207016) to SS; Sanming Project of Medicine in Shenzhen (SZSM202311009) to GZ; Shenzhen Science and Technology Program (JCYJ20220530152406015) to HX; Shenzhen Medical Research Fund (A2303043) and National Natural Science Foundation of China (82203477) to RX. Computational analysis of this work was supported by the Center for Computational Science and Engineering at Southern University of Science and Technology.

Author contributions

Conceptualization, YZ, SS; data curation, YZ, YR; formal analysis, YZ; investigation, YZ, RX, YF, YR, CM; methodology, YZ, SS; project administration, YZ, SS; resources, YF, HJ, TW, ZZ, JT, HX, GZ; software, YZ, JL; funding acquisition, GZ, SS; supervision, GZ, SS; validation, RX, JY, RZ; visualization, YZ; writing - original draft, YZ, SS; writing - review & editing, YZ, GZ, SS.

Declaration of interests

The patent application related to the article was submitted to the State Intellectual Property Office of China (Application No.: 202511878789.5) for YZ, GZ, and SS.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Gemini (Google) in order to improve the readability, clarity, and language quality of the manuscript. After using this tool or service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Supplemental information

Document S1. Figures S1–S13.

Table S1. Patient information of this study cohort.

Table S2. Public scRNA-seq datasets used in this study.

Table S3. Single-cell sequencing information of this study cohort.

Table S4. Bulk transcriptome datasets used in this study.

Table S5. Top 30 DEGs for each minor cell subpopulations. Related to Figure 1.

Table S6. Tissue remodeling analysis results. Related to Figure 1.

Table S7. Tissue ecotype analysis data and results. Related to Figure 2.

Table S8. Selected gene pairs for the classifier. Related to Figure 3.

Table S9. Tumor ecotype and clinical information of public patient cohorts. Related to Figure 3.

Table S10. Sample information for trajectory inference of oral carcinogenesis using bulk data. Related to Figure 4.

Table S11. Public genomic data used in this study. Related to Figure 5.

Table S12. Ecotype-specific SNVs and SCNVs. Related to Figure 5.

Table S13. Immunomodulatory genes list and ecotype-specific immunoregulatory genes. Related to Figure 6.

Table S14. Ecotype-specific intercellular communication signals. Related to Figure 7.

Figure legends

Figure 1. A multi-modal atlas reveals the cellular landscape of oral carcinogenesis.

(A) Overview of the dataset composition across the disease spectrum, including normal tissue (NT), adjacent normal (AJ), oral leukoplakia (OL), squamous cell carcinoma (CA), and lymph nodes with (LN_P) or without (LN_N) metastasis.

(B) Integrated UMAP visualization of ~1 million single cells from 227 samples.

(C) Boxplots showing the relative abundance of major cell lineages across tissue states. Cell proportions are calculated as a percentage of total cells in the tumor microenvironment (TME) per sample. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns = not significant by Wilcoxon test.

(D) Comparative analysis of immune and stromal cell proportions across anatomical sites (tongue, gingiva and buccal) during oral carcinogenesis. Statistical significance was the same as in C.

Figure 2. Unsupervised analysis identifies four core tissue ecotypes.

(A) Consensus hierarchical clustering of 192 tissue samples based on the relative abundance of 50 immune and stromal cell subpopulations. The analysis identifies four robust ecotypes: Immune-Quiescent (IQ), Lymphoid-Enriched (LE), Infiltrated-Exhausted (IE), and Fibrotic-Excluded (FE). Top annotations denote cluster identify, tissue pathology, clinical stage, risk factors, anatomical site and cohort identity.

(B) Network diagrams visualizing the defining cellular architectures of each ecotype. Nodes represent cell subsets; edge thickness indicates the strength of co-occurrence in the ecotype (Jaccard index).

(C) Sankey diagram illustrating the mapping of tissue types to the four ecotypes.

(D) Comparison of cellular composition between the inflamed (IE.1) and exhausted (IE.2) states of the IE ecotype. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by Wilcoxon test.

Figure 3. Spatial validation and clinical relevance of IE and FE tumor ecotypes.

(A-B) Representative immunohistochemistry (IHC) staining of IE (A) and FE (B) human OSCC tumors with matched scRNA-Seq. Markers: CD3 (T cells), CD68 (Macrophages), and Picro-Sirius Red (PSR, type I collagen).

(C) Quantification of immune and stromal markers from the IHC cohort ($n = 9$ fields /ecotype). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by Wilcoxon test.

(D-E) Spatial transcriptomics validation of the IE (D) and FE (E) ecotypes. Slide visualizations show H&E staining, pathological annotations and the spatial expression of key cell-type markers (*KRT14*, *CD3D*, *CD8A*, *CD68*, *COL1A1*).

(F) Kaplan-Meier survival analysis of overall survival (OS) in four independent OSCC cohorts stratified by the IE vs. FE classifier. P-values from log-rank test.

(G) Heatmap summarizing the prognostic association (Hazard Ratio, HR) of individual immune and stromal cell subpopulations across multiple datasets. Green: favorable prognosis ($HR < 1$); Purple: poor prognosis ($HR > 1$). Opacity indicates statistical significance (p-values from log rank tests). Key prognostic cell names are highlighted.

(H) Sankey diagram illustrating ecotypes associated with established OSCC classifications, clinical stages and anatomical sites. P-values from Fisher's exact test.

Figure 4. Dynamic remodeling of cell compositions during tumor evolution.

(A) Paired analysis of differential cell proportions across the AJ-OL-CA continuum in patients classified into IE and FE ecotypes. Only stromal and immune cells are shown. Colors indicate fold-change; intensity indicates statistical significance (Wilcoxon test, p-values adjusted by the Benjamini-Hochberg procedure).

(B) Summary of ecotype-specific cell dynamics across the AJ-OL-CA continuum. Cell subpopulations are categorized by their temporal patterns of change. Only lineages displaying divergent trajectories between IE and FE ecotypes are shown (shared lineages are shown in Figure S8A).

(C) Unsupervised clustering heatmap of bulk transcriptome-based deconvolution for OL samples across public datasets. Precancerous lesions stratify into groups resembling either the IE-like or FE-like ecotype. Colors indicate z-scores of cell abundance.

(D) Pseudotime trajectory analysis based on inferred immune and stromal cell composition profiles of bulk transcriptome samples. Left: points colored by tissue type (OL vs CA). Right: samples labeled by ecotype (IE-like vs FE-like).

(E) Concordance analysis of ecotype assignments between paired OL and CA samples ($n = 11$), demonstrating stability of the ecotype during transformation.

(F) Schematic model showing the gradual emergence of microenvironmental divergence during oral carcinogenesis.

Figure 5. Distinct genomic alterations underlie IE and FE evolutionary trajectories.

(A) Genome-wide SCNV profiles inferred from WGS samples.

(B) Quantification of total SCNV scores in paired OL and CA samples across IE and FE tumors. Left: our WGS cohort. Right: TCGA validation. P-values from Wilcoxon test.

(C-D) Single-cell CNV inference and clonal phylogeny for a representative IE tumor (C) and FE tumor (D). Distinct clones were reconstructed based on CNV patterns (upper); matched CNV profiles from WGS are shown below. Pie charts depict the clonal composition across tissue types, showing linear vs. branched evolution.

(E) Tumor mutation burden (TMB) comparison between paired OL and CA samples from IE and FE ecotypes. Left: our WGS cohort. Right: TCGA validation. P-values from Wilcoxon tests.

(F) Chromosomal instability (CIN) index between paired OL and CA samples. HGI: high genomic instability; MA: mutational accumulation.

(G) Comparison of the CIN index between established IE and FE ecotypes. P-value from Wilcoxon test.

(H) Heatmap of genomic regions with significant SCNV differences between IE-like and FE-like subtypes. Key driver genes are highlighted. P-values from Fisher's exact test.

(I) Boxplots comparing gene expression of selected SCNV-associated genes between IE-like and FE-like subtypes. P-values from Wilcoxon test.

(J) UMAP showing single-cell expression of selected ecotype-associated drivers (*EGFR*, *CD74*, *HMGB1*, *TMEM184B*) in tumor cells from this study cohort.

(K) Oncoplot showing enrichment of somatic mutations in specific driver genes in IE-like (n = 180) vs. FE-like (n = 143) OSCC tumors in TCGA. P-values from Fisher's exact test.

Figure 6. Ecotype-specific immune evasion strategies and the *TNFRSF12A* axis.

(A) Schematic of the strategy to identify key immunomodulatory genes distinguishing IE and FE ecotypes. L-R: Ligand-Receptor; TAA: tumor-associated antigen.

(B) Volcano plot of differentially expressed immunomodulatory genes between IE and FE tumors. Genes are annotated by functional category.

(C) Multi-layer integrative analysis of key genes. Top: Expression bubble plot across single-cell populations. Middle: Prognostic Hazard Ratio (HR) across four datasets (colored by direction and log-rank p-value). Bottom: Expression trajectories across the AJ-OL-CA transition (p-value from one-way ANOVA). Genes with prognostic associations in at least three cohorts are in purple.

(D) Single-cell expression of the FE-specific checkpoints *TNFRSF12A* and *CD276* across epithelial and stromal subpopulations.

(E-F) Validation of *TNFRSF12A* and *CD276* upregulation during oral carcinogenesis in two independent bulk cohorts. $**p < 0.01$, $***p < 0.001$; Wilcoxon test.

(G) Pearson correlation analysis between *TNFRSF12A* expression and the abundance of major cell populations in the OSCC TME.

(H) IHC validation of TNFRSF12A expression in representative IE and FE tumors.

(I) Quantification of TNFRSF12A protein levels from IHC (n = 9 fields /ecotype). $**p < 0.01$; Wilcoxon test.

(J) Spatial transcriptomics visualization of *TNFRSF12A* expression in representative IE-like and FE-like tumors.

(K) Pearson correlation between *TNFRSF12A* expression and macrophage markers in bulk RNA-seq data.

(L) Spatial co-localization analysis in an FE tumor. Left: Predicted spatial abundance of all selected cell subsets using cell2location is depicted by color intensity. Middle: Density plot showing the Kullback-Leibler (KL) divergence between cell distributions compared to a null distribution. Empirical p-values determined by permutation testing and $p < 0.05$ indicates significant spatial co-localization. Right: The spatial abundance of selected cell susbets.

Figure 7. Ecotype-specific intercellular signaling networks shape the tumor niche.

(A) Distribution of tumor cell subsets (Tu_C1-Tu_C4) across IE.1, IE.2, and FE ecotypes.

(B) Functional annotation of tumor cell subsets based on hallmark gene sets and epithelial differentiation programs.

(C) Transcriptomic similarity between tumor and stromal cell subpopulations measured by the Jaccard index of top 100 marker genes.

(D) Trajectory analysis of tumor cells and fibroblasts inferred by RNA velocity and PAGA. Arrows denote predicted directionality.

(E) Somatic mutation burden estimated from single-cell SNV inference across tumor and stromal subpopulations in FE samples.

(F) Pathway-level comparison of tissue- and ecotype-specific ligand–receptor (L–R) signaling activity.

(G) Cell-cell communication networks revealing upregulated crosstalk among tumor cells, macrophages, and fibroblasts in FE tumors.

(H) Dot plot showing the expression across key cell subsets of key ligand-receptor pairs driving the macrophage-fibroblast-tumor signaling axis in FE tumors.

Figure 8. *TNFRSF12A* blockade suppresses tumor growth and enhances antitumor immunity *in vivo*.

(A) Schematic of the experimental design. Wild-type (WT) and Fn14 (Tnfrsf12a) knockout (KO) MOC2 cells were injected into C57BL/6J mice. Tumors were harvested at Day 14 for scRNA-seq.

(B) Immunoblot confirming loss of Fn14 protein in KO cells.

(C) Representative images of excised WT and Fn14-KO tumors.

(D) Quantification of tumor weight (left), tumor volume (middle), and body weight (right). Data: mean ± SEM; ***$p < 0.001$ by Wilcoxon test. N = 4 mice inoculated bilaterally per group.

(E) UMAP visualization of scRNA-seq data from murine tumors (n = 2 per group), annotated by major cell population.

(F) Bar plots showing the shift in relative proportions of T cells and fibroblasts upon Fn14 knockout.

(G) IHC validation of $CD8^+$ T-cell infiltration in WT vs. KO tumors. Right: Quantification of $CD8^+$ density. ***$p < 0.001$, Wilcoxon test. N = 3 tumors per group, 5 random fields per tumor.

(H) Dot plot showing the average transcription activity of functional gene sets in Fn14 KO and WT tumor cells across two biological replicates.

(I) UMAP of macrophages in murine tumors, highlighting the expression of polarization markers.

(J) Quantification of the *Cxcl9:Spp1* expression ratio in macrophages.

(K) Expression of cytotoxicity and exhaustion markers in T cells from WT vs. KO tumors.

(L) Schematic of the anti-Fn14 (L524-0366) and anti-PD-1 combination therapy trial.

(M) Efficacy of combination therapy. Quantification of tumor weight (left), tumor volume (middle) and body weight (right). Data: mean ± SEM; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Wilcoxon rank-sum test). N = 3 mice inoculated bilaterally per group.

(N) Representative images of excised murine tumors from the combination therapy trial.

Methods

Human Specimens

This study included 30 patients diagnosed with oral squamous cell carcinoma (OSCC), confirmed by pathological examination. All patients were recruited from Shenzhen People's Hospital (Shenzhen, China). Prior to sample collection, written informed consent was obtained from all participants in accordance with ethical guidelines. The sample collection protocols were reviewed and approved by the Ethics Committees of Southern University of Science and Technology (Approval No. 20220115) and Shenzhen People's Hospital (Approval No. LL-KY-2022170). This study uniquely focused on synchronously collected paired samples of tumor tissue, precancerous lesions (oral leukoplakia), and adjacent normal tissues from surgical resection specimens. Each sample was carefully dissected and weighed immediately after collection. Based on experimental requirements, samples were divided into multiple portions for single-cell RNA sequencing (scRNA-seq), bulk RNA sequencing, whole-genome sequencing (WGS), and validation assays. In addition to this cohort, publicly available datasets were integrated, comprising 328 genomic samples, 822 bulk transcriptomics samples, 197 scRNA-seq samples, and 12 spatial transcriptome samples from OSCC tissue (see **Table S2**,**S4, and S11** for accession numbers of public data).

Cell Lines

The murine OSCC cell line MOC2 used in this study was obtained from Bluefbio (Shanghai) Biology Technology Development Co., Ltd. Cells were cultured in DMEM (high glucose) supplemented with 10% FBS and 1% penicillin/streptomycin at 37°C in a humidified 5% CO_2 atmosphere. MOC2 were routinely authenticated by short tandem repeat (STR) profiling and confirmed to be free of mycoplasma contamination using the MycoAlert™ Mycoplasma Detection Kit.

Mice

All animal experiments were conducted in accordance with institutional guidelines and approved by the Institutional Animal Care and Use Committee (IACUC) of the Southern University of Science and Technology. Mice were housed under specific pathogen-free (SPF) conditions with a 12-hour light/dark cycle and *ad libitum* access to food and water. Mice were monitored for signs of morbidity, including excessive tumor volume (maximum size permitted by the IACUC), hunched posture, ruffled fur, weight loss, paralysis, dyspnea, or inability to groom/feed. Animals exhibiting these signs were humanely euthanized.

Single-Cell RNA Sequencing and Preprocessing

Fresh tissue samples were dissociated into single-cell suspensions using the GEXSCOPE Tissue Dissociation Mix. Two droplet-based scRNA-seq platforms were utilized: five samples were prepared using the GEXSCOPE® Single Cell RNA Library Kit, and 25 samples were prepared using the MobiNova®-100 system with the MobiCube® High-Throughput 3' Single-Cell Transcriptome Kit. All procedures followed the manufacturers' standard protocols. Purified libraries were sequenced on the DNBSEQ-T7 platform (150 bp paired-end), generating ~100 Gb of raw data per sample.

Raw data were processed using MOBIDROP’s proprietary MobiVision platform (v3.0) or Singleron’s CeleScope (v1.17.0), depending on the platform. To remove technical artifacts and background noise, CellBender[37] was applied to the raw count matrix with a learning rate of 2.5×10^{-5} and an expected cell number of 10,000. The filtered matrix then underwent rigorous quality control: 1) empty droplet removal: cells with <200 detected genes or <1,000 unique molecular identifiers (UMIs) were excluded; 2) viability screening: cells with >15% mitochondrial content or >5% hemoglobin expression were discarded; 3) multiplet removal: putative multuplets were identified and removed using a combination of library complexity screening (>6,000 genes) and DoubletFinder (v2.0.4)[38]. The filtered gene count matrix was normalized, scaled and subjected to principal component analysis (PCA) using the Seurat R package (v4.4.0)[39]. Publicly available datasets were processed using this identical pipeline to ensure comparability.

Bulk RNA Sequencing and Processing

Total RNA was extracted from tissue samples using a mechanical fragmentation device (XM-26A) following the manufacturer's protocol and subjected to ribosomal RNA depletion (Ribo-off rRNA Depletion Kit, Vazyme) through sequence-specific probe hybridization (from 95°C to 22°C at a ramp rate of 0.1°C/sec) followed by RNase H and DNase I digestion. Libraries were prepared using the Hieff NGS® EvoMax RNA Library Prep Kit (dUTP) (Yeasen Biotechnology), quantified via Qubit 4.0 (Thermo Fisher) and size-validated by agarose electrophoresis. Libraries were sequenced on the DNBSEQ-T7 platform (150 bp paired-end), yielding ~30 million raw reads per sample.

Gene transcription quantification followed the ENCODE RNA-seq guidelines[40]. Briefly, raw sequencing data were quality-controlled (FastQC) and trimmed (TrimGalore) using default parameters. Clean reads were then aligned to the human GRCh38 reference genome using STAR[41]. Gene expression was quantified using RSEM[42] and normalized to Counts Per Million (CPM). Differential expression analysis was conducted using DESeq2[43]. Genes with a Benjamini-Hochberg adjusted p-value < 0.05 and absolute fold-change > 1.5 were considered differentially expressed. Batch effects between cohorts were corrected using the ComBat function in the sva R package[44].

Whole Genome Sequencing and Processing

Genomic DNA was extracted from tissue samples using mechanical fragmentation (XM-26A). Whole-genome sequencing (WGS) libraries were prepared by the Hieff NGS® DNA Library Prep Kit (Yeasen Biotechnology). Libraries were amplified for 7 cycles (Canace® Pro Amplification Mix) and quantified via Qubit, and sequenced on the DNBSEQ-T7 platform with 150 bp paired-end reads, generating ~90 Gb of raw data per sample.

Raw sequencing reads were QC-filtered (FastQC and TrimGalore) and aligned to GRCh38 using BWA-MEM2. Somatic variants were called using the GATK Best Practices pipeline[45], using matched adjacent normal tissues as the reference. Functional annotation was performed with Funcotator from GATK. Somatic copy number variations (SCNVs) were identified using CNVkit[46] and gene-level SCNVs were annotated and quantified with GISTIC2[47].

Generation of *Tnfrsf12a* Knockout Cell Lines

A lentiviral CRISPR/Cas9 system was used to target *Tnfrsf12a* (Fn14). An sgRNA sequence (5’-AGGAGCTGCCGCTAGAGCAT-3’) targeting *Tnfrsf12a* was cloned into lentiCRISPRv2 lentiviral vector (Addgene, #52961) and confirmed by sequencing. Next, MOC2 cells were co-transfected with the transfer plasmid (lentiCRISPRv2), psPAX2 (Addgene #12260), and pVSVg (Addgene #8454), using PEI. Viral supernatant was harvested, concentrated, and used to transduce MOC2 cells. Positive clones were selected with puromycin (1 μg/mL) for two weeks. Single colonies were screened by immunoblotting and validated by PCR/Sanger sequencing (Forward 5’

-CGCAAGCTTCTGACCTGC-3’; Reverse 5’-ACACACCCTGATAGAAGCCC-3’).

In Vivo Mouse Tumor Model

Wild-type and *Tnfrsf12a* (Fn14)-KO MOC2 cells were resuspended in PBS (5×10^5 cells in 100 µL) and subcutaneously injected into two flanks of female C57BL/6 or BALB/c nude mice (age 4-6 weeks, Gempharmatech). For combination therapy, mice bearing wild-type tumors were randomized into four groups: Vehicle, Fn14 antagonist (L524-0366) alone, anti-PD-1 alone, or in combination. L524-0366 (6.25mg/kg per dose) was injected intraperitoneally (i.p.) every other day. Anti-PD-1 (BioXCell #BE0146) or IgG2a isotype control (BioXCell #BP0089) was injected i.p. for a total of three doses (150 µg per 20 g body weight per dose). Each mouse was inoculated bilaterally in the abdominal flanks with MOC2 or FN14-KO tumor cells. Tumor volume ($0.5 \times Length \times Width^2$) and body weight were monitored every other day. Mice were euthanized when tumor burden reached 1,500 mm^3, per IACUC-approved endpoints.

Histological Validation

For immunohistochemistry (IHC), tumor specimens were fixed in 10% neutral-buffered formalin, embedded in paraffin, and sectioned at 4 µm. Sections underwent heat-mediated antigen retrieval, permeabilization (0.5% PBST), hydrogen peroxide treatment, and blocking. Sections were then incubated overnight at 4°C with primary antibody against CD3 (MXB, MAB-0740), CD8a (Servicebio, GB15068), CD68 (Gene, sc-20060), and CD266/Fn14 (MedChemExpress, HY-P82931). Detection was performed using an appropriate secondary antibody and DAB chromogen. For Picro-Sirius Red staining, sections were dewaxed, rehydrated, and stained using a kit (Servicebio, G1078) according to the manufacturer's protocol. Slides were dehydrated, cleared in xylene, and mounted.

Stained sections were imaged using an Aperio VERSA 8 microscope. For quantification, three random fields of view per section were captured at 20x magnification. The percentage of Picro-Sirius Red positive area (collagen deposition) or the density of positive cells (for IHC markers) was quantified using ImageJ software[48].

Public scRNA-seq Data Collection and Integration

We systematically collected scRNA-Seq datasets from Gene Expression Omnibus (GEO), Genome Sequence Archive (GSA), and Human Cell Atlas (HCA). Datasets encompassing diverse oral sites and disease states were processed to obtain quantitative matrices (metadata available in **Table S2**).

Large-scale integration was performed using Seurat v5. Each dataset underwent independent quality control, and top 2,000 highly variable genes (HVG) identification. SketchData was used to subsample 10,000 representative cells per dataset to facilitate efficient integration via IntegrateLayers (using Harmony[49]). The full dataset was mapped to this reference using ProjectIntegration and ProjectData. Clustering and dimensionality reduction (UMAP) were performed to delineate distinct cell populations and visualize the integrated landscape.

Cell Annotation

Cells were annotated in a two-step process: major lineage clustering followed by detailed sub-clustering. For T/B cells, mitochondrial, ribosomal and TCR/BCR genes were excluded from HVGs to prevent clustering bias. Cell identities were assigned based on the top 50 DEGs (FindMarkers: $FC > 1.5$, $FDR < 0.05$) and canonical markers from literature[6,9,20,50-52]. Annotations for cross-checked using functional scoring of published marker lists. The independent single-cell validation dataset (GSE234933[19]) were annotated using the main atlas as the reference via the FindTransferAnchors and TransferData functions from Seurat.

Gene Signature Score Analysis

Functional signatures were scored using AddModuleScore (Seurat) for single-cell data or ssGSEA (GSVA[53]) for bulk data. Metabolic pathways and Hallmark gene sets were collected from the Molecular Signature Database (MSigDB[54]). Inflammation-related signaling pathways and other functional gene sets were collected from literature[55-57].

Identification of Ecotypes

We analyzed 192 samples containing ⩾500 cells. The relative proportions of 50 cell types (excluding epithelial cells and DNT cells) were used for consensus hierarchical clustering (ConsensusClusterPlus[58]) to identify robust ecotypes. The dendrogram revealed five primary branches: IQ, LE, IE.1, IE.2, and FE. Since IE.1 and IE.2 were both tumor-enriched and characterized by high lymphocyte infiltration, they were merged into a broader Infiltrated-Exhausted ecotype for downstream classification. Co-occurrence patterns were verified by computing the Jaccard index between cell subpopulations (using samples in the top 25% of relative abundance as the set).

Ecotype Classifier Construction

A classifier to distinguish IE from FE tumors was built for bulk transcriptomic data. Feature selection involved: 1) Identification of DEGs between IE and FE pseudo-bulk profiles ($padj < 0.01$, $|\log2FC| > 1$); 2) Intersection with the top 6,000 protein-coding genes ranked by pseudo-bulk expression levels; 3) Selection of gene pairs with expression ratios achieving $AUC > 0.95$ across four annotated single-cell cohorts; and 3) Spatial verification using dataset GSE208253, requiring gene pair ratios to correlate with lymphocyte-enriched spatial regions ($|r| \geq 0.5$, $p < 0.1$). Selected gene pairs are listed in **Table S8**. Six machine learning classifiers (SVM, Decision Tree, Random Forest, Naïve Bayes, kNN and Neural Network) were benchmarked using a 70/30 train/test split and 5-fold cross-validation. Performance was evaluated by accuracy, sensitivity, specificity, and AUC.

Cell Population Composition Analysis

We quantified the enrichment or depletion of each cell population across different tissues using the ratio of observed to expected (Ro/e) cell numbers as described in literature[59]. $Ro/e > 1$ indicates that the cell population is enriched in the tissue; $Ro/e < 1$ implies depletion.

Single-Cell Copy Number and Mutation Analysis

Single-cell CNVs were inferred using inferCNV[60] under a clustering model, with paired normal epithelial or stromal cells serving as the reference for normalization. A gene expression cutoff of 0.1 and denoising were applied to minimize technical noise. CNV detection accuracy was further improved using a hidden Markov model (HMM). Genomic instability was quantified via single-cell CNV scores, calculated as the sum of squared copy number amplitudes.

To infer the clonal structure and phylogeny, we screened patients with paired adjacent normal samples and performed copy number inference using inferCNV. All cells from non-adjacent normal samples were grouped into different subclones. We merged CNV signals into a unified subclonal consensus set and identified meta clones (Clone A, B, C, and D) across different patients using adjacency matrix. We visualized the phylogenetic tree for each patient and assigned a numerical identifier to each meta clone node. Based on the defined clonal branching distance and the proportion of cells from paired precancerous samples (OL) and tumor samples (four tumor cell subpopulations) within different subclones of each patient, patients were further divided into two major evolutionary patterns: linear evolution and branching evolution. The branching evolution pattern is characterized by a clear monophyletic structure in the patient's phylogenetic tree.

Single-cell SNVs were identified from scRNA-seq data using SComatic[61]. Firstly, we prepared a cell label annotation file of high-quality cells in each sample to infer the number of SNVs in each cell subpopulation. Next, scRNA-seq alignment files were split into cell-type-specific BAM files, followed by the collection of base counts at each genomic position for each cell type. Potential somatic mutations were detected based on a beta-binomial distribution modeling the background error rate. To minimize artifacts and germline polymorphisms, putative SNVs were further filtered by excluding those located in repetitive or homopolymer-adjacent regions, proximal variants (<5 bp apart, except validated dinucleotide substitutions), and recurrent artifacts identified by a panel of normals (PON) constructed from non-tumor datasets. Additionally, putative SNVs were intersected with high-confidence regions of the human genome, as recommended. Mutation burden was calculated as the total number of somatic mutations detected in each cell type divided by the total number of callable sites in all cells of the same type.

Single-Cell Trajectory and RNA Velocity Analysis

RNA velocity was inferred using velocyto (v0.17.17)[62] and scVelo (v0.3.1)[63] using default parameters. Partition-based graph abstraction (PAGA) was used to infer the direction. Visualization was performed with scvelo.pl.velocity_embedding_stream (velocity stream plots) and scvelo.pl.paga (directed PAGA graphs).

Cell-Cell Communication Analysis

Using the annotated cell types and processed expression matrix, we applied CellChat[64] (v1.6.1) with its integrated database to infer signaling pathways mediating intercellular interactions with default parameters. Communication probabilities were computed to quantify interaction strength and delineate key signaling networks. Similarly, the CellChat v2.2.0 was used to infer communication between cell subpopulations in ST data based on default parameters.

Genomic Analysis

Somatic mutations in OL and CA samples were analyzed using Mutation Annotation Format (MAF) files generated from WGS alignment and variant calling. The R package maftools[65] was used to process MAF data, summarize mutational profiles, and generate oncoplots.

Genomic instability was quantified using the chromosome instability (CIN) index[66], which is calculated as the difference between scaled somatic mutation burden (SMB) and scaled copy number burden (CNB). Briefly, SMB was defined as the number of nonsynonymous mutations detected per million bases of the genome, whereas CNB was calculated as the proportion of protein-coding genes affected by copy number alterations (amplifications or deletions). Both SMB and CNB values were normalized to the range [0,1] using min–max scaling. Samples with CIN index > 0 were classified as exhibiting chromosomal instability dominance, whereas those with CIN index < 0 were considered to display a hypermutator phenotype.

Spatial Transcriptomics Data Analysis

scRNA-seq data and ST data were integrated using cell2location[67] to map cell population proportion to spatial spots. To define the spots containing cell types of interest, we first calculated the global mean abundance of each cell type across all spots. A cell type was designated as a candidate identity for a given spot if its abundance in that spot exceeded the corresponding global mean. For spots harboring multiple candidate identities, the cell type with the highest abundance among candidates was assigned as the final unique identity. Spots lacking any candidate identity were labeled as “Other Cells”.

To investigate the spatial co-localization of paired cell types, we utilized the deconvolution results from cell2location as input for the calculate_Density function in the Nebulosa R package[68]. This function computes the two-dimensional density distribution of each cell subtype across the ST slide. We then quantified the dissimilarity between the spatial distributions of any two cell subsets using the Kullback-Leibler (KL) divergence, implemented in the philentropy R package[69]. To assess the statistical significance of observed co-localization patterns, we generated a null distribution of KL divergence values by repeatedly shuffling the cell type deconvolution matrix and recalculating the density distributions and KL divergence for randomly selected subsets of spots (80% of total spots per iteration). Each process was repeated 1,000 times to construct the null distribution. The empirical p-value was calculated by dividing the number of permutations in which the observed KL divergence exceeded null distribution values by the total permutations.

Public Bulk Transcriptome Data Collection and Analysis

To validate the dynamic transcriptomic changes during oral carcinogenesis and to explore the prognostic relevance of cell subpopulations or genes, we performed bulk RNA sequencing on 49 paired OSCC samples (AJ, OL, CA, and LN) and integrated eight public datasets (Table S1 and S4). Normalized bulk RNA-seq data and corresponding clinical information of head and neck squamous cell carcinoma (HNSCC) were downloaded from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). Molecular subtype annotations were obtained from the TCGA Pan-Cancer Atlas[23]. Additionally, two large-scale primary OSCC microarray datasets (GSE42743 and GSE41613), which include matched clinical information, were retrieved from the Gene Expression Omnibus (GEO). To investigate gene expression alterations during oral carcinogenesis, we obtained four GEO microarray datasets (GSE30784, GSE85195, GSE131568, and GSE246050) containing precancerous lesion samples. For microarray data, probe-level intensities mapping to the same gene were averaged to generate unified gene expression values. To test whether the ecotype classifier can be used for other gastrointestinal cancers, we downloaded bulk RNA-seq data and clinical information of colon cancer (COAD), esophageal cancer (ESCA), rectal cancer (READ), stomach cancer (STAD) and bile duct cancer (CHOL) generated by TCGA.

To investigate the biological significance of ecotypes using public data, we first assigned ecotype labels to each sample. Briefly, we inferred cell type composition for each sample by combining a cell type-specific gene list with the ssGSEA algorithm, which provided relative activity scores. To calculate cell type-specific genes of minor cell subsets, we initially calculated DEGs relative to their major cell type and then de-intersected the resulting gene sets by removing overlap with other minor cell types under the same major lineage. The top 30 genes from this filtered list were selected as markers for each minor cell subset. Finally, unsupervised hierarchical clustering was performed to categorize the bulk samples into IE-like and FE-like groups.

Inference of Malignant Transformation Trajectory with bulk transcriptomes

To reveal the underlying structure of malignant transformation from OL tissue to OSCC, we applied manifold learning-based dimensionality reduction techniques and minimum spanning trees to the cell subpopulation composition matrix using R package Monocle2[70]. Dimensionality reduction was performed with the DDRTree algorithm via the reduceDimension function, followed by cell ordering and branch point detection with the orderCells function. We then examined the cellular states and branching trajectories in the reduced space and visualized them with the plot_cell_trajectory function.

Survival analysis

In all survival analyses, we used the log-rank test to calculate the p-value between groups, and the Kaplan-Meier method to plot survival curves. For survival association analysis of cell composition, the samples were divided into high-activity and low-activity groups based on the optimal cut-point calculated by the surv_cutpoint function in the survminer R package[71]. The risk ratio was calculated using a univariate Cox proportional hazards model by survival R package[72].

Statistical Analysis

All analyses were performed in R (v4.3.0) and Python (v3.10.3). Comparisons of cell type proportions between two tissue groups were conducted using the Wilcoxon rank-sum test, while paired Wilcoxon rank-sum tests were applied for matched samples from the same patient. Associations between cell ecotypes and histological or molecular subtypes, as well as enrichment of genetic variations across ecotypes, were examined using two-sided Fisher's exact tests.

References

1 Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229-263, doi:10.3322/caac.21834 (2024).

2 Zheng, R. *et al.* Cancer incidence and mortality in China, 2016. *J Natl Cancer Cent* **2**, 1-9, doi:10.1016/j.jncc.2022.02.002 (2022).

3 Chi, A. C., Day, T. A. & Neville, B. W. Oral cavity and oropharyngeal squamous cell carcinoma--an update. *CA Cancer J Clin* **65**, 401-421, doi:10.3322/caac.21293 (2015).

4 Ferris, R. L. *et al.* Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* **375**, 1856-1867, doi:10.1056/NEJMoa1602252 (2016).

5 Burtness, B. *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* **394**, 1915-1928, doi:10.1016/s0140-6736(19)32591-7 (2019).

6 Hu, S. *et al.* TDO2+ myofibroblasts mediate immune suppression in malignant transformation of squamous cell carcinoma. *J Clin Invest* **132**, doi:10.1172/jci157649 (2022).

7 Cillo, A. R. *et al.* Immune Landscape of Viral- and Carcinogen-Driven Head and Neck Cancer. *Immunity* **52**, 183-199.e189, doi:10.1016/j.immuni.2019.11.014 (2020).

8 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, doi:10.1038/s41467-021-27619-4 (2021).

9 Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624.e1624, doi:10.1016/j.cell.2017.10.044 (2017).

10 Liu, Z. L. *et al.* Single cell deciphering of progression trajectories of the tumor ecosystem in head and neck cancer. *Nat Commun* **15**, 2595, doi:10.1038/s41467-024-46912-6 (2024).

11 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, doi:10.1038/s41467-023-40271-4 (2023).

12 Sun, L. *et al.* Single-cell and spatial dissection of precancerous lesions underlying the initiation process of oral squamous cell carcinoma. *Cell Discov* **9**, 28, doi:10.1038/s41421-023-00532-4 (2023).

13 Mello, F. W. *et al.* Prevalence of oral potentially malignant disorders: A systematic review and meta-analysis. *J Oral Pathol Med* **47**, 633-640, doi:10.1111/jop.12726 (2018).

14 Evren, I. *et al.* Annual malignant transformation rate of oral leukoplakia remains consistent: A long-term follow-up study. *Oral Oncol* **110**, 105014, doi:10.1016/j.oraloncology.2020.105014 (2020).

15 Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576-582, doi:10.1038/nature14129 (2015).

16 Siddiqui, I. *et al.* Intratumoral Tcf1(+)PD-1(+)CD8(+) T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. *Immunity* **50**, 195-211.e110, doi:10.1016/j.immuni.2018.12.021 (2019).

17 Ayers, M. *et al.* IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* **127**, 2930-2940, doi:10.1172/jci91190 (2017).

18 Sade-Feldman, M. *et al.* Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* **175**, 998-1013.e1020, doi:10.1016/j.cell.2018.10.038 (2018).

19 Bill, R. *et al.* CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. *Science* **381**, 515-524, doi:10.1126/science.ade2292 (2023).

20 Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809.e723, doi:10.1016/j.cell.2021.01.010 (2021).

21 Kono, M., Saito, S., Egloff, A. M., Allen, C. T. & Uppaluri, R. The mouse oral carcinoma (MOC) model: A 10-year retrospective on model development and head and neck cancer investigations. *Oral Oncol* **132**, 106012, doi:10.1016/j.oraloncology.2022.106012 (2022).

22 Johnson, D. E. *et al.* Head and neck squamous cell carcinoma. *Nature Reviews Disease Primers* **6**, 92, doi:10.1038/s41572-020-00224-3 (2020).

23 Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e296, doi:10.1016/j.cell.2018.03.022 (2018).

24 Chung, C. H. *et al.* Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* **5**, 489-500, doi:10.1016/s1535-6108(04)00112-6 (2004).

25 Walter, V. *et al.* Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PLoS One* **8**, e56823, doi:10.1371/journal.pone.0056823 (2013).

26 Sun, S. *et al.* The role of neoantigens and tumor mutational burden in cancer immunotherapy: advances, mechanisms, and perspectives. *Journal of hematology & oncology* **18**, 84, doi:10.1186/s13045-025-01732-z (2025).

27 Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, doi:10.1126/science.aaf8399 (2017).

28 Bakhoum, S. F. & Landau, D. A. Chromosomal Instability as a Driver of Tumor Heterogeneity and Evolution. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a029611 (2017).

29 Choi, J. H. *et al.* Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat Commun* **14**, 1055, doi:10.1038/s41467-023-36691-x (2023).

30 Gao, Y. *et al.* Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation. *Cancer Cell* **42**, 1764-1783.e1710, doi:10.1016/j.ccell.2024.08.020 (2024).

31 Dohi, T. & Burkly, L. C. The TWEAK/Fn14 pathway as an aggravating and perpetuating factor in inflammatory diseases: focus on inflammatory bowel diseases. *J Leukoc Biol* **92**, 265-279, doi:10.1189/jlb.0112042 (2012).

32 Dohi, T. *et al.* TWEAK/Fn14 pathway: a nonredundant role in intestinal damage in mice through a TWEAK/intestinal epithelial cell axis. *Gastroenterology* **136**, 912-923, doi:10.1053/j.gastro.2008.11.017 (2009).

33 Shaikh, F. S. *et al.* TWEAK receptor (Fn14) exacerbates TNF-α-induced inflammation in rheumatoid arthritis synovial fibroblasts and influences response to anti-TNF-α therapy. *Cell Mol Immunol* **23**, 315-328, doi:10.1038/s41423-026-01386-y (2026).

34 Lam, E. T. *et al.* Phase I Study of Enavatuzumab, a First-in-Class Humanized Monoclonal Antibody Targeting the TWEAK Receptor, in Patients with Advanced Solid Tumors. *Mol*

Cancer Ther **17**, 215-221, doi:10.1158/1535-7163.Mct-17-0330 (2018).

35 Butterfield, L. H. & Najjar, Y. G. Immunotherapy combination approaches: mechanisms, biomarkers and clinical observations. *Nature reviews. Immunology* **24**, 399-416, doi:10.1038/s41577-023-00973-8 (2024).

36 Goswami, S., Pauken, K. E., Wang, L. & Sharma, P. Next-generation combination approaches for immune checkpoint therapy. *Nature immunology* **25**, 2186-2199, doi:10.1038/s41590-024-02015-4 (2024).

37 Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods* **20**, 1323-1335, doi:10.1038/s41592-023-01943-7 (2023).

38 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e324, doi:10.1016/j.cels.2019.03.003 (2019).

39 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529, doi:10.1016/j.cell.2021.04.048 (2021).

40 Hitz, B. C. *et al.* The ENCODE Uniform Analysis Pipelines. *bioRxiv*, 2023.2004.2004.535623, doi:10.1101/2023.04.04.535623 (2023).

41 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

42 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

43 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

44 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).

45 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

46 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873, doi:10.1371/journal.pcbi.1004873 (2016).

47 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).

48 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

49 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

50 Chu, Y. *et al.* Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nature Medicine* **29**, 1550-1562, doi:10.1038/s41591-023-02371-y (2023).

51 Guo, X. *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nature Medicine* **24**, 978-985, doi:10.1038/s41591-018-0045-3 (2018).

52 Kieffer, Y. *et al.* Single-Cell Analysis Reveals Fibroblast Clusters Linked to

Immunotherapy Resistance in Cancer. *Cancer Discovery* **10**, 1330-1351, doi:10.1158/2159-8290.Cd-19-1384 (2020).

53 Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).

54 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).

55 Kieffer, Y. *et al.* Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer. *Cancer Discov* **10**, 1330-1351, doi:10.1158/2159-8290.Cd-19-1384 (2020).

56 Zheng, L. *et al.* Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474, doi:10.1126/science.abe6474 (2021).

57 Ma, J. *et al.* A blueprint for tumor-infiltrating B cells across human cancers. *Science* **384**, eadj4857, doi:10.1126/science.adj4857 (2024).

58 Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572-1573, doi:10.1093/bioinformatics/btq170 (2010).

59 Wang, R. *et al.* Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. *Nat Med* **27**, 141-151, doi:10.1038/s41591-020-1125-8 (2021).

60 *inferCNV of the Trinity CTAT Project*, <https://github.com/broadinstitute/inferCNV> (2019).

61 Muyas, F. *et al.* De novo detection of somatic mutations in high-throughput single-cell profiling data sets. *Nat Biotechnol* **42**, 758-767, doi:10.1038/s41587-023-01863-z (2024).

62 La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498, doi:10.1038/s41586-018-0414-6 (2018).

63 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414, doi:10.1038/s41587-020-0591-3 (2020).

64 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088, doi:10.1038/s41467-021-21246-9 (2021).

65 Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**, 1747-1756, doi:10.1101/gr.239244.118 (2018).

66 Heiser, C. N. *et al.* Molecular cartography uncovers evolutionary and microenvironmental dynamics in sporadic colorectal tumors. *Cell* **186**, 5620-5637.e5616, doi:10.1016/j.cell.2023.11.006 (2023).

67 Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* **40**, 661-671, doi:10.1038/s41587-021-01139-4 (2022).

68 Alquicira-Hernandez, J. & Powell, J. E. Nebulosa recovers single-cell gene expression signals by kernel density estimation. *Bioinformatics* **37**, 2485-2487, doi:10.1093/bioinformatics/btab003 (2021).

69 Terry M Therneau, T. L., Atkinson Elizabeth, Crowson Cynthia. Philentropy: Information Theory and Distance Quantification with R. *Journal of Open Source Software* **3**, doi:10.21105/joss.00765 (2018).

70 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by

pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

71 Alboukadel Kassambara, M. K., Przemyslaw Biecek, Scheipl Fabian. *survminer: Drawing Survival Curves using 'ggplot2'*, <https://rpkgs.datanovia.com/survminer/index.html> (2025).

72 Terry M Therneau, T. L., Atkinson Elizabeth, Crowson Cynthia. *survival: A package for survival analysis in R*, <https://github.com/therneau/survival> (2024).

Figure 1

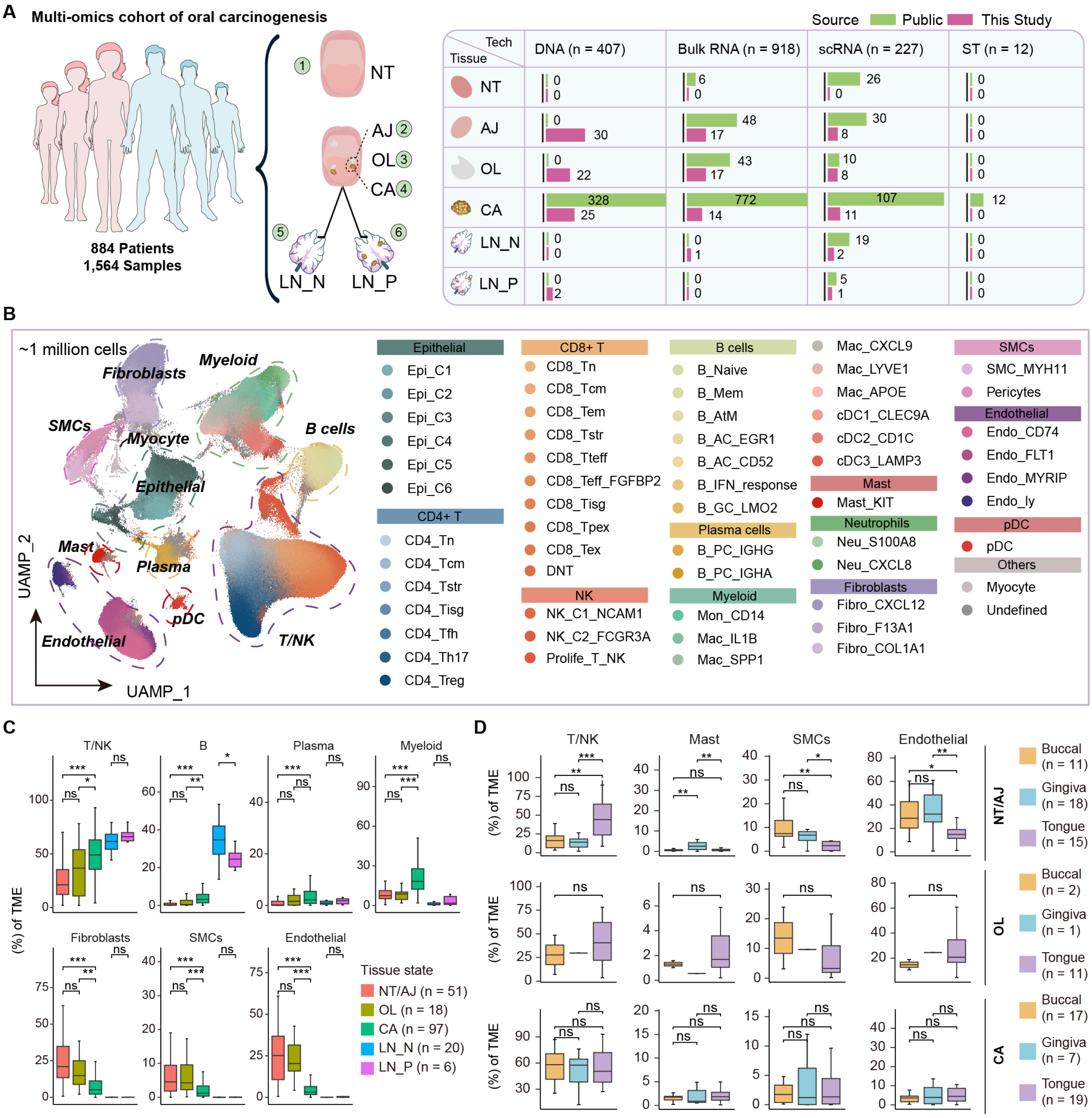

A
Multi-omics cohort of oral carcinogenesis
884 Patients
1,564 Samples
① NT
AJ ②
OL ③
CA ④
⑤ LN_N
⑥ LN_P
Source Public This Study
Tech
Tissue
DNA (n = 407)
Bulk RNA (n = 918)
scRNA (n = 227)
ST (n = 12)
NT 0 0 6 0 26 0 0 0
AJ 0 30 48 17 30 8 0 0
OL 0 22 43 17 10 8 0 0
CA 328 25 772 14 107 11 12 0
LN_N 0 0 0 1 19 2 0 0
LN_P 0 2 0 0 5 1 0 0
B
~1 million cells
Fibroblasts
Myeloid
SMCs
Myocyte
B cells
Epithelial
Mast
Plasma
pDC
Endothelial
T/NK
UAMP_2
UAMP_1
Epithelial
Epi_C1
Epi_C2
Epi_C3
Epi_C4
Epi_C5
Epi_C6
CD4+ T
CD4_Tn
CD4_Tcm
CD4_Tstr
CD4_Tisg
CD4_Tfh
CD4_Th17
CD4_Treg
CD8+ T
CD8_Tn
CD8_Tcm
CD8_Tem
CD8_Tstr
CD8_Tteff
CD8_Teff_FGFBP2
CD8_Tisg
CD8_Tpex
CD8_Tex
DNT
NK
NK_C1_NCAM1
NK_C2_FCGR3A
Prolife_T_NK
B cells
B_Naive
B_Mem
B_AtM
B_AC_EGR1
B_AC_CD52
B_IFN_response
B_GC_LMO2
Plasma cells
B_PC_IGHG
B_PC_IGHA
Myeloid
Mon_CD14
Mac_IL1B
Mac_SPP1
Mac_CXCL9
Mac_LYVE1
Mac_APOE
cDC1_CLEC9A
cDC2_CD1C
cDC3_LAMP3
Mast
Mast_KIT
Neutrophils
Neu_S100A8
Neu_CXCL8
Fibroblasts
Fibro_CXCL12
Fibro_F13A1
Fibro_COL1A1
SMCs
SMC_MYH11
Pericytes
Endothelial
Endo_CD74
Endo_FLT1
Endo_MYRIP
Endo_ly
pDC
pDC
Others
Myocyte
Undefined
C
T/NK
B
Plasma
Myeloid
Fibroblasts
SMCs
Endothelial
(%) of TME
*** * ** ns
Tissue state
NT/AJ (n = 51)
OL (n = 18)
CA (n = 97)
LN_N (n = 20)
LN_P (n = 6)
D
T/NK
Mast
SMCs
Endothelial
(%) of TME
*** ** * ns
NT/AJ
Buccal (n = 11)
Gingiva (n = 18)
Tongue (n = 15)
OL
Buccal (n = 2)
Gingiva (n = 1)
Tongue (n = 11)
CA
Buccal (n = 17)
Gingiva (n = 7)
Tongue (n = 19)

Figure 2

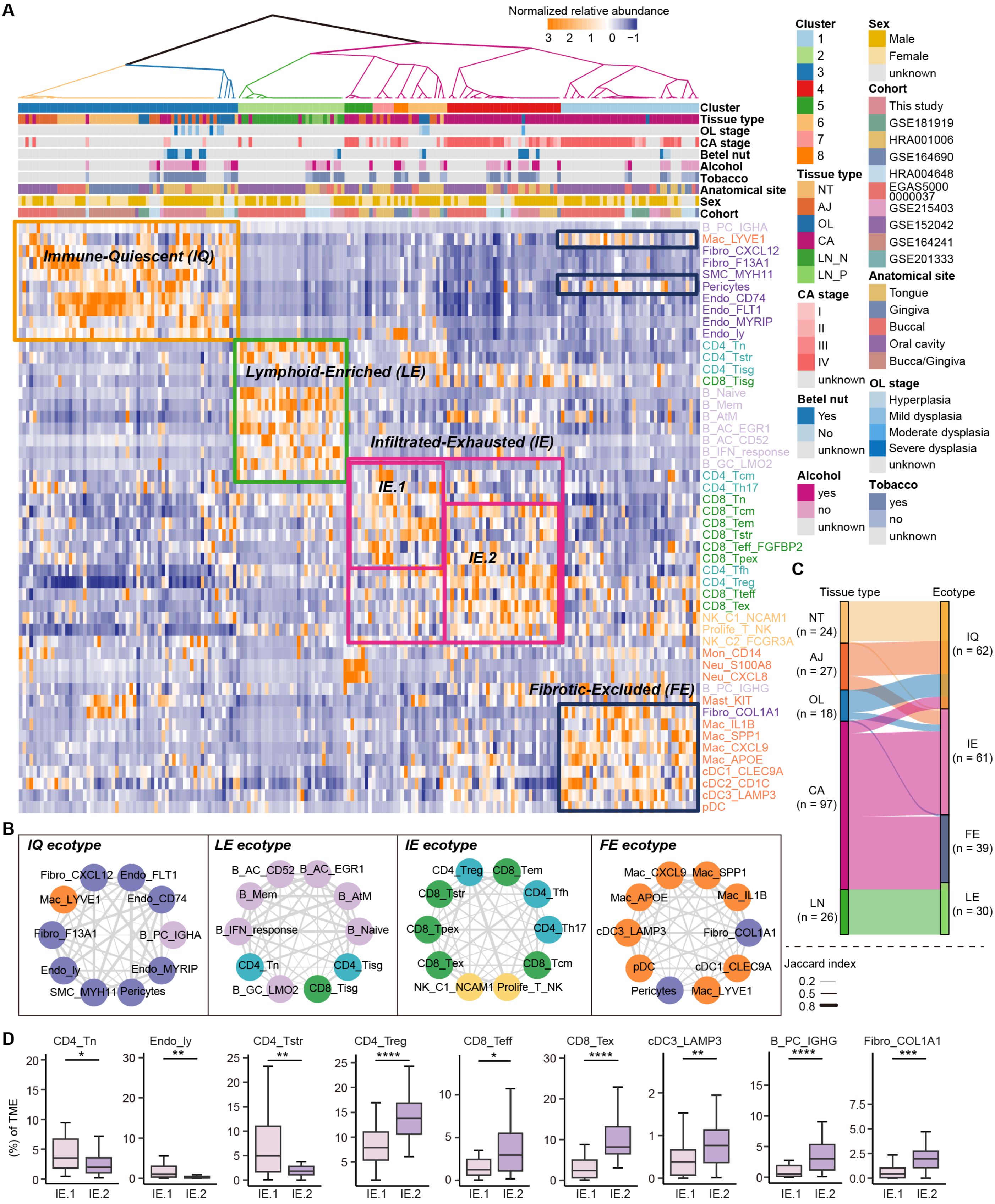
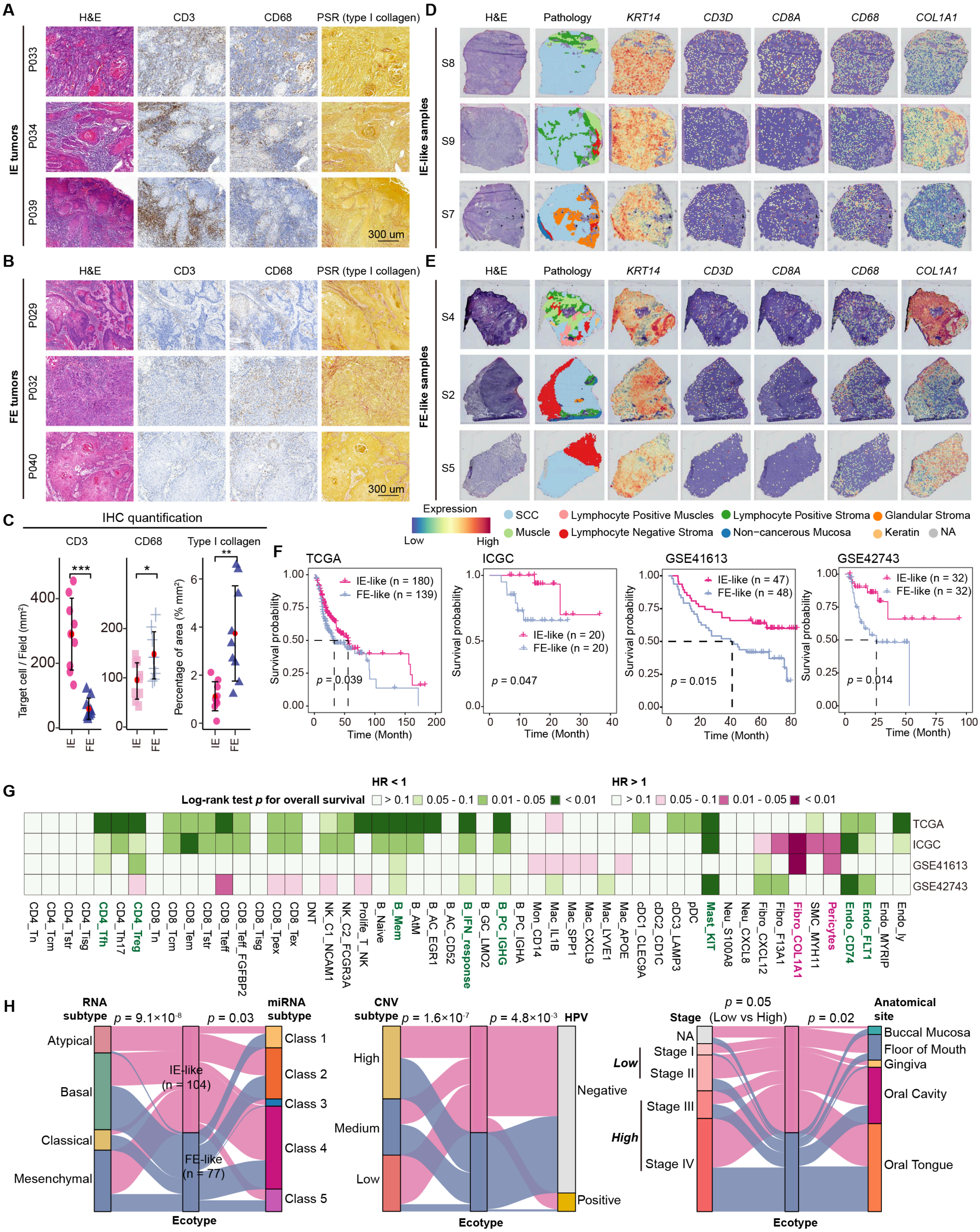

A
Normalized relative abundance
3 2 1 0 −1
Cluster
Tissue type
OL stage
CA stage
Betel nut
Alcohol
Tobacco
Anatomical site
Sex
Cohort
Immune-Quiescent (IQ)
Lymphoid-Enriched (LE)
Infiltrated-Exhausted (IE)
IE.1
IE.2
Fibrotic-Excluded (FE)
B_PC_IGHA
Mac_LYVE1
Fibro_CXCL12
Fibro_F13A1
SMC_MYH11
Pericytes
Endo_CD74
Endo_FLT1
Endo_MYRIP
Endo_ly
CD4_Tn
CD4_Tstr
CD4_Tisg
CD8_Tisg
B_Naive
B_Mem
B_AtM
B_AC_EGR1
B_AC_CD52
B_IFN_response
B_GC_LMO2
CD4_Tcm
CD4_Th17
CD8_Tn
CD8_Tcm
CD8_Tem
CD8_Tstr
CD8_Teff_FGFBP2
CD8_Tpex
CD4_Tfh
CD4_Treg
CD8_Tteff
CD8_Tex
NK_C1_NCAM1
Prolife_T_NK
NK_C2_FCGR3A
Mon_CD14
Neu_S100A8
Neu_CXCL8
B_PC_IGHG
Mast_KIT
Fibro_COL1A1
Mac_IL1B
Mac_SPP1
Mac_CXCL9
Mac_APOE
cDC1_CLEC9A
cDC2_CD1C
cDC3_LAMP3
pDC
Cluster
1
2
3
4
5
6
7
8
Tissue type
NT
AJ
OL
CA
LN_N
LN_P
CA stage
I
II
III
IV
unknown
Betel nut
Yes
No
unknown
Alcohol
yes
no
unknown
Sex
Male
Female
unknown
Cohort
This study
GSE181919
HRA001006
GSE164690
HRA004648
EGAS50000000037
GSE215403
GSE152042
GSE164241
GSE201333
Anatomical site
Tongue
Gingiva
Buccal
Oral cavity
Bucca/Gingiva
OL stage
Hyperplasia
Mild dysplasia
Moderate dysplasia
Severe dysplasia
unknown
Tobacco
yes
no
unknown
B
IQ ecotype
Fibro_CXCL12
Endo_FLT1
Mac_LYVE1
Endo_CD74
Fibro_F13A1
B_PC_IGHA
Endo_ly
Endo_MYRIP
SMC_MYH11
Pericytes
LE ecotype
B_AC_CD52
B_AC_EGR1
B_Mem
B_AtM
B_IFN_response
B_Naive
CD4_Tn
CD4_Tisg
B_GC_LMO2
CD8_Tisg
IE ecotype
CD4_Treg
CD8_Tem
CD8_Tstr
CD4_Tfh
CD8_Tpex
CD4_Th17
CD8_Tex
CD8_Tcm
NK_C1_NCAM1
Prolife_T_NK
FE ecotype
Mac_CXCL9
Mac_SPP1
Mac_APOE
Mac_IL1B
cDC3_LAMP3
Fibro_COL1A1
pDC
cDC1_CLEC9A
Pericytes
Mac_LYVE1
C
Tissue type
Ecotype
NT (n = 24)
AJ (n = 27)
OL (n = 18)
CA (n = 97)
LN (n = 26)
IQ (n = 62)
IE (n = 61)
FE (n = 39)
LE (n = 30)
Jaccard index
0.2
0.5
0.8
D
(%) of TME
CD4_Tn *
Endo_ly **
CD4_Tstr **
CD4_Treg ****
CD8_Teff *
CD8_Tex ****
cDC3_LAMP3 **
B_PC_IGHG ****
Fibro_COL1A1 ***
IE.1
IE.2
0
5
10
15
20
25
30
0.0
2.5
5.0
7.5

Figure 3



A
H&E
CD3
CD68
PSR (type I collagen)
IE tumors
P033
P034
P039
300 um
B
H&E
CD3
CD68
PSR (type I collagen)
FE tumors
P029
P032
P040
300 um
C
IHC quantification
CD3
CD68
Type I collagen
***
*
**
Target cell / Field (mm²)
Percentage of area (% mm²)
400
200
0
100
6
4
2
IE
FE
D
H&E
Pathology
KRT14
CD3D
CD8A
CD68
COL1A1
IE-like samples
S8
S9
S7
E
FE-like samples
S4
S2
S5
Expression
Low
High
SCC
Lymphocyte Positive Muscles
Lymphocyte Positive Stroma
Glandular Stroma
Muscle
Lymphocyte Negative Stroma
Non−cancerous Mucosa
Keratin
NA
F
TCGA
IE-like (n = 180)
FE-like (n = 139)
p = 0.039
ICGC
IE-like (n = 20)
FE-like (n = 20)
p = 0.047
GSE41613
IE-like (n = 47)
FE-like (n = 48)
p = 0.015
GSE42743
IE-like (n = 32)
FE-like (n = 32)
p = 0.014
Survival probability
Time (Month)
1.00
0.75
0.50
0.25
0.00
0
50
100
150
200
1.0
0.5
0.0
10
20
30
40
60
80
25
75
G
HR < 1
HR > 1
Log-rank test p for overall survival
> 0.1
0.05 - 0.1
0.01 - 0.05
< 0.01
TCGA
ICGC
GSE41613
GSE42743
CD4_Tn
CD4_Tcm
CD4_Tstr
CD4_Tisg
CD4_Tfh
CD4_Th17
CD4_Treg
CD8_Tn
CD8_Tcm
CD8_Tem
CD8_Tstr
CD8_Teff
CD8_Teff_FGFBP2
CD8_Tisg
CD8_Tpex
CD8_Tex
DNT
NK_C1_NCAM1
NK_C2_FCGR3A
Prolife_T_NK
B_Naive
B_Mem
B_AtM
B_AC_EGR1
B_AC_CD52
B_IFN_response
B_GC_LMO2
B_PC_IGHG
B_PC_IGHA
Mon_CD14
Mac_IL1B
Mac_SPP1
Mac_CXCL9
Mac_LYVE1
Mac_APOE
cDC1_CLEC9A
cDC2_CD1C
cDC3_LAMP3
pDC
Mast_KIT
Neu_S100A8
Neu_CXCL8
Fibro_CXCL12
Fibro_F13A1
Fibro_COL1A1
SMC_MYH11
Pericytes
Endo_CD74
Endo_FLT1
Endo_MYRIP
Endo_ly
H
RNA subtype
p = 9.1×10⁻⁸
p = 0.03
miRNA subtype
Atypical
Basal
Classical
Mesenchymal
IE-like (n = 104)
FE-like (n = 77)
Class 1
Class 2
Class 3
Class 4
Class 5
Ecotype
CNV subtype
p = 1.6×10⁻⁷
p = 4.8×10⁻³
HPV
High
Medium
Low
Negative
Positive
Ecotype
Stage
p = 0.05 (Low vs High)
p = 0.02
Anatomical site
NA
Stage I
Stage II
Stage III
Stage IV
Low
High
Buccal Mucosa
Floor of Mouth
Gingiva
Oral Cavity
Oral Tongue
Ecotype

Figure 4

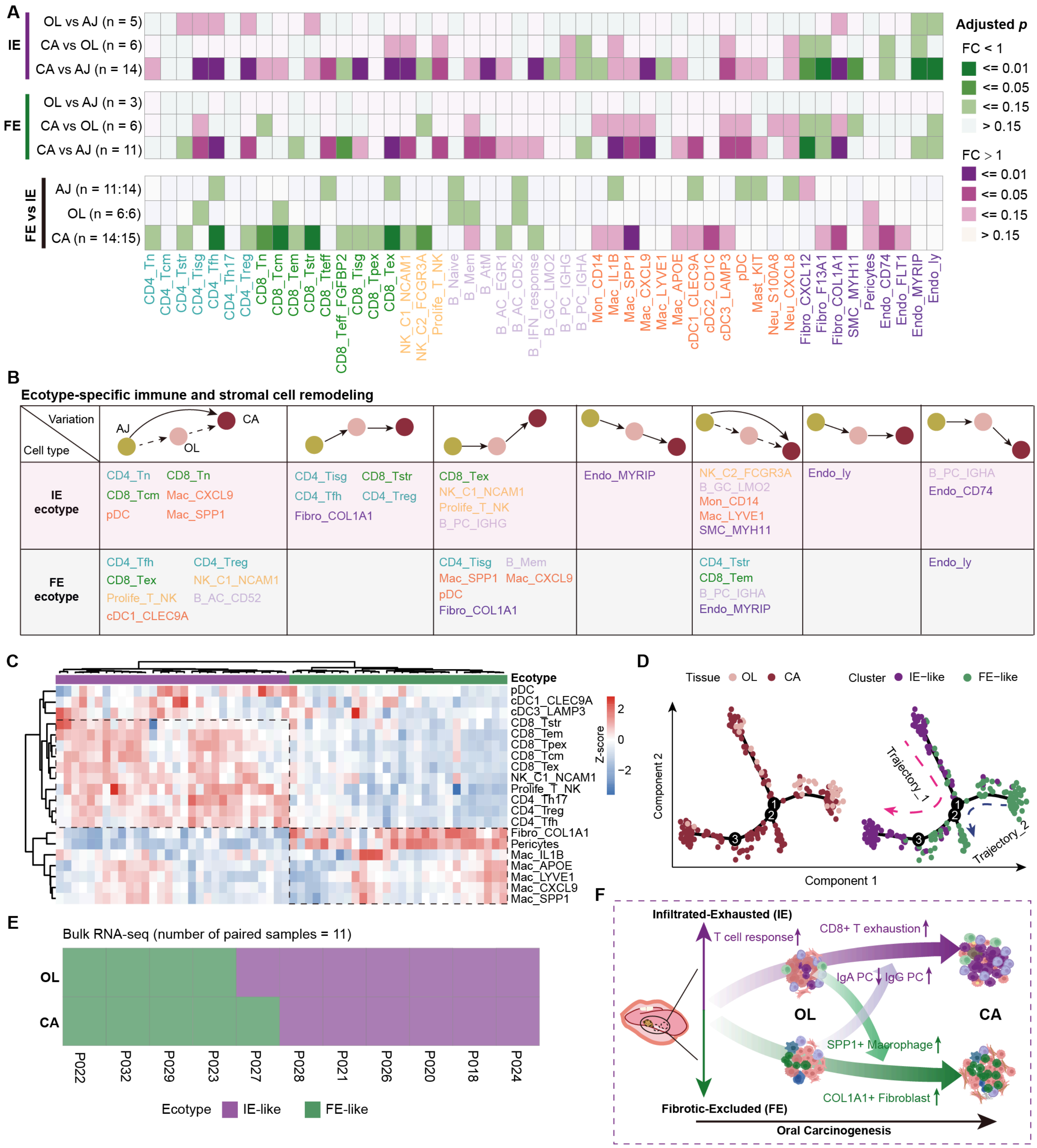

A
IE
OL vs AJ (n = 5)
CA vs OL (n = 6)
CA vs AJ (n = 14)
FE
OL vs AJ (n = 3)
CA vs OL (n = 6)
CA vs AJ (n = 11)
FE vs IE
AJ (n = 11:14)
OL (n = 6:6)
CA (n = 14:15)
Adjusted p
FC < 1
<= 0.01
<= 0.05
<= 0.15
> 0.15
FC > 1
<= 0.01
<= 0.05
<= 0.15
> 0.15
CD4_Tn
CD4_Tcm
CD4_Tstr
CD4_Tisg
CD4_Tfh
CD4_Th17
CD4_Treg
CD8_Tn
CD8_Tcm
CD8_Tem
CD8_Tstr
CD8_Teff
CD8_Teff_FGFBP2
CD8_Tisg
CD8_Tpex
CD8_Tex
NK_C1_NCAM1
NK_C2_FCGR3A
Prolife_T_NK
B_Naive
B_Mem
B_AtM
B_AC_EGR1
B_AC_CD52
B_IFN_response
B_GC_LMO2
B_PC_IGHG
B_PC_IGHA
Mon_CD14
Mac_IL1B
Mac_SPP1
Mac_CXCL9
Mac_LYVE1
Mac_APOE
cDC1_CLEC9A
cDC2_CD1C
cDC3_LAMP3
pDC
Mast_KIT
Neu_S100A8
Neu_CXCL8
Fibro_CXCL12
Fibro_F13A1
Fibro_COL1A1
SMC_MYH11
Pericytes
Endo_CD74
Endo_FLT1
Endo_MYRIP
Endo_ly
B
Ecotype-specific immune and stromal cell remodeling
Variation
Cell type
AJ
OL
CA
IE ecotype
CD4_Tn
CD8_Tn
CD8_Tcm
Mac_CXCL9
pDC
Mac_SPP1
CD4_Tisg
CD8_Tstr
CD4_Tfh
CD4_Treg
Fibro_COL1A1
CD8_Tex
NK_C1_NCAM1
Prolife_T_NK
B_PC_IGHG
Endo_MYRIP
NK_C2_FCGR3A
B_GC_LMO2
Mon_CD14
Mac_LYVE1
SMC_MYH11
Endo_ly
B_PC_IGHA
Endo_CD74
FE ecotype
CD4_Tfh
CD4_Treg
CD8_Tex
NK_C1_NCAM1
Prolife_T_NK
B_AC_CD52
cDC1_CLEC9A
CD4_Tisg
B_Mem
Mac_SPP1
Mac_CXCL9
pDC
Fibro_COL1A1
CD4_Tstr
CD8_Tem
B_PC_IGHA
Endo_MYRIP
Endo_ly
C
Ecotype
Z-score
2
0
−2
pDC
cDC1_CLEC9A
cDC3_LAMP3
CD8_Tstr
CD8_Tem
CD8_Tpex
CD8_Tcm
CD8_Tex
NK_C1_NCAM1
Prolife_T_NK
CD4_Th17
CD4_Treg
CD4_Tfh
Fibro_COL1A1
Pericytes
Mac_IL1B
Mac_APOE
Mac_LYVE1
Mac_CXCL9
Mac_SPP1
D
Tissue
OL
CA
Cluster
IE−like
FE−like
Component 2
Component 1
1
2
3
Trajectory_1
Trajectory_2
E
Bulk RNA-seq (number of paired samples = 11)
OL
CA
P022
P032
P029
P023
P027
P028
P021
P026
P020
P018
P024
Ecotype
IE-like
FE-like
F
Infiltrated-Exhausted (IE)
T cell response↑
CD8+ T exhaustion↑
IgA PC↓ IgG PC↑
OL
CA
SPP1+ Macrophage↑
COL1A1+ Fibroblast↑
Fibrotic-Excluded (FE)
Oral Carcinogenesis

Figure 5

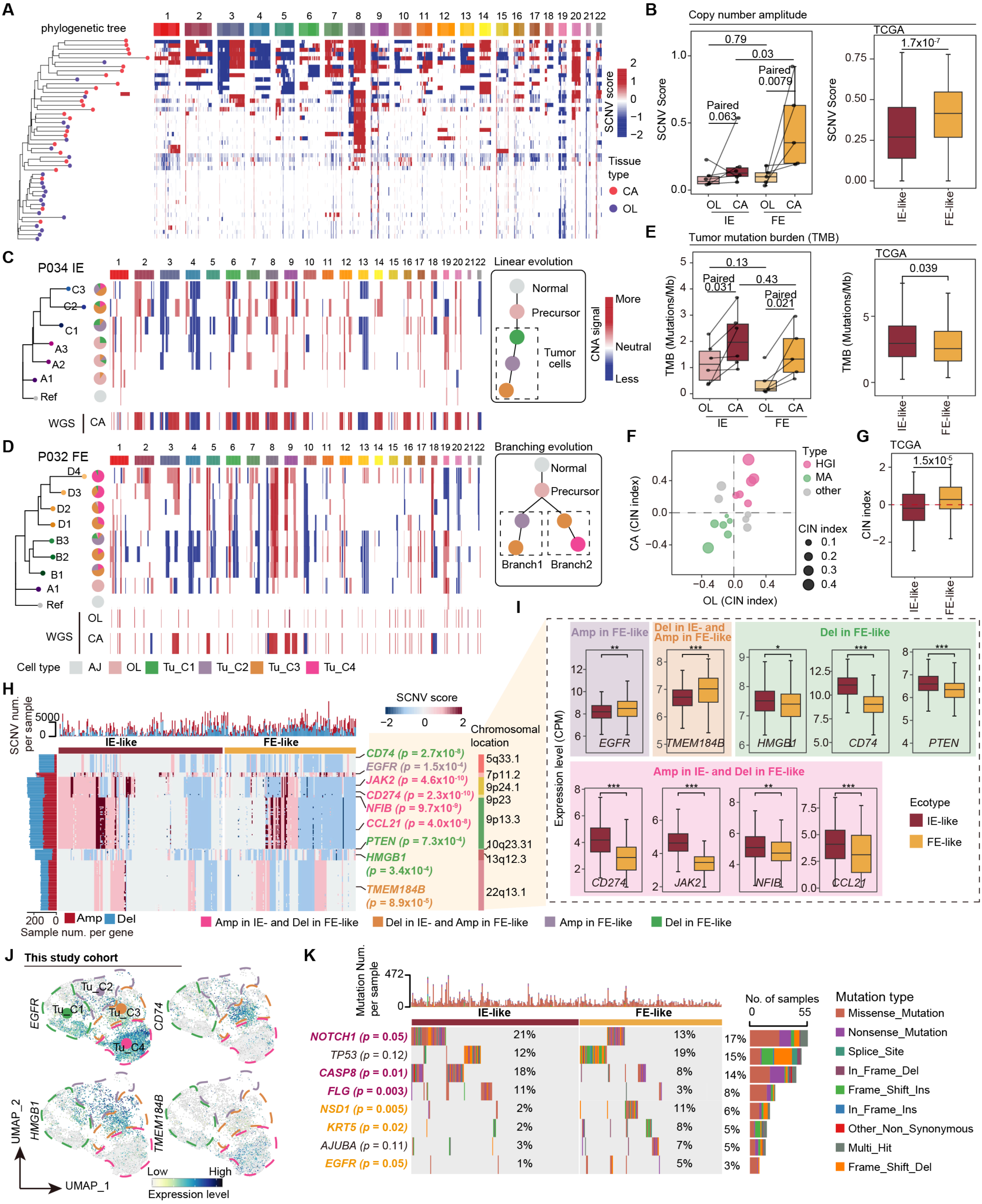

A
phylogenetic tree
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
SCNV score
2
1
0
−1
−2
Tissue type
CA
OL
B
Copy number amplitude
0.79
0.03
Paired 0.0079
Paired 0.063
SCNV Score
1.0
0.5
0.0
OL CA OL CA
IE FE
TCGA
1.7x10⁻⁷
0.75
0.50
0.25
0.00
IE-like
FE-like
C
P034 IE
C3
C2
C1
A3
A2
A1
Ref
WGS
CA
Linear evolution
Normal
Precursor
Tumor cells
CNA signal
More
Neutral
Less
D
P032 FE
D4
D3
D2
D1
B3
B2
B1
A1
Ref
WGS
OL
CA
Branching evolution
Normal
Precursor
Branch1
Branch2
Cell type
AJ
OL
Tu_C1
Tu_C2
Tu_C3
Tu_C4
E
Tumor mutation burden (TMB)
0.13
Paired 0.031
0.43
Paired 0.021
TMB (Mutations/Mb)
5
4
3
2
1
0
OL CA OL CA
IE FE
TCGA
0.039
F
Type
HGI
MA
other
CIN index
0.1
0.2
0.3
0.4
CA (CIN index)
OL (CIN index)
0.4
0.0
−0.4
G
TCGA
1.5x10⁻⁵
CIN index
2
0
−2
IE-like
FE-like
H
SCNV num. per sample
5000
0
SCNV score
−2 0 2
IE-like
FE-like
Chromosomal location
CD74 (p = 2.7x10⁻⁸)
EGFR (p = 1.5x10⁻⁴)
JAK2 (p = 4.6x10⁻¹⁰)
CD274 (p = 2.3x10⁻¹⁰)
NFIB (p = 9.7x10⁻⁹)
CCL21 (p = 4.0x10⁻⁸)
PTEN (p = 7.3x10⁻⁴)
HMGB1 (p = 3.4x10⁻⁴)
TMEM184B (p = 8.9x10⁻⁵)
5q33.1
7p11.2
9p24.1
9p23
9p13.3
10q23.31
13q12.3
22q13.1
200 0
Amp
Del
Sample num. per gene
Amp in IE- and Del in FE-like
Del in IE- and Amp in FE-like
Amp in FE-like
Del in FE-like
I
Amp in FE-like
Del in IE- and Amp in FE-like
Del in FE-like
Amp in IE- and Del in FE-like
Expression level (CPM)
EGFR
TMEM184B
HMGB1
CD74
PTEN
CD274
JAK2
NFIB
CCL21
*
**
***
Ecotype
IE-like
FE-like
J
This study cohort
EGFR
CD74
HMGB1
TMEM184B
Tu_C1
Tu_C2
Tu_C3
Tu_C4
UMAP_1
UMAP_2
Low
High
Expression level
K
Mutation Num. per sample
472
0
IE-like
FE-like
NOTCH1 (p = 0.05) 21% 13% 17%
TP53 (p = 0.12) 12% 19% 15%
CASP8 (p = 0.01) 18% 8% 14%
FLG (p = 0.003) 11% 3% 8%
NSD1 (p = 0.005) 2% 11% 6%
KRT5 (p = 0.02) 2% 8% 5%
AJUBA (p = 0.11) 3% 7% 5%
EGFR (p = 0.05) 1% 5% 3%
No. of samples
0 55
Mutation type
Missense_Mutation
Nonsense_Mutation
Splice_Site
In_Frame_Del
Frame_Shift_Ins
In_Frame_Ins
Other_Non_Synonymous
Multi_Hit
Frame_Shift_Del

Figure 6

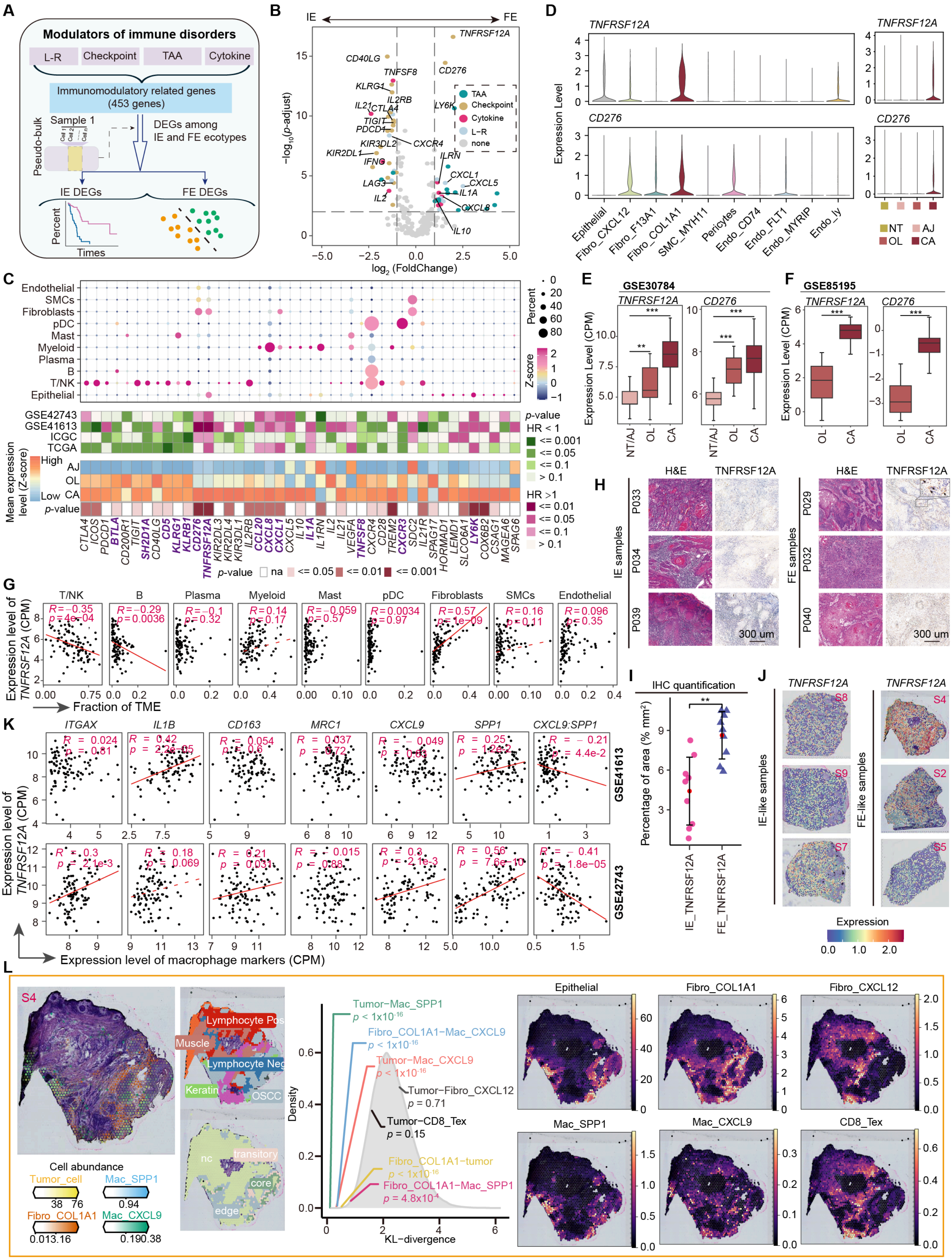

A
Modulators of immune disorders
L-R
Checkpoint
TAA
Cytokine
Immunomodulatory related genes (453 genes)
Pseudo-bulk
Sample 1
DEGs among IE and FE ecotypes
IE DEGs
FE DEGs
Percent
Times
B
IE
FE
−log10(p-adjust)
log2 (FoldChange)
TNFRSF12A
CD40LG
CD276
TNFSF8
KLRG1
IL2RB
IL21
CTLA4
LY6K
TIGIT
PDCD1
KIR3DL2
CXCR4
KIR2DL1
IL1RN
IFNG
CXCL1
CXCL5
LAG3
IL1A
IL2
CXCL8
IL10
TAA
Checkpoint
Cytokine
L−R
none
C
Endothelial
SMCs
Fibroblasts
pDC
Mast
Myeloid
Plasma
B
T/NK
Epithelial
Percent
Z-score
GSE42743
GSE41613
ICGC
TCGA
Mean expression level (Z-score)
High
Low
AJ
OL
CA
p-value
HR < 1
<= 0.001
<= 0.05
<= 0.1
> 0.1
HR >1
<= 0.01
<= 0.05
<= 0.1
> 0.1
CTLA4
ICOS
PDCD1
BTLA
CD200R1
TIGIT
SH2D1A
CD40LG
CD5
KLRG1
KLRB1
CD276
TNFRSF12A
KIR2DL3
KIR2DL4
KIR3DL1
IL2RB
CCL20
CXCL8
CXCL1
CXCL5
IL10
IL1A
IL1RN
IL2
IL21
VEGFA
TNFSF8
CXCR4
CD28
TREM2
CXCR3
SDC2
IL21R
SPAG17
HORMAD1
LEMD1
SLCO6A1
LY6K
COX6B2
CSAG1
MAGEA6
SPAG6
p-value
na
<= 0.05
<= 0.01
<= 0.001
D
TNFRSF12A
CD276
Expression Level
Epithelial
Fibro_CXCL12
Fibro_F13A1
Fibro_COL1A1
SMC_MYH11
Pericytes
Endo_CD74
Endo_FLT1
Endo_MYRIP
Endo_ly
NT
AJ
OL
CA
E
GSE30784
TNFRSF12A
CD276
Expression Level (CPM)
NT/AJ
OL
CA
F
GSE85195
TNFRSF12A
CD276
Expression Level (CPM)
OL
CA
G
T/NK
B
Plasma
Myeloid
Mast
pDC
Fibroblasts
SMCs
Endothelial
R = −0.35 p = 4e−04
R = −0.29 p = 0.0036
R = −0.1 p = 0.32
R = 0.14 p = 0.17
R = −0.059 p = 0.57
R = 0.0034 p = 0.97
R = 0.57 p = 1e−09
R = 0.16 p = 0.11
R = 0.096 p = 0.35
Expression level of TNFRSF12A (CPM)
Fraction of TME
H
H&E
TNFRSF12A
IE samples
P033
P034
P039
FE samples
P029
P032
P040
300 um
I
IHC quantification
Percentage of area (% mm²)
IE_TNFRSF12A
FE_TNFRSF12A
J
TNFRSF12A
IE-like samples
FE-like samples
S8
S9
S7
S4
S2
S5
Expression
K
ITGAX
IL1B
CD163
MRC1
CXCL9
SPP1
CXCL9:SPP1
R = 0.024 p = 0.81
R = 0.42 p = 2.2e-05
R = 0.054 p = 0.6
R = 0.037 p = 0.72
R = −0.049 p = 0.63
R = 0.25 p = 1.2e-2
R = −0.21 p = 4.4e-2
R = 0.3 p = 2.1e-3
R = 0.18 p = 0.069
R = 0.21 p = 0.031
R = −0.015 p = 0.88
R = 0.3 p = 2.1e-3
R = 0.56 p = 7.6e-10
R = −0.41 p = 1.8e-05
GSE41613
GSE42743
Expression level of TNFRSF12A (CPM)
Expression level of macrophage markers (CPM)
L
S4
Lymphocyte Pos
Muscle
Lymphocyte Neg
Keratin
OSCC
nc
transitory
core
edge
Cell abundance
Tumor_cell
Mac_SPP1
Fibro_COL1A1
Mac_CXCL9
Tumor−Mac_SPP1 p < 1x10−16
Fibro_COL1A1−Mac_CXCL9 p < 1x10−16
Tumor−Mac_CXCL9 p < 1x10−16
Tumor−Fibro_CXCL12 p = 0.71
Tumor−CD8_Tex p = 0.15
Fibro_COL1A1−tumor p < 1x10−16
Fibro_COL1A1−Mac_SPP1 p = 4.8x10−4
Density
KL−divergence
Epithelial
Fibro_COL1A1
Fibro_CXCL12
Mac_SPP1
Mac_CXCL9
CD8_Tex

Figure 7

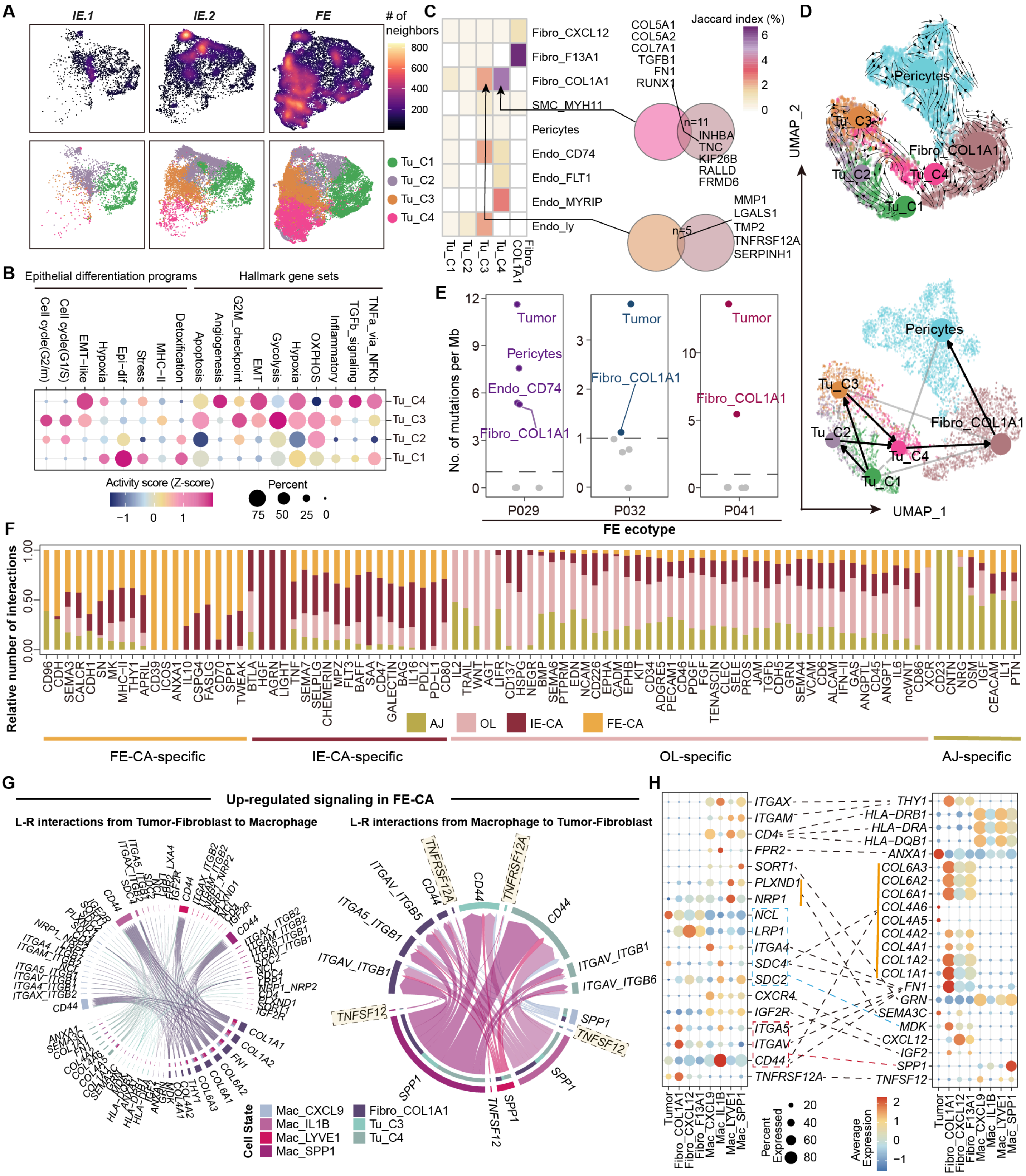
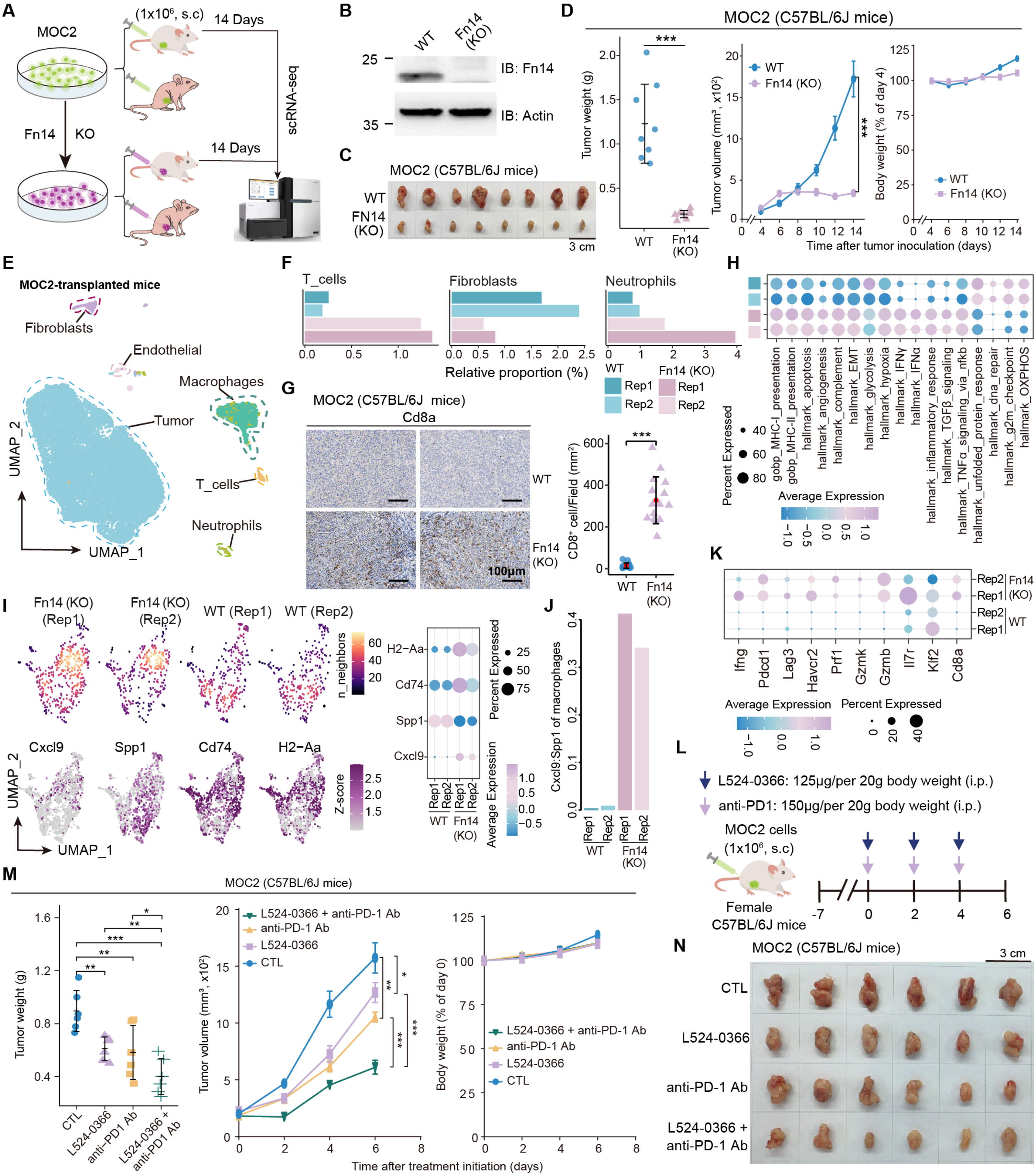

A
IE.1
IE.2
FE
# of neighbors
800
600
400
200
Tu_C1
Tu_C2
Tu_C3
Tu_C4
B
Epithelial differentiation programs
Hallmark gene sets
Cell cycle(G2/m)
Cell cycle(G1/S)
EMT-like
Hypoxia
Epi-dif
Stress
MHC-II
Detoxification
Apoptosis
Angiogenesis
G2M_checkpoint
EMT
Glycolysis
Hypoxia
OXPHOS
Inflammatory
TGFb_signaling
TNFa_via_NFKb
Activity score (Z-score)
−1
0
1
Percent
75
50
25
0
C
Fibro_CXCL12
Fibro_F13A1
Fibro_COL1A1
SMC_MYH11
Pericytes
Endo_CD74
Endo_FLT1
Endo_MYRIP
Endo_ly
Jaccard index (%)
6
4
2
0
COL5A1
COL5A2
COL7A1
TGFB1
FN1
RUNX1
n=11
INHBA
TNC
KIF26B
RALLD
FRMD6
n=5
MMP1
LGALS1
TMP2
TNFRSF12A
SERPINH1
D
UMAP_2
UMAP_1
Pericytes
Fibro_COL1A1
E
No. of mutations per Mb
Tumor
Pericytes
Endo_CD74
Fibro_COL1A1
P029
P032
P041
FE ecotype
F
Relative number of interactions
0.00
0.50
1.00
CD96
CDH
SEMA3
CALCR
CDH1
SN
MK
MHC−II
THY1
APRIL
CD39
ICOS
ANXA1
IL10
CSPG4
FASLG
CD70
SPP1
TWEAK
BTLA
HGF
AGRN
LIGHT
TNF
SEMA7
SELPLG
CHEMERIN
MPZ
FLT3
BAFF
SAA
CD47
GALECTIN
BAG
IL16
PDL2
PD−L1
CD80
IL2
TRAIL
WNT
AGT
LIFR
CD137
HSPG
NEGR
BMP
SEMA6
PTPRM
EDN
NCAM
CD226
EPHA
CADM
EPHB
KIT
CD34
ADGRE5
PECAM1
CD46
PDGF
FGF
TENASCIN
CLEC
SELE
PROS
JAM
TGFb
CDH5
GRN
SEMA4
VCAM
CD6
ALCAM
IFN−II
GAS
ANGPTL
CD45
ANGPT
IL6
ncWNT
CD86
XCR
CD23
CNTN
NRG
OSM
IGF
CEACAM
IL1
PTN
AJ
OL
IE-CA
FE-CA
FE-CA-specific
IE-CA-specific
OL-specific
AJ-specific
G
Up-regulated signaling in FE-CA
L-R interactions from Tumor-Fibroblast to Macrophage
L-R interactions from Macrophage to Tumor-Fibroblast
CD44
ITGAV_ITGB1
ITGA5_ITGB1
ITGAV_ITGB5
ITGAV_ITGB6
TNFRSF12A
TNFSF12
SPP1
COL1A1
COL1A2
FN1
COL6A2
COL6A1
COL6A3
Cell State
Mac_CXCL9
Mac_IL1B
Mac_LYVE1
Mac_SPP1
Fibro_COL1A1
Tu_C3
Tu_C4
H
ITGAX
ITGAM
CD4
FPR2
SORT1
PLXND1
NRP1
NCL
LRP1
ITGA4
SDC4
SDC2
CXCR4
IGF2R
ITGA5
ITGAV
CD44
TNFRSF12A
THY1
HLA−DRB1
HLA−DRA
HLA−DQB1
ANXA1
COL6A3
COL6A2
COL6A1
COL4A6
COL4A5
COL4A2
COL4A1
COL1A2
COL1A1
FN1
GRN
SEMA3C
MDK
CXCL12
IGF2
SPP1
TNFSF12
Tumor
Fibro_COL1A1
Fibro_CXCL12
Fibro_F13A1
Mac_CXCL9
Mac_IL1B
Mac_LYVE1
Mac_SPP1
Percent Expressed
20
40
60
80
Average Expression
2
1
0
−1

Figure 8


A
MOC2
(1x10⁶, s.c)
14 Days
Fn14 KO
14 Days
scRNA-seq
B
WT
Fn14 (KO)
25
35
IB: Fn14
IB: Actin
C
MOC2 (C57BL/6J mice)
WT
FN14 (KO)
3 cm
D
MOC2 (C57BL/6J mice)
Tumor weight (g)
2.0
1.5
1.0
0.5
***
WT
Fn14 (KO)
Tumor volume (mm³, x10²)
20
15
10
5
0
WT
Fn14 (KO)
***
0 4 6 8 10 12 14
Body weight (% of day 4)
125
100
75
50
25
0
WT
Fn14 (KO)
0 4 6 8 10 12 14
Time after tumor inoculation (days)
E
MOC2-transplanted mice
Fibroblasts
Endothelial
Macrophages
Tumor
T_cells
Neutrophils
UMAP_2
UMAP_1
F
T_cells
0.0 0.5 1.0
Fibroblasts
0.0 0.5 1.0 1.5 2.0 2.5
Neutrophils
0 1 2 3 4
Relative proportion (%)
WT
Rep1
Rep2
Fn14 (KO)
Rep1
Rep2
G
MOC2 (C57BL/6J mice)
Cd8a
WT
Fn14 (KO)
100μm
CD8⁺ cell/Field (mm²)
600
400
200
0
***
WT
Fn14 (KO)
H
Percent Expressed
40
60
80
Average Expression
−1.0 −0.5 0.0 0.5 1.0
gobp_MHC-I_presentation
gobp_MHC-II_presentation
hallmark_apoptosis
hallmark_angiogenesis
hallmark_complement
hallmark_EMT
hallmark_glycolysis
hallmark_hypoxia
hallmark_IFNγ
hallmark_IFNα
hallmark_inflammatory_response
hallmark_TGFβ_signaling
hallmark_TNFα_signaling_via_nfkb
hallmark_unfolded_protein_response
hallmark_dna_repair
hallmark_g2m_checkpoint
hallmark_OXPHOS
I
Fn14 (KO) (Rep1)
Fn14 (KO) (Rep2)
WT (Rep1)
WT (Rep2)
n_neighbors
60
40
20
Cxcl9
Spp1
Cd74
H2−Aa
Z-score
2.5
2.0
1.5
UMAP_2
UMAP_1
H2−Aa
Cd74
Spp1
Cxcl9
Rep1 Rep2 Rep1 Rep2
WT Fn14 (KO)
Percent Expressed
25
50
75
Average Expression
1.0
0.5
0.0
−0.5
J
Cxcl9:Spp1 of macrophages
0.4
0.3
0.2
0.1
0.0
Rep1 Rep2 Rep1 Rep2
WT Fn14 (KO)
K
Rep2
Rep1
Fn14 (KO)
Rep2
Rep1
WT
Ifng
Pdcd1
Lag3
Havcr2
Prf1
Gzmk
Gzmb
Il7r
Klf2
Cd8a
Average Expression
−1.0 0.0 1.0
Percent Expressed
0
20
40
L
L524-0366: 125μg/per 20g body weight (i.p.)
anti-PD1: 150μg/per 20g body weight (i.p.)
MOC2 cells (1x10⁶, s.c)
Female C57BL/6J mice
-7 0 2 4 6
M
MOC2 (C57BL/6J mice)
Tumor weight (g)
1.6
1.2
0.8
0.4
*
**
***
**
**
CTL
L524-0366
anti-PD1 Ab
L524-0366 + anti-PD1 Ab
Tumor volume (mm³, x10²)
20
15
10
5
0
L524-0366 + anti-PD-1 Ab
anti-PD-1 Ab
L524-0366
CTL
**
*
***
***
0 2 4 6 8
Body weight (% of day 0)
125
100
75
50
25
0
L524-0366 + anti-PD-1 Ab
anti-PD-1 Ab
L524-0366
CTL
0 2 4 6 8
Time after treatment initiation (days)
N
MOC2 (C57BL/6J mice)
3 cm
CTL
L524-0366
anti-PD-1 Ab
L524-0366 + anti-PD-1 Ab