The Immunoregulatory Architecture of the Adult Oral Cavity

Bruno F. Matuck[1,2]*, Khoa L. A. Huynh[3]*, Diana Pereira[4]*, XiuYu Zhang[3], Meik Kunz[5], Nikhil Kumar[1,6], Quinn T. Easter[1,2], Alexandre Fernandes[4], Ameer Ghodke[7], Alexander V. Predeus[8], Lili Szabó[9], Nadja Harnischfeger[9], Zohreh Khavandgar[10,11], Margaret Beach[10,11], Paola Perez[10], Benedikt Nilges[12], Maria M. Moreno[12], Kang I. Ko[13], Sarah A. Teichmann[14], Adam Kimple[7,15], Sarah Pringle[6], Kai Kretzschmar[9], Blake M. Warner[10,11], Inês Sequeira[4#], Jinze Liu[3,17#], Kevin M. Byrd[1,2, 6,11,17#]

[1] Lab of Oral & Craniofacial Innovation (LOCI), Department of Innovation & Technology Research, ADA Science & Research Institute, Gaithersburg, MD, USA.

[2] Department of Oral and Craniofacial Molecular Biology, Philips Institute for Oral Health Research, Virginia Commonwealth University, Richmond, VA, USA.

[3] Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA.

[4] Center for Oral Immunobiology and Regenerative Medicine, Barts Centre for Squamous Cancer, Institute of Dentistry, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

[5] The Bioinformatics CRO, FL, USA.

[6] Division of Oral and Craniofacial Health Sciences, Adams School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[7] Department of Otolaryngology-Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, NC, USA.

[8] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK.

[9] Mildred-Scheel Early Career Centre (MSNZ) for Cancer Research, University Hospital Wuerzburg, Wuerzburg, DE.

[10] Sjögren's Clinical Investigations Team, National Institutes of Dental and Craniofacial Research. Bethesda, Maryland, USA.

[11] Salivary Disorders Unit, National Institutes of Dental and Craniofacial Research. Bethesda, Maryland, USA.

[12] OMAPiX, Inc. Leuven, Leuven, BE.

[13] Department of Periodontics, School of Dental Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[14] Department of Physics, Cavendish Laboratory, University of Cambridge, UK.

[15] Marsico Lung Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[16] University of Groningen and University Medical Center, Groningen, NL.

[17] VCU Massey Comprehensive Cancer Center, Bioinformatics Shared Resource Core, Virginia Commonwealth University, Richmond, VA, USA.

CONTRIBUTIONS:

* Each author contributed equally.

#CORRESPONDING AUTHOR:

Correspondence to:

Kevin Matthew Byrd: kevinmbyrd@gmail.com

Jinze Liu: liuj15@vcu.edu

Inês Sequeira: i.sequeira@qmul.ac.uk

KEYWORDS: single-cell, atlas, spatial biology, multiomics, cell-cell communication, oral mucosa, salivary glands, keratinocytes, fibroblasts, vasculature, ligand-receptor activity, tissue microenvironment, neighborhoods

ABSTRACT (203):

The oral cavity is a critical barrier with immunosurveillance capabilities. A detailed understanding of its cellular, molecular, and spatial architecture is essential for advancing precision medicine across aerodigestive tissues. Here, we present the first integrated atlas of human adult oral and craniofacial tissues, derived from single-cell RNA sequencing of ~250,000 cells from 70 samples across 13 niches, including salivary glands and oral mucosae. Complementary spatial atlases using a 40-plex protein and 300-plex transcriptomics panels were developed, covering >2 million cells. Using our new AstroSuite package, we identified 8 tissue cellular neighborhoods (TCNs) and 17 multicellular interaction modules (MCIMs) across gland and mucosae in health. Diverse fibrovascular TCNs emerged as centralized hubs, coordinating immune cell migration and activation. Mucosal hubs within stroma displayed higher immune cell density, diversity, and activation than glands, underscoring differences in functional architecture. In keratinized and nonkeratinized mucosae, common fibrovascular hubs, and MCIMs revealed a specialized biogeography of immunoregulation. Under chronic inflammation, mucosal TCNs were fragmented, and MCIMs involving fibroblasts and immune cells were reorganized and expanded. Focused analysis of tertiary lymphoid structures (TLS) and fibroblasts identified TLS-associated MCIMs and drug candidates targeting these spatial interactions, offering a foundational framework for broadly dissecting mucosal immunity to guide precision medicine approaches.

KEY TERMS:

- *AstroSuite*: A suite of bioinformatics tools, including *TACIT, Constellation, STARComm*, and *Astrograph*, used for identifying and analyzing spatial features, cell types, and interactions in tissues.
- *Astrograph*: A visualization tool within AstroSuite for exploring cellular neighborhoods and spatial relationships within tissues.
- Cell Ontology: A structured, hierarchical classification system that provides a controlled vocabulary for cell types based on distinct properties like morphology, lineage, function, and molecular markers.
- Cell Recipe: A pre-defined set of markers used to specify cell identities and states by combining proteins or transcripts.
- *Constellation*: An integration and clustering algorithm within *AstroSuite* that identifies Tissue Cellular Neighborhoods by dividing tissue into subregions, aiding in scalable whole-slide imaging analyses.
- Fibrovascular Hubs: Centralized regions where fibroblasts and vascular endothelial cells within defined Tissue Cellular Neighborhoods coordinate with immune cells.
- Heterotypic Interactions: Predicted receptor-ligand interactions occurring between different cell types, such as immune cells and structural cells.
- Homotypic Interactions: Predicted receptor-ligand interactions between cells of the same type.
- Multicellular Interaction Modules (MCIMs): Spatial clusters where multiple cell types engage in receptor-ligand interactions—conceptually akin to Tissue Cellular Neighborhoods.
- Receptor-Ligand Density Scoring: A method in *STARComm* used to measure the density of receptor-ligand interactions within tissue regions, predicting areas of high cell-cell communication.
- Spatial Drug Targeting: The approach of identifying druggable targets within specific tissue regions, guided by spatial analysis of receptor-ligand interactions within identified regions of interest.
- *STARComm*: A computational tool in *AstroSuite* for analyzing spatial data, focusing on quantifying receptor-ligand interactions and communication densities in tissues to discover single or Multicellular Interaction Modules.
- Structural Immunity: The concept that structural cells, like fibroblasts, contribute to immune regulation through interactions with immune cells to maintain homeostasis.
- *TACIT*: A tool within *AstroSuite* used to identify cell types and interactions in spatial data, supporting neighborhood analysis and visualization.
- Tiered Cell Annotation: A classification system that organizes cells into broad groups, refining to specific cell types for in-depth tissue analysis—i.e., higher tiered ontologies are more generally labeled.
- Tissue Cellular Neighborhoods (TCNs): Spatially organized groups of cells within a tissue that are predicted to interact with one another in specific patterns or proximities. These neighborhoods represent the local microenvironments where different cell types coordinate biological processes—conceptually akin to or Multicellular Interaction Modules.

INTRODUCTION (910):

The oral cavity is both barrier and beacon, positioned at the crossroads of external exposure and internal defense where morphologically and functionally unique tissues unite to support communication, aid digestion, and safeguard the body[1,2]. It is comprised of distinct niches such as the major and minor salivary glands, zones of masticatory (keratinized) and lining (non-keratinized) mucosae, as well as specialized organs like the tongue and teeth. Despite regional differences and morphological adaptations, mucosae are thought to be generally consistently constructed of only a few key tissue types, including stratified squamous epithelia predominated by keratinocytes, and stroma predominated by fibroblasts, neurovascular units, muscle and various immune cells—even in health[3,4]. The salivary glands are embedded in and interconnected to oral mucosae and include the parotid, submandibular, sublingual, and numerous minor salivary glands. These glands secrete saliva that is essential for lubrication, digestion, and antimicrobial defense[5]; however, these glands are also structurally and functionally distinct, with their epithelia dominated by ductal and acinar cells, along with similar supporting stromal and immune cells to mucosae[6]. While diseases with specific spatial patterns within oral and craniofacial tissues have long been observed[7,8], the underlying cellular and molecular mechanisms behind these predilections remain poorly understood.

Within the oral cavity, soft tissues like the oral mucosa and glands not only function as physical barriers but also play a significant role as sensory organs[9-11] and immune-regulatory interfaces[12,13]. Each oral macroniche is interconnected and supported by epithelial and stromal domains, embedded within interlacing extracellular matrix, enabling communication between structural cell types (keratinocytes, fibroblasts, and vascular endothelial cells) and immune cells[14-16]. Traditionally, these structural cells have been viewed as passive support cells, but growing evidence highlights their role in “structural immunity”, referring to the immune functions performed

by structural cells to actively regulate immune responses to maintain tissue homeostasis, supporting tissue regeneration, and repair and interact with immune cells to defend against pathogens[17,18]. For example, fibroblasts not only maintain tissue homeostasis by producing extracellular matrix components but also secrete cytokines that mediate immune responses[19-22]. Likewise, vascular endothelial cells regulate immune cell trafficking, controlling how immune cells enter and exit tissues during inflammation or injury[23,24]. While an emerging concept, these immunoregulatory functions illustrate that structural cells are active participants in immune defense[6,21,23,25-27].

Despite the growing recognition of the critical role of structural immunity across the body, the specific mechanisms by which this applies to tissues of the oral cavity remain poorly characterized at a single-cell or spatial biological level. Tissue cellular neighborhoods (TCNs), defined as spatially distinct groups of individual cells organized within the tissue's structural meta-architecture, have been recently shown to act as regulatory hubs that modulate immune responses[28,29]. How the oral cavity is organized into the functional tissue units remains to be defined[17]. Furthermore, the patterns of ligand-receptor signaling between structural and immune cells are not fully understood, but toolkits to scalably resolve this at single-cell resolution have also recently emerged[30,31]. Mapping these cell-cell interactions within TCNs and across evolving definitions of tissue-specific cellular communities is crucial to understanding how immune responses are regulated within oral tissues.

To address these knowledge gaps, we developed the first integrated single cell RNA sequencing atlas of adult oral and craniofacial tissues, analyzing nearly 250,000 cells from 70 samples across 13 distinct niches, including the oral mucosa, salivary glands, and pulp—to date the largest reference dataset compiled in humans, building off of recent integration efforts[17,32]. Using this published and unpublished data, we identified key cell types in each niche and characterized their specific cellular heterogeneity and gene expression profiles, creating a detailed cellular landscape

atlas of the oral cavity. To further investigate these interactions, we created two complementary spatial atlases using a 40-plex protein (multiplex immunofluorescence; Multi-IF) and a 300-plex transcriptomics panel (MERFISH) designed using known 50 receptor-ligand pairs within the panel. These were applied to 21 unique samples, approaching 2 million cells from 6 distinct niches, a first-in-class multiomics dataset for adult oral and craniofacial tissues. We then developed and applied advanced bioinformatics tools as part of spatial analysis package called *AstroSuite*, including 1) *TACIT* for spatial cell identification and state assignment, 2) *Constellation* for integrated and scalable cellular neighborhood identification, 3) *STARComm* for spatially resolved receptor-ligand analyses, and 4) *Astrograph* for interactive spatial visualization[33].

Applying these tools for integrated spatial multiomics analyses, we identified distinct TCNs and established spatially resolved, multicellular interaction modules (MCIMs) using receptor-ligand pairs across oral mucosa and salivary glands. This meta-architecture of molecules, cells, TCNs and MCIMs consistently revealed fibrovascular neighborhoods as centralized and coordinating hubs, predicted to orchestrate immune cell tropism and activation. To assess the value of understanding structural immunity considering niche we first compared various oral mucosae from masticatory and lining mucosal niches, finding shared and distinct TCNs and MCIMs. Considering this heterogeneity, we examined an inflammatory disease of the oral cavity in one mucosal niche, comparing it to our healthy human oral cavity atlases to identify disease-related changes, finding stromal TCN fragmenting, hub rearrangement, and additional and rearranged MCIMs. By applying the spatial drug discovery algorithm *Drug2Cell* to specific fibroblasts linked with within and surrounding immune foci, we generated a targeted list of therapeutic candidates, filtered based on their spatial context around ectopic inflammatory regions. Taken together, our study presents the first comprehensive, spatial map of cells and their interactions in the adult oral cavity. These findings have important implications for understanding the role of the oral mucosa as an innate immune organ and developing precision medicine strategies to precisely treat diseases.

RESULTS: (4773)

The Landscape of Human Adult Oral and Craniofacial Cell Types

The diversity of human cell types and states remains unknown across the many niches and domains within the oral cavity. To address this, we first compiled 11 single cell RNA sequencing (scRNAseq) studies with samples from across oral mucosa[2,22,34-37], salivary glands[36,38-40], and pulp[37,41,42] (Figure 1a,b). We then harmonized clinical and sample-level metadata (Supplementary Table 1). We then generated and integrated new three additional scRNAseq datasets: labial (lip) mucosa, soft palate mucosa, and parotid salivary glands to cover more mucosal and glandular niches (Figure 1b). Finally, we established a biorepository of formalin-fixed, paraffin-embedded samples (Figure 1c); these were used for two complementary, paired atlases of spatial transcriptomics (MERSCOPE; Vizgen) and spatial proteomics (Phenocycler-Fusion 2.0, Akoya Biosciences with the goal of integrating analyses (Figure 1d).

For scRNAseq, we only included healthy adults, totaling 246,102 cells from 70 samples and 13 oral “macroniches”. After batch corrections and integration of single-cell data, cells ontologies were annotated in a tiered approach with a broad categorization of structural and immune cell types in Tier 0 (T0) further divided into 8 cell types at T1 (Figure 1e). This data was deposited to CELLxGENE for public exploration (https://cellxgene.cziscience.com/collections/065ad318-59fd-4f8c-b4b1-66caa7665409). Within the dataset, there were more female among those analyzed, more often were younger subjects (>50% from <50 years-old). Most of the published datasets are from white/Caucasian donors, presenting limited racial and ethnic diversity. In an effort to create a more inclusive Oral Craniofacial Atlas, we are actively working to increase diversity in sample ancestry by including African American, Asian, Latin, and American Indian donors.

(Figure 1f). Samples were normalized and dimensionally reduced (Figure 1g; Extended Data 1a). Tier1 annotations of the oral and craniofacial atlas were completed with the main structural and immune cell populations: epithelial cells, fibroblasts, vasculature, neural crest-derived,

mineralized (i.e., odontoblasts), muscle, and immune: myeloid and lymphoid-derived cells. Most cell types were capture in all samples, with interesting differences observed between the same niche collected in different studies showing slightly different Tier 1 cellular proportions. The oral and craniofacial atlas comprises three main tissue groups: the glands, the mucosal tissues and the dental pulp which vary in cell type composition. Epithelial cells were proportionally overrepresented in salivary glands and tongue; neural-crest derived populations in the dental pulp; and vasculature in buccal mucosa and gingiva (Figure 1h; Extended Data 1b,c). Fibroblasts proportions were present in nearly all samples, but proportions varied.

Structural-Immune Axes Within Oral Tissues

Oral tissues are a primary point of interaction between the body and external elements; we hypothesized that oral-resident cells would interact via receptor-ligand (R-L) pairs—even in "disease-free" samples. *CellPhoneDB* provided first evidence of elevated fibroblast and vascular activity in the atlas, followed by epithelial and neural populations, and then lowest for immune populations (*Methods*; Figure 1i; Supplementary Table 2). These findings further focused our study on cell-cell interactions, primarily on the structural-immune axis. Further exploring uni- and bidirectional interactions, we first assessed general patterns of activity using *CellChat*, which revealed some shared pathways among epithelia, fibroblasts, and vasculature, such as *MIF-CD74/CXCR4* that can guide immune cells to sites of inflammation or damage[43] (*Methods*; Figure 1j). More broadly, top interactions were found between structural (primarily fibroblasts, vasculature) and immune (both myeloid and lymphoid) via extracellular matrix (ECM) molecules and various receptors (Extended Data 1d). Pathway analyses confirmed the observed trends for both outgoing ("ligand") and incoming signaling patterns ("receptor"; Figure 1k, l); fibroblasts were found to be particularly involved in both sending and receiving signals through IGFBP and CXC-

motif chemokine ligands[44]; in parallel, vasculature utilized CC-motif chemokines, supporting the concept of structural cell coordination for immunoregulation[18].

Oral mucosa and salivary glands share roles for immune defense and tissue homeostasis, but they differ in their primary functions: the oral mucosa more directly supports immune barrier functions; instead, salivary glands on secretion to facilitate digestion, pathogen defense, speech, and taste[45]. Our aggregated single cell data identified a “fibrovascular” signaling axis within oral tissues (Figure 1i-l); to understand this we used *MultiNicheNet* (see *Methods*) to analyze differential cell-cell communication across multiple samples and conditions simultaneously (Extended Data 1e,f; Supplementary Table 2). Predicted R-L interactions between structural cells with immune cell types varied significantly between niches. Mucosa displayed more pronounced myeloid-fibroblast signaling via *TGFB1, TGFBR3, SDC2, PDGFRA, ACVRL1, TNFSF14*, which can be both anti- and proinflammatory[46,47] (Extended data 1e, f). VECs showed enhanced immune signaling via *CD40-CD40L*—costimulatory molecules known to sustain immunoreactivity[48].

Integrated Spatial Proteomic Atlas of Oral Mucosa and Salivary Glands

We hypothesized that the structural organization of the oral mucosa and salivary glands support and amplify R-L interactions. Informed by scRNAseq, we designed a panel of 40 oligo-conjugated primary antibodies for highly multiplexed immunofluorescence (Multi-IF; Figure 2a; panel: Extended Data 2a). This spatial proteomic atlas included three salivary glands (parotid, submandibular, minors) and three mucosal niches (tongue, gingiva, and buccal; Extended Data 2b). Most samples were from >50-year-olds, white, and female subjects (Supplementary Table 2). Ages between glands and mucosae were similar (Mean+/-SD; 56.2+/-13.9 years, 54.3+/-13.2 years, respectively). The antibody panel was extensively validated using the Akoya Phenocycler-Fusion-2.0 system in health and inflammatory disease states by this project and others[17,33,49]. Cellpose3 used post-assay H&E stains and these Multi-IF images for creating cell segmentation

via human-in-the-loop training sessions (Extended Data 2c,d). Cursory analyses of structural and immune cell types revealed dramatic differences in cellular and metacellular architecture (Extended Data 2e); however, the diversity and proportion of cell types, along with cell-cell co-occurrence patterns, required innovative approaches for spatial analyses following human-in-the-loop cell segmentation (*Cellpose*).

To integrate and scale analyses, we developed an interconnected series of algorithms, collectively called *AstroSuite,* to support integrated and scalable spatial biology analyses, which includes 1) *TACIT* (Threshold-based Assignment of Cell Types from Multiplexed Imaging DaTa), a deep learning tool for spatial cell identification and cell state quantification of whole slide images (WSIs), 2) *Constellation* for integrated neighborhood analyses (Extended Data 3a), and 3) *Astrograph* for spatial data visualization[33]. *TACIT* was specifically developed to support the identification of cell types with a predefined recipe, requiring no pre-labeled training sets; moreover, *TACIT* allowed us to group cells into any tiered class from structural and immune cell types (Figure 2c) all the way to cell state annotated lower tier cell types (i.e., Activated Tregs are $CD45^+$ $CD4^+$;$CD3^+$ $FOXP3^+$ $ICOS^+$).

Our final cell numbers for proteomics alone were >1,000,000 cells from the 18 healthy tissues, in which we first identified 15 unique cell types, 10 cell states, and 8 tissue cellular neighborhoods (Figure 2a-d; Extended Data 2a). As predicted, the proportion of immune cells were distinct in glands versus mucosa, with mucosa supporting more innate cell populations and glands, more adaptive ones (Figure 2e). We identified statistically significant increases in the proportion of macrophages in mucosa (p=0.012), while CD8+ T were more prevalent in glands (p=0.006). We then developed spoke-and-wheel plots to analyze the average distance (μm) between any structural cell type (aggregated among epithelial and stromal populations) and immune cells (Figure 2f). As predicted, immune cell types were found in closer proximity to mucosal structural cells compared to glands (p-values; Figure 2 legend).

Already recognizing differences in immune cell residency and occurrence, we next focus on the discovery of tissue cellular neighborhoods (TCNs), defined as distinct clusters or groups of cells within a tissue that are spatially organized and functionally related (Figure 2g). We developed *Constellation* to identify TCNs in atlas-scale datasets, which differs from other tools designed to handle smaller spatial datasets, such as *CytoCommunity* (Extended Data 3a).[28]. *Constellation* initially identifies local TCNs within overlapping sub-regions by employing various partitioning strategies, such as vertical stripes, horizontal stripes, and square grids. Following this, a merging process is applied to integrate these local patterns into a consensus set of TCNs. We applied *Constellation* in our spatial proteomics atlas, identifying 8 TCNs among mucosa and glands (TCN^{A-H}), representing grossly different proportions within tissues.

When assessing TCN^{A-H}, we broke them down by cell types, revealing relationships of structural and immune cell enrichment and supporting for neighborhood annotation (Figure 2g; Extended Data 4a,b). Mucosa displayed all 8 identified neighborhoods. Fibroblasts were shared among 4 ($TCN^{B, C, D, H}$) vasculature endothelial cells (VECs) were enriched among 4 ($TCN^{D, E, G, H}$), two of them also enriched with fibroblasts (fibrovascular). Innate immune cells were found in nearly every TCN whereas adaptive immune cell types were concentrated near fibrovascular TCNs (TCN^{D-G}). Glands only displayed 5 TCNs ($TCN^{A, C-F}$). The same trend of innate (all TCNs) and adaptive immune cell divergence ($TCN^{D, E}$) was also observed. Even though detected in both niches, mucosa exhibits an eightfold increase in TCN^{C} identified as ***Fibro-myeloid 2***, compared to the glands, indicating a higher predicted tendency for fibroblast-innate interactions.

Comparatively, glands exhibited an average 58% of TCNs composed of solely epithelial (acinar and ductal epithelial cells) and VEC-associated lymphocytes (Extended Data 4). We performed TCN connection analysis to identify the spatial adjacency relationship between different TCNs (network) analyses (Figure 2h) and plotted the aggregated TCN network diagrams comparing

mucosal and glandular tissues. We labeled any TCN connecting at least 3 other TCNs as a “hub”. TCN[A, D, G] emerged as mucosal hubs (*Mixed-immune 1*, *Fibro-lymphoid 1*, and *Vascular-immune*, respectively); TCN[C,F], as glandular hubs (*Fibro-myeloid 2* and *Mixed Immune 2*, respectively). Within mucosae and glands, fibrovascular hubs predominate; however, mucosal fibrovascular TCNs are notably more fragmented compared to glandular ones (Figure 2i).

Tier 2 Ontologies Reveal Immunoregulatory Signaling of Fibrovascular Cell Types

We hypothesized that this may be reactionary to barrier stresses but also a phenotype that supports more immune cell tropism and activation. Due to the diversity of signals and unique cell types known among structural cells, we set out to refine annotations within the scRNAseq with Tier 2 ontologies: this particularly expanded epithelial and immune cell types (Figure 3a,b). UMAPs were also generated for glands and mucosa and deposited to CELLxGENE (see *Data Availability*). As expected, some gland samples had mucosal epithelial cells captured; same for mucosal samples and gland epithelia (Figure 3c). In *CellPhoneDB* analysis, we observed again distinct interaction patterns between structural and immune cells (19 Tier 2 cell types in glands; 22, in mucosa; Figure 3d; Supplementary Table 2). Even with this more detailed annotation, fibroblasts and VECs again emerged as major cells in both glands and mucosal tissues. In glands, fibroblasts exhibited strongest interactions with epithelia; VECs showed robust interactions with fibroblasts and immune cells. Mucosal fibroblasts were also prominent interactors, with increased signaling to keratinocytes and myeloid cells; VECs exhibited fewer interactions with immune cells compared to glands. We again used *CellChat* to analyze the interactions between structural cell types within mucosae (Figure 3e; Supplementary Table 2); this revealed deeper insights into the structural cell-immune cell communication patterns (Extended Data 5a). Quantifying significant R-L pairs, we see few shared between mucosa and glands; however, most R-L interactions was structural to immune (~2:1), and about 20% of those were shared in both organs.

Mucosal tissues displayed 120 unique R-L interactions emanating from structural cells (Extended Data 5b). Pathway enrichment analysis revealed known CXC-motif chemokines *CXCL8* and *CXCL12* (Extended Data 5c). Again, fibroblasts and VECs but also lymphatic endothelial cells (LECs) and mural cells also were shown to contribute to chemotaxis and immune responses. Oral fibroblasts are predicted contributed to the recruitment of monocytes and dendritic cells through *CCL2-CCR2* signaling; mast cells and NK cells via *CCL5-CCR1* and *CXCL12-CXCR4*. Fibroblasts were prominently involved in *CCL19-CCR7* signaling. Further extending their role, fibroblasts actively engaged in MHC-I and CD99 pathways. VECs demonstrated high activity in *VEGF, ANGPT,* and *CDH5* signaling pathways, crucial for angiogenesis and vascular stability; however, mucosal VECs also favored *CXCL8* and CCL pathways. LECs were predominantly involved in *CCL21-CCR7* signaling, essential for movement of dendritic cells and T cells[50]. Mural cells also contributed to CCL and CXC signaling. Keratinocytes and other epithelial-resident cells (Merkel cells and Langerhans cells) participated in MHC-I and APP pathways. Overall, this illustrates the critical role of structural cells in coordinating immune cell recruitment through CCL and CXC-motif chemokine pathways.

Spatial Immunophenotyping Confirms Fibrovascular Neighborhoods as Immunoregulatory Niches

We then hypothesized that not only would we see overall more and increasingly diverse immune cells in mucosa (as seen in Multi-IF; Figure 2), but that immune cells would display more diverse cell states. Using our proteomics atlas, we visualized 10 individual cell states across gland and mucosal samples (Figure 3f). The various markers (e.g., HLA-A, IDO1, PD-L1, PD-1) correspond to immune checkpoint markers and immune activation states. In mucosa, we see a notably higher presence of HLA-A^{+}, PD-L1^{+} and CD45RO^{+} cells in dendritic cells (DCs) (p-values=<0.0001; p-values=0.0034, p-values=0.0003, respectively; Figure 3g). Generally, CD4+ T and CD4+; FOXP3+ Regulatory T cells (Treg) also show higher expression of ICOS, a key marker for T cell

activation (p=0.0001; p=0.008, respectively). Moreover, mucosa shows increased expression of HLA-A on Tregs (p=0.07). We immunophenotyped TCNs identified in Figure 2 (expanded in Figure 3h, i). Fibrovascular hubs are predicted to function as critical coordination zones, spatially linking fibroblasts and vascular cells with both innate and adaptive immune cells, forming a diverse fibrovascular corridor. TCN^A (*Mixed-Immune 1*) was identified as a mucosal hub, enriched with innate immune cells such as dendritic cells (DCs) and markers including IDO1, HLA-A, and PD-1 (p=<0.0001 for each). TCN^B (*Fibro-Myeloid 1*), found only in mucosae, is populated by innate immune cells and shows enrichment for PD-L1 and ICOS. TCN^D (*Fibro-Lymphoid 1*) also serve as hubs, with higher concentrations of adaptive immune cells enriched for IDO1 and ICOS (p=0.0001, respectively). In contrast, *Mixed-Immune 1* and *Fibro-Myeloid 1* are glandular hubs primarily enriched for IFNG/PD-1 as well as HLA-DR/PD-L1, respectively. While TCNs are similar in cell composition across tissues, they often differ in immune states.

We then subclustered immune cells from scRNAseq to orthogonally confirm the differential gene expression profiles of specific immune cells across these organs (Figure 3j,k). As expected, distinct immune activation profiles were found. In mucosae, immune cells such as neutrophils, monocyte-macrophages, and dendritic cells display elevated expression of inflammatory genes, including *S100A9*, *CCL4*, and *CXCL8*, the latter of which was also higher in fibroblasts and vasculature. Additionally, NK/T cells exhibited genes linked to cytotoxicity and activation (*PIP, IL7R*, and *TNFAIP3)*. In contrast, the immune cells in the glands exhibit a more immune-regulatory and homeostatic profile as seen in TCNs. Neutrophils, monocyte-macrophages, and glandular DCs show upregulation of genes like *TMSB4X, FTH1*, and *AIF1*; CD8+ T and plasma cells, *HLA-A, CCL5,* and *GZMB*. This divergence of immune cell states matched our proteomic findings, suggesting that oral mucosal tissues are generally more immunoreactive than glands.

Spatial Receptor-Ligand Atlas Reveals Multicellular Interaction Modules of Mucosa and Glands

Using sequential sections from the integrated spatial proteomic atlas blocks (Figure 2), we developed a custom 300-plex MERSCOPE (Vizgen) panel, leveraging the MERFISH assay. The panel design was informed by the scRNAseq data, focusing on cell identity, cell state, and receptor-ligand pairs such as interleukins and C-C and C-X-C motif chemokines (~20% of the 300-plex panel). This data was processed similarly using *TACIT* for cell type annotation (Figure 2, 3; Extended Data 2-4; Figure 4a). We then developed *STARComm (SpaTiAl Receptor-Ligand Analysis for Cell-Cell COMMunication; Extended Data 3b)*, that uses the same receptor-ligand database as the *CellChat* framework but also takes into consideration the spatial distribution of the transcript pairs, to conduct in situ analyses of receptor-ligand (R-L) interactions. This adaptation enabled *STARComm* to quantify R-L pairs based on local proximity. By locking R-L pairs within a defined spatial radius, *STARComm* builds a network of all receptor ligands communications revealed within MERFISH data and identifies clusters of spatially dense R-L activities, which we conceptualized this as “multicellular interaction modules” (MCIMs; see *Methods*). In scRNAseq analyses, one observation was that structural cell ligand signatures were enriched in only one of the tissue compartments—either epithelial or stomal; this was observed for glands and mucosa in scRNAseq, again supporting the idea of structural cooperativity in homeostasis (Figure 4b). This followed for MERFISH datasets (Figure 4c). To lock R-L in spatial context, we first used *TACIT* for cell type annotation and assessed distribution as done for Multi-IF (Figure 4d; Extended Data 6). We again assessed TCNs, this time using the MERFISH panel instead of Multi-IF, finding similar trends of TCNs by type of distribution (Extended Data 7a). We then compared locked R-L pairs, finding slightly more predicted interaction within the glands (Extended Data 7b), and unique enrichment of cell types most likely to be interacting (Extended 7c). Comparatively, mucosal fibroblasts we much more actively involved in R-L signaling compared to glands across a wide array of R-L pairs.

To understand 1) whether there were local patterns of MCIMs like TCNs, and 2) to identify the architectures and/or cell types that may support MCIMs within tissues, we first started by analyzing MCIMs across glands and mucosa agnostic to specific cell types. In this analysis, we identified 17 multicellular interaction modules (MCIMs) across the mucosa and glands in health (Figure 4e; Extended Data 7d,e; Supplementary Table 4). Some of the modules were only found in one organ system (i.e., mucosa over glands) or even a single niche like Module 14 in minor salivary glands (Extended Data 7d). Mucosae supported 15/17 modules (missing 2, 13); glands only contained 10/17. Modules 3 and 7, defined by key ligand pairs such as *CXCL12-CXCR4*, *CCL14-CCR1*, and *CXCL16-CXCR6*, were present in 26% of all mucosae, while Modules 3 and 4, defined by ligand pairs such as *IL1-IL1R*, *CCL5-CCR5*, and *CXCL1-CXCR2*, occupied 38% of tissues in the glands. Module 3 was dominated by fibroblast-immune signaling in the mucosa through *CXCL12-CXCR4*, *CXCL16-CXCR6*, and *CCL14-CCR1*, predicted to be pathways support T cell migration and neutrophil recruitment. Module 5 and 7 revealed significant fibroblast-vascular-endothelial interactions. Module 5 highlighted *IL10-IL10RA* and *EGF-EGFR*. Module 7, observed predominantly in the mucosa, featured *CXCL8-CXCR1* and *IL15-IL15RA* signaling.

Using a similar approach to TCN connection analysis, we identified hubs of MCIMs that share spatial distribution with at least three modules (Figure 4f). Module 3 was a hub in both glands and mucosa, comprised of antigen-presenting cells along with lymphatic endothelial and immune cell types. This co-segregation of R-L was unexpected and is likely tied to structural architectural differences of the tissue types as well as transcriptional programs still to be elucidated within and between niches (Extended Data 7e). Within each module, we then went back to the distribution of cell types to look for cell type signatures that may be more associated with certain MCIMs (Figure 4g). We noticed that structural cells were the higher signaling cell types within MCIMs. Lymphoid cells were highly associated with, module 5, while the myeloid cell types were more broadly represented across MCIMs.

Oral Mucosae from Distinct Niches Are Architecturally and Functionally Diverse.

We aimed to dissect the architectural, cellular, molecular, and MCIMs across the distinct oral mucosal niches. We hypothesized that the oral cavity, with its widely differing mucosal types and morphologies[1,51,52], would support unique and niche-specific programs of structural immunoregulation. To achieve this, we subclustered our spatial proteomics and transcriptomics data, focusing on specialized, masticatory mucosa (i.e. tongue, parts of gingiva) and lining mucosa, classified as "non-keratinized" (i.e., buccal, gingiva); Figures 2-4). Doing this, we identified shared cell types and neighborhoods (Figure 5a-c) alongside distinct hubs (Figure 5d-e). The tongue and buccal mucosa exhibited high cellular diversity, with more complex TCNs compared to the gingiva. These regions shared fibrovascular-immune-enriched hubs (TCN[A, D, G]: *Mixed-Immune 1*, *Fibro-lymphoid 1*, and *Vascular-immune*, respectively). In contrast, the gingiva presented a higher diversity of hubs (TCN[A, C, D, F]: *Mixed-immune 1*, *Fibro-myeloid 2*, *Fibro-lymphoid 1*, *Mixed-immune 2*). Using fibroblasts as an anchor point and measuring the proximity to the nearest immune cell, the gingiva showed immune cells in closer proximity compared to the buccal mucosa and tongue. This relationship is further emphasized by the presence of neutrophils, which are found at half the distance from the fibroblast in the gingiva compared to the distance between neutrophils and fibroblasts in other tissues, supporting findings in health and disease[17,21,22]. Statistically, we observed significantly higher concentrations of innate immune cells in the gingiva compared to the tongue and buccal mucosa (p-value=0.0007; Figure 5f).

When assessing same-cell and different cell (homo- and heterotypic cell interactions) in all six niches, homotypic interactions between structural cell types dominated (Figure 5g). Within the stroma, fibroblasts again emerged as the primary cell types responsible for R-L activity compared to vascular cell types (Figure 5h). Overall, we observed that both homotypic interactions are 18 times more frequent than heterotypic R-L pairs. Mucosal niches such as the gingiva and buccal

mucosa show six times more interactions than the glands. *CXCL12-CXCR4* and *CCL5-CCR5* pairs were found across all three mucosal sites, suggesting core *CXCL12-CXCR4* and *CCL5-CCR5* activity in homeostasis that likely supports immune readiness, tissue surveillance, and local maintenance in mucosae. In the tongue, fibroblasts were enriched in R-L interactions via *CXCL16.* In the buccal mucosa, both fibroblasts and VECs engaging in a mix of immune and structural signaling via *CCL2-CCR4* and *CXCL8-CXCR2*.

Unique R-L pairs were more prominent in the gingiva, with fibroblasts engaging in stronger interactions with pro-inflammatory cytokines, such as *TNF-TNFRSF1A* and *IL1B-IL1R*—key pathways involved in gingivitis and periodontitis pathophysiology[53]. We again utilized cell-agnostic, R-L analysis to examine the 17 MCIMs (Figure 4). These modules captured unique and shared receptor-ligand interactions across distinct tissue sites. Module 7 was enriched for *CXCL12-CXCR4* and *CXCL8-CXCR1* signaling, particularly prevalent in the tongue, where fibroblast-vascular interactions were abundant. Similarly, Module 3, a key hub shared between glands and mucosa (Figure 4), contained *IL5-IL5RA* and *CCL5-CCR1*. Furthermore, Modules 1, 4, 6, and 9 showed varying degrees of fibroblast and VEC involvement. Surprisingly, several hubs, in the tongue (Modules 1, 2, 4, 6, 7, 9, and 10) were identified among fibroblasts and VECs. This diversity of MCIMs highlights the unique architectural and signaling dynamics across different oral niches and emphasizes the need to understand the oral cavity considering tissue origin.

Understanding the Spatial Immunoregulation of Fibrovascular Hubs in Chronic Inflammation

Understanding each mucosal niche is likely unique, we sought to explore how structural-immune interactions and spatial architectures are altered in chronic inflammation when compared to a healthy state within the same niche. We selected gingival mucosa for this study due to its unique morphology and susceptibility to both autoimmune[54] and dysbiosis-driven, chronic inflammatory[55]

conditions. By employing the same advanced scRNAseq as well as spatial proteomics and transcriptomics approach (Figure 1), we integrated both healthy and diseased datasets[22,35] (see v1 periodontitis atlas for clinical metadata[17]; Figure 6a), revealing increased and diversified immune populations (Figure 6b). Subclustering epithelia and stromal structural populations from this atlas revealed a substantial increase in all populations but most dramatically for keratinocytes and fibroblasts in chronic inflammation as published recently (Figure 6c,d)[17,21,56]. We then applied our Mult-IF and MERFISH assays to these tissues to understand if we could understand how the tissue responds in concert to chronically inflammatory stresses.

Firstly, we used Multi-IF to understand TCNs, which were integrated with healthy tissues for an integrated neighborhood analysis, again finding all 8 TCNs in health and disease. With this focused analysis, we observed a significant polarization of the lesion toward regions of known tooth-associated microbial dysbiosis (Figure 6e). Notably, the fibrovascular corridor displayed substantial fracturing into distinctly smaller TCNs compared to health, which were unique in cell-type enrichment (Figure 6f,g). This was wound closer to the tooth crown/enamel ("peri-epithelial") and deeper in the tissues near the alveolar bone ("peri-osteal") In healthy gingiva, fibrovascular cell types were found within TCN[A, C, F] (*Mixed-immune 1*, *Fibro-myeloid 2*, *Mixed-immune 2* TCNs) New hubs emerged in inflamed tissue, particularly *Fibro-myeloid 2*, *Fibro-lymphoid 1*, *Vascular-immune* TCNs. These new TCNs were characterized by the upregulation of immune checkpoint molecules like IDO1 (*Fibro-myeloid 2* TCN) and PD-L1 (*Mixed-immune, Fibro-myeloid* 2, *Vascular-immune*, and *Fibro-vascular* TCNs), alongside the pro-survival molecule BCL-2 (*Vascular-immune* TCN; Figure 6d). *Fibro-myeloid 2*, *Mixed-immune 2*, and *Fibro-vascular* TCNs were shown to downregulate HLA-A. This downregulation suggests a potentially significant reduction in TCNs enriched for fibroblasts, shifting immunoregulatory potential and arrangement of the fibrovascular corridor (Figure 6h). As such, only *Fibro-immune 1* and *Vascular-immune* TCNs emerged as the most interconnected hubs in networks analysis of inflammation (Figure 6i).

Overall, the stromal tissue nearest to the dysbiosis has completely restructured itself in response to dysbiosis, doing so in a limited volume of tissue to address the barrier challenge.

Decoding MCIMs During Chronic Inflammation

Building on this, we then examined the MERFISH data, where we observed similar shifts in fibroblast signaling as scRNAseq (Figure 6a-d) but focused on the immune activated fibroblasts at the epithelial-stromal border where TCNs were most fragmented in the Multi-IF analysis (Figure 6e,f; Figure 7a). In both periodontitis and another mucosal inflammatory condition of this niche (peri-implantitis, included to observe inflammation of the niche in another way), fibroblasts showed expanded signaling pathways, with increased expression of *TNF*, *IL1B*, *CXCL8*, and *AREG* (Extended Data 7f). Assessing R-L locked pairs, we observed a substantial upregulation of fibroblast signaling pathways in periodontitis compared to VECs, mural cells, and LECs across the various pathways (chemokines, interleukins, EGFR; Extended Data 7g). Furthermore, when assessing MCIMs, 20 were detected in disease (Extended Data 7h), compared to the 17 total in health across glands and mucosa and the 15 of which were detected in mucosae (Figure 4h).

As TCNs, the spatial arrangement of MCIMs between health and disease was substantially reorganized (Figure 7b). In healthy tissue, we observed a cohesive arrangement of modules, where distinct hubs maintained (i.e., Module 9 in health). Yet, as seen for TCNs, there is a fracturing of the MCIMs, as seen in the emergence of new hubs (Modules 3 and 7). These disease-associated hubs are marked by more homotypic interactions within fibroblasts (autocrine and paracrine), as well as an increase in heterotypic interactions between fibroblasts and CD4+ and CD8+ T (Figure 7c). This shift in fibroblast activity, from primarily maintaining structural and homotypic interactions in health, to driving inflammation and tissue degradation in disease, is further evidenced by the increased receptor-ligand pairs involving inflammatory cytokines (e.g., *IL1B, TNF*) and chemokines (e.g., *CXCL12, CCL5*).

MCIMs to Create Filtered List of Potential Therapeutics using Drug2Cell

We aimed to delve deeper into identified tertiary lymphoid structures (TLS) observed in the gingival mucosa during inflammation. TLS are ectopic lymphoid tissues that arise in response to chronic inflammation and serve as a local site for immune activation, fostering antigen presentation and lymphocyte priming[57]. Using *TACIT* and *Astrograph*, we precisely mapped the stromal architecture around the TLS, identifying fibroblasts and neighboring immune cells (Figure 7d). We observed an enriched structural-immune interaction zone around and within the TLS, as is known for reticular fibroblast subtypes in secondary lymphoid tissues like palatine tonsils[58]. We sought to understand the relationship between various clusters and the specific roles of homotypic and heterotypic R-L interactions. Modules 1, 7, 8, 13, 14, and 17 were primarily localized within the TLS itself (enriched for receptor engagement by *CCL2, CCL3, CCL5, CX3CL1 IL1B, IL3)*, while Modules 2, 3, 4, 5, 6, and 20 (enriched for receptor engagement by cytokines: *CCL2, CCL4, CCL5, CCL17, CCL14, CCL22, CCL23, CXCL1, CXCL8, CXCL10, CXCL3, CXCL9, CXCL12, IL1A, IL6, IL10, IL12B, IL15, and TNF* as well as EGFR ligands: *EGF, AREG, and HBEGF*) surrounded the TLS in the broader stromal environment (Figure 7d). We grouped R-L interactions, aiming to distinguish structural-immune (heterotypic), immune-immune (homotypic), and structural-structural (also homotypic) interactions. Our findings revealed a significant increase in heterotypic interactions (p-value<0.0001), particularly between structural and immune cells (p-value<0.0001; Figure 7e). Immune-immune (homotypic) interactions also increased appreciably within the TLS; however, structural-structural interactions, which often maintain tissue integrity, did not increase.

This analysis enabled us to filter specific R-L interactions that contribute to the disease origin and progression, offering potential targets for therapeutic intervention. We then adapted and applied *Drug2cell* in spatial context, focusing on the TLS. We aimed to identify cell-cell interactions

between fibroblast-immune cells within the TLS niche that could be therapeutically targeted. The analysis revealed that nearly all interactions, particularly those between fibroblasts and immune cells, were targetable by existing pharmacological agents (Figure 7f; Supplementary Table 4). To visualize these findings, we plotted druggable targets specific to the TLS (Figure 7g). The results pointed to several key drugs, such as Sargramostim and Satralizumab, known for immune modulation, emerged as potential candidates for targeting inflammatory fibroblast-immune interactions (Figure 7h). Additionally, immune checkpoint inhibitors such as Basiliximab and Muromonab-CD3 appeared among the top targetable agents, aligning with the upregulation of immune checkpoints in inflamed regions. This methodology opens new avenues for therapeutic intervention in chronic inflammatory diseases, enabling more precise targeting of specific cellular interactions within distinct tissue microenvironments and regions of interest where fibroblasts are predicted to coordinate and control immune activity within distinct TCNs and MCIMs.

DISCUSSION (931):

This study presents the first comprehensive single-cell and spatial exploration of oral and craniofacial tissues as part of the Human Cell Atlas Initiative[4,59,60]. Here, by focusing on structural cells and their roles in immune regulation and tissue homeostasis within a spatially resolved, single-cell context, we reveal the intricate architecture of cellular interactions specific to the oral organ system of glands and mucosae. This work establishes a foundation for an oral and craniofacial common coordinate framework (CCF)—a standardized reference that enables consistent mapping and analysis of cellular and molecular data across diverse samples, tissues, and organisms (Figure 1a)[61]; this builds off our CCF for periodontium[17]. Such a framework is essential for understanding the spatial dynamics of cellular communities, revealing reproducible patterns in cell distributions and signaling interactions, especially as this atlas grows and similar atlases develop across other tissues in 2D and 3D[62,63]. By linking single-cell and spatial data to anatomical landmarks through the CCF, our approach enables reproducible mapping of immune cell distributions, cellular communities, and signaling interactions across studies. This framework facilitates cross-study comparisons and integration of findings, enhancing our ability to study the spatial dynamics of structural cells in different tissue conditions. Our study demonstrates that the CCF can serve as a foundational tool for spatially aware research in both healthy and diseased tissue contexts, allowing researchers to gain consistent insights across different labs and models. This facilitates scientific progress by enabling reproducible and comparable findings across studies.

Our findings also support and extend the emerging concept of "structural immunity," in which structural cells like fibroblasts are active participants in immune modulation within glandular and mucosal tissues, rather than passive components of tissue architecture.[64] Previous studies have demonstrated the role of epithelial cells and fibroblasts in this process[21,22,35], but this present work provides new insights into how broader tissue architecture supports immunoregulation and

homeostasis. Importantly, within these tissue cellular neighborhoods (TCNs) and multicellular interaction modules (MCIMs), fibroblasts and vascular components, such as vascular endothelial cells, mural cells, and lymphatic endothelial cells, appear to play prominent roles in guiding immune cell localization and immunoregulation, underscoring their influence among structural cells within soft tissue niches common to the oral cavity—but also the lower airways, gastrointestinal and genitourinary tracts.

The implications of this work are significant for advancing our understanding of fibroblast heterogeneity and the impact of spatial organization on immune responses. Our analysis highlights the need to further investigate the diversity of fibroblast subtypes across the oral cavity, including peri-epithelial, perivascular/adventitial, peri-neural fibroblasts, and myofibroblasts[19,65], as well as disease-associated fibroblasts such as cancer-associated fibroblasts (CAFs)[66] and reticular fibroblast subtypes that support the emergence of tertiary lymphoid structures (TLSs) seen in head and neck cancers, Sjogren's Disease, Graft-versus-host-disease, and periodontitis[17,33,39,58,67]. These fibroblast subtypes still need discovery but are likely conserved. In other tissues, these demonstrate distinct immune-regulating functions, suggesting that fibroblasts actively contribute to localized immune responses by creating immunological niches . By influencing the activity and distribution of immune cells within tissue ecotypes, fibroblasts help maintain balanced immune responses in both health and disease, underscoring their critical, context-dependent roles.

This study's use of advanced multiomics techniques, including single-cell RNA sequencing, spatial proteomics, and spatial transcriptomics, allows unprecedented observation of cellular interactions within complex tissue environments. Combined with our comprehensive analysis suite—comprising *TACIT, Constellation, StarComm,* and *Astrograph*[33]—and specialized interaction tools like *CellPhoneDB, CellChat, MultiNicheNet,* and a spatially adapted *Drug2Cell*[68-71], this approach facilitates detailed mapping of cell-cell communication networks across various

tissue conditions. The ability to map these interactions in a spatially defined framework enables researchers to track how fibroblasts and other structural cells contribute to immune processes in both healthy and diseased states. This engineering will also support the study of TCNs and ecotypes, advancing knowledge of how spatial arrangements in tissues influence cellular behavior and tissue resilience. Although our platform allows detailed examination of cellular neighborhoods and their signaling networks, certain limitations exist. For example, variability among tissue samples could impact reproducibility. Additionally, while we identify key fibroblast TCNs and signaling pathways, further work is needed to unravel the exact mechanisms by which specific fibroblast subsets influence immune cell behaviors in disease. Future research will need to address these limitations, particularly as new tools improve resolution and expand our ability to study cellular ecotypes within varied tissue landscapes.

The therapeutic implications of this work are significant, particularly for diseases where fibroblast-immune interactions play central roles, such as cancer, autoimmune disorders, and chronic infections. By identifying distinct fibroblast-driven signaling pathways, such as IL1B or TNF signaling, as well as PD-L1-mediated immune checkpoint pathways, our study opens new possibilities for targeting fibroblast-immune crosstalk specifically. In chronic inflammatory diseases, for instance, modulating these fibroblast-specific pathways could help mitigate immune-mediated tissue damage. In cancer, targeting CAFs and the immune-exclusion zones within the tumor microenvironment could improve immune infiltration and enhance response to immunotherapies by targeting mechanisms of fibrosis simultaneously to immunomodulation. Our study provides a spatially aware framework for identifying niche-specific therapeutic targets, potentially leading to more precise interventions that address localized immune dysregulation while maintaining tissue homeostasis.

Overall, this study contributes a foundational understanding of the roles of structural cells, particularly fibroblasts, in immune regulation within oral and craniofacial tissues. By integrating

spatial and single-cell multiomics data into a common coordinate framework, we offer a robust model for examining the spatial dynamics of structural immunity and for advancing targeted therapies. As the field evolves, the tools and frameworks developed here can extend to other tissues and diseases, paving the way for future research into how spatial biology shapes immune responses across tissue ecotypes, communities, and states of health and disease.

FIGURE LEGENDS:

Figure 1. The First Integrated Atlas of Oral and Craniofacial Tissues. | (a) The common coordinate framework of integrated niches in the oral cavity analyzed, comprehending 13 macroniches in the oral cavity. (**b)** Publicly available studies (blue) and newly generated single-cell RNA sequencing (scRNAseq) data (red) were integrated into the study. (**c)** A niche-matched FFPE collection was curated for spatial evaluation, representing each niche in hematoxylin & eosin staining. (**d)** Integration of multi-modal data to construct a harmonized cell atlas for spatial multiomics and single-cell transcriptomics analysis. The scRNAseq Atlas (gray) integrates both publicly available datasets and newly generated data. The Multiplex IF Atlas (red) employs a 40-antibody panel to capture cell types and neighborhood architectures, while the Multiplexed ISH Atlas (green) applies a 300-plex panel to map cell types and cell-cell communication networks within tissues. Integrated analyses (blue) focus on cell-cell communication. (**e)** Tiered annotation across scRNAseq data: Tier (T) 1: structural cell types and immune cell types and T1: structural cells are categorized into epithelial, stromal, vascular, lymphatic, muscle, and neural crest; immune cells are characterized as myeloid and lymphoid. (**f)** Sex distribution (male, female, N/a), age groups (18-39, 40-49, 50-59, 60-69), and race/ethnicity (White, Asian, Black, Indian, Latino/a, N/a) of patients included in the scRNAseq Atlas. (**g)** The integrated OCF Atlas (Tier 1), with UMAP plots depicting cell clustering. The top UMAP plot illustrates clusters based on cell types, the bottom plot shows clustering according to niche. The dataset includes 246,102 cells from 70 samples, spanning 14 studies and representing 13 distinct niches (available at CELLxGENE). The color-coded legend indicates the different tissue types: glands (pink), mucosa (green), and pulp (blue). (**h)** The heterogeneity of T1 annotated cell types across samples, each bar representing a different patient/sample. The samples are grouped by tissue niches, including anatomical regions such as the salivary glands, mucosa, and dental pulp. Each color in the stacked bars represents distinct cell types, categorized using the tiered annotation framework. **(i)**

Receptor-ligand analysis using *CellPhoneDB*, showing communication between structural cells and immune cells. The heatmap highlights interaction, with the strongest communication indicated by (+). Notably, the highest interaction is observed from and between vascular cells and fibroblasts. (**j)** Chord plot showing the receptor-ligand statistical differences in the top 50 interactions across all niches, where communication was inferred between structural and immune cell types. **(k,l)** Pathway analysis validated the observed trends for both outgoing (ligands) and incoming signaling patterns (receptors) using *CellChat*, showing the statistical differences between receptors and ligands. *Abbreviations: FFPE* = formalin-fixed paraffin embedded, *OCF* = Oral and Craniofacial, *T1* = Tier one. Scale bars: Glands: 100 µm Mucosal 250 µm; pulp 50 µm (c).

Figure 2. Spatial Proteomics Atlas Across Mucosal and Glands Niches. | (a) Comparison of Tier 2 cell type annotations from scRNAseq data and the multi-IF (PhenoCycler-Fusion 2.0; Akoya Biosciences) annotation. (**b)** Voronoi masks of representative glands and mucosa showing cell types annotated using *TACIT* (part of *AstroSuite*; see *Methods*). Inset shows a representative peri-epithelial area with vascular endothelial cells (VEC) in the mucosa and the stromal region of a major salivary gland. (**c)** Cell type co-occurrence between structural and immune cell types, demonstrating the heterogeneity of cell co-localization in glands and mucosa. (**d)** Integrated definition of tissue cellular neighborhoods (TCNs) defined by *Constellation* in glands and mucosa. (**e)** Bar plot showing the proportion differences of immune cell types between glands and mucosa. Immune cells were separated from structural cells for this analysis; p-values for each comparison are macrophages are higher in the mucosa (p=0.014); CD8 T cells were higher in the glands (p=0.002). (**f)** Spoke-and-wheel plots were used to analyze the co-occurrence and average proximity of structural and Tier 2 immune cells. In these plots, the central point represents a structural cell, and the cell-cell distance, measured in microns, indicates the distance to the nearest immune cell from the structural cell at the center. The thickness of the spokes reflects the

number of interactions identified between the structural cell type and immune cells, averaged across glands and mucosa. (**g)** Heatmaps showing TCN cell type enrichment, with TCN annotations based on the percentage of cells distributed within each neighborhood. The *red* and *blue* regions highlight hubs, which represent cell types that interact with more than two other neighborhoods. (**h)** Graph representation of the co-occurrence of neighborhoods across glands and mucosa, highlighting the fibrovascular TCN as hubs of interaction with more than two other TCNs. (**i)** Spatial distribution of TCN hubs across glands and mucosa, highlighting the heterogeneity of TCN co-occurrence in these tissues. Abbreviations: *TCN* = Tissue cellular neighborhoods. Scale bars *Insets 50* µm: (a-d). Statistical test: One-way ANOVA and T-test.

Figure 3. Structural Immune Analysis of Salivary Glands and Oral Mucosa. | (a) Tier 2 annotation of structural and immune cells. (**b)** All integrated glands datasets were subclustered from the integrated oral craniofacial atlas (Figure 1), annotations were carried across the subclustering. Cluster colors match with the cell types of colors presented in a. (**c)** Proportion of total cells reveal epithelial-rich samples from glands and more balanced structural and immune cell types in mucosa. **(d,e)** Receptor ligands analysis using *CellPhoneDB* and Tier 2 annotations across glands and mucosa showing high signaling from structural-to-structural cells and structural-to-immune cells in glands and mucosa. Cell-cell communication in structural cell types shows enrichment of CXCL, CCL, MHC, and MIF immune-related pathways. (**f,g)** Spatial validation of cell immunophenotyping in glands and mucosa showing co-occurrence and different cell states Heatmap presenting the relative proportion of cell states distribution of various immune checkpoint markers and immune activation states, including HLA-A, IDO1, PD-L1, and PD-1. In the mucosa, there is a significantly higher presence of HLA-A+, PD-L1+, and CD45RO+ cells within dendritic cells. (**h,i)** Spatial distribution of the TCN co-occurrence. The proportion of cell states distribution in each of the neighborhoods in glands and mucosa showing increased expression of ICOS in CD4+ T cells and CD4+; FOXP3+ Regulatory T cells (Tregs). (**j)**

Subclustered immune cells in glands and mucosae and cell type proportions in each niche. (**k)** Differential gene expression analysis by each of the immune cell types in glands and mucosa. *Abbreviations:* CXCL = Chemokines, CCL = Chemokines Ligands, MHC = Major Histocompatibility Complex, MIF = Macrophage Migration Inhibitory Factor, TCN = Tissue cellular neighborhoods. Statistical test: T-test.

Figure 4. Spatially Resolved Cell-Cell Communication Networks within the Oral Cavity | (a) The *AstroSuite* pipeline used for the MERFISH 300-plex spatial transcriptomics assay, now incorporating *STARComm*. Annotated cell coordinates and co-occurrences were extracted based on the spatial distribution of cell types, and receptor-ligand pairs retrieved from a database were used to identify matched pairs within a 50-micron radius, allowing for comparisons across macroniches. (**b)** scRNAseq analysis revealed structural ligand signatures enriched exclusively in one structural compartment—either epithelial or stromal—within glands and mucosa and in distinct cell types. (**c)** MERFISH single-cell spatial analysis revealed similar structural ligand signatures enriched one cell type and validating our scRNAseq receptor-ligands signatures. (**d)** Voronoi mask reconstruction of glands and mucosa using *TACIT* on MERFISH data to assign the cells of glands and mucosa (**e)** Spatial distribution of the multicellular interaction modules (MCIMs) across the mucosa and glands. (**f)** The identification of MCIM hubs with spatial overlap in at least three modules, Module 3 emerged as a key hub in glands and mucosa. (**g)** The co-occurrence of MCIMs again created interaction hubs, here defined by the R-L module. *Abbreviations:* MCIM = Multicellular Interaction Modules. Statistical test: One-way ANOVA and T-test.

Figure 5. Exploring Differences of TCNs and MCIMs Across Oral Mucosal Types. | (a) Tiered cell annotation and neighborhood (TCN) classification for this mucosal niche specific analysis using two mix keratinized sites (dorsal tongue, gingiva) and non-keratinized (buccal mucosa). Each is functionally and biogeographically distinct within the oral cavity. (**b)** Cell type distribution

of the three mucosal niches. (**c)** TCNs distribution of the three mucosa niches showing the share and unique TCNs in the neighborhood cell type distribution (see Figure 2 for TCN names). (**d-f)** Cell type proportion distribution of each TCN across mucosal niches shows that the tongue and buccal mucosa exhibited high cellular diversity, with more complex TCNs compared to the gingiva. (**e**) Within the fibrovascular corridor of each mucosal type, buccal and tongue exhibited the same TCN pattern whereas gingiva displayed a unique hub arrangement. (**f**) Spoke-and-wheel plots support enrichment of diverse innate and adaptive populations across using structural cell types as the center point. (**g)** Heterotypic interactions in healthy tissues are 18 times higher than homotypic interactions, occurring expectedly between structural cell types over immune. (**h)** Using the *STARComm* pipeline we analyzed each of the receptor-ligands pair enriched in each of the three niches to highlight specific structural-immune interactions. (**I, j)** Spatial distribution of receptor-ligand pairs across the mucosal niches, showing the co-occurrence of MCIMs. The tongue and buccal mucosa share one of the hubs, with CCL-CXCL and IL1 pathway enrichment observed across the stroma of the mucosal niches. *Abbreviations:* TCN = Tissue cellular neighborhoods, . Statistical test: One-way ANOVA and T-test.

Figure 6. Reorganization of Fiborvascular TCNs and MCIMs upon Inflammatory Stress | (**a**) UMAP of the integrated scRNAseq periodontitis atlas, highlighting different cell types across both conditions, with a total of 86,174 cells from 24 samples across two studies and one niche. (**b)** Bar plot compares the proportion of cell types in healthy and periodontitis samples, illustrating an increase in immune cells in periodontitis compared to healthy gingiva. (**c)** Spatial transcriptomics using our 300-plex MERSCOPE, with UMAPs depicting the distribution of cell populations in health (*top*) and periodontitis (*bottom*). (**d)** Representative structural cell signatures, matched to Figure 4b,c. In periodontitis, there is a notable increase in IL signaling pathways within the epithelial and stromal compartments, indicating heightened inflammatory activity in disease. **(e,f)** Images of chronic inflammation from Multi-IF (e) and *Constellation*-generated TCNs (f) in chronic

inflammation (periodontitis) with lesion polarity focused on the right side of the image where dysbiosis originates near the tooth surface. The biogeographical distribution of TCN clusters B, D, F, and G is more prominent in chronically inflamed areas (see Figure 2 for TCN names). (**g)** Heatmaps of cell type proportions for each TCN across periodontitis and healthy gingiva, highlighting the the fibrovascular shifts in periodontitis. (**h)** Volcano plot showing the upregulation of inflammatory activation markers in the TCNs more when compared with healthy gingiva. (**i)** The TCN co-ocurrance (hub analysis) showing changes of the fibrovascular corridor in periodontitis. Abbreviations: *TCN* = Tissue cellular neighborhoods. Statistical test: One-way ANOVA and T-test.

Figure 7. Targeting Focal Fibroblast-Immune Axes for Precision Therapeutics | (a)) Images of chronic inflammation from MERFISH, a sequential secton from Figure 6 (a). **(b)** Co-occurrence of receptor-ligand motifs, highlighting a cohesive module arrangement in healthy tissue, particularly Module 9, which acts as a hub. In disease, new hubs (Modules 3 and 7) emerge. (**c**) Preferential cell interactions are marked by increased homotypic interactions within fibroblasts and heterotypic interactions between fibroblasts and CD4+ and CD8+ T cells. (**d)** Spatial distribution of stromal cell types and receptor-ligand motifs, identifying structural responses around the tertiary lymphoid structure (TLS), with fibroblasts and immune cells forming interaction zones .(**e)** Homotypic and heterotypic interactions in health and disease, revealing a significant increase in heterotypic structural-immune interactions in disease, while structural-structural interactions remain statistically unchanged. The MCIMs in the same area presented a high prevalence of interactions in the TLS area. (**f)** Drug expression comparisons within TLS regions in periodontitis, focusing on three groups: fibroblast (ligands) to immune (receptor), immune (ligands) to fibroblast (receptor), and immune (ligands) to immune (receptor). Each point represents a single drug, with its position indicating the log2 fold change (log2FC) of that group compared to the others. Points closer to or oriented towards a specific group indicate higher enrichment or drug expression within that group. In this analysis, most points (drugs) are directed

towards the fibroblast (ligands) to immune (receptor) group, indicating that drug expression in this group is higher than in the other two groups within TLS regions**. (g)** Spatially-defined therapeutic enrichment analysis, highlighting potential drugs targeting structural-immune signaling pathways, particularly within TLS regions using *Drug2Cell*. Drugs such as Sargramostim, Lifitegrast, and Satralizumab show potential efficacy in modulating key pathways within diseased tissues regions of interest—here the ROI of the TLS. **(h)** Confirming the predicted spatial targeting of fibroblasts around the TLS. *Abbreviations:* TCN = Tissue cellular neighborhoods, MCIM = Multicellular Interaction Modules, TLS = Tertiary Lymphoid Structure. Statistical test: One-way ANOVA and T-test.

EXTENDED DATA FIGURES:

Extended Data 1: **Niche Diversity and Cell Types within the Oral and Craniofacial Atlas | (a)** The integrated scRNAseq atlas was subclustered into major and minor glands, oral mucosae, and studies involve pulp (all available at CELLxGENE). (**b)** Cell type proportion in each of the subcluster UMAPs. (**c)** DEGs of Tier (T) 1 cell type annotation in the subclustered UMAPs. (**d)** CellPhoneDB plot showing the receptor-ligands pairs between structural cell types and immune cell types. (**e,f)** The *MultiNicheNet* plot card showing the top 50 up-down regulation of receptor ligands pairs, communication subtraction from glands to mucosa (*left)* and mucosa to glands (*right).* Abbreviations: DEG = Differential Expression Genes

Extended Data 2: Building and Validating a Spatial Proteomics Atlas of the Oral Cavity. | (**a-c)** By utilizing Phenocycler Fusion (PCF) technology we created a 40-antibody panel (a) to analyze the tropism and colocalization of innate and adaptive immune cells within specific neighborhoods and structures of the 6 distinct tissues (b) using 2 slides at a time (c). (**d)** To extract single-cell spatial data we trained a segmentation model using tissues from the oral cavity using *Cellpose3.* (**e)** Multiplex IF of a representative buccal mucosa and glands tissue from each marker used on the Phenocycler Fusion.

Extended Data 3: Overview of *Constellation* and *STARComm*, from *AstroSuite*. | (**a**) The proposed method, *Constellation*, is designed to identify and classify Tissue cellular neighborhoods (TCNs) within Whole Slide Images (WSIs) by leveraging diverse spatial partitioning strategies and unsupervised learning techniques. It involves four primary stages: (1) WSI Partitioning, (2) Sub-graph Construction, (3) Collective Unsupervised Training, and (4) Consensus-Based TCN Assignment. In Step 1, the input WSI, consisting of cells characterized by their x-y coordinates and cell types, is partitioned into a set of disjoint subsets using three strategies: vertical stripes,

horizontal stripes, and square grids. These partitions are recursively refined to ensure an appropriate number of cells for effective analysis. In Step 2, sub-graphs are constructed for each subset, where nodes represent cells with cell type information, and edges capture spatial relationships based on k-nearest neighbors. Step 3 involves collective unsupervised training of these sub-graphs across all partitioning strategies, utilizing a MinCut-based loss for enhanced clustering and orthogonality to enforce unique TCN assignments. Sub-graphs are randomly shuffled in every epoch to promote robust learning. In Step 4, a consensus-based TCN assignment is derived through majority voting over the TCN assignments from the different partitioning methods, ensuring robust TCN classifications. Cells without consensus across partitioning strategies are filtered to enhance the reliability of the final analysis. This final stage identifies robust community structures across the tissue, enhancing the detection of spatial cell interactions through adaptive partitioning and graph representation. (**b**) The *STARCOMM* analysis pipeline begins with input data, including X,Y coordinates for each cell, a CELLxGENE matrix (containing expression data for ligands and receptors per cell), and a curated list of ligand-receptor (LR) pairs. In Step 1, the nearest neighboring cells within a 50-micron radius are identified for each cell. A connection is established if one cell expresses a ligand gene, and its neighbor expresses the corresponding receptor gene (with a count value > 0). Step 2 divides the tissue image into a grid, and for each LR pair, a spatial kernel density score is calculated, identifying regions with dense LR interactions. In Step 3, this process is repeated across all LR pairs, and the Louvain clustering algorithm is applied to the density scores, revealing clusters of spatially co-located LR pairs. Abbreviations: WSI = Whole Slide Imaging, TCN = Tissue cellular neighborhoods, MCIM = Multicellular Interaction Modules**,** RL = Receptors - Ligands

Extended Data 4: Building and Validating a Spatial Proteomics Atlas of the Oral Cavity to Understand Multicellular Interaction Modules. | (a) Cell type annotation across all niches using spatial proteomics. *TACIT* was used to self-threshold all slides, generating Voronoi masks and

slide reconstructions for each annotated cell at its respective coordinates. (**b)** TCN annotation was performed across all niches in glands and mucosae in a single batch to evaluate the different structures and cell proportions within each cluster. *Abbreviations* = TCN = Tissue cellular neighborhoods.

Extended Data 5: Cell-cell communication patterns of structural cell types in glands and mucosae. | (**a)** *MultiNicheNet* chord plots showing the top 50 interactions between tier one annotated cells, on *left* the top 50 interactions on glands, and on *right* top 50 interactions in mucosa. (**b)** Venn diagram showing the overlap of receptor-ligand inference in between glands and mucosa, and structural to immune and immune to structural cells (**c)** Heatmap illustrates pathway enrichment across mucosal niches, illustrating the critical role of innate cell recruitment in mucosal cell-cell communication. This recruitment significantly contributes to the inferred interactions between various cell types, emphasizing the importance of innate immune responses in maintaining mucosal tissue homeostasis and in the development of disease-related processes.

Extended Data 6: Macroniche Regulation in Chronic Inflammation | Representative samples across glands and mucosae were assigned using *TACIT* to create the Voronoi masks similar to Multi-IF datasets.

Extended Data 7: Patterns of TCNs and MCIMs as defined by MERFISH. (**a)** Cell proportion distribution in each of the TCNs in glands and mucosa, with TCN^E being the only similar proportion between the two niches. The heterogeneity of TCN distributions highlights the differential arrangements of cell types when analyzed through spatial proteomics assays. (**b)** The total number of interactions for each MCIM across glands and mucosa, with both niches combined into a single analysis, shows a higher number of interactions in glands. (**c)** Heatmap displays the top 50 interactions by cell type, showing that in both glands and mucosa, structural cell types—such

as epithelial cells (acinar, basal, suprabasal keratinocytes) and fibroblasts—are the primary signaling cell types. The interaction signatures of each cell type vary when analyzing glands and mucosa separately. (**d)** The proportion of each cluster distributed across macroniches reveals that some MCIMs are exclusive to specific niches, underscoring the unique communication characteristics of each niche and emphasizing the need to consider cell-cell communication as distinct within each context. The presence of MCIMs also varies between glands and mucosa, with approximately half being exclusive to each tissue type. (**e)** MCIMs contains groups of R-L pairs co-occurring in the same tissue space. (**f)** Spatial Transcriptomics (MERFISH) of healthy (*top*) and periodontitis (*bottom*) samples, showing shifts in ligand expression related to structural cell types across health, periodontitis, and peri-implantitis. The ligand signatures found in health are now mixed in disease. (**g)** Receptor-ligand upregulation of fibroblast signaling pathways in periodontitis compared to VECs, mural cells, and LECs across the pathways. (**g)** Heatmap of the MCIMs clustering across the MCIMs in periodontitis. *Abbreviations.* TCN = Tissue cellular neighborhoods, MCIM = Multicellular Interaction Modules, R-L = Receptor-Ligands, VEC = Vascular Endothelial Cells

SUPPLEMENTAL TABLES

Supplemental Tables 1. Phenocycler-Fusion Antibody Panel

PCF Antibody	Wavelength	Clone	Barcode/Reporter	Concentration	Catalog #
CD8A	Atto550	AKYP0028	BX/RX026	1/200	4250012
CD4	AF647	AKYP0048	BX/RX003	1/200	4550112
CD20	AF750	AKYP0049	BX/RX020	1/200	4450018
FOXP3	AF647	AKYP0102	BX/RX031	1/200	4550071
Ki67	Atto550	AKYP0052	BX/RX047	1/200	4450096
PHH3	AF647	AKYP0060	BX/RX030	1/200	4550115
EPCAM	AF750	AKYP0119	BX/RX091	1/200	4450088
HLA-A	AF750	AKYP0078	BX/RX004	1/200	4450046
Galectin-3	Atto550	AKYP0067	BX/RX035	1/200	4450034
CD3E	AF647	AKYP0062	BX/RX045	1/200	4550119
CD45RO	Atto550	AKYP0059	BX/RX017	1/200	4550127
CD45	AF647	AKYP0074	BX/RX021	1/200	4550121
CD21	Atto550	AKYP0061	BX/RX032	1/200	4450027
PD-L1	AF647	AKYP0070	BX/RX043	1/200	4550128
CD14	Atto550	AKYP0079	BX/RX037	1/200	4450047
PD-1	AF647	AKYP0070	BX/RX046	1/200	4550038
MPO	Atto550	AKYP0113	BX/RX098	1/200	4250083
CD68	AF647	AKYP0050	BX/RX015	1/200	4550113
IDO1	AF647	AKYP0084	BX/RX027	1/200	4550123
CD31	AF750	AKYP0047	BX/RX001	1/200	4450017
KRT14	Atto550	AKYP0064	BX/RX002	1/200	4450031
CD107a	AF647	AKYP0004	BX/RX006	1/200	4550098
KRT8/18	AF750	AKYP0112	BX/RX081	1/200	4450082
CD141	Atto550	AKYP0124	BX/RX087	1/200	4250097
ICOS	AF647	AKYP0090	BX/RX054	1/200	4550117
SMA	AF750	AKYP0081	BX/RX013	1/200	4450049
PDPN	Atto550	AKYP0007	BX/RX023	1/200	4250094
COL_IV	AF647	AKYP0083	BX/RX042	1/200	4550122
CD34	Atto550	AKYP0088	BX/RX025	1/200	4250057
HLA-DR	AF647	AKYP0063	BX/RX033	1/200	4550118
CD38	Atto550	AKYP0110	BX/RX089	1/200	4250080
Bcl2	AF647	AKYP0120	BX/RX085	1/200	4550089
IFNG	Atto550	AKYP0093	BX/RX020	1/200	4250062
CD66A/C/E	AF647	AKYP0080	BX/RX016	1/200	4550001
Vimentin	AF750	AKYP0082	BX/RX022	1/200	4450050
CD11c	AF647	AKYP0147	BX/RX024	1/200	4550135
PanCK	AF750	AKYP0053	BX/RX019	1/200	4550020
Caveolin	Atto550	AKYP0086	BX/RX086	1/200	4550084
CD79a	AF647	AKYP0092	BX/RX092	1/200	4550103
GP100	AF750	AKYP0123	BX/RX123	1/200	4550091

Supplemental Tables 2. Metadata Associated with the scRNAseq Data and Spatial Data

Sample ID	Study	Specimen_from_organism.organ	Specimen_from_organism.organ_part	Donor	Age range	Donor_Sex	ancestry_harmonised	Ethnicity / ancestry	Race	Tobacco (> 20 pack-years)	Alcohol
B1	Williams *et al.* 2021	Buccal Mucosa	Not provided	1	20-29	Female	Asian		Asian	Not provided	Not provided
B2	Williams *et al.* 2021	Buccal Mucosa	Not provided	2	20-29	Female	Asian		Asian	Not provided	Not provided
B3	Williams *et al.* 2021	Buccal Mucosa	Not provided	3	30-39	Female	Black		African American	Not provided	Not provided
B4	Williams *et al.* 2021	Buccal Mucosa	Not provided	4	20-29	Female	White		Caucasian	Not provided	Not provided
B5	Williams *et al.* 2021	Buccal Mucosa	Not provided	5	20-29	Female	White		Caucasian	Not provided	Not provided
B6	Williams *et al.* 2021	Buccal Mucosa	Not provided	6	20-29	Male	White		Caucasian	Not provided	Not provided
B7	Williams *et al.* 2021	Buccal Mucosa	Not provided	7	20-29	Male	Asian		Asian	Not provided	Not provided
B8	Williams *et al.* 2021	Buccal Mucosa	Not provided	8	20-29	Male	White		Caucasian	Not provided	Not provided
L1	Unpublished, Warner Lab	Labial mucosa	incisional biopsy	3629	30-39	Female	White	White	Caucasian	none	none/social
L2	Unpublished, Warner Lab	Labial mucosa	incisional biopsy	3630	60-69	Male	White	White	Caucasian	none	none/social
L3	Unpublished, Warner Lab	Labial mucosa	incisional biopsy	3631	50-59	Male	White	White	Caucasian	none	none/social
L4	Unpublished, Warner Lab	Labial mucosa	incisional biopsy	3632	60-69	Female	White	White	Caucasian	none	none/social
L5	Unpublished, Warner Lab	Labial mucosa	incisional biopsy	3633	60-69	Female	White	White	Caucasian	none	none/social
SP1	Kretzschmar et al.	Soft Palate	palate	1	60-69	Male	White	White	Caucasian	Not provided	Sometimes
MSG1	Huang *et al.* 2021	Minor Salivary Gland	Mandibular labial minor salivary glands	1	30-39	Female	Not provided				
MSG2	Huang *et al.* 2021	Minor Salivary Gland	Mandibular labial minor salivary glands	2	50-59	Female	Not provided				
MSG3	Huang *et al.* 2021	Minor Salivary Gland	Mandibular labial minor salivary glands	3	40-49	Female	Not provided				
MSG4	Huang *et al.* 2021	Minor Salivary Gland	Mandibular labial minor salivary glands	4	50-59	Male	Not provided				
MSG5	Huang *et al.* 2021	Minor Salivary Gland	Mandibular labial minor salivary glands	5	40-49	Male	Not provided				
MSG6	2022	Gland	gland	1	20-29	Male		non-Hispanic, 1 White Hispanic and 1 African. No			
MSG7	2022	Gland	gland	2	20-29	Male					
MSG8	2022	Gland	gland	3	30-39	Female					
MSG9	2022	Gland	gland	4	20-29	Female					
P1	Consortium 2022	gland	Parotid salivary gland	1	60-69	Female	White	White	European	No	Not provided
P2	Consortium 2022	gland	Parotid salivary gland	2	50-59	Male	White	White	Not provided	Yes	Yes
P3	Pringle	gland	Parotid salivary gland	1	50-59	Male	White	White	Not provided		
P4	Pringle	gland	Parotid salivary gland	2	30-39	Male	White	White	Not provided		
P5	Chen *et al.* 2022	Major salivary gland	Parotid salivary gland. Expanded resection was used to remove part of	1	60-69	Male	Not provided		Not provided		
SMG1	Horeth et al. 2021	gland	Submandibular gland	1	40-49	Male	Not provided				
SMG2	Horeth et al. 2021	gland	Submandibular gland	2	60-69	Female	Not provided				
SLG1	Consortium 2022	gland	Sublingual*	1	50-59	Male	White	White	Not provided	Yes	Yes
Tong1	Consortium 2022	Dorsal Tongue	dorsum	1	60-69	Female	White	White	European	No	Not provided
Tong2	Consortium 2022	Dorsal Tongue	circumvallate papilla	1	60-69	Female	White	White	European	No	Not provided
Tong3	Consortium 2022	Dorsal Tongue	dorsum	2	50-59	Male	White	White	Not provided	Yes	Yes
Tong4	Consortium 2022	Dorsal Tongue	circumvallate papilla	2	50-59	Male	White	White	Not provided	Yes	Yes
DP1	Pagella *et al.* 2021	Dental Pulp	Not provided	1	18-35	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP2	Pagella *et al.* 2021	Dental Pulp	Not provided	2	18-35	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP3	Pagella *et al.* 2021	Dental Pulp	Not provided	3	18-35	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP4	Pagella *et al.* 2021	Dental Pulp	Not provided	4	18-35	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP5	Pagella *et al.* 2021	Dental Pulp	Not provided	5	18-35	Not provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP6	Krivanek *et al.* 2020	Dental Pulp	Third molar	1	18–31	provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP7	Krivanek *et al.* 2020	Dental Pulp	Third molar	2	18–31	provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP8	Krivanek *et al.* 2020	Dental Pulp	Third molar	3	18–31	provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP9	Krivanek *et al.* 2020	Dental Pulp	Third molar	4	18–31	provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP10	Krivanek *et al.* 2020	Dental Pulp	Third molar	5	20-29	provided	Not provided	Not provided	Not provided	Not provided	Not provided
DP11	Opasawatchai *et al.* 2022	Dental Pulp	Tooh number 18 (upper maxillary right third molar)	1	20-29	Female	Not provided	Not provided	Not provided	Not provided	Not provided
DP12	Opasawatchai *et al.* 2022	Dental Pulp	Tooh number 18 (upper maxillary right third molar)	2	20-29	Male	Not provided	Not provided	Not provided	Not provided	Not provided
HP1	Ko et al. 2023	Hard Palate	Rugous Area anterior Hard palate	1	30-39	Not provided	Not provided	Not provided			
HP2	Ko et al. 2023	Hard Palate	Hard palate	2	30-39	provided	Not provided	Not provided			
G1	Williams *et al.* 2021	Gingival Mucosa	14-16	1	30-39	Female	Latino		Latinx	Not provided	Not provided
G2	Williams *et al.* 2021	Gingival Mucosa	Not provided	2	50-59	Female	Black		African American	Not provided	Not provided
G3	Williams *et al.* 2021	Gingival Mucosa	16-17	3	60-69	Female	Asian		Asian	Not provided	Not provided
G4	Williams *et al.* 2021	Gingival Mucosa	15-16	4	30-39	Female	White		Caucasian	Not provided	Not provided
G5	Williams *et al.* 2021	Gingival Mucosa	13-14	5	50-59	Female	White		Caucasian	Not provided	Not provided
G6	Williams *et al.* 2021	Gingival Mucosa	26-27	6	20-29	Male	White		Caucasian	Not provided	Not provided
G7	Williams *et al.* 2021	Gingival Mucosa	Not provided	7	30-39	Female	Asian		Asian	Not provided	Not provided
G8	Williams *et al.* 2021	Gingival Mucosa	Not provided	8	20-29	Female	Indian		American Indian	Not provided	Not provided
G9	Williams *et al.* 2021	Gingival Mucosa	Not provided	9	20-29	Female	White		Caucasian	Not provided	Not provided
G10	Williams *et al.* 2021	Gingival Mucosa	Not provided	10	30-39	Female	White		Caucasian	Not provided	Not provided
G11	Williams *et al.* 2021	Gingival Mucosa	Not provided	11	20-29	Female	Asian		Asian	Not provided	Not provided
G12	Williams *et al.* 2021	Gingival Mucosa	Not provided	12	20-29	Male	Asian		Asian	Not provided	Not provided
G13	Williams *et al.* 2021	Gingival Mucosa	Not provided	13	30-39	Male	Asian		Asian	Not provided	Not provided
G14	Caetano *et al.* 2021	Gingival Mucosa	Buccal gingival margin UR6	1	50-59	Male	White	Not provided	Caucasian	Non-smoker	Not provided
G15	Caetano *et al.* 2021	Gingival Mucosa	Buccal gingival margin UR3/UL3	4	50-59	Male	White	Not provided	Caucasian	Non-smoker	Not provided
G16	Huang *et al.* 2021	Gingival Mucosa	Palatal side, interdental papilla #14 distal/15 mesial	1	20-29	Female	White	Non-Hispanic	Caucasian		
G17	Huang *et al.* 2021	Gingival Mucosa	Palatal side, interdental papilla #14 distal/15 mesial	2	20-29	Male	Black	Non-Hispanic	African American		
G18	Huang *et al.* 2021	Gingival Mucosa	Palatal side, interdental papilla #14 distal/15 mesial	3	30-39	Male	White	Non-Hispanic	Caucasian		
G19	Huang *et al.* 2021	Gingival Mucosa	Palatal side, interdental papilla #14 distal/15 mesial	4	20-29	Female	White	Non-Hispanic	Caucasian		
Pd1	Pagella *et al.* 2021	Periodontium	Scraping of the surface o	1	18-35	provided	Not provided	Not provided	Not provided	Not provided	Not provided
Pd2	Pagella *et al.* 2021	Periodontium	Scraping of the surface o	2	18-35	provided	Not provided	Not provided	Not provided	Not provided	Not provided
Pd3	Pagella *et al.* 2021	Periodontium	Scraping of the surface o	3	18-35	provided	Not provided	Not provided	Not provided	Not provided	Not provided
Pd4	Pagella *et al.* 2021	Periodontium	Scraping of the surface o	4	18-35	provided	Not provided	Not provided	Not provided	Not provided	Not provided
Pd5	Pagella *et al.* 2021	Periodontium	Scraping of the surface o	5	18-35	provided	Not provided	Not provided	Not provided	Not provided	Not provided

Block Designation	Niche	Age (time of bx)	Sex	Race	Disease biopsied	Patient	Fixation
A1	Parotid	72	Female	Caucasian	solitary fibrous tumor	Healthy	PFA 4%
A8	Parotid	59	Male	Caucasian	Warthin's tumor	Healthy	PFA 4%
A6	Parotid	66	Female	Caucasian	squamous cell carcinoma	Healthy	PFA 4%
C1-3	Parotid	52	Male	Caucasian	Warthin's tumor	Healthy	PFA 4%
B1-2	Submandibular Gland	63	Female	Caucasian	SCC	Healthy	PFA 4%
G1, H1	Submandibular Gland	75	Male	Caucasian	SCC	Healthy	PFA 4%
C1-3	Submandibular Gland	62	Female	Caucasian	papillary thyroid carcinoma	Healthy	PFA 4%
H1	Submandibular Gland	65	Female	Caucasian	invasive carcinoma of squamous phenotype	Healthy	PFA 4%
A1	Minor Salivary Gland	37	Male	Caucasian	OSA	Healthy	PFA 4%
C1	Minor Salivary Glands	37	Male	Caucasian	OSA	Healthy	PFA 4%
H1	Minor Salivary Gland	35	Female	Caucasian	Squamous papilloma	Healthy	PFA 4%
H7	Minor Salivary Gland	65	Male	Caucasian	Skin adnexal carcinoma	Healthy	PFA 4%
E1-E2	Buccal Mucosa	65	Female	Caucasian	SCC	Healthy	PFA 4%
E13	Buccal/Oral mucosa	69	Male	Caucasian	SCC	Healthy	PFA 4%
H1	Buccal Mucosa	35	Female	Caucasian	Squamous papilloma	Healthy	PFA 4%
H7	Buccal Mucosa	65	Male	Caucasian	Skin adnexal carcinoma	Healthy	PFA 4%
D17,D18	Anterior/Lateral Tongue	44	Female	Caucasian	SCC	Healthy	PFA 4%
D13-18	Lateral Tongue	44	Female	Caucasian	SCC	Healthy	PFA 4%
F1-33	Lateral Tongue	58	Female	Caucasian	SCC	Healthy	PFA 4%
ID 1	Gingiva	45	Male	Caucasian	Third Molar Extraction	Healthy	PFA 4%
ID 5	Gingiva	68	8	Caucasian	third Molar Extraction	Healthy	PFA 4%

Supplemental Tables 3. MERFISH panel designed.

Gene Name	Ensemble Gene	ENSEMBL Transcript
ACKR1	ENSG00000213088	ENST00000368121
ACTA2	ENSG00000107796	ENST00000224784
APCDD1	ENSG00000154856	ENST00000355285
APOC1	ENSG00000130208	ENST00000588750
AQP3	ENSG00000165272	ENST00000297991
AREG	ENSG00000109321	ENST00000395748
ASCL3	ENSG00000176009	ENST00000531618
ATOH1	ENSG00000172238	ENST00000306011
AXL	ENSG00000167601	ENST00000301178
BATF3	ENSG00000123685	ENST00000243440
BPIFB2	ENSG00000078898	ENST00000170150
CALD1	ENSG00000122786	ENST00000361901
CCL14	ENSG00000276409	ENST00000618404
CCL17	ENSG00000102970	ENST00000616880
CCL2	ENSG00000108691	ENST00000225831
CCL21	ENSG00000137077	ENST00000259607
CCL22	ENSG00000102962	ENST00000219235
CCL23	ENSG00000274736	ENST00000615050
CCL3	ENSG00000277632	ENST00000614051
CCL4	ENSG00000275302	ENST00000613947

CCL5	ENSG00000271503	ENST00000603197
CCND1	ENSG00000110092	ENST00000227507
CCNE1	ENSG00000105173	ENST00000357943
CCR1	ENSG00000163823	ENST00000296140
CCR10	ENSG00000184451	ENST00000591568
CCR2	ENSG00000121807	ENST00000400888
CCR3	ENSG00000183625	ENST00000395940
CCR4	ENSG00000183813	ENST00000330953
CCR5	ENSG00000160791	ENST00000292303
CCR6	ENSG00000112486	ENST00000341935
CCR7	ENSG00000126353	ENST00000246657
CCR8	ENSG00000179934	ENST00000326306
CCR9	ENSG00000173585	ENST00000395963
CD14	ENSG00000170458	ENST00000401743
CD19	ENSG00000177455	ENST00000324662
CD1A	ENSG00000158477	ENST00000289429
CD1C	ENSG00000158481	ENST00000368170
CD1D	ENSG00000158473	ENST00000368171
CD207	ENSG00000116031	ENST00000410009
CD209	ENSG00000090659	ENST00000315599
CD22	ENSG00000012124	ENST00000085219
CD28	ENSG00000178562	ENST00000324106
CD34	ENSG00000174059	ENST00000356522
CD38	ENSG00000004468	ENST00000226279
CD3D	ENSG00000167286	ENST00000300692
CD3E	ENSG00000198851	ENST00000528600
CD4	ENSG00000010610	ENST00000011653
CD40LG	ENSG00000102245	ENST00000370629
CD68	ENSG00000129226	ENST00000584180
CD79A	ENSG00000105369	ENST00000221972
CD83	ENSG00000112149	ENST00000612003
CD8A	ENSG00000153563	ENST00000352580
CDH1	ENSG00000039068	ENST00000261769
CDH3	ENSG00000062038	ENST00000429102
CDKN2A	ENSG00000147889	ENST00000579755
CEACAM8	ENSG00000124469	ENST00000244336
CFD	ENSG00000197766	ENST00000327726
CFTR	ENSG00000001626	ENST00000003084
CLEC10A	ENSG00000132514	ENST00000576617
CLEC9A	ENSG00000197992	ENST00000355819
CNN1	ENSG00000130176	ENST00000252456
COL15A1	ENSG00000204291	ENST00000375001
COL1A1	ENSG00000108821	ENST00000225964
CPA3	ENSG00000163751	ENST00000296046
CSF2RB	ENSG00000100368	ENST00000403662

CST3	ENSG00000101439	ENST00000376925
CST7	ENSG00000077984	ENST00000480798
CX3CL1	ENSG00000006210	ENST00000006053
CX3CR1	ENSG00000168329	ENST00000399220
CXCL1	ENSG00000163739	ENST00000395761
CXCL10	ENSG00000169245	ENST00000306602
CXCL12	ENSG00000107562	ENST00000343575
CXCL16	ENSG00000161921	ENST00000293778
CXCL17	ENSG00000189377	ENST00000601181
CXCL3	ENSG00000163734	ENST00000296026
CXCL5	ENSG00000163735	ENST00000296027
CXCL6	ENSG00000124875	ENST00000226317
CXCL8	ENSG00000169429	ENST00000307407
CXCL9	ENSG00000138755	ENST00000264888
CXCR1	ENSG00000163464	ENST00000295683
CXCR2	ENSG00000180871	ENST00000318507
CXCR3	ENSG00000186810	ENST00000373693
CXCR4	ENSG00000121966	ENST00000241393
CXCR5	ENSG00000160683	ENST00000292174
CXCR6	ENSG00000172215	ENST00000458629
DCN	ENSG00000011465	ENST00000552962
DIO2	ENSG00000211448	ENST00000438257
DPP4	ENSG00000197635	ENST00000360534
EBI3	ENSG00000105246	ENST00000221847
EGF	ENSG00000138798	ENST00000265171
EGFR	ENSG00000146648	ENST00000275493
ENTPD1	ENSG00000138185	ENST00000453258
ENTPD2	ENSG00000054179	ENST00000355097
EPGN	ENSG00000182585	ENST00000413830
ERBB2	ENSG00000141736	ENST00000541774
ERBB3	ENSG00000065361	ENST00000549832
EREG	ENSG00000124882	ENST00000244869
FABP5	ENSG00000164687	ENST00000396359
FBLN1	ENSG00000077942	ENST00000327858
FCN1	ENSG00000085265	ENST00000371806
FN1	ENSG00000115414	ENST00000456923
FOXI1	ENSG00000168269	ENST00000449804
FOXP3	ENSG00000049768	ENST00000376199
GATA3	ENSG00000107485	ENST00000379328
GFAP	ENSG00000131095	ENST00000253408
GFRA3	ENSG00000146013	ENST00000274721
GNLY	ENSG00000115523	ENST00000263863
GZMA	ENSG00000145649	ENST00000274306
GZMB	ENSG00000100453	ENST00000216341
HBEGF	ENSG00000113070	ENST00000230990

HLA-DRA	ENSG00000204287	ENST00000395388
ICAM1	ENSG00000090339	ENST00000264832
ICAM2	ENSG00000108622	ENST00000579788
ICOS	ENSG00000163600	ENST00000435193
IFNLR1	ENSG00000185436	ENST00000327535
IGF1	ENSG00000017427	ENST00000307046
IGF2	ENSG00000167244	ENST00000381395
IGFBP2	ENSG00000115457	ENST00000456764
IGFBP3	ENSG00000146674	ENST00000381086
IGFBP4	ENSG00000141753	ENST00000269593
IGFBP5	ENSG00000115461	ENST00000233813
IGFBP6	ENSG00000167779	ENST00000549628
IGFBP7	ENSG00000163453	ENST00000295666
IGHD	ENSG00000211898	ENST00000390556
IGHM	ENSG00000211899	ENST00000390559
IKZF3	ENSG00000161405	ENST00000346872
IL10	ENSG00000136634	ENST00000423557
IL10RA	ENSG00000110324	ENST00000227752
IL11	ENSG00000095752	ENST00000264563
IL11RA	ENSG00000137070	ENST00000602473
IL12A	ENSG00000168811	ENST00000305579
IL12B	ENSG00000113302	ENST00000231228
IL12RB1	ENSG00000096996	ENST00000322153
IL13	ENSG00000169194	ENST00000304506
IL13RA1	ENSG00000131724	ENST00000371666
IL15	ENSG00000164136	ENST00000320650
IL15RA	ENSG00000134470	ENST00000379977
IL16	ENSG00000172349	ENST00000394652
IL17A	ENSG00000112115	ENST00000648244
IL17B	ENSG00000127743	ENST00000261796
IL17C	ENSG00000124391	ENST00000244241
IL17D	ENSG00000172458	ENST00000304920
IL17F	ENSG00000112116	ENST00000336123
IL17RB	ENSG00000056736	ENST00000288167
IL18	ENSG00000150782	ENST00000280357
IL18R1	ENSG00000115604	ENST00000233957
IL19	ENSG00000142224	ENST00000340758
IL1A	ENSG00000115008	ENST00000263339
IL1B	ENSG00000125538	ENST00000263341
IL1F10	ENSG00000136697	ENST00000341010
IL1R1	ENSG00000115594	ENST00000410023
IL1R2	ENSG00000115590	ENST00000393414
IL1RN	ENSG00000136689	ENST00000354115
IL2	ENSG00000109471	ENST00000226730
IL20	ENSG00000162891	ENST00000367098

IL20RA	ENSG00000016402	ENST00000316649
IL21	ENSG00000138684	ENST00000611104
IL21R	ENSG00000103522	ENST00000337929
IL22	ENSG00000127318	ENST00000328087
IL22RA1	ENSG00000142677	ENST00000270800
IL23A	ENSG00000110944	ENST00000228534
IL23R	ENSG00000162594	ENST00000637002
IL24	ENSG00000162892	ENST00000611909
IL25	ENSG00000166090	ENST00000329715
IL26	ENSG00000111536	ENST00000229134
IL27	ENSG00000197272	ENST00000356897
IL27RA	ENSG00000104998	ENST00000263379
IL2RB	ENSG00000100385	ENST00000216223
IL2RG	ENSG00000147168	ENST00000374202
IL3	ENSG00000164399	ENST00000296870
IL32	ENSG00000008517	ENST00000440815
IL33	ENSG00000137033	ENST00000381434
IL34	ENSG00000157368	ENST00000429149
IL36A	ENSG00000136694	ENST00000259211
IL36G	ENSG00000136688	ENST00000259205
IL37	ENSG00000125571	ENST00000263326
IL3RA	ENSG00000185291	ENST00000331035
IL4	ENSG00000113520	ENST00000350025
IL4R	ENSG00000077238	ENST00000170630
IL5	ENSG00000113525	ENST00000231454
IL5RA	ENSG00000091181	ENST00000418488
IL6	ENSG00000136244	ENST00000258743
IL6R	ENSG00000160712	ENST00000368485
IL6ST	ENSG00000134352	ENST00000336909
IL7	ENSG00000104432	ENST00000263851
IL7R	ENSG00000168685	ENST00000303115
INSR	ENSG00000171105	ENST00000341500
IRF7	ENSG00000185507	ENST00000397566
IRF8	ENSG00000140968	ENST00000268638
ISG15	ENSG00000187608	ENST00000624652
ITGAL	ENSG00000005844	ENST00000356798
IVL	ENSG00000163207	ENST00000368764
KLF4	ENSG00000136826	ENST00000374672
KLK5	ENSG00000167754	ENST00000593428
KLRB1	ENSG00000111796	ENST00000229402
KLRG1	ENSG00000139187	ENST00000266551
KLRK1	ENSG00000213809	ENST00000540267
KRT15	ENSG00000171346	ENST00000393976
KRT18	ENSG00000111057	ENST00000388835
KRT19	ENSG00000171345	ENST00000361566

KRT20	ENSG00000171431	ENST00000167588
KRT4	ENSG00000170477	ENST00000551956
KRT6A	ENSG00000205420	ENST00000330722
KRT7	ENSG00000135480	ENST00000331817
KRT76	ENSG00000185069	ENST00000332411
LEP	ENSG00000174697	ENST00000308868
LGALS3	ENSG00000131981	ENST00000254301
LGR4	ENSG00000205213	ENST00000379214
LGR5	ENSG00000139292	ENST00000266674
LGR6	ENSG00000133067	ENST00000255432
LILRA4	ENSG00000239961	ENST00000291759
LRIG1	ENSG00000144749	ENST00000496559
LRIG2	ENSG00000198799	ENST00000466161
LRIG3	ENSG00000139263	ENST00000379141
LUM	ENSG00000139329	ENST00000548071
LYZ	ENSG00000090382	ENST00000261267
MAL	ENSG00000172005	ENST00000309988
MBP	ENSG00000197971	ENST00000397860
MFAP5	ENSG00000197614	ENST00000359478
MGP	ENSG00000111341	ENST00000228938
MKI67	ENSG00000148773	ENST00000368653
MLANA	ENSG00000120215	ENST00000381476
MPO	ENSG00000005381	ENST00000225275
MS4A1	ENSG00000156738	ENST00000532073
MUC5B	ENSG00000117983	ENST00000526859
MUC7	ENSG00000171195	ENST00000413702
MYH11	ENSG00000133392	ENST00000573908
MYOC	ENSG00000034971	ENST00000037502
NCAM1	ENSG00000149294	ENST00000316851
NCR1	ENSG00000189430	ENST00000357397
NCR3	ENSG00000204475	ENST00000340027
NGF	ENSG00000134259	ENST00000369512
NGFR	ENSG00000064300	ENST00000172229
NKG7	ENSG00000105374	ENST00000221978
NOTCH1	ENSG00000148400	ENST00000651671
NOTCH2	ENSG00000134250	ENST00000256646
NOTCH3	ENSG00000074181	ENST00000263388
NPPC	ENSG00000163273	ENST00000409852
PAPPA	ENSG00000182752	ENST00000328252
PAX5	ENSG00000196092	ENST00000358127
PAX7	ENSG00000009709	ENST00000420770
PCLAF	ENSG00000166803	ENST00000300035
PCNA	ENSG00000132646	ENST00000379143
PDCD1	ENSG00000188389	ENST00000334409
PDGFRA	ENSG00000134853	ENST00000257290

PDGFRB	ENSG00000113721	ENST00000261799
PDPN	ENSG00000162493	ENST00000621990
PIP	ENSG00000159763	ENST00000291009
PITX1	ENSG00000069011	ENST00000265340
PLCB2	ENSG00000137841	ENST00000260402
PLVAP	ENSG00000130300	ENST00000252590
PODXL	ENSG00000128567	ENST00000378555
POSTN	ENSG00000133110	ENST00000541179
PRDM1	ENSG00000057657	ENST00000369089
PROX1	ENSG00000117707	ENST00000261454
PRR4	ENSG00000111215	ENST00000542658
PTN	ENSG00000105894	ENST00000348225
PTPRC	ENSG00000081237	ENST00000348564
RGCC	ENSG00000102760	ENST00000379359
RGS5	ENSG00000143248	ENST00000313961
S100A2	ENSG00000196754	ENST00000368710
S100B	ENSG00000160307	ENST00000291700
SDC1	ENSG00000115884	ENST00000254351
SELE	ENSG00000007908	ENST00000333360
SELL	ENSG00000188404	ENST00000236147
SERPINE2	ENSG00000135919	ENST00000409304
SFRP1	ENSG00000104332	ENST00000379845
SFRP2	ENSG00000145423	ENST00000274063
SLC5A5	ENSG00000105641	ENST00000222248
SLPI	ENSG00000124107	ENST00000338380
SOSTDC1	ENSG00000171243	ENST00000307068
SOX10	ENSG00000100146	ENST00000396884
SOX14	ENSG00000168875	ENST00000306087
SOX6	ENSG00000110693	ENST00000316399
SOX9	ENSG00000125398	ENST00000245479
SYCP1	ENSG00000198765	ENST00000477590
TGFA	ENSG00000163235	ENST00000295400
TNF	ENSG00000232810	ENST00000449264
TNFRSF1A	ENSG00000067182	ENST00000162749
TNFRSF1B	ENSG00000028137	ENST00000376259
TP53	ENSG00000141510	ENST00000620739
TP63	ENSG00000073282	ENST00000354600
TP73	ENSG00000078900	ENST00000378288
TRAC	ENSG00000277734	ENST00000611116
TRDC	ENSG00000211829	ENST00000390477
TUBB3	ENSG00000258947	ENST00000553967
TWIST1	ENSG00000122691	ENST00000242261
TWIST2	ENSG00000233608	ENST00000612363
TXLNA	ENSG00000084652	ENST00000373610
TYRP1	ENSG00000107165	ENST00000388918

WFDC2	ENSG00000101443	ENST00000372676
WNT10A	ENSG00000135925	ENST00000258411
WNT2B	ENSG00000134245	ENST00000369686
WNT4	ENSG00000162552	ENST00000290167
WNT5A	ENSG00000114251	ENST00000474267
WNT5B	ENSG00000111186	ENST00000310594
WNT6	ENSG00000115596	ENST00000233948
ZBTB16	ENSG00000109906	ENST00000335953
ZG16B	ENSG00000162078	ENST00000382280
ZNF683	ENSG00000176083	ENST00000349618

Supplemental Tables 4 and 5 Cell type annotation across the TCNs (4) and MCIMs (5).

Tables Attached.

REFERENCES:

1 Squier, C. A. & Kremer, M. J. Biology of oral mucosa and esophagus. *J Natl Cancer Inst Monogr*, 7-15 (2001). https://doi.org:10.1093/oxfordjournals.jncimonographs.a003443

2 Huang, N. *et al.* SARS-CoV-2 infection of the oral cavity and saliva. *Nature medicine* **27**, 892-903 (2021).

3 Moutsopoulos, N. M. & Konkel, J. E. Tissue-specific immunity at the oral mucosal barrier. *Trends in immunology* **39**, 276-287 (2018).

4 Caetano, A. J., Sequeira, I. & Byrd, K. M. A Roadmap for the Human Oral and Craniofacial Cell Atlas. *J Dent Res* **101**, 1274-1288 (2022). https://doi.org:10.1177/00220345221110768

5 Chibly, A. M., Aure, M. H., Patel, V. N. & Hoffman, M. P. Salivary gland function, development, and regeneration. *Physiol Rev* **102**, 1495-1552 (2022). https://doi.org:10.1152/physrev.00015.2021

6 Verstappen, G. M., Pringle, S., Bootsma, H. & Kroese, F. G. M. Epithelial-immune cell interplay in primary Sjögren syndrome salivary gland pathogenesis. *Nat Rev Rheumatol* **17**, 333-348 (2021). https://doi.org:10.1038/s41584-021-00605-2

7 Easter, Q. T., Matuck, B. F., Warner, B. M. & Byrd, K. M. Biogeographical Impacts of Dental, Oral, and Craniofacial Microbial Reservoirs. *J Dent Res* **102**, 1303-1314 (2023). https://doi.org:10.1177/00220345231191115

8 Regezi, J. A., Sciubba, J. & Jordan, R. C. *Oral pathology: clinical pathologic correlations*. (Elsevier Health Sciences, 2016).

9 Moayedi, Y., Michlig, S., Park, M., Koch, A. & Lumpkin, E. A. Somatosensory innervation of healthy human oral tissues. *Journal of Comparative Neurology* **529**, 3046-3061 (2021).

10 Koehler, M. *et al.* Biophysical investigations using atomic force microscopy can elucidate the link between mouthfeel and flavour perception. *Nat Food* **5**, 281-287 (2024). https://doi.org:10.1038/s43016-024-00958-3

11 Byrd, K. M. *et al.* Heterogeneity within Stratified Epithelial Stem Cell Populations Maintains the Oral Mucosa in Response to Physiological Stress. *Cell Stem Cell* **25**, 814-829.e816 (2019). https://doi.org:10.1016/j.stem.2019.11.005

12 Challacombe, S. J., Shirlaw, P. J. & Thornhill, M. H. in *Mucosal immunology* 1943-1983 (Elsevier, 2015).

13 Wu, R.-Q., Zhang, D.-F., Tu, E., Chen, Q.-M. & Chen, W. The mucosal immune system in the oral cavity—an orchestra of T cell diversity. *International journal of oral science* **6**, 125-132 (2014).

14 Glim, J. E., Everts, V., Niessen, F. B., Ulrich, M. M. & Beelen, R. H. Extracellular matrix components of oral mucosa differ from skin and resemble that of foetal skin. *Arch Oral Biol* **59**, 1048-1055 (2014). https://doi.org:10.1016/j.archoralbio.2014.05.019

15 Sutherland, T. E., Dyer, D. P. & Allen, J. E. The extracellular matrix and the immune system: A mutually dependent relationship. *Science* **379**, eabp8964 (2023). https://doi.org:10.1126/science.abp8964

16 Cavagnero, K. J. & Gallo, R. L. Essential immune functions of fibroblasts in innate host defense. *Front Immunol* **13**, 1058862 (2022). https://doi.org:10.3389/fimmu.2022.1058862

17 Easter, Q. T. *et al.* Single-cell and spatially resolved interactomics of tooth-associated keratinocytes in periodontitis. *Nat Commun* **15**, 5016 (2024). https://doi.org:10.1038/s41467-024-49037-y

18 Krausgruber, T. *et al.* Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296-302 (2020). https://doi.org:10.1038/s41586-020-2424-4

19 Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575-579 (2021). https://doi.org:10.1038/s41586-021-03549-5

20 Salminen, A., Kaarniranta, K. & Kauppinen, A. Tissue fibroblasts are versatile immune regulators: An evaluation of their impact on the aging process. *Ageing Res Rev* **97**, 102296 (2024). https://doi.org:10.1016/j.arr.2024.102296
21 Caetano, A. J. *et al.* Spatially resolved transcriptomics reveals pro-inflammatory fibroblast involved in lymphocyte recruitment through CXCL8 and CXCL10. *Elife* **12** (2023). https://doi.org:10.7554/eLife.81525
22 Williams, D. W. *et al.* Human oral mucosa cell atlas reveals a stromal-neutrophil axis regulating tissue immunity. *Cell* **184**, 4090-4104.e4015 (2021). https://doi.org:10.1016/j.cell.2021.05.013
23 Amersfoort, J., Eelen, G. & Carmeliet, P. Immunomodulation by endothelial cells - partnering up with the immune system? *Nat Rev Immunol* **22**, 576-588 (2022). https://doi.org:10.1038/s41577-022-00694-4
24 Mai, J., Virtue, A., Shen, J., Wang, H. & Yang, X. F. An evolving new paradigm: endothelial cells--conditional innate immune cells. *J Hematol Oncol* **6**, 61 (2013). https://doi.org:10.1186/1756-8722-6-61
25 Shao, Y. *et al.* Vascular endothelial cells and innate immunity. *Arteriosclerosis, thrombosis, and vascular biology* **40**, e138-e152 (2020).
26 Slany, A. *et al.* Contribution of human fibroblasts and endothelial cells to the hallmarks of inflammation as determined by proteome profiling. *Molecular & Cellular Proteomics* **15**, 1982-1997 (2016).
27 Whitsett, J. A. & Alenghat, T. Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nat Immunol* **16**, 27-35 (2015). https://doi.org:10.1038/ni.3045
28 Hu, Y. *et al.* Unsupervised and supervised discovery of tissue cellular neighborhoods from cell phenotypes. *Nat Methods* **21**, 267-278 (2024). https://doi.org:10.1038/s41592-023-02124-2
29 Mayer, A. T. *et al.* A tissue atlas of ulcerative colitis revealing evidence of sex-dependent differences in disease-driving inflammatory cell types and resistance to TNF inhibitor therapy. *Sci Adv* **9**, eadd1166 (2023). https://doi.org:10.1126/sciadv.add1166
30 Cang, Z. *et al.* Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* **20**, 218-228 (2023). https://doi.org:10.1038/s41592-022-01728-4
31 Dimitrov, D. *et al.* LIANA+ provides an all-in-one framework for cell-cell communication inference. *Nat Cell Biol* **26**, 1613-1622 (2024). https://doi.org:10.1038/s41556-024-01469-w
32 Hermans, F. *et al.* Establishment of inclusive single-cell transcriptome atlases from mouse and human tooth as powerful resource for dental research. *Front Cell Dev Biol* **10**, 1021459 (2022). https://doi.org:10.3389/fcell.2022.1021459
33 Huynh, K. L. A. *et al.* Spatial Deconvolution of Cell Types and Cell States at Scale Utilizing TACIT. *bioRxiv* (2024). https://doi.org:10.1101/2024.05.31.596861
34 Ko, K. I. *et al.* Distinct fibroblast progenitor subpopulation expedites regenerative mucosal healing by immunomodulation. *J Exp Med* **220** (2023). https://doi.org:10.1084/jem.20221350
35 Caetano, A. J. *et al.* Defining human mesenchymal and epithelial heterogeneity in response to oral inflammatory disease. *Elife* **10** (2021). https://doi.org:10.7554/eLife.62810
36 Jones, R. C. *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022). https://doi.org:10.1126/science.abl4896
37 Pagella, P., Stadlinger, B. & Mitsiadis, T. A. Isolation of dental pulp and periodontal cells from human teeth for single-cell RNA sequencing. *STAR Protoc* **2**, 100953 (2021). https://doi.org:10.1016/j.xpro.2021.100953

38 Costa-da-Silva, A. C. *et al.* Salivary ZG16B expression loss follows exocrine gland dysfunction related to oral chronic graft-versus-host disease. *iScience* **25**, 103592 (2022). https://doi.org:10.1016/j.isci.2021.103592
39 Horeth, E. *et al.* High-Resolution Transcriptomic Landscape of the Human Submandibular Gland. *J Dent Res* **102**, 525-535 (2023). https://doi.org:10.1177/00220345221147908
40 Chen, M. *et al.* Transcriptomic Mapping of Human Parotid Gland at Single-Cell Resolution. *J Dent Res* **101**, 972-982 (2022). https://doi.org:10.1177/00220345221076069
41 Krivanek, J. *et al.* Dental cell type atlas reveals stem and differentiated cell types in mouse and human teeth. *Nat Commun* **11**, 4816 (2020). https://doi.org:10.1038/s41467-020-18512-7
42 Opasawatchai, A. *et al.* Single-cell transcriptomic profiling of human dental pulp in sound and carious teeth: a pilot study. *Frontiers in Dental Medicine* **2**, 806294 (2022).
43 Farr, L., Ghosh, S. & Moonah, S. Role of MIF Cytokine/CD74 Receptor Pathway in Protecting Against Injury and Promoting Repair. *Front Immunol* **11**, 1273 (2020). https://doi.org:10.3389/fimmu.2020.01273
44 Smolgovsky, S., Theall, B., Wagner, N. & Alcaide, P. Fibroblasts and immune cells: At the crossroad of organ inflammation and fibrosis. *American Journal of Physiology-Heart and Circulatory Physiology* **326**, H303-H316 (2024).
45 Marsh, P. D., Do, T., Beighton, D. & Devine, D. A. Influence of saliva on the oral microbiota. *Periodontol 2000* **70**, 80-92 (2016). https://doi.org:10.1111/prd.12098
46 Deng, Z. *et al.* TGF-β signaling in health, disease, and therapeutics. *Signal transduction and targeted therapy* **9**, 61 (2024).
47 Ware, C. F., Croft, M. & Neil, G. A. Realigning the LIGHT signaling network to control dysregulated inflammation. *Journal of Experimental Medicine* **219**, e20220236 (2022).
48 Elgueta, R. *et al.* Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunological reviews* **229**, 152-172 (2009).
49 Pranzatelli, T. J. F. *et al.* GZMK+CD8+ T cells Target A Specific Acinar Cell Type in Sjögren's Disease. *Res Sq* (2024). https://doi.org:10.21203/rs.3.rs-3601404/v1
50 Farnsworth, R. H., Karnezis, T., Maciburko, S. J., Mueller, S. N. & Stacker, S. A. The Interplay Between Lymphatic Vessels and Chemokines. *Front Immunol* **10**, 518 (2019). https://doi.org:10.3389/fimmu.2019.00518
51 Groeger, S. & Meyle, J. Oral Mucosal Epithelial Cells. *Front Immunol* **10**, 208 (2019). https://doi.org:10.3389/fimmu.2019.00208
52 Jones, K. B. & Klein, O. D. Oral epithelial stem cells in tissue maintenance and disease: the first steps in a long journey. *Int J Oral Sci* **5**, 121-129 (2013). https://doi.org:10.1038/ijos.2013.46
53 Cheng, R., Wu, Z., Li, M., Shao, M. & Hu, T. Interleukin-1β is a potential therapeutic target for periodontitis: a narrative review. *Int J Oral Sci* **12**, 2 (2020). https://doi.org:10.1038/s41368-019-0068-8
54 Saccucci, M. *et al.* Autoimmune Diseases and Their Manifestations on Oral Cavity: Diagnosis and Clinical Management. *J Immunol Res* **2018**, 6061825 (2018). https://doi.org:10.1155/2018/6061825
55 Kinane, D. F., Stathopoulou, P. G. & Papapanou, P. N. Periodontal diseases. *Nature reviews Disease primers* **3**, 1-14 (2017).
56 Easter, Q. T. *et al.* Polybacterial Intracellular Macromolecules Shape Single-Cell Epikine Profiles in Upper Airway Mucosa. *bioRxiv*, 2024.2010. 2008.617279 (2024).
57 Gago da Graça, C., van Baarsen, L. G. M. & Mebius, R. E. Tertiary Lymphoid Structures: Diversity in Their Development, Composition, and Role. *J Immunol* **206**, 273-281 (2021). https://doi.org:10.4049/jimmunol.2000873
58 Fletcher, A. L., Acton, S. E. & Knoblich, K. Lymph node fibroblastic reticular cells in health and disease. *Nat Rev Immunol* **15**, 350-361 (2015). https://doi.org:10.1038/nri3846

59 Rood, J. E., Maartens, A., Hupalowska, A., Teichmann, S. A. & Regev, A. Impact of the Human Cell Atlas on medicine. *Nature medicine* **28**, 2486-2496 (2022).
60 Regev, A. *et al.* The human cell atlas. *elife* **6**, e27041 (2017).
61 Rood, J. E. *et al.* Toward a common coordinate framework for the human body. *Cell* **179**, 1455-1467 (2019).
62 Kronman, F. N. *et al.* Developmental mouse brain common coordinate framework. *Nature Communications* **15**, 9072 (2024).
63 Wang, Q. *et al.* The Allen mouse brain common coordinate framework: a 3D reference atlas. *Cell* **181**, 936-953. e920 (2020).
64 Minton, K. A gene atlas of 'structural immunity'. *Nat Rev Immunol* **20**, 518-519 (2020). https://doi.org:10.1038/s41577-020-0398-y
65 Gao, Y. *et al.* Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation. *Cancer Cell* **42**, 1764-1783.e1710 (2024). https://doi.org:10.1016/j.ccell.2024.08.020
66 Chen, Y., McAndrews, K. M. & Kalluri, R. Clinical and therapeutic relevance of cancer-associated fibroblasts. *Nat Rev Clin Oncol* **18**, 792-804 (2021). https://doi.org:10.1038/s41571-021-00546-5
67 Wang, M. *et al.* Tertiary lymphoid structures in head and neck squamous cell carcinoma improve prognosis by recruiting CD8(+) T cells. *Mol Oncol* **17**, 1514-1530 (2023). https://doi.org:10.1002/1878-0261.13403

METHODS

Published scRNAseq harmonization: Raw FASTQ files from published single-cell RNA sequencing projects were retrieved using scripts from [this GitHub repository](https://github.com/cellgeni/reprocess_public_10x). First, metadata was collected from the GEO soft family file, and the ENA web API was used to gather data format details (SRR/ERR) and link samples to runs. Raw reads were downloaded as SRA archives, 10X BAM files, or gzipped paired-end FASTQ files, and then converted to FASTQ using fastq-dump (SRA tools v2.11.0) or bamtofastq (v1.3.2). The raw reads were processed using STARsolo for mapping and quantification, with wrapper scripts from [this GitHub repository](https://github.com/cellgeni/STARsolo/) to identify kit versions and sample specifics. The human reference genome matched Cell Ranger 2022-A standards. For 10x samples, STARsolo settings were optimized to mimic Cell Ranger v6 outputs, including UMI/barcode processing and paired-end alignment. Cell filtering was performed with EmptyDrops (Cell Ranger v4+), producing both exon-only and full-length gene counts as well as RNA velocity matrices.

Newly generated scRNAseq data*: Soft Palate (MSNZ Würzburg)*: Resected soft palate oral mucosa samples were placed in ice-cold Advanced DMEM+++ medium (Advanced DMEM F12 (Gibco) supplemented with 10 mmol/L HEPES (Gibco), 100 U/mL penicillin/streptomycin (Gibco), 1 x GlutaMax (Gibco), and 10 µM of the Rho kinase (ROCK) inhibitor compound Y-2763) and transferred to the laboratory on ice as soon as possible. **For histology**, representative pieces of tissue were embedded in Tissue-Tek® O.C.T. compound (Sakura) and placed directly on dry ice. The OCT-embedded tissues were stored at -80°C for long-term storage. OCT blocks were cut into 10 mm sections and stained with hematoxylin and eosin (H&E, Morphisto) after PFA fixation. **For cryopreservation**, the tissues were cut into smaller pieces and were incubated in CryoStor® CS10 (Stem Cell Technologies) freezing medium for 10 minutes on ice. Then, the samples were

transferred to a -80°C freezer. Frozen samples were stored in -150°C long-term. Frozen tissue samples were thawed at 37°C and washed in pre-warmed Advanced DMEM+++ medium. After rinsing, the tissue pieces were incubated in Advanced DMEM+++ for 10 min at room temperature. Samples were minced to a size of 0.5-1 mm3 and digested in Advanced DMEM+++ supplemented with 2% fetal bovine serum (Gibco), 2 mg/ml collagenase-P (Sigma-Aldrich), 1 mg/ml collagenase-D (Sigma-Aldrich), 0.5 mg/ml DNase-I (Sigma-Aldrich) and 10 µmol/mL Y-27632 (Tocris) for 30 minutes in a thermomixer (Eppendorf) at 37 °C, 1000 rpm. During incubation, the digestion mixture was vigorously resuspended every 10 minutes. Samples were further digested into single cells in 0.2% trypsin in PBS supplemented with 10 µmol/mL Y-27632 (Tocris). The digestion was stopped with EDTA. Single cell suspension was filtered through a pre-wetted Pluristrainer with a pore size of 70 µm. **Antibody staining and sorting:** Cells were stained with Zombie Violet Fixable Viability Stain (1:1000 in PBS, Thermo Fisher), and blocked with Human TruStain FcX™ Fc Receptor Blocking Solution (1:10 in PBS, Biolegend). Then, cells were stained with TotalSeqTM-B0251 and B0252 (1:50, Biolegend), and alive cells were sorted and collected into a clean tube. Sorted samples were pooled, counted with Trypan Blue, and loaded onto a 10X Chromium Controller (10X Genomics). Gene expression and cell surface protein library preparation was performed according to the manufacturer's instructions for the 10X Chromium Next GEM Single Cell Library kit v3.1 Dual Index (10X Genomics). The libraries were sequenced on a NextSeq Illumina sequencer. *Labial Mucosa (National Institutes of Health):* Samples were collected from patients who provided informed consent under the NIH Central IRB Protocol 15-D-0051 (Principal Investigator: Warner). The sections were immediately placed in ice-cold RPMI, dissected into 1–2-mm pieces, and dissociated using a Miltenyi Multi-tissue Dissociation Kit A. It was then placed in 10% formaldehyde for a 16-24h fixation at 4°C during the fixation step and subsequently placed in a -80 fridge. To initiate the thawing process, the Thaw Enhancer (10x Genomics PN-2000482) was incubated at 65°C for 10 minutes, followed by vortexing and a brief centrifugation to ensure no precipitate was present. The reagent was kept warm, and the absence

of precipitate was verified before use. Thawed Enhancer was not kept on ice to prevent precipitation, and once thawed, it was maintained at 42°C for up to 10 minutes. Next, 0.1 volume of pre-warmed Enhancer was added to the fixed sample in Quenching Buffer. For example, 100µl of Enhancer was added to 1,000µl of fixed sample in Quenching Buffer, followed by pipetting to mix. Alternatively, to conserve Enhancer volume, cells were centrifuged at 850 RCF, 500µl of Quenching Buffer was removed, and 50µl of Enhancer was added to the sample. 50% glycerol was introduced to achieve a final concentration of 10%. For instance, 275µl of 50% glycerol was added to 1,100µl of fixed sample in Quenching Buffer and Enhancer, followed by pipetting to mix. The samples could be stored at -80oC for up to 6 months. In the post-storage processing phase of our study, all the samples were processed at the same time and, the steps outlined below were followed: Samples were thawed at room temperature until no ice was present. The samples were then centrifuged at 850rcf for 5 minutes at room temperature. The supernatant was carefully removed without disturbing the pellet. The cell pellet was resuspended in 1 ml of 0.5X PBS + 0.02% BSA* (optionally supplemented with 0.2 U/µl RNase Inhibitor) or Quenching Buffer and kept on ice. *RNase-free BSA was used at this step. The dissociation process was conducted at 37 °C in an OctoMACS tissue disruptor using heated sleeves. Single-cell suspensions underwent serial filtration through 70- and 30-µm filters and were rinsed with 1× Hanks' buffered salt solution. Cell counting and viability assessments were performed using a Trypan blue exclusion assay, with suspensions having greater than 35% viability chosen for subsequent sequencing. For single-cell capture, library preparation, and sequencing, approximately 10,000 cells were targeted and loaded onto a 10x Genomics Chromium Next GEM Chip B. Following cell capture, library preparation was performed using the 10x Chromium Next GEM Single Cell 3′ kit v3, and the libraries were sequenced on a NextSeq500 sequencer (Illumina). Cell processing was undertaken using the 10x Genomics Chromium Controller and the Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3 (PN-1000075), adhering to the guidelines provided in the manufacturer's user guide. Viability and concentration assessments were conducted by staining aliquots of sorted cells

with acridine orange and propidium iodide, followed by analysis using a fluorescent Cell Counter. Each sample was loaded with approximately 3,200 cells onto the Chromium Chip B, aiming for a recovery of 2,000 cells per sample during library preparation. The creation of gel beads in emulsion (GEMs) involved encapsulating single cells, reverse transcription reagents, and gel beads coated with barcoded oligos within an oil droplet. Reverse transcription was executed using a C1000 thermal cycler (Bio-Rad), resulting in complementary DNA (cDNA) libraries tagged with a cell barcode and unique molecular index (UMI). GEMs were then broken, and the cDNA libraries purified using Dynabeads MyOne SILANE (Invitrogen) with subsequent 12 amplification cycles. Amplified libraries underwent purification with SPRIselect magnetic beads (Beckman Coulter) and quantification using an Agilent Bioanalyzer High Sensitivity DNA chip (Agilent Technologies). Following steps included fragmentation, end repair, A-tailing, and double-sided size selection with SPRIselect beads. Ligation of Illumina-compatible adapters onto size-selected cDNA fragments occurred. Adapter-ligated cDNA was purified using SPRIselect beads, and uniquely identifiable indexes were incorporated during 12 amplification cycles. The finalized sequencing libraries underwent purification with SPRIselect beads, visualization using the Bioanalyzer High Sensitivity DNA chip, quantification with the KAPA SYBR FAST Universal qPCR Kit for Illumina (Roche) and StepOnePlus Real-Time PCR System (Applied Biosystems), normalization to 4nM, and pooling. Sequencing took place on a NextSeq 500 machine (Illumina), with libraries denatured and diluted following the standard Illumina protocol. A 1% PhiX sequencing control (Illumina) was spiked in, and the libraries were loaded onto the flow cell at 1.8pM. *Parotid Salivary Glands (University of Groningen):* Biopsy samples of parotid SG tissue were obtained from donors after institutional review board [IRB] approval, (METc 2008.078, METc 2023.281) who were treated for a squamous cell carcinoma of the oral cavity. In these patients, an elective head and neck dissection procedure was performed. During this procedure, the parotid SG is exposed and removed. This tissue does not contain malignant cells, as oral squamous cell carcinoma does not disseminate to the parotid SG. Salivary gland tissue was processed into a cell suspension according to our previously

published method (PMID: 29984480), and CS10 cryopreservation media until use. After thawing, separate cells were processed using 10X Genomics protocol version 3.1 chemistry (PN 2000121) accompanying the Version 3.1 kit, with a targeted cell yield of 8000 per lane. Samples were not multiplexed per 10X machine lane. Sample indexes from the Chromium i7 Sample Index Kit (PN-120262) were used for identification of pooled sampled on sequencing chip. Sequencing was performed using the Next Seq2500 (RapidRun) sequencer for a total of 127 cycles.

Published scRNAseq Integration, pre-process and visualization, and spatial single-cell annotation using Trailmaker: *scRNAseq:* The single-cell RNA sequencing dataset was analyzed and visualized using Trailmaker®, accessible on the Biomage community platform ([Biomage](https://biomage.net/) at https://scp.biomage.net/), provided by Biomage and developed by Parse Biosciences. Pre-filtered count matrices were loaded onto the platform for additional processing. Barcodes were filtered in four steps: first, barcodes with fewer than 500 UMIs and those indicating low quality or signs of apoptosis (over 15% mitochondrial content) were excluded. A robust linear regression model from the `MASS` package was then used to predict expected gene counts per barcode, applying a tolerance based on sample size to exclude outliers outside this range. Doublets were identified and removed using the `scDblFinder` R package, filtering out barcodes with a score over 0.5. Following filtering, between 300 and 8,000 barcodes per sample remained. Data was normalized with a log transformation, and the top 2,000 most variable genes were identified using variance-stabilizing transformation. PCA was applied, with the first 40 components (capturing 95.65% of the variance) then batch-corrected using the `Harmony` package. Louvain clustering and UMAP were conducted in Seurat for dimensionality reduction and cluster identification. Differentially expressed genes were identified for each cluster using the Wilcoxon rank-sum test (`presto` package). These samples were then imported into Trailmaker®, bypassing initial filtering since pre-processing had been completed. The same analytical pipeline was applied to this subset as

to the complete dataset. Cells were annotated based on literature and further classified with the CellTypist tool. For MERFISH data from MERSCOPE, segmented data—including barcodes, features, and matrices—was processed within Trailmaker®. Cells with no transcripts were removed based on transcript count per cell, and cell types were annotated using the lasso tool, focusing on the most highly expressed genes within each Louvain cluster.

Receptor-ligand analyses via *Cellphone DB*[69], *CellChat*[70], and *MultiNicheNet*[71]: We employed *CellPhoneDB* (v5, https://github.com/ventolab/CellphoneDB), leveraging its curated database of ligand-receptor interactions to analyze cell-to-cell signaling interactions at a single-cell level. Using default parameters, we calculated the likelihood of interactions between pairs of cell types based on the expression of known ligand-receptor pairs, resulting in probabilistic scores that highlight statistically significant intercellular communication. Additionally, we applied CellChat (v2, https://doi.org/10.1101/2023.11.05.565674) to quantify signaling probabilities between manually defined sender and receiver cell groups. Both methods were applied to the same datasets, providing a comprehensive communication profile across different niches and enabling cross-validation of key interactions. We also employed *MultiNicheNet*, a novel framework available at https://github.com/saeyslab/multinichenetr, to enhance the analysis of cell-cell communication within multi-sample multi-condition single-cell transcriptomics data, with a focus on oral mucosal tissues. The primary objectives of MultiNicheNet were to infer differentially expressed and active ligand-receptor pairs between conditions of interest and predict the putative downstream target genes of these pairs. The pipeline was applied in two different approaches, one related to a Tier 1 annotation dataset into the integrated oral craniofacial atlas. Statistical difference was employed to determine the top 50 interactions between receptor and ligands in glands versus mucosa. A second approach was made to determine the difference within each of the major niches, such as Salivary Glands and all Mucosa. Cells annotations were carried from Tier 2 to run downstream analysis. In the purpose of setting the threshold of cell, only cell types

with more than 10 cells in all niches were included in the analysis. p-values were extracted using muscat framework for DEG and testing which is the top highest expressed based on mixed linear models.

Sectioning for spatial biology: Samples from the pathology archive of UNC medicine school (UNC IRB 22-1786) were selected with two trained oral pathologists, only samples not containing architectural or cytologic alterations were included. FFPE Blocks were sectioned in 5 microns sequential sections and mounted in SuperFrost Plus (Thermo Fisher) slides for Phenocycler-Fusion 2.0, H&E and Stains. Coverslips (MERSCOPE slides) from MERSCOPE (Vizgen) were also used to mounting the sequential sections from each slide that went through other Phenocycler experiment using the '91600112 MERSCOPE User Guide Formalin-Fixed Paraffin-Embedded Tissue Sample Preparation RevB'. MERSCOPE TM Slides have a defined imageable of a 2X1.5cm rectangular where two sections from the same tissue were placed at time. All slides preparation was performed using RNA free water and RNAse protocol from Leica autotec system.

Multiplex Protein Immunofluorescence (Formerly, CODEX; now, PhenoCycler Fusion 2.0, Akoya Biosciences): For multiplexed immunofluorescence (Multi-IF), we used the PhenoCycler Fusion 2.0 from Akoya Biosciences. All samples underwent deparaffinization using a gradual reduction of alcohol concentrations from 100% to 30%. Antigen retrieval was performed using AR9 (EDTA) buffer from Akoya Biosciences in a low-pressure cooker for 15 minutes. After cooling for 1 hour, samples were hydrated with ethanol for 2 minutes followed by a 20-minute incubation in staining buffer. The antibody cocktail buffer was prepared according to the Phenocycler Fusion Manual using four blockers and nuclease-free water in addition to the staining buffer. Antibodies (Supplementary Table 2) were diluted to a 1:200 concentration in the antibody cocktail buffer. A mixed solution containing all primary antibodies was then prepared. Slides were incubated overnight at 4°C in a humidity chamber (Sigma-Aldrich) with the primary antibodies. After primary

incubation the slides were placed in staining buffer for 2 min, followed by post-stain fixing solution (10% PFA in staining buffer) for 10 min. Following 3 baths of 2 min washing in 1X PBS, the slide was immersed in ice-cold MeOH for 5 min. The final fixative solution (FFS) was prepared in accordance with the Phenocycler manual. Slides were then incubated with the FFS for 20 minutes at room temperature. Following this, the slides underwent a PBS wash and were promptly transferred to the FCAD Machine (Akoya Biosciences) for flow cell mounting. Flow cell stitching was achieved through a 30-second application of high pressure to the top of the slides. Once mounted, the flow cells with the slides and sectioning were immersed in PCF buffer for 10 minutes before being transferred to the Phenocycler Fusion. Reporters for each antibody were prepared using a Report Stock solution along with 5490 µL of nuclease-free water, 675 µL of 10X PCF buffer, 450 µL of PCF assay reagent, and 27 µL of in-house prepared concentrated DAPI to achieve a final DAPI concentration of 1:1000 per cycle. This process was repeated for a total of two slides per time. Reporters were included into a 1:50 dilution for each cycle, including the respective channels. A 250 microliters of each specific report stock solution were pipetted into a 96 well-plate that was sealed using an aluminum foil provided by Akoya Biosciences. Two slides were included into the fluid equipment at time, using the Phenocycler Fusion 2.0. Manual mapping was used to the scanning area to be scanned by the PhenoImager using brightfield. Low DMSO and High DMSO was prepared as suggested by the Index B from the Phenocycler Fusion 2.0 manual of operations.

Multiplexed error-robust fluorescence in situ hybridization (Vizgen MERSCOPE): For multiplexed error-robust fluorescence in situ hybridization (MERFISH), we used MERSCOPE from Vizgen. To ensure maximum sensitivity and consistency, MERSCOPE verification was performed on representative samples from all oral niches using the MERSCOPE Verification Kit (Vizgen) according to the manufacturer's instructions (91600004 Rev D). Briefly, verification mimics all sample processing steps but uses a single-round smFISH readout of the EEF2

housekeeping gene to determine signal brightness, background fluorescence and consistent adhesion. Verification results with increased background in several tissue types made alterations to the manufacturer's protocol necessary. A harsh clearing protocol was developed and verified on additional samples in alignment with the Vizgen team. It replaced the usual digestion followed by single stage clearing of clearing resistant tissues with a multi-step protocol with increased Proteinase K (ProtK, NEB, Cat P8107S) concentrations. After gel embedding, samples were initially cleared with 1.25 ml clearing premix (Vizgen) and 120 µl ProtK for 4 hours at 47 °C. After incubation another 3.75 ml of clearing premix without ProtK was added to the samples and incubated over night at 47 °C. Samples were washed three times for 5 minutes in Sample Prep Wash Buffer (Vizgen) to remove the clearing solution and digested in Digestion Mix (Vizgen) for 2 hours at 37 °C. 5 ml fresh Clearing Premix with 50 µl ProtK were added directly to the digestion reaction and incubated for 3 days at 37°C before continuing with autofluorescence quenching according to the unmodified MERSCOPE protocol. These modifications to the FFPE MERSCOPE protocol (91600112 Rev B) were used to generate all MERSCOPE results otherwise following the manufacturer's instructions. Samples were hybridized with MERSCOPE probes for the RNAs listed in Supplementary table 3 and smFISH probes for KRT14 (ENSG00000186847), KRT1 (ENSG00000167768), KRT6B (ENSG00000185479), APOD (ENSG00000189058), LORICRIN (ENSG00000203782) and B2M (ENSG00000166710) that were all part of the custom designed panel (Vizgen panel ID: BP0983). Next to transcript locations, tiff images were generated for DAPI and polyT as well as three proprietary cell boundary stains (Vizgen, PN 10400009) at 0.108 µm pixel size and the smFISH probes listed above.

Cell Segmentation for Multi-IF and MERFISH Images using *Cellpose3*[72]: Cell segmentation was carried out using *Cellpose3*. We employed a pipeline that integrated both denoising and segmentation steps to improve the quality of input images and the accuracy of the segmentation outputs based in nuclei expansion. The model was custom-trained on H&E-stained tissue

sections from the niches of the oral cavity, allowing for optimized performance on this tissue type, which often presents significant histological heterogeneity. Prior to segmentation, a denoising step was applied to reduce image noise and artifacts. The cell expansion was based on the auto-calibration tool from *Cellpose3* and the starting trained model used was the *Cyto2*. For this analysis, we used the *QuPath* extension for *Cellpose*, allowing us to streamline the workflow by combining *QuPath's* visualization tools with the *Cellpose* segmentation model trained. Annotations created in *QuPath* were exported as a CSV file. The segmentation results were then used as input for further downstream analyses. Specifically, the segmented data were inputted into Phenocycler Fusion (PCF) and MERSCOPE platforms for spatial transcriptomics and spatial proteomics analysis, enabling the identification of spatial patterns and cell-to-cell interactions at single-cell resolution. We Inputted the segmentation mode in the Vizualizer tool (Vizgen), which provided an intuitive interface for examining the quality and accuracy of the segmentations. This visualization step was critical for ensuring that the segmentations aligned with biological expectations and tissue architecture.

***AstroSuite for* Auto Assignment of Cell Identities and States, TCNs, and MCIMs:** *AstroSuite* contains interconnected algorithms for analyzing spatial biology; two of the algorithms, *TACIT* and *Astrograph* have been published previously[33]. *TACIT (Threshold-based Assignment of Cell Types from Multiplexed Imaging Data) for cell type annotation*. We used *TACIT* (Threshold-based Assignment of Cell Types from Multiplexed Imaging DaTa) for cell type annotation in our spatial omics dataset. *TACIT* is an unsupervised algorithm designed for single-cell resolved spatial modalities, including spatial transcriptomics and proteomics. *TACIT* takes as input the normalized CELLxFEATURE matrix as a result of *Cellpose3* cell segmentation and quantification of individual cell features such as probe intensity or count values. Additionally, a TYPExMARKER matrix, informed by expert knowledge, is used to indicate the relevance of specific markers for defining cell types. The annotation process occurs in two stages. In the first stage, cells are grouped into

Microclusters (MCs) using the Louvain algorithm to capture highly homogeneous cell populations. Simultaneously, Cell Type Relevance scores (CTRs) are calculated for each cell through a linear combination of its normalized marker intensity vector with predefined cell type signatures. A higher CTR indicates a stronger association with a specific cell type and vice versa. Next, TACIT determines a threshold that distinguishes true positive cell type signals from background noise. CTRs are ranked, and segmental regression is applied to divide the CTR growth curve into 2 to 4 segments, identifying high and low relevance clusters. Unavoidably, there exists cells with low relevance score present in high relevance group and vice versa. An absolute CTR threshold is then determined to minimize such inconsistent assignments. Subsequently, cells with CTRs exceeding this threshold are labeled as positive for the associated cell type. In cases where cells are labeled with multiple cell types, *TACIT* employs a deconvolution step using the k-nearest neighbors (k-NN) algorithm and cells with unique cell type assignment to resolve such ambiguity.

Constellation for Tissue Cellular Neighborhood Assignment in Whole Slide Imaging: We developed *Constellation* to address the challenges of unsupervised Tissue Cellular Neighborhood (TCN) assignment in Whole Slide Imaging (WSI) data, building upon the original CytoCommunity framework. While CytoCommunity effectively clusters cells within smaller tissue slides, it struggles with scalability for large WSIs or multiple slides. Constellation introduces novel partitioning strategies and a consensus-based assignment approach, enabling efficient processing of large datasets. The first step is WSI partitioning, where tissues are divided into smaller, non-overlapping sub-regions to reduce computational complexity. The tissue, represented as a set of cells W, with spatial coordinates and phenotypic labels, is partitioned using three strategies: vertical stripes (based on x-coordinate boundaries), horizontal stripes (based on y-coordinate boundaries), and square grids (dividing along both axes). Each partition contains a minimum and maximum number of cells to ensure computational efficiency and biological relevance, where partitions exceeding the maximum cell count are subdivided, and smaller ones are merged. Formally, let $W = \bigcup_{i=1}^{K} S_i$,

where S_i representing the $i-th$ subset of cells, and K is the total number of partitions such that $S_i \cap S_j = \emptyset$ for all $i \neq j$. Next, sub-graph $G_i = (V_i, E_i)$ are constructed for each partition S_i capturing both phenotypic and spatial relationships among the cells. Here, V_i denotes the set of nodes (cells) within S_i, where each node $v_n \in V_i$ is represented as a one-hot encoded phenotype vector:

$$v_n = [t_{n1}, t_{n2}, \dots, t_{nT}] \text{ with } t_{ni} = 1 \text{ if phenotype = } i,$$

and T is the total number of phenotypes. The edge set E_i connects nodes based on the k-nearest neighbors determined by Euclidean distance in the spatial coordinate space. The sub-graph is represented by an adjacency matrix $A_{G_i} \in \mathbb{R}^{|V_i| \times |V_j|}$, where entries are binary (1 for an edge, 0 otherwise), and a feature matrix $F_{G_i} \in \mathbb{R}^{|V_i| \times T}$, representing cell phenotypes. The collective unsupervised training stage processes these sub-graphs in a graph neural network (GNN). Each sub-graph is treated as an independent training sample, and the model processes them in mini-batches with a gradient accumulation technique over $N_{acc} = 32$ batches. The GNN minimizes a combined loss function for training:

$$L_t = \frac{1}{N_{acc}} (L_{MinCut} + L_{orth}),$$

where L_{MinCut} encourages separation between different clusters and L_{orth} promotes orthogonality, ensuring that each cell is confidently assigned to distinct TCNs. Finally, TCN assignment is performed using a consensus-based approach that aggregates results across the partitioning strategies. For each cell v_n, TCN assignments from vertical stripes, horizontal stripes, and square grid partitioning methods are compared, with the final assignment determined by majority voting:

$$c_{final}^n = mode(c_1^n, c_2^n, c_3^n),$$

where c_m^n denotes the TCN assignment from the m-th partitioning method. In rare cases where all three assignments differ, the cell is excluded to maintain result reliability. This consensus-based approach ensures robust TCN assignments by integrating diverse spatial perspectives within the

WSI. *STARComm for spatial-resolved cell-cell communication:* In STARComm (*SpaTiAl Receptor-Ligand Analysis for Cell-Cell COMMunication)* analyses, we focus on identifying and characterizing co-location communication between cells based on 50 ligand-receptor (LR) interactions in spatial transcriptomics datasets, such as MERSCOPE. The process begins by identifying cell pairs where communication may occur. Specifically, a cell that expresses a ligand gene (with value above 0 count) interacts with a nearby cell that expresses the corresponding receptor gene (also above 0 count), provided they are within 50 microns of each other in the spatial tissue layout. This process is repeated for all LR pairs (50 pairs in total), resulting in a network of interactions where each pair of cells that meets the ligand-receptor criteria is connected by an edge. Once all the LR interactions are established, we organize the spatial region into a grid of bins size 100 microns x 100 microns. Each bin in this grid captures the spatial region, and we ensure that each bin contains a very small number of edges representing ligand-receptor pairs. For each bin in the grid, we tally the count of individual LR pairs that fall within that spatial region, providing a cumulative measure of communication strength in that area. To analyze the distribution of these interactions more deeply, we calculate the kernel density for each ligand-receptor pair.

$$\hat{f}(x,\ y)\ =\ \frac{1}{n\ *h_x*h_y}\sum_{(x_i,y_i)\in n} K(\frac{x-x_i}{h_x})\,K(\frac{y-y_i}{h_y})$$

Where:

$\hat{f}(x,\ y)$: the density at the grid $(x,\ y)$

n: the total number of grids

h_x and h_y : the bandwidths in the x and y directions (default 10)

$(x_i,\ y_i)$: the coordinates of the grid.

$K(.)$: kernel function - $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$

The kernel density estimation reveals how communication is spatially distributed—areas where the interactions between cells are more frequent will have higher density values, indicating that those locations are hubs of ligand-receptor communication. Finally, we apply the Louvain clustering algorithm to the grid, clustering the bins based on their communication density vectors. This clustering process groups spatial regions that exhibit similar co-location of LR pairs meaning that the LR interactions are concentrated in specific areas of the tissue. By clustering the bins, we capture and visualize how cell communication is spatially organized, highlighting areas where ligand-receptor interactions are occurring at higher densities, providing insight into the spatial dynamics of cellular communication across the tissue. *Astrograph* for spatial data visualization. We developed Astrograph, an R Shiny application that enables interactive visualization and analysis of spatial cell type data. The app allows users to customize cell type annotations with user-defined colors and select specific tissues or regions for focused visualization. It supports the generation of heatmaps that display cell type distributions based on marker expression, which can be tailored to reflect either cell type-specific markers or cell state markers. Beyond visualizing cell types, Astrograph also facilitates the exploration of cell states, providing insights into the proportions of various cell types and states across tissues or regions. Additionally, the app offers Voronoi plot functionality to depict spatial relationships between cells, thereby visualizing cell neighborhoods and interactions. Furthermore, Astrograph supports quantitative analysis of cell type interactions by counting neighboring relationships and identifying patterns in spatial cell dynamics, offering a comprehensive approach to spatial omics data exploration.

Adapting *Drug2Cell*[68] for spatial datasets. In the Spatial Drug2Cell analysis, we apply the Drug2Cell approach to the MERSCOPE data after the cells have been annotated with their respective cell types. Once cell type annotation is completed, we proceed by extracting the drug scores for each cell after we run drug2cell packages. These drug scores quantify the response or relevance of individual cells to specific drugs, based on expression profiles and other molecular

characteristics. After obtaining the drug scores, we map these scores onto the spatial coordinates of the cells, allowing us to visualize the distribution of drug responses across the tissue landscape. This spatial mapping enables us to analyze how drug sensitivity or resistance varies based on the physical location of the cells within the tissue. Furthermore, we can compare the drug responses between different cell types and regions, highlighting variations in drug effectiveness or targeting based on cellular identity and spatial context. This comparison can reveal how certain cell types, such as immune cells, fibroblasts, or tumor cells, may exhibit distinct drug responses depending on their location within the tissue microenvironment. The combination of spatial information with drug scores provides a comprehensive view of how drugs interact with different cell populations, aiding in targeted therapeutic strategies.

Motif graphs: The motif graph is a network-based representation designed to capture spatial adjacency relationships between groups of functional neighborhoods (FNs). A functional neighborhood can represent a specific TCN from Mega-CytoCommunity or an MCIM from StarComm. In this graph, each node corresponds to an FN, and an edge exists between two nodes if there exist neighboring cells linking two different FNs, i.e., a cell from one FN is within 50 micrometers of a cell in the other FN. Since FNs may appear in multiple locations on a slide as instances, the motif graph effectively summarizes the interaction relationships between these types of FNs, with repeated FN motifs forming a subgraph structure. For each pair of connected functional neighborhoods (FNs), we compute the proportion of neighboring cells linking them relative to the total number of cells in both FNs. This metric reflects the frequency of neighboring relationships, and it is visually represented by the thickness of the edges in the motif graph. Note that in the visualization of motif graph, we highlight only the top three most frequent interactions between clusters to emphasize key connections.

STATISTICAL METHODS

General methods. A variety of tools and statistical approaches were used for data analysis, as appropriate for the data type and study objectives. Non-sequencing-based data were primarily analyzed using QuPath and/or Prism 9. The choice of statistical tests is detailed in the main text and figure legends. All statistical analyses were two-sided, and significance was determined at a threshold of (p-value<0.05). For comparisons between groups, a t-test was applied when normality assumptions were met. If the data did not meet these assumptions, a Wilcoxon rank-sum test (Mann-Whitney U test) was used instead. All visualizations, including graphs and supplementary figures, were created using Prism 9/10, a community instance of Cellenics® (hosted by Biomage), and the `ggplot2`, `pheatmap`, `ggforce` package in R unless explicitly stated otherwise. Venn Diagrams were generated using http://www.interactivenn.net. In spatial multiomics analyses, workflows for cell segmentation, cell type annotation, and tissue structure annotation were tailored to the data and the biological questions of interest. Cell segmentation was performed using the advanced capabilities of Cellpose3, which excels at robust segmentation of complex tissue images. For annotation, AstroSuite was employed, providing an automated pipeline for assigning cell identities, functional states, tissue structure neighborhoods, and multiplexed cell interaction maps (MCIMs). These methods are described in detail in the " **Cell Segmentation for Multi-IF and MERFISH Images using *Cellpose3*** and ***AstroSuite for* Auto Assignment of Cell Identities and States, TCNs, and MCIMs** " section. Heatmaps for spatial data were generated by summarizing marker expression. For PhenoCycler datasets, the mean expression levels were used to capture trends across the data, while for MERFISH datasets, the median count values were used to better handle the sparsity inherent in these datasets. When alternative methods for summarization were used, these are specified in the corresponding figure legends to ensure transparency and reproducibility. Differential expression analyses for genes, markers, and drug targets were conducted using the Seurat pipeline (version 5.1.0). This pipeline integrates preprocessing, normalization, and differential expression analysis, allowing us to compare expression profiles across groups effectively.

DATA AVAILABILITY

All data, including links to original raw data from each of the 11 studies (oral mucosa[2,22,34-37], salivary glands[36,38-40], and pulp[37,41,42]) can be found at GEO: https://www.ncbi.nlm.nih.gov/geo/. Original raw data for soft palate, labial mucosa, and parotid salivary gland can be also found at GEO (TBD). The data can also be analyzed at: https://cellxgene.cziscience.com/collections/065ad318-59fd-4f8c-b4b1-66caa7665409

CODE AVAILABILITY

Analysis notebooks and CELLxFEATURE matrices for MERSCOPE and Phenocycler-Fusion 2.0 data are available at: https://github.com/Loci-lab/Oral-Craniofacial-Atlas

METHODS REFERENCES:

68 Kanemaru, K. *et al.* Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801-810 (2023). https://doi.org:10.1038/s41586-023-06311-1

69 Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506 (2020). https://doi.org:10.1038/s41596-020-0292-x

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

71 Browaeys, R. *et al.* MultiNicheNet: a flexible framework for differential cell-cell communication analysis from multi-sample multi-condition single-cell transcriptomics data. *BioRxiv*, 2023.2006. 2013.544751 (2023).

72 Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. *bioRxiv*, 2024.2002. 2010.579780 (2024).

ACKNOWLEDGEMENTS

IS and KMB want to extend a heartfelt thanks to the Human Cell Atlas (HCA), specifically Sarah Teichman, Aviv Regev, Ellen Todres, and Lucia Robson, as well as the HCA Oral & Craniofacial Bionetwork, for supporting this and ongoing project to map the oral and craniofacial tissues in health and diseases. We also want to thank OMAPiX, Inc, especially Nachiket Kashikar and Mridula Iyer, for their thoughtful support of this and future projects. Furthermore, we acknowledge that this article has become stronger and more comprehensive through conversations with Katarzyna M Tyc (VCU), Di Wu, and Ji-Eun Park—both at UNC. Figure 1d, Extended Data 1c, and Figure 4a were adapted using images from BioRender.com and released under a Creative Commons Attribution-NonCommercial-NoDerivs (CC-BY-NC-ND) 4.0 International license. Figure 1a, Extended Data 1b utilized and adapted the "salivary-glands" icon by Servier (https://smart.servier.com/), which is licensed under CC-BY 3.0 (https://creativecommons.org/licenses/by/3.0/.) Data services in support of the research project were provided by the VCU Massey Comprehensive Cancer Center Bioinformatics Shared Resource. We would also like to thank the support and the staff of the NIDCR Office of the Clinical Director, NIDCR Dental Clinic, the National Eye Institute Ophthalmology Clinic, and the NCI Center for Cancer Research Anatomic Pathology. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We would like to thank the following NIH core facilities which enabled this study: NIDCR/NICHD Computational Biology and Genomics Core, the NIDCR Microscopy Core, the NIDCR Combined Technical Research Core, and the Laboratory of Genitourinary Cancer Pathogenesis (LCGP) Microscopy Core Facility. We would also like to thank the generous support of the NIH/UMD Graduate Partnership Program for TJFP. This research was principally supported through research awards to BMW from the Division of Intramural Research (DIR) Program of the National Institute of Dental and Craniofacial Research of the National Institutes of Health (NIH/NIDCR ZIA: DE000704). Massey is supported in part with

funding from NIH-NCI Cancer Center Support Grant P30CA016059. D.P. is supported by Fundação para a Ciência e a Tecnologia (2020.08715.BD). IS is funded by Barts Charity (MGU045), the Royal Society (RGS/R2/202291), and the British Skin Foundation (004/RA/23). This work was supported by generous start-up funds from the ADA Science & Research Institute (Volpe Research Scholar Award) to KMB. This publication is part of the Human Cell Atas: http://www.humancellatlas.org/publications.

CONTRIBUTIONS

BFM, KLAH, DP, IS, JL, and KMB conceptualized the project.

BFM, KLAH, DP, XZ, MK, SAT, JL, KMB developed methods for data analysis.

BFM, DP, NK, QTE, AG, KIK, AK, SP, KK, BMW, IS, KMB supported sample collection.

BFM, KLAH, DP, XZ, MK, AVP, BN, MMM, IS, JL, KMB performed experimental analysis.

BFM, KMB wrote the original draft.

BFM, KLAH, DP, XZ, QTE, SAT, IS, JL, KMB reviewed and edited the final manuscript.

CORRESPONDING AUTHORS

Kevin Matthew Byrd: kevinmbyrd@gmail.com

Jinze Liu: liuj15@vcu.edu

Inês Sequeira: i.sequeira@qmul.ac.uk

COMPETING INTERESTS

The authors had access to the study data and reviewed and approved the final manuscript. Although the authors view each of these as noncompeting financial interests, BFM, KLAH, DP, QTE, AVP, KIK, SAT, AK, SP, KK, BWM, IS, JL, and KMB are all active members of the Human

Cell Atlas; furthermore, SAT. has consulted for Roche and Genentech and is a scientific advisor for Biogen, GlaxoSmithKline, and Foresite Labs. IS a consultant for L'Oréal Research and Innovation; KMB is a scientific advisor at Arcato Laboratories; KMB and JL are co-founder of Stratica Biosciences, Inc. All other authors declare no competing interests.

Figure 1

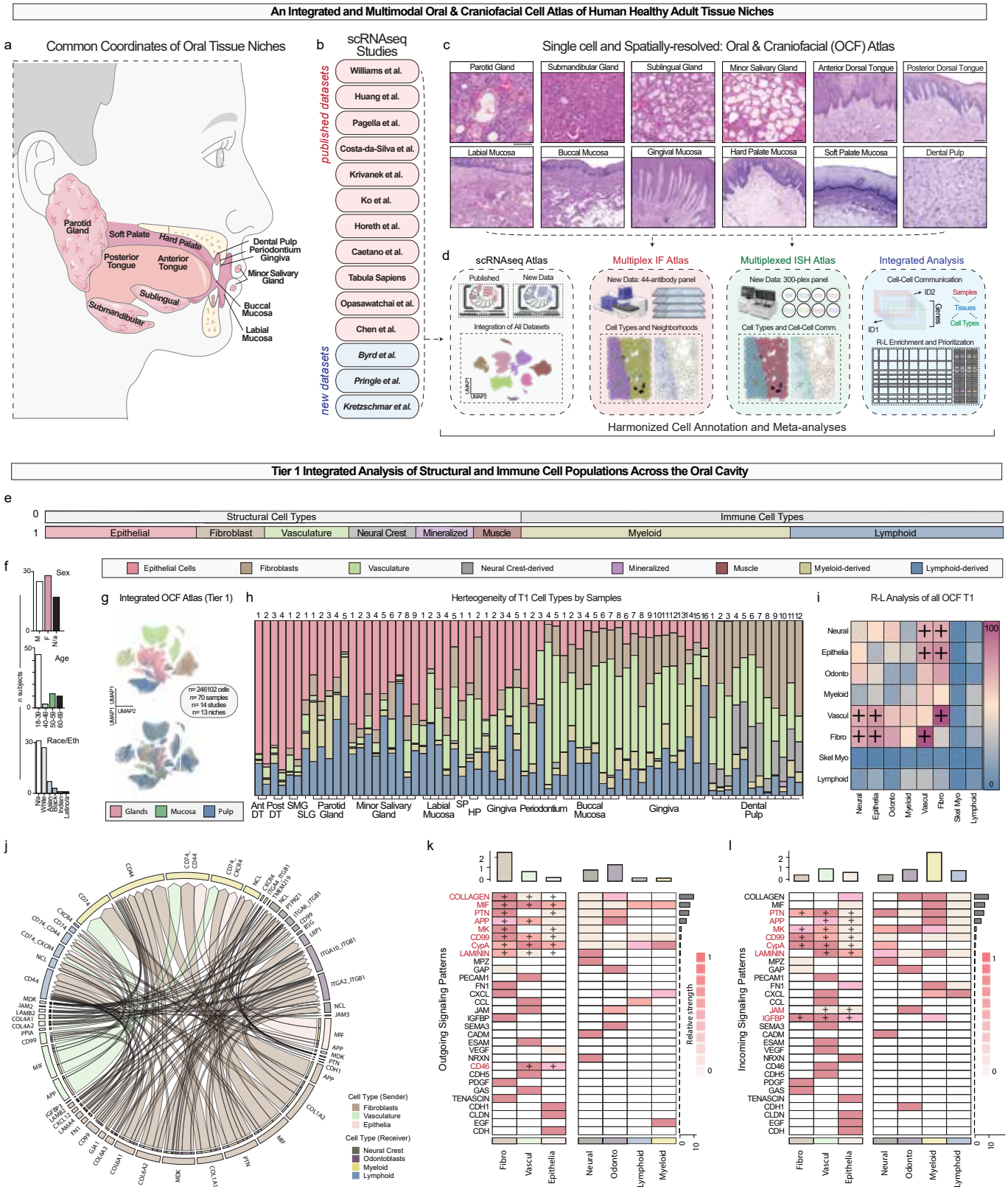

An Integrated and Multimodal Oral & Craniofacial Cell Atlas of Human Healthy Adult Tissue Niches
a
Common Coordinates of Oral Tissue Niches
Parotid Gland
Soft Palate
Hard Palate
Posterior Tongue
Anterior Tongue
Sublingual
Submandibular
Dental Pulp
Periodontium
Gingiva
Minor Salivary Gland
Buccal Mucosa
Labial Mucosa
b
scRNAseq Studies
published datasets
Williams et al.
Huang et al.
Pagella et al.
Costa-da-Silva et al.
Krivanek et al.
Ko et al.
Horeth et al.
Caetano et al.
Tabula Sapiens
Opasawatchai et al.
Chen et al.
new datasets
Byrd et al.
Pringle et al.
Kretzschmar et al.
c
Single cell and Spatially-resolved: Oral & Craniofacial (OCF) Atlas
Parotid Gland
Submandibular Gland
Sublingual Gland
Minor Salivary Gland
Anterior Dorsal Tongue
Posterior Dorsal Tongue
Labial Mucosa
Buccal Mucosa
Gingival Mucosa
Hard Palate Mucosa
Soft Palate Mucosa
Dental Pulp
d
scRNAseq Atlas
Published
New Data
Integration of All Datasets
UMAP1
UMAP2
Multiplex IF Atlas
New Data: 44-antibody panel
Cell Types and Neighborhoods
Multiplexed ISH Atlas
New Data: 300-plex panel
Cell Types and Cell-Cell Comm.
Integrated Analysis
Cell-Cell Communication
ID2
ID1
Samples
Tissues
Cell Types
Genes
R-L Enrichment and Prioritization
Harmonized Cell Annotation and Meta-analyses
Tier 1 Integrated Analysis of Structural and Immune Cell Populations Across the Oral Cavity
e
0
Structural Cell Types
Immune Cell Types
1
Epithelial
Fibroblast
Vasculature
Neural Crest
Mineralized
Muscle
Myeloid
Lymphoid
f
Sex
Age
Race/Eth
n subjects
Epithelial Cells
Fibroblasts
Vasculature
Neural Crest-derived
Mineralized
Muscle
Myeloid-derived
Lymphoid-derived
g
Integrated OCF Atlas (Tier 1)
n= 246102 cells
n= 70 samples
n= 14 studies
n= 13 niches
Glands
Mucosa
Pulp
h
Herteogeneity of T1 Cell Types by Samples
Ant DT
Post DT
SMG SLG
Parotid Gland
Minor Salivary Gland
Labial Mucosa
SP HP
Gingiva
Periodontium
Buccal Mucosa
Gingiva
Dental Pulp
i
R-L Analysis of all OCF T1
Neural
Epithelia
Odonto
Myeloid
Vascul
Fibro
Skel Myo
Lymphoid
j
Cell Type (Sender)
Fibroblasts
Vasculature
Epithelia
Cell Type (Receiver)
Neural Crest
Odontoblasts
Myeloid
Lymphoid
k
Outgoing Signaling Patterns
COLLAGEN
MIF
PTN
APP
MK
CD99
CypA
LAMININ
MPZ
GAP
PECAM1
FN1
CXCL
CCL
JAM
IGFBP
SEMA3
CADM
ESAM
VEGF
NRXN
CD46
CDH5
PDGF
GAS
TENASCIN
CDH1
CLDN
EGF
CDH
Relative strength
Fibro
Vascul
Epithelia
Neural
Odonto
Lymphoid
Myeloid
l
Incoming Signaling Patterns

Figure 2

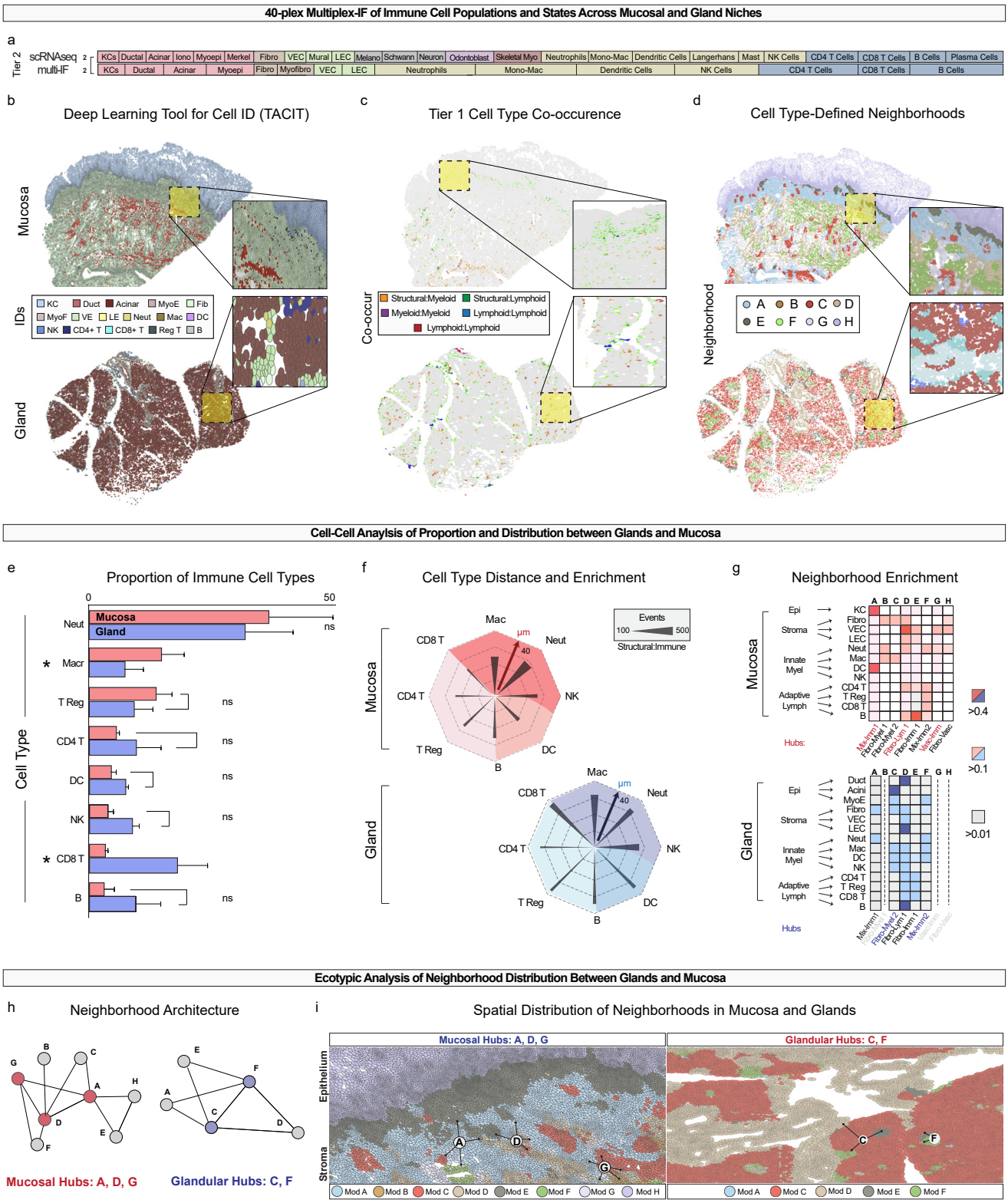

40-plex Multiplex-IF of Immune Cell Populations and States Across Mucosal and Gland Niches
a
Tier 2
scRNAseq
multi-IF
KCs Ductal Acinar Iono Myoepi Merkel Fibro VEC Mural LEC Melano Schwann Neuron Odontoblast Skeletal Myo Neutrophils Mono-Mac Dendritic Cells Langerhans Mast NK Cells CD4 T Cells CD8 T Cells B Cells Plasma Cells
KCs Ductal Acinar Myoepi Fibro Myofibro VEC LEC Neutrophils Mono-Mac Dendritic Cells NK Cells CD4 T Cells CD8 T Cells B Cells
b
Deep Learning Tool for Cell ID (TACIT)
Mucosa
IDs
Gland
KC Duct Acinar MyoE Fib
MyoF VE LE Neut Mac DC
NK CD4+ T CD8+ T Reg T B
c
Tier 1 Cell Type Co-occurence
Co-occur
Structural:Myeloid Structural:Lymphoid
Myeloid:Myeloid Lymphoid:Lymphoid
Lymphoid:Lymphoid
d
Cell Type-Defined Neighborhoods
Neighborhood
A B C D
E F G H
Cell-Cell Anaylsis of Proportion and Distribution between Glands and Mucosa
e
Proportion of Immune Cell Types
0
50
Cell Type
Neut
Mucosa
Gland
ns
Macr
T Reg
CD4 T
DC
NK
CD8 T
B
f
Cell Type Distance and Enrichment
Mucosa
Gland
Mac
Neut
NK
DC
B
T Reg
CD4 T
CD8 T
μm
40
Events
100
500
Structural:Immune
g
Neighborhood Enrichment
Mucosa
Gland
Epi
Stroma
Innate Myel
Adaptive Lymph
KC
Fibro
VEC
LEC
Neut
Mac
DC
NK
CD4 T
T Reg
CD8 T
B
Duct
Acini
MyoE
Hubs:
>0.4
>0.1
>0.01
Ecotypic Analysis of Neighborhood Distribution Between Glands and Mucosa
h
Neighborhood Architecture
Mucosal Hubs: A, D, G
Glandular Hubs: C, F
i
Spatial Distribution of Neighborhoods in Mucosa and Glands
Mucosal Hubs: A, D, G
Glandular Hubs: C, F
Epithelium
Stroma
Mod A Mod B Mod C Mod D Mod E Mod F Mod G Mod H
Mod A Mod C Mod D Mod E Mod F

Figure 3

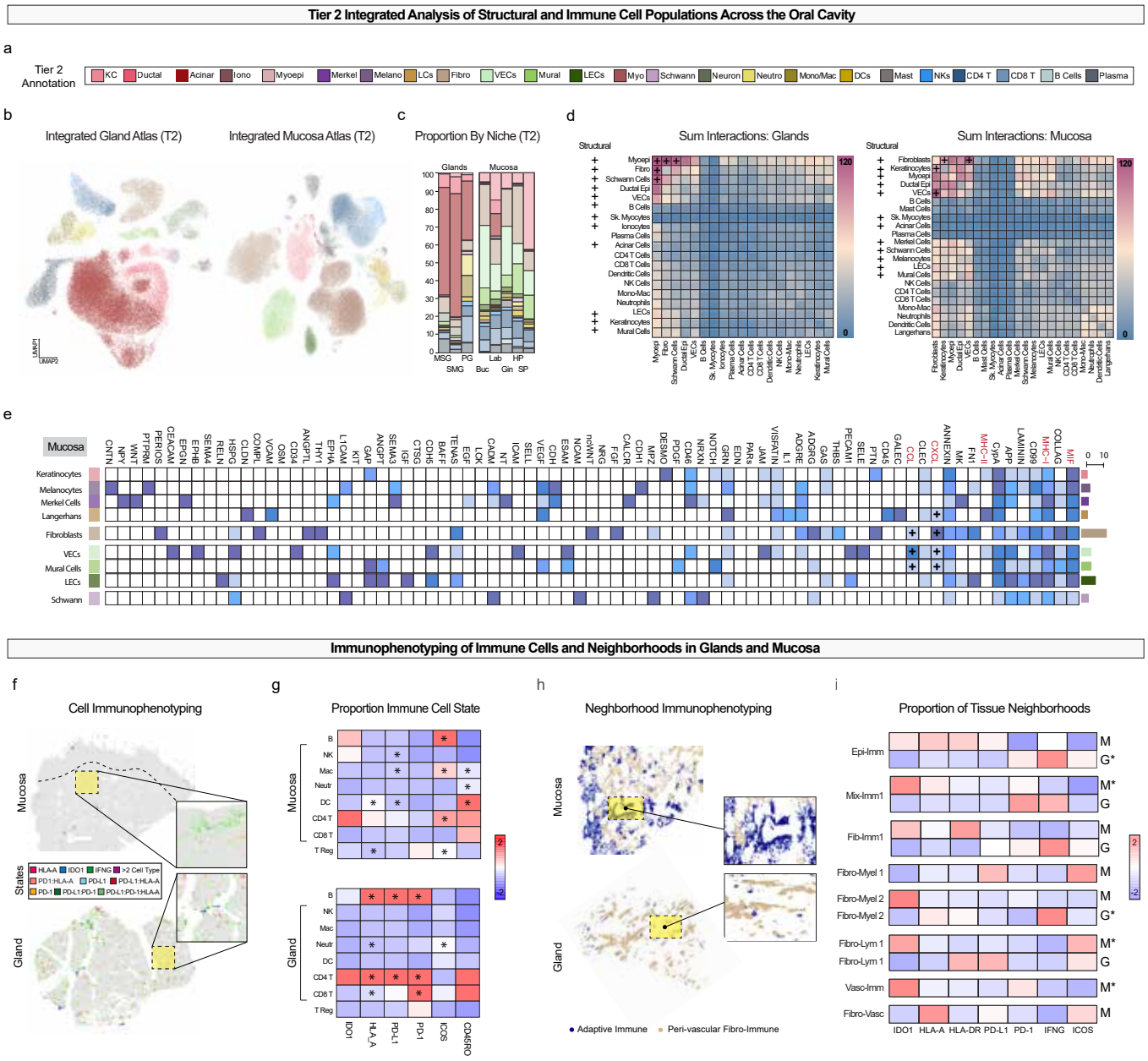
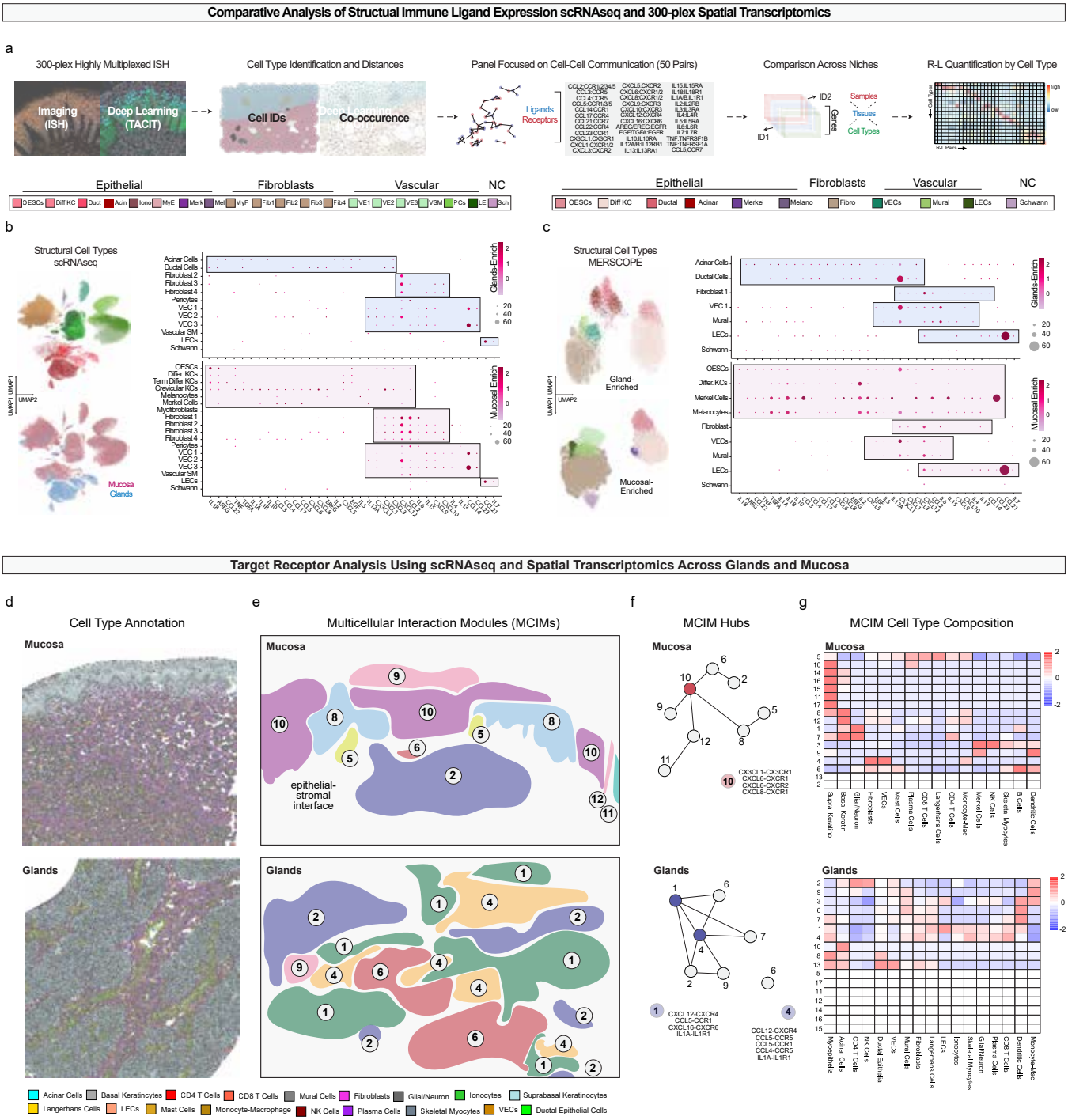

Tier 2 Integrated Analysis of Structural and Immune Cell Populations Across the Oral Cavity
a
Tier 2 Annotation
KC Ductal Acinar Iono Myoepi Merkel Melano LCs Fibro VECs Mural LECs Myo Schwann Neuron Neutro Mono/Mac DCs Mast NKs CD4 T CD8 T B Cells Plasma
b
Integrated Gland Atlas (T2)
Integrated Mucosa Atlas (T2)
c
Proportion By Niche (T2)
Glands
Mucosa
MSG PG Lab HP
SMG Buc Gin SP
d
Sum Interactions: Glands
Sum Interactions: Mucosa
Structural
e
Mucosa
Immunophenotyping of Immune Cells and Neighborhoods in Glands and Mucosa
f
Cell Immunophenotyping
Mucosa
Gland
States
g
Proportion Immune Cell State
h
Neghborhood Immunophenotyping
Adaptive Immune
Peri-vascular Fibro-Immune
i
Proportion of Tissue Neighborhoods
IDO1 HLA-A HLA-DR PD-L1 PD-1 IFNG ICOS
Immunophenotyping of Immune Cells Comparing Glands and Mucosa using scRNAseqq
j
Oral Immune Subcluster
Mucosal Immune
Glandular Immune
Proportion By Niche
k
DEGs of Innate and Adaptive Immune Cells
Neutrophils
Monocyte-Macrophage
Dendritic Cells
Natural Killer Cells
CD4 T Cells
CD8 T Cells
Plasma Cells
DEGs

Figure 4



Comparative Analysis of Structual Immune Ligand Expression scRNAseq and 300-plex Spatial Transcriptomics
a
300-plex Highly Multiplexed ISH
Imaging (ISH)
Deep Learning (TACIT)
Cell Type Identification and Distances
Cell IDs
Co-occurence
Panel Focused on Cell-Cell Communication (50 Pairs)
Ligands
Receptors
Comparison Across Niches
ID2
ID1
Genes
Samples
Tissues
Cell Types
R-L Quantification by Cell Type
Epithelial
Fibroblasts
Vascular
NC
b
Structural Cell Types scRNAseq
Mucosa
Glands
Glands-Enrich
Mucosal Enrich
c
Structural Cell Types MERSCOPE
Gland-Enriched
Mucosal-Enriched
Target Receptor Analysis Using scRNAseq and Spatial Transcriptomics Across Glands and Mucosa
d
Cell Type Annotation
Mucosa
Glands
e
Multicellular Interaction Modules (MCIMs)
epithelial-stromal interface
f
MCIM Hubs
g
MCIM Cell Type Composition
Acinar Cells
Basal Keratinocytes
CD4 T Cells
CD8 T Cells
Mural Cells
Fibroblasts
Glial/Neuron
Ionocytes
Suprabasal Keratinocytes
Langerhans Cells
LECs
Mast Cells
Monocyte-Macrophage
NK Cells
Plasma Cells
Skeletal Myocytes
VECs
Ductal Epithelial Cells

Figure 5

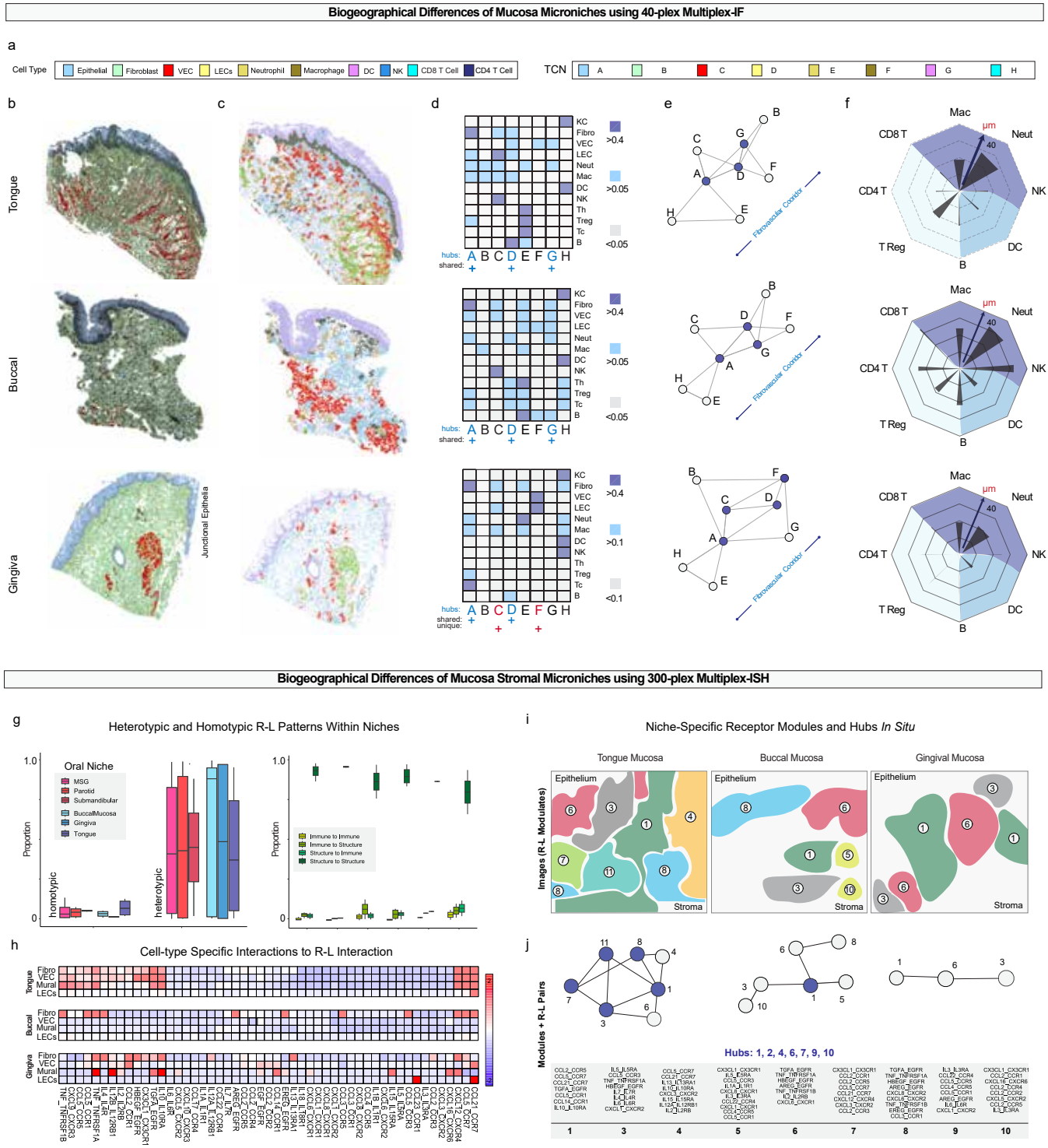

Biogeographical Differences of Mucosa Microniches using 40-plex Multiplex-IF
a
Cell Type
Epithelial
Fibroblast
VEC
LECs
Neutrophil
Macrophage
DC
NK
CD8 T Cell
CD4 T Cell
TCN
A
B
C
D
E
F
G
H
b
c
d
e
f
Tongue
Buccal
Gingiva
Junctional Epithelia
KC
Fibro
VEC
LEC
Neut
Mac
DC
NK
Th
Treg
Tc
B
>0.4
>0.05
<0.05
>0.1
<0.1
hubs:
shared:
unique:
Fibrovascular Corridor
Mac
Neut
NK
DC
B
T Reg
CD4 T
CD8 T
μm
40
Biogeographical Differences of Mucosa Stromal Microniches using 300-plex Multiplex-ISH
g
Heterotypic and Homotypic R-L Patterns Within Niches
Oral Niche
MSG
Parotid
Submandibular
BuccalMucosa
Gingiva
Tongue
Proportion
homotypic
heterotypic
Immune to Immune
Immune to Structure
Structure to Immune
Structure to Structure
h
Cell-type Specific Interactions to R-L Interaction
Tongue
Buccal
Gingiva
Fibro
VEC
Mural
LECs
i
Niche-Specific Receptor Modules and Hubs In Situ
Tongue Mucosa
Buccal Mucosa
Gingival Mucosa
Images (R-L Modulates)
Epithelium
Stroma
j
Modules + R-L Pairs
Hubs: 1, 2, 4, 6, 7, 9, 10

Figure 6

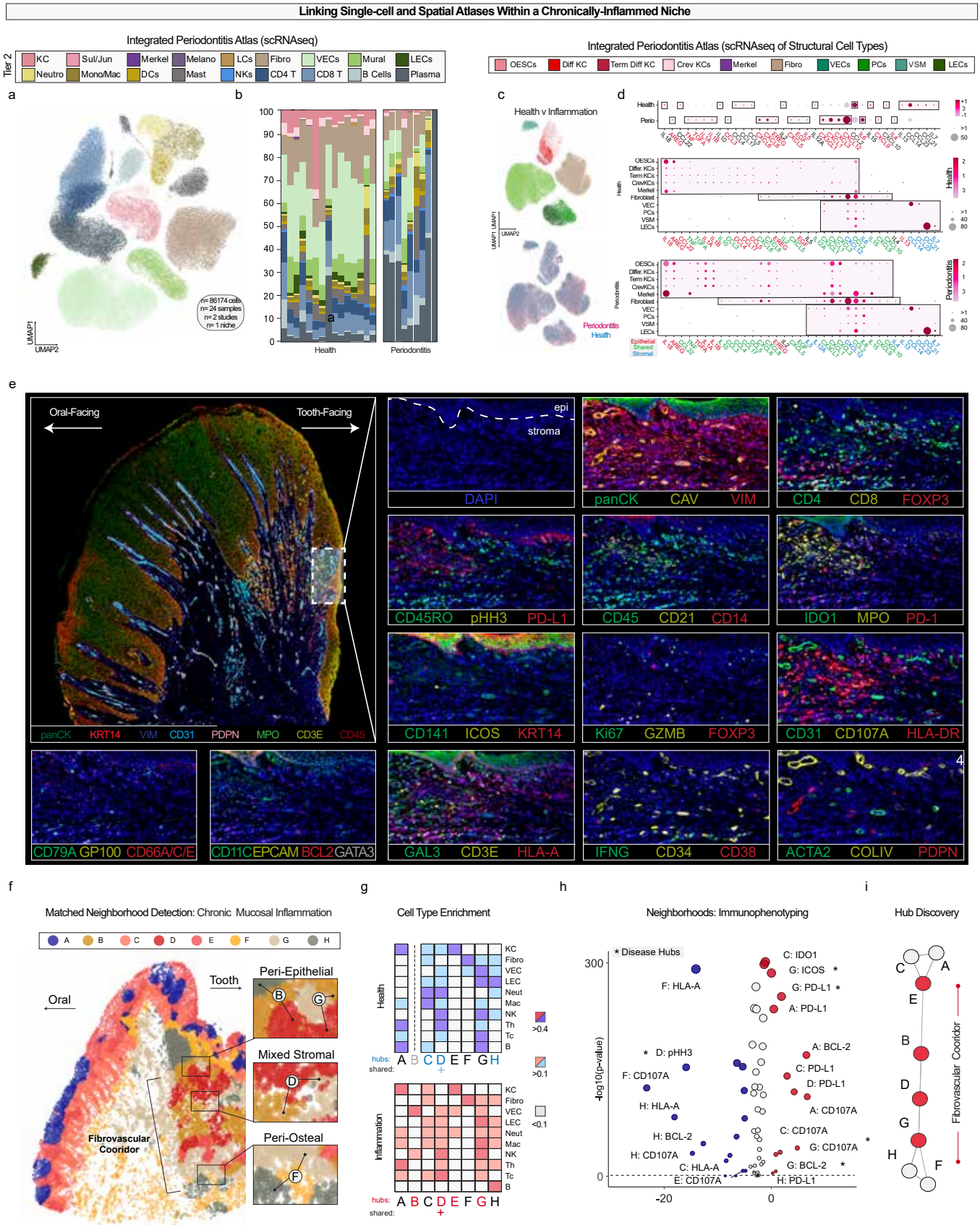
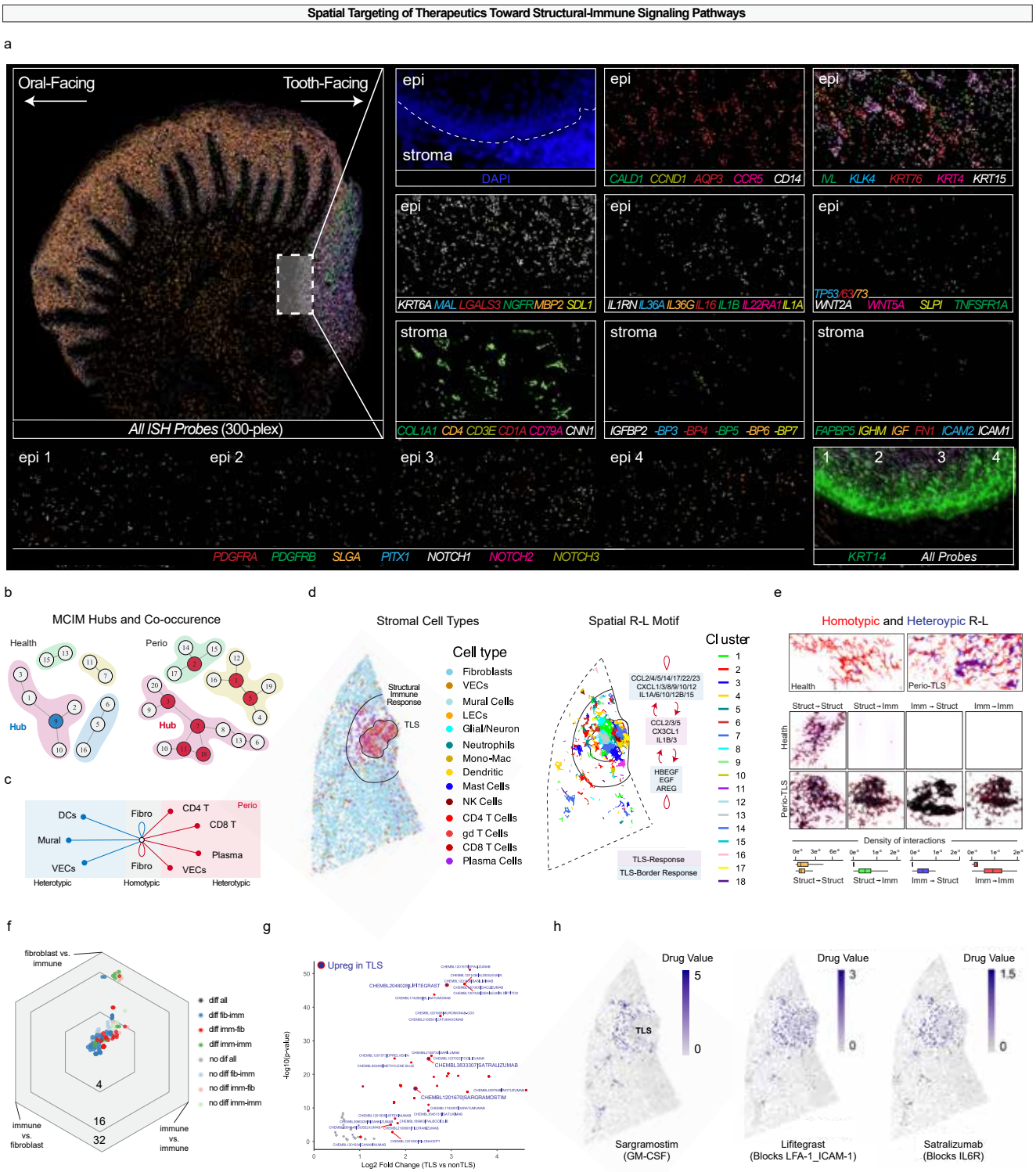
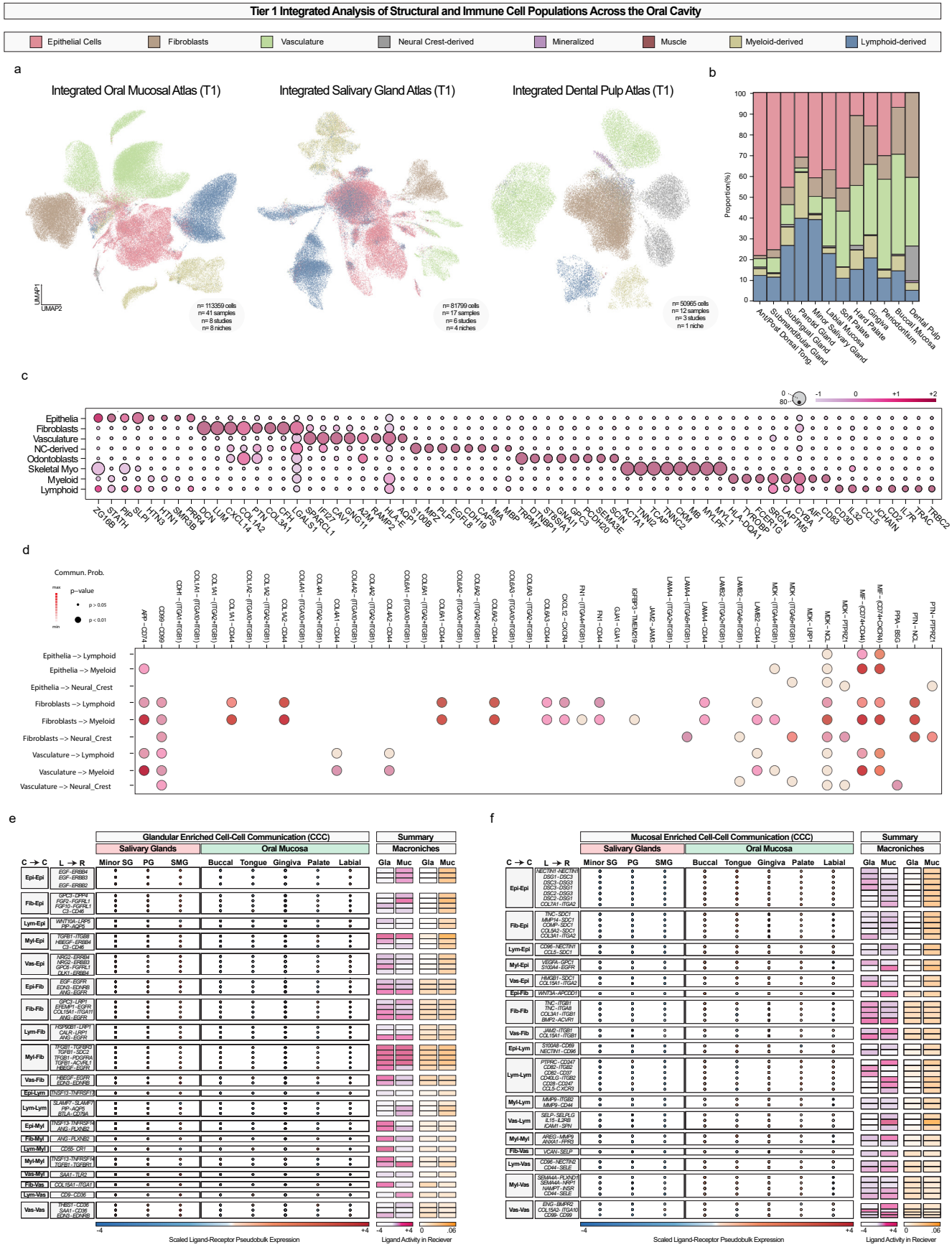

Linking Single-cell and Spatial Atlases Within a Chronically-Inflammed Niche
Integrated Periodontitis Atlas (scRNAseq)
Tier 2
KC
Sul/Jun
Merkel
Melano
LCs
Fibro
VECs
Mural
LECs
Neutro
Mono/Mac
DCs
Mast
NKs
CD4 T
CD8 T
B Cells
Plasma
n= 86174 cells
n= 24 samples
n= 2 studies
n= 1 niche
UMAP1
UMAP2
Health
Periodontitis
Integrated Periodontitis Atlas (scRNAseq of Structural Cell Types)
OESCs
Diff KC
Term Diff KC
Crev KCs
Merkel
Fibro
VECs
PCs
VSM
LECs
Health v Inflammation
Epithelial
Shared
Stromal
Oral-Facing
Tooth-Facing
panCK
KRT14
VIM
CD31
PDPN
MPO
CD3E
CD45
epi
stroma
DAPI
panCK
CAV
VIM
CD4
CD8
FOXP3
CD45RO
pHH3
PD-L1
CD45
CD21
CD14
IDO1
MPO
PD-1
CD141
ICOS
KRT14
Ki67
GZMB
FOXP3
CD31
CD107A
HLA-DR
CD79A
GP100
CD66A/C/E
CD11C
EPCAM
BCL2
GATA3
GAL3
CD3E
HLA-A
IFNG
CD34
CD38
ACTA2
COLIV
PDPN
Matched Neighborhood Detection: Chronic Mucosal Inflammation
Oral
Tooth
Peri-Epithelial
Mixed Stromal
Peri-Osteal
Fibrovascular Cooridor
Cell Type Enrichment
Health
Inflammation
hubs:
shared:
Neighborhoods: Immunophenotyping
Disease Hubs
-log10(p-value)
Hub Discovery
Fibrovascular Cooridor

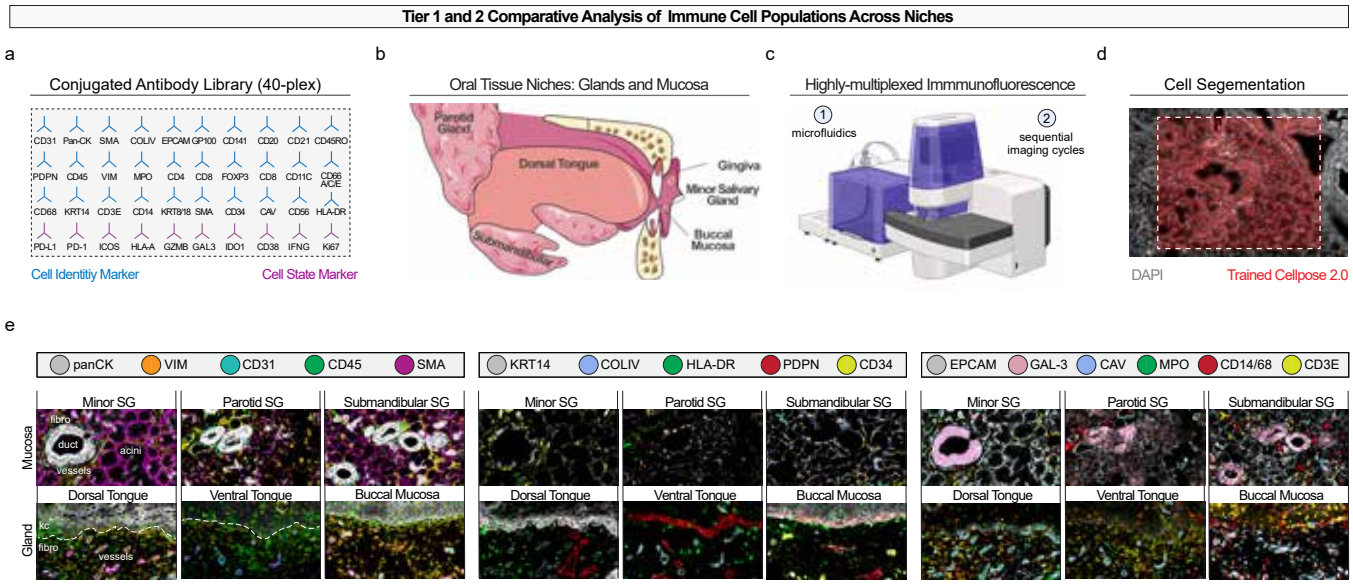
Figure 7


Spatial Targeting of Therapeutics Toward Structural-Immune Signaling Pathways
a
Oral-Facing
Tooth-Facing
All ISH Probes (300-plex)
epi
stroma
DAPI
CALD1 CCND1 AQP3 CCR5 CD14
IVL KLK4 KRT76 KRT4 KRT15
KRT6A MAL LGALS3 NGFR MBP2 SDL1
IL1RN IL36A IL36G IL16 IL1B IL22RA1 IL1A
TP53 63/73
WNT2A WNT5A SLPI TNFSFR1A
COL1A1 CD4 CD3E CD1A CD79A CNN1
IGFBP2 -BP3 -BP4 -BP5 -BP6 -BP7
FAPBP5 IGHM IGF FN1 ICAM2 ICAM1
epi 1
epi 2
epi 3
epi 4
PDGFRA PDGFRB SLGA PITX1 NOTCH1 NOTCH2 NOTCH3
KRT14 All Probes
b
MCIM Hubs and Co-occurence
Health
Perio
Hub
c
DCs
Fibro
CD4 T
CD8 T
Mural
Plasma
VECs
Heterotypic
Homotypic
d
Stromal Cell Types
Structural Immune Response
TLS
Cell type
Fibroblasts
VECs
Mural Cells
LECs
Glial/Neuron
Neutrophils
Mono-Mac
Dendritic
Mast Cells
NK Cells
CD4 T Cells
gd T Cells
CD8 T Cells
Plasma Cells
Spatial R-L Motif
Cluster
CCL2/4/5/14/17/22/23 CXCL1/3/8/9/10/12 IL1A/6/10/12B/15
CCL2/3/5 CX3CL1 IL1B/3
HBEGF EGF AREG
TLS-Response
TLS-Border Response
e
Homotypic and Heteroypic R-L
Health
Perio-TLS
Struct→Struct
Struct→Imm
Imm→Struct
Imm→Imm
Density of interactions
f
fibroblast vs. immune
immune vs. fibroblast
immune vs. immune
diff all
diff fib-imm
diff imm-fib
diff imm-imm
no dif all
no diff fib-imm
no diff imm-fib
no diff imm-imm
g
Upreg in TLS
-log10(p-value)
Log2 Fold Change (TLS vs nonTLS)
CHEMBL2048028|LIFITEGRAST
CHEMBL383307|SATRALIZUMAB
CHEMBL1201670|SARGRAMOSTIM
h
Drug Value
TLS
Sargramostim (GM-CSF)
Lifitegrast (Blocks LFA-1_ICAM-1)
Satralizumab (Blocks IL6R)

Extended Data 1



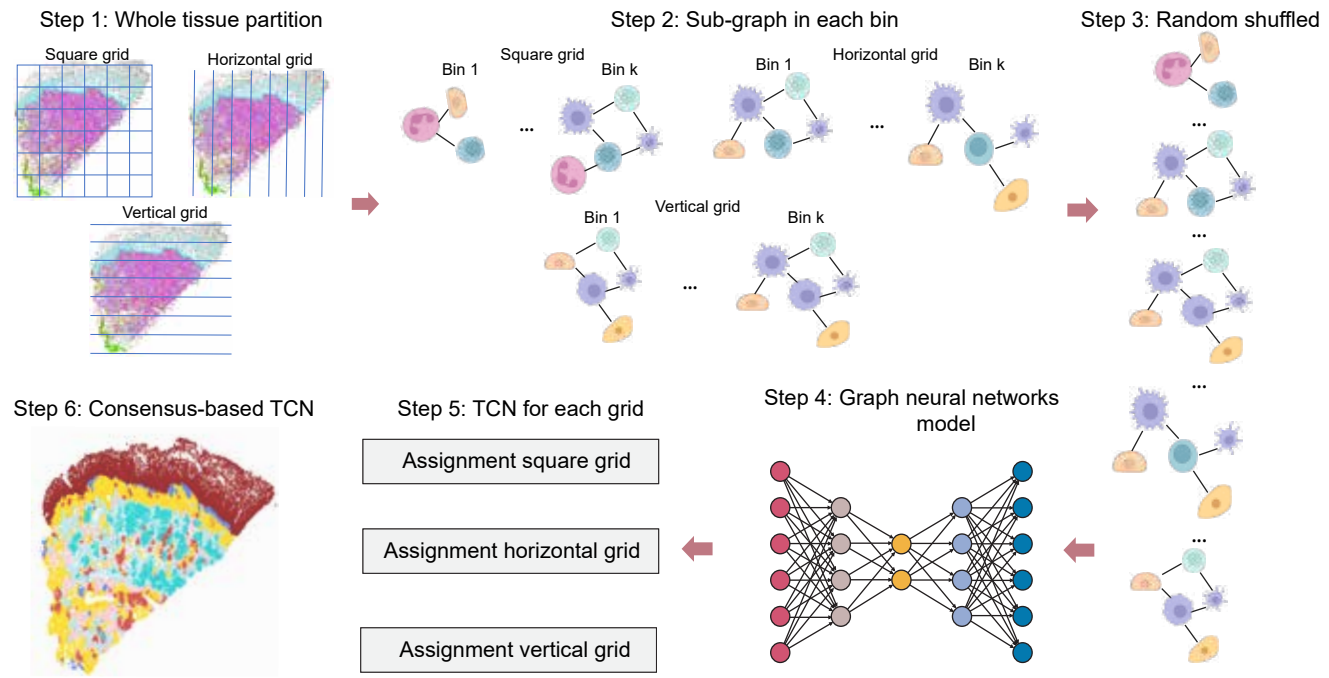
Extended Data 2



Extended Data 3

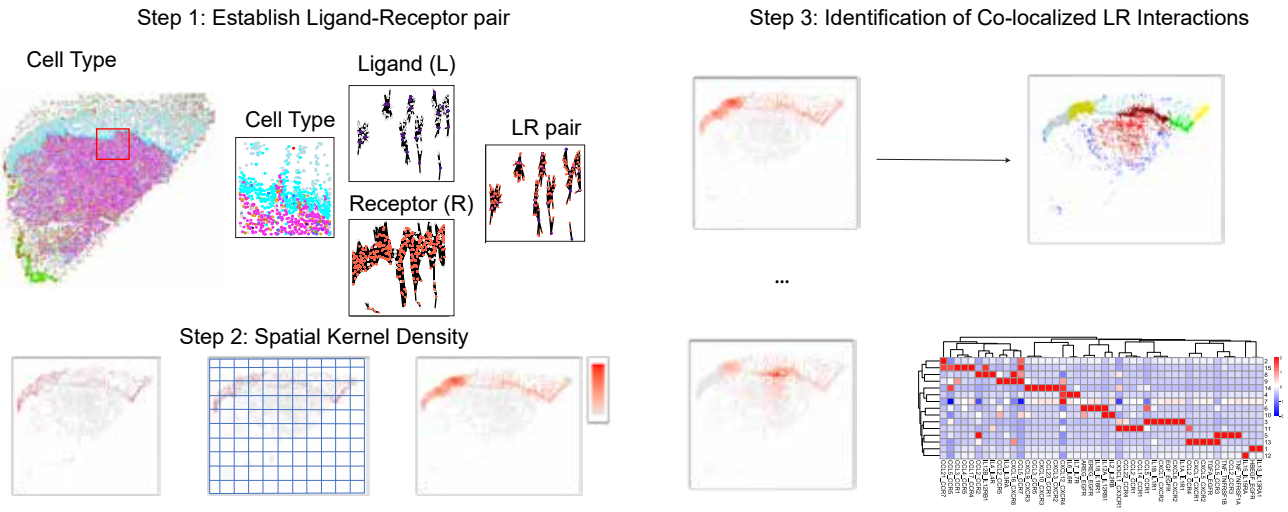
Constellation

a



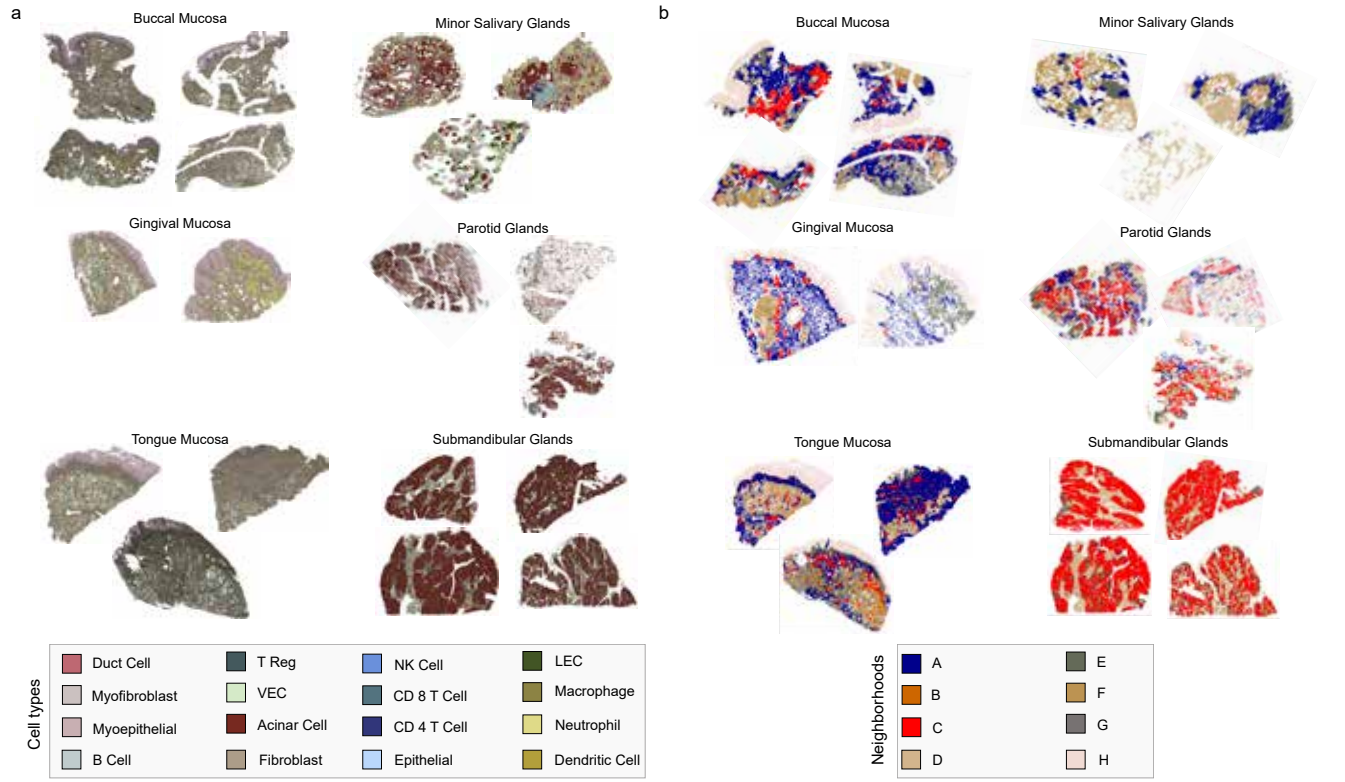
SpaTiAl Receptor-Ligand Analysis for Cell-Cell COMMunication

b



Extended Data 4

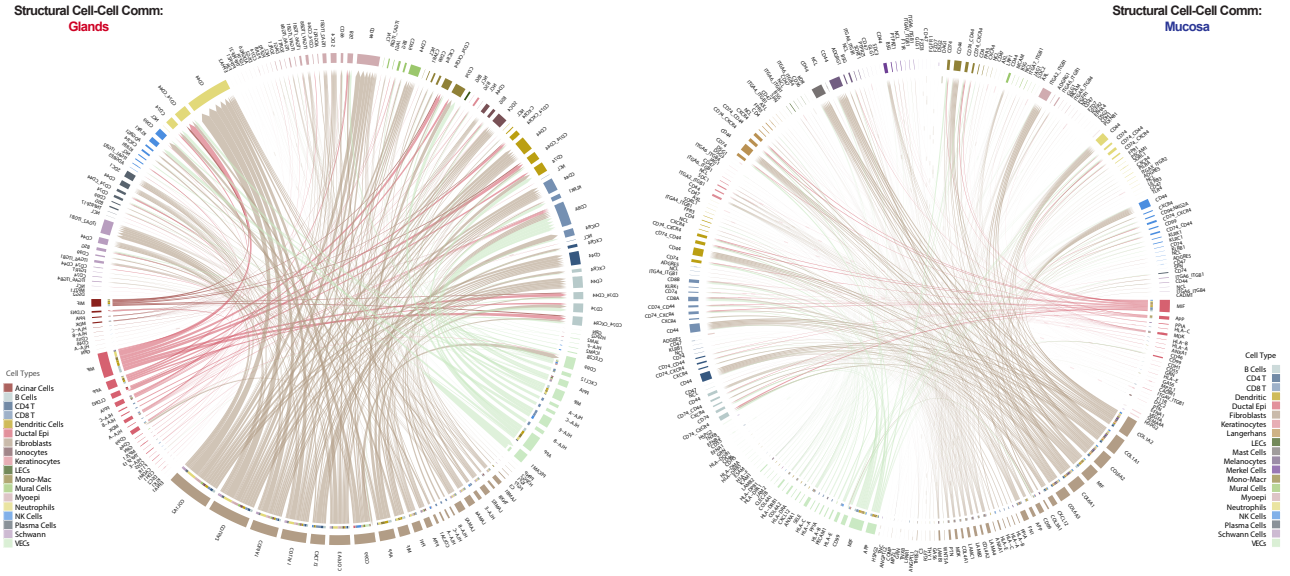
TACIT Cell type Annotation of Oral Cavity Niches using Multiplex IF



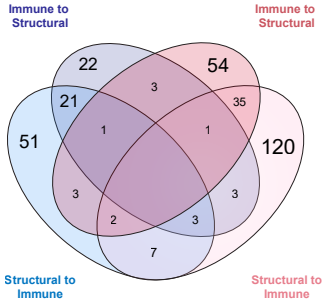
Extended Data 5

Tier 2 Integrated Analysis of Structural and Immune Cell Populations Across the Oral Cavity

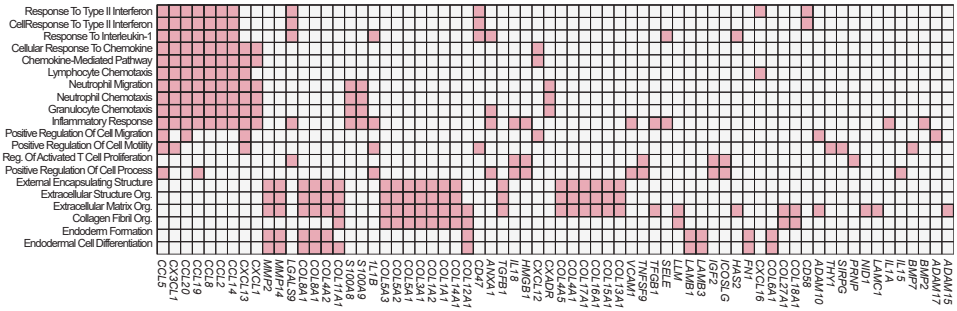
a



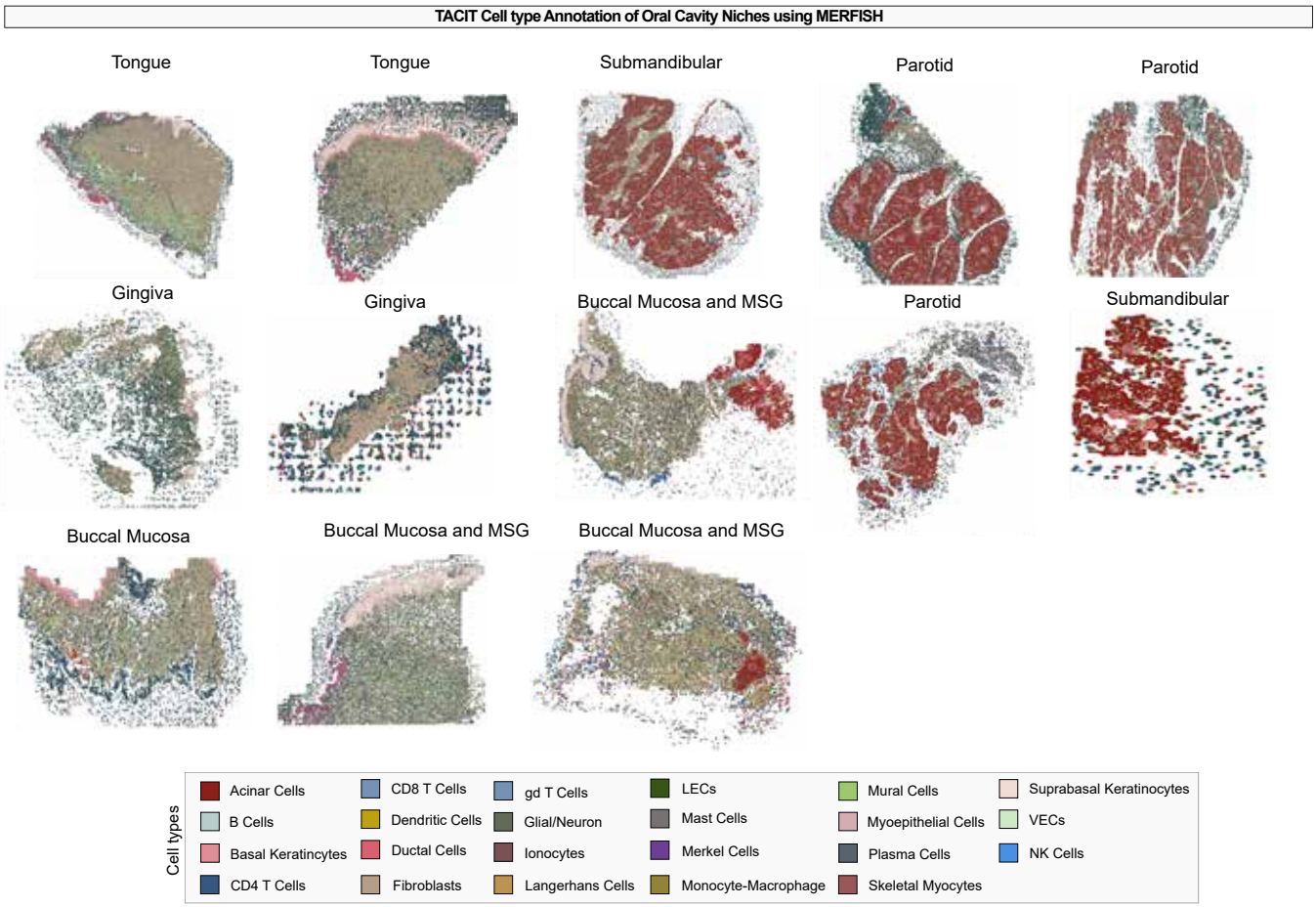
b



c

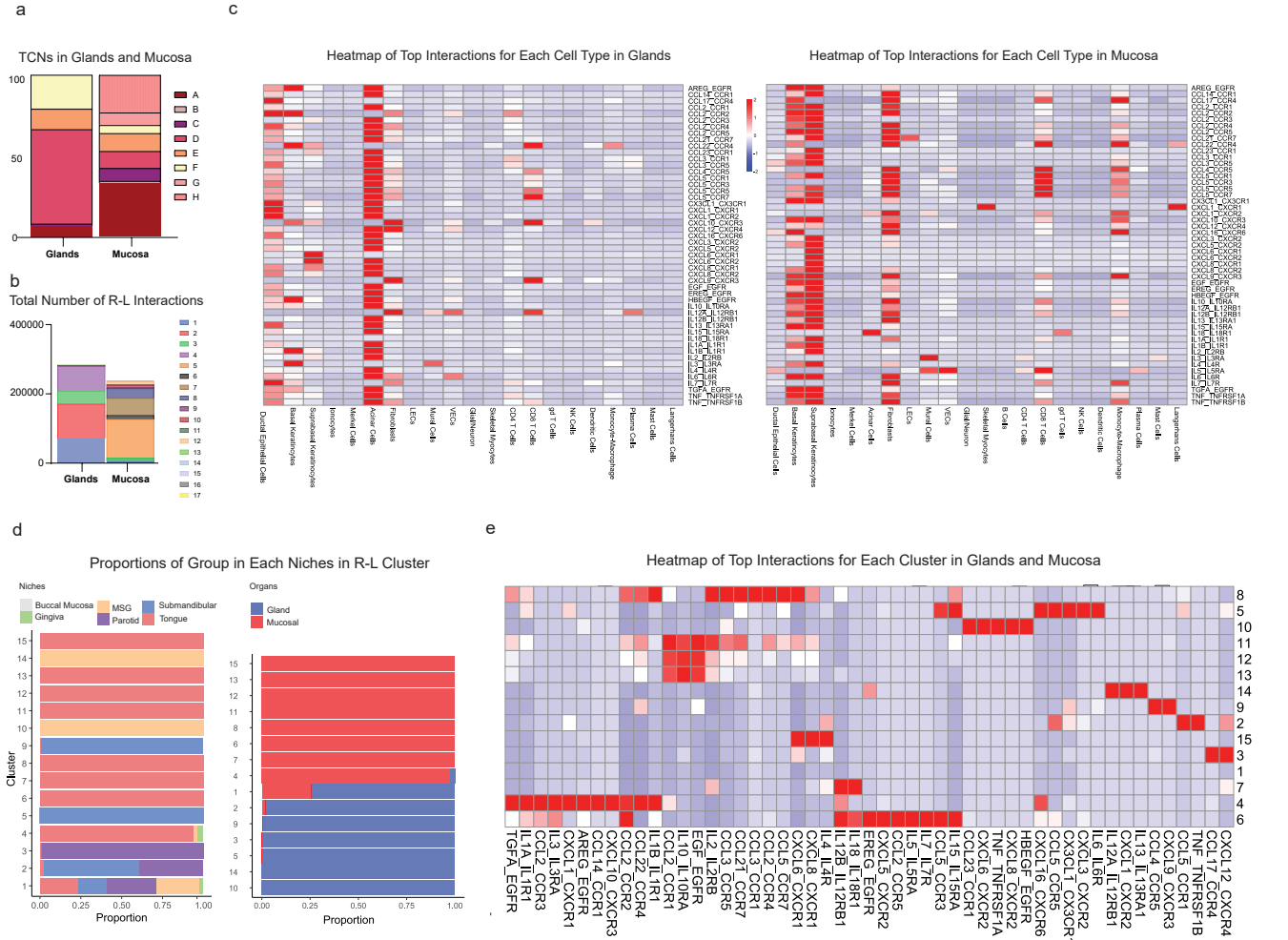
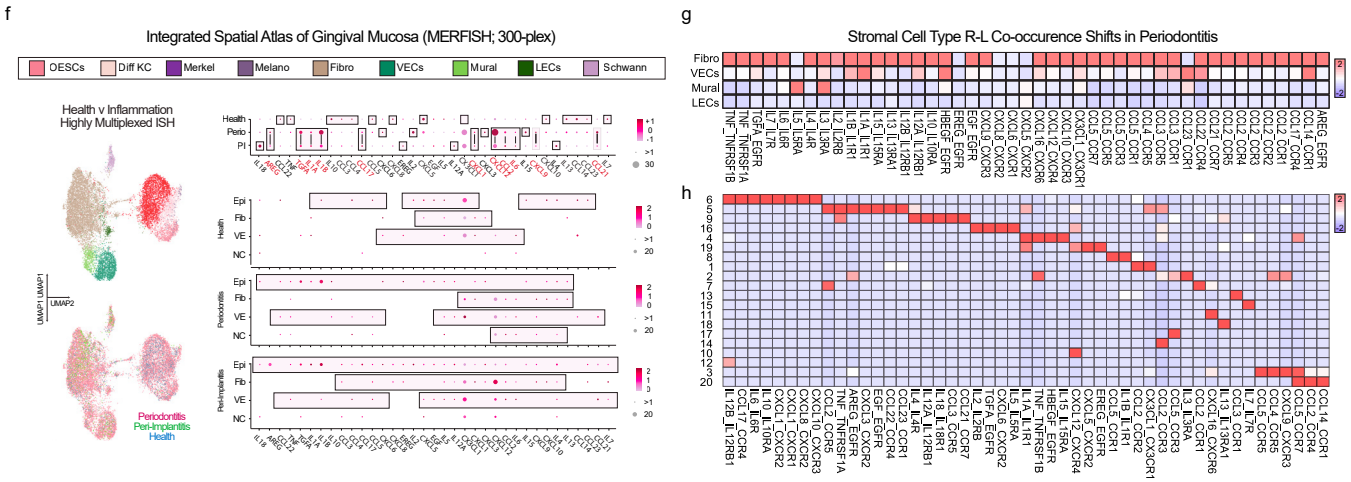


Extended Data 6



Extended Data 7

TCNs and MCIMs Across Glands and Mucosa



Spatial Cell-Cell Communication of Stromal Cell Types in Chronic Infammation: MERFISH