Supplemental Information

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

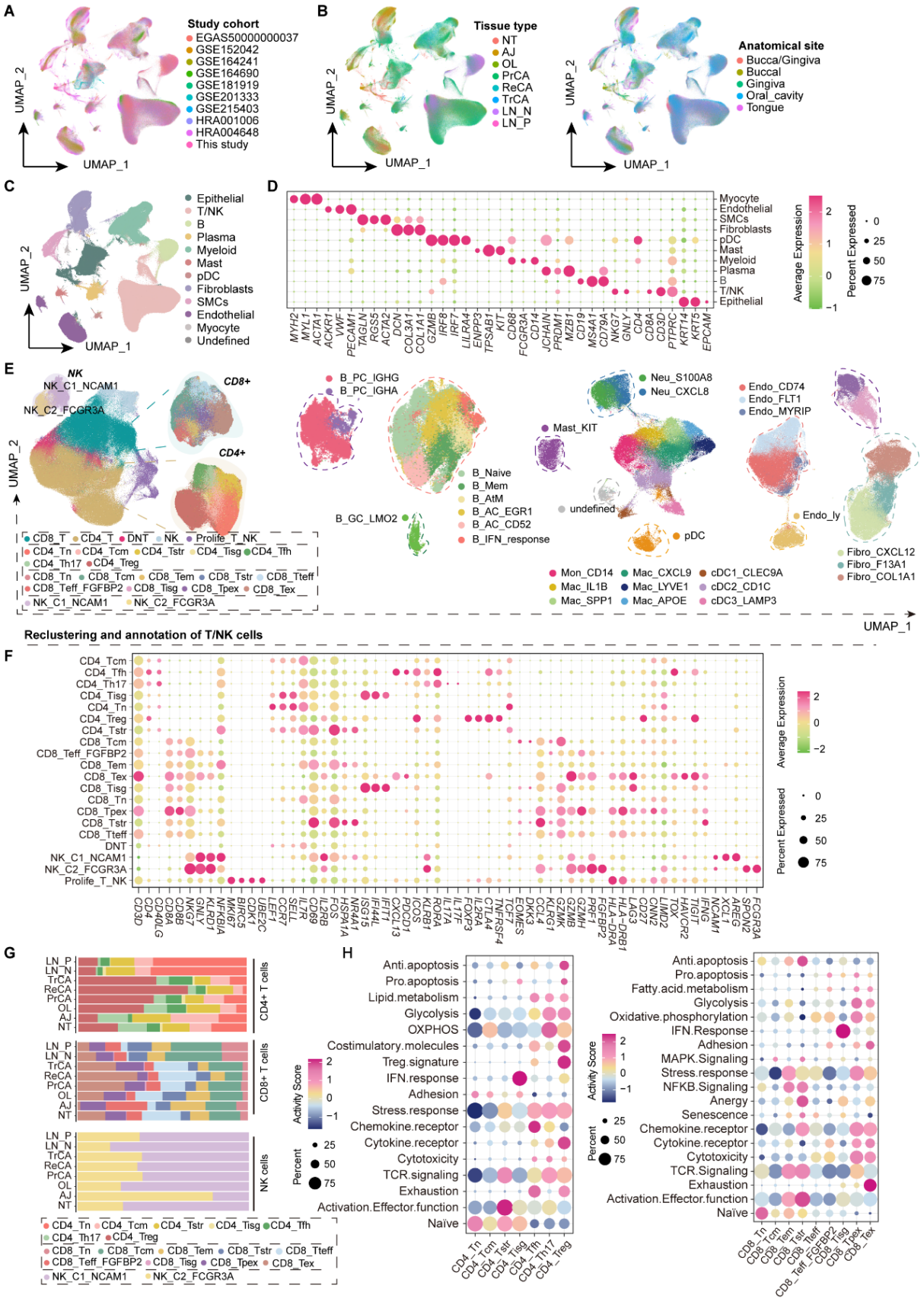

Figure S1. Constructing a single-cell atlas of oral carcinogenesis through multi-level cell annotation. Related to Figure 1.

(A) Integrated UMAP projection of ~1 million single-cell transcriptomes, colored by cohort origin (integrating 10 public datasets with this study).
(B) UMAP colored by tissue type (left) and anatomical site (right). PrCA: primary tumor; ReCA: recurrent tumor; TrCA: post-treatment tumor.
(C) UMAP colored by 11 major cell lineages.
(D) Dot plot of canonical marker expression across major lineages. Color: average expression (z-score); Size: percent of expressing cells.
(E) UMAPs displaying the lineage-specific re-clustering and annotation of immune and stromal populations.
(F) Dot plot of marker genes for annotated T and NK cell subsets.
(G) Stacked bar plots showing the proportional composition of $CD4^+$ T, $CD8^+$ T, and NK cell subsets across tissue states.
(H) Functional activity scores of $CD4^+$ T (left) and $CD8^+$ T (right) subsets.

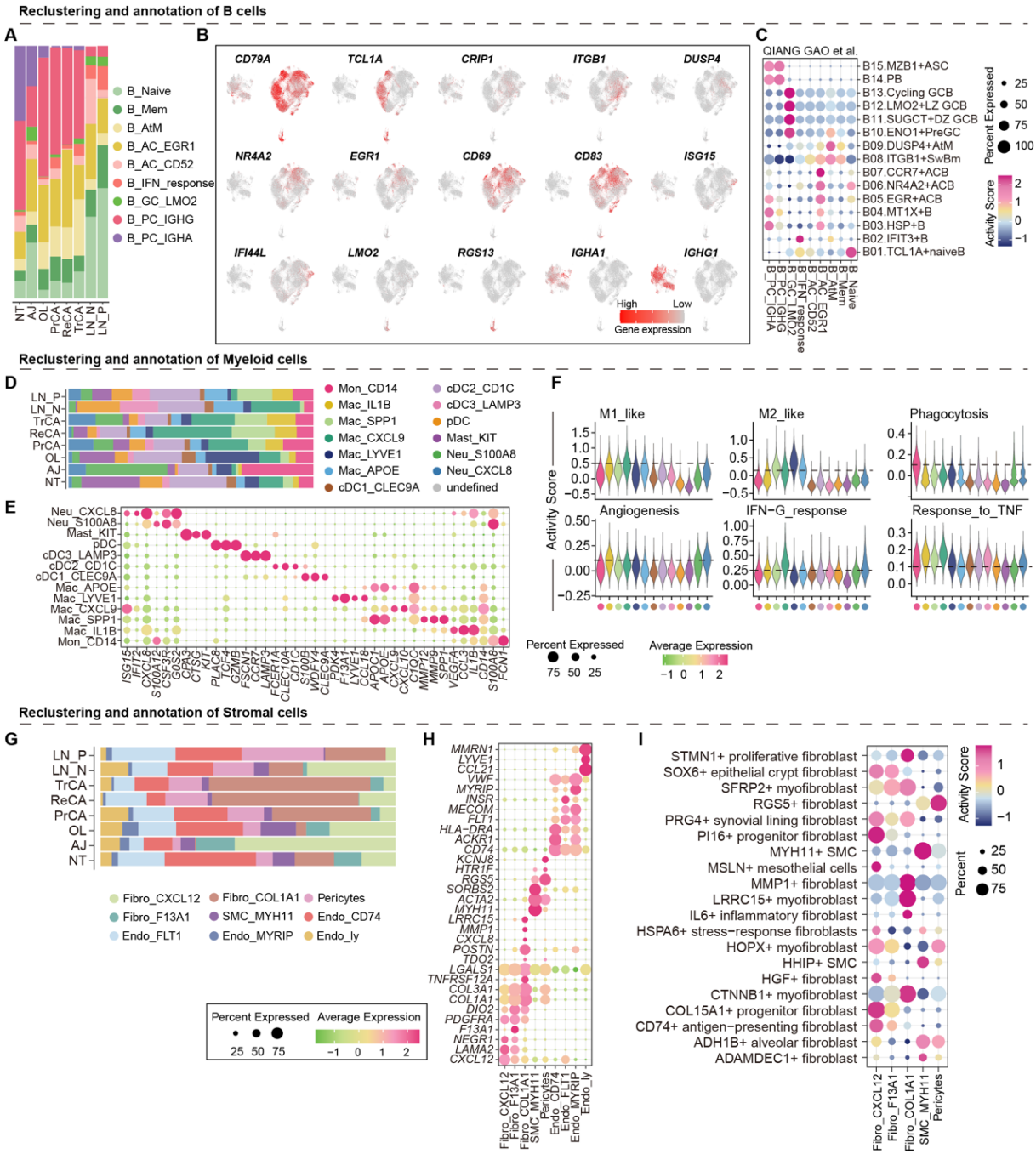

Figure S2. Reclustering and functional annotation of B cell, myeloid, and stromal subsets. Related to Figure 1.

(A) Proportional distribution of B cell subpopulations across tissue states.

(B) Feature plots of representative markers for B cell subsets.

(C) Canonical B cell signature scores across subsets.

(D) Proportional distribution of myeloid cell subpopulations across tissue states.

(E) Dot plot of key markers for myeloid subpopulations.

(F) Functional myeloid signature scores across subsets.

(G) Proportional distribution of stromal subsets across tissue states.

(H) Dot plot of markers for stromal subsets.

(I) Functional fibroblast signature scores across subsets.

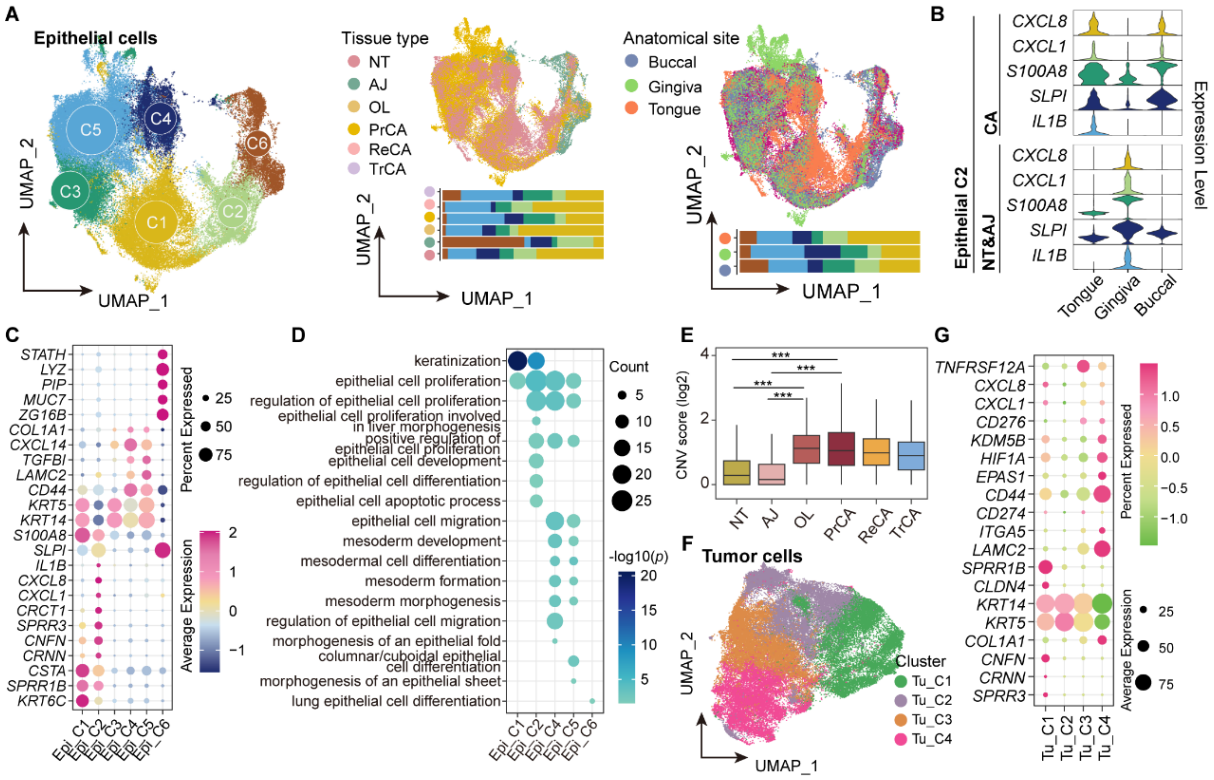

Figure S3. Molecular and functional characterization of epithelial and tumor cell subpopulations. Related to Figure 1.

(A) UMAP of all epithelial cells colored by subcluster (C1-C6), tissue type, and anatomical site. Bar plots quantify composition.

(B) Violin plots of markers in the C2 epithelial subset across anatomical sites and tissues.

(C) Dot plot of markers across epithelial subpopulations (C1-C6).

(D) Enrichment of GO biological processes in epithelial subclusters.

(E) Box plots comparing inferred single-cell CNV scores of epithelial cells across tissue types. ***$p < 0.001$, Wilcoxon test.

(F) UMAP of malignant epithelial (tumor) cells, colored by functional subcluster (Tu_C1-Tu_C4).

(G) Dot plot of immune-related and epithelial lineage markers across tumor cell subclusters (Tu_C1-Tu_C4).

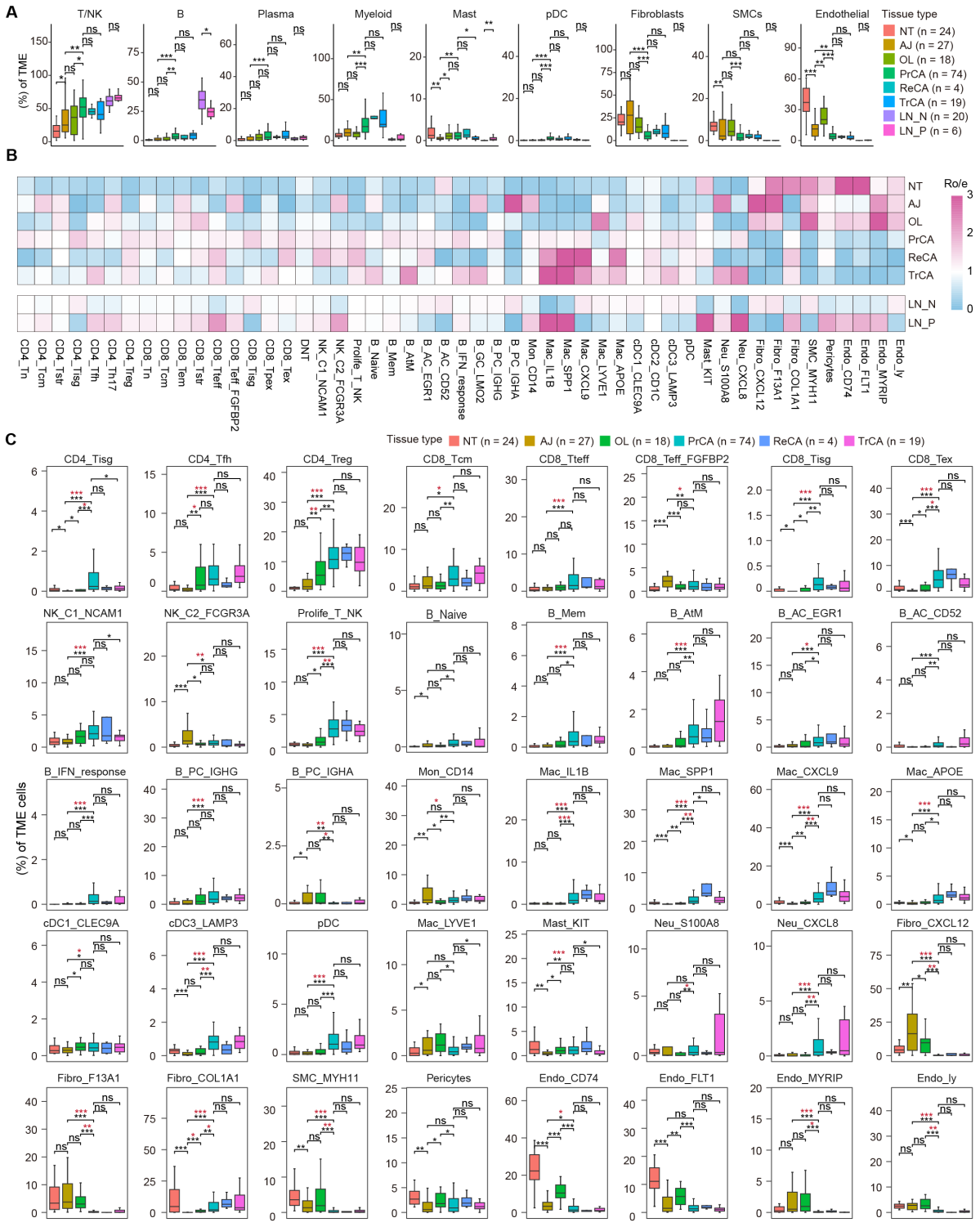

Figure S4. Comparative analysis of cell composition dynamics. Related to Figure 1.

(A) Relative abundance of major cell lineages across tissue types (% of total cells). *p < 0.05, **p < 0.01, ***p < 0.001, ns = not significant, Wilcoxon test.

(B) Heatmap showing the relative enrichment of minor cell subsets across tissue states. Ro/e: ratio of observed to expected cell numbers.

(C) Box plots detailing proportional differences in key minor cell populations across the AJ-OL-CA continuum. *p < 0.05, **p < 0.01, ***p < 0.001, ns = not significant; Wilcoxon test. Red asterisk indicates significance in paired Wilcoxon test.

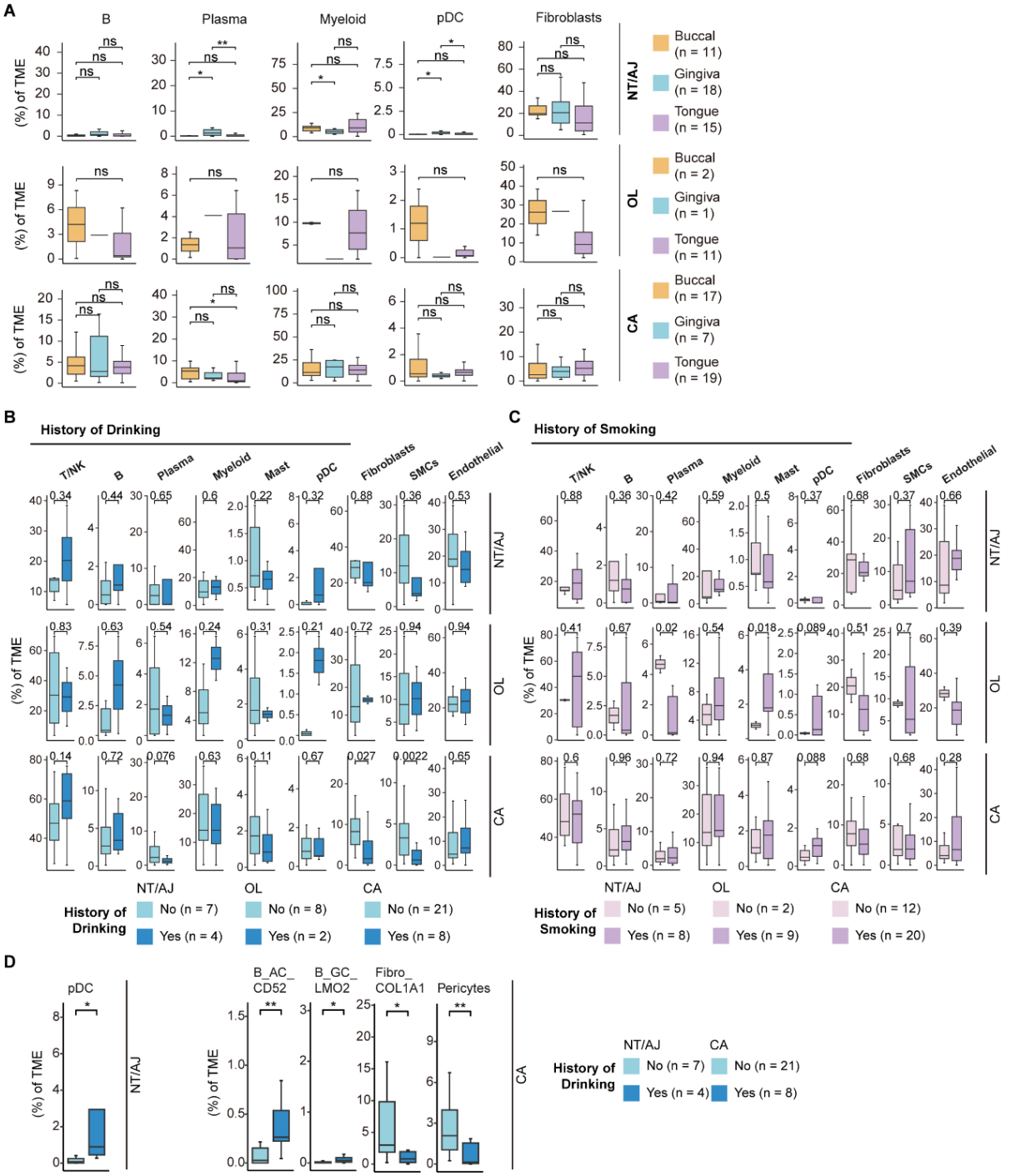

Figure S5. Impact of anatomical sites and risk factors on tissue composition. Related to Figure 1.

(A) Comparative analysis of major immune and stromal cell proportions across anatomical sites by tissue type. *p < 0.05, **p < 0.01, ns = not significant; Wilcoxon test.

(B-C) Box plots comparing major cell abundance in individuals with vs. without a history of alcohol consumption (B) or smoking (C), stratified by tissue type. P-values from Wilcoxon test.

(D) Box plots showing significant differences in selected minor cell subpopulations between drinkers and non-drinkers. *p < 0.05, **p < 0.01; Wilcoxon test.

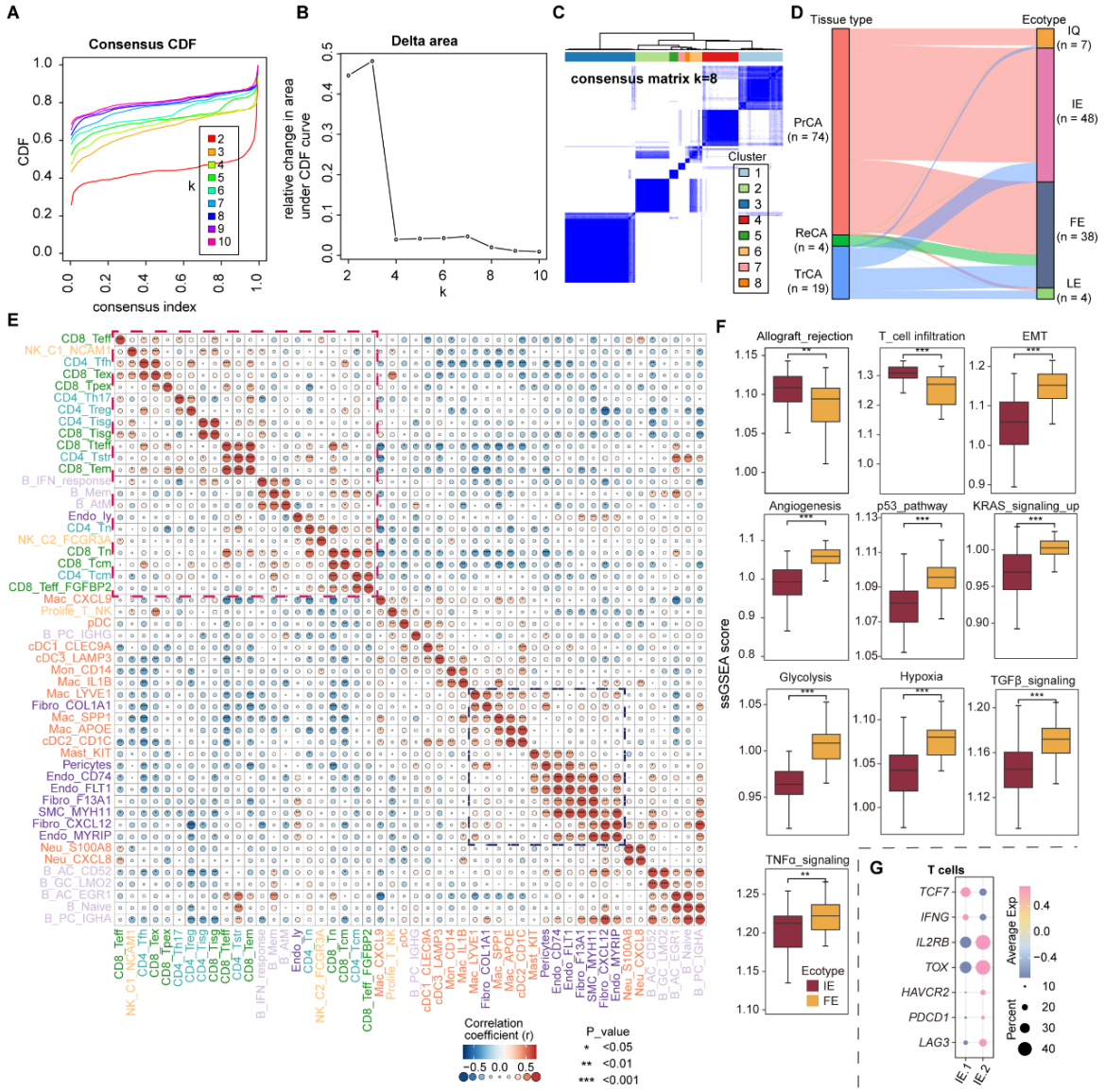

Figure S6. Clustering and functional profiling of ecotypes. Related to Figure 2.

(A) Consensus cumulative distribution function (CDF) plot for cluster numbers k = 2 to 10. The area under the CDF curve reflects relative consensus and supports cluster number determination.

(B) Delta area plot quantifying the relative change in area under the CDF curve. The plateau indicates k = 8 as the optimal cluster number.

(C) Consensus matrix heatmap for k = 8, showing robust cluster separation.

(D) Sankey diagram mapping PrCA, ReCA, and TrCA tumor samples to the identified ecotypes.

(E) Pearson correlation matrix of relative abundance of cell subpopulations. Unsupervised clustering highlighted two co-associated modules: lymphocytes (purple) and stromal cells, along with several macrophage subpopulations (blue).

(F) Box plots of hallmark pathway activity scores (ssGSEA) comparing IE and FE ecotypes. **p < 0.01, ***p < 0.001; Wilcoxon test.

(G) Dot plot comparing immune checkpoint and exhaustion gene expression between T cells from the inflamed (IE.1) and exhausted (IE.2) states of the IE ecotype.

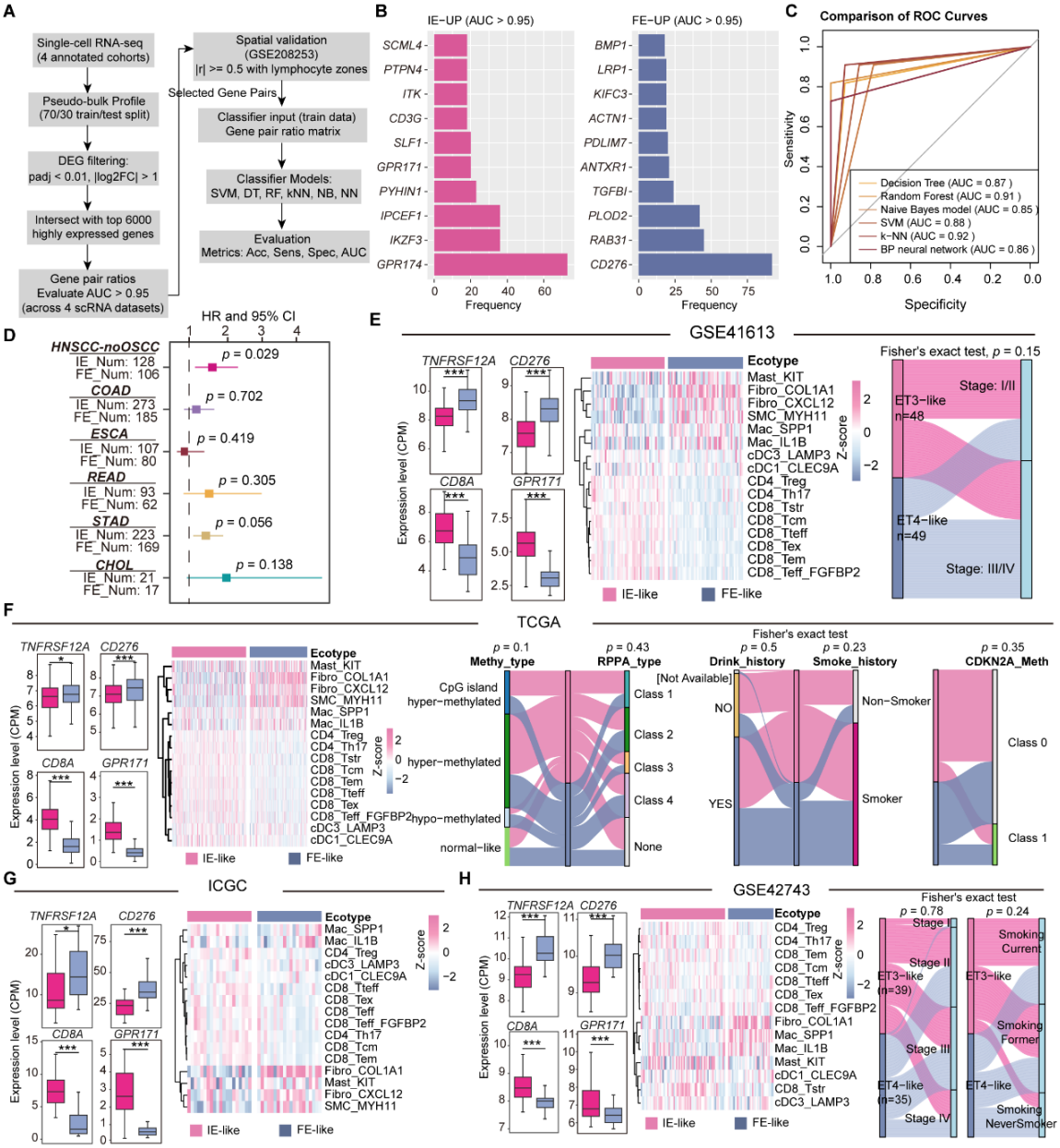

Figure S7. Classifier development and validation of tumor ecotypes. Related to Figure 3.

(A) Flowchart for the ecotype classifier development using pseudo-bulk profiles.

(B) Feature robustness plot, measured by the frequency of a gene appearing in high-performing gene pairs (AUC > 0.95 across four datasets).

(C) ROC curves benchmarking six machine learning models.

(D) Forest plot showing impact of IE/FE ecotypes on overall survival across diverse gastrointestinal cancers (TCGA). HNSCC: head and neck squamous cell carcinoma; COAD: colon adenocarcinoma; ESCA: esophageal carcinoma; READ: rectal adenocarcinoma; STAD: stomach adenocarcinoma; CHOL: cholangiocarcinoma. P-values from Cox proportional hazards regression.

(E-H) Validation in four independent OSCC cohorts. Panels show: marker expression (left); heatmap of ecotype-defining cell abundance (middle); Sankey diagram of clinical characteristics (Right). Significance: Wilcoxon test (*p < 0.1, ***p < 0.001, left) and Fisher's exact test (right).

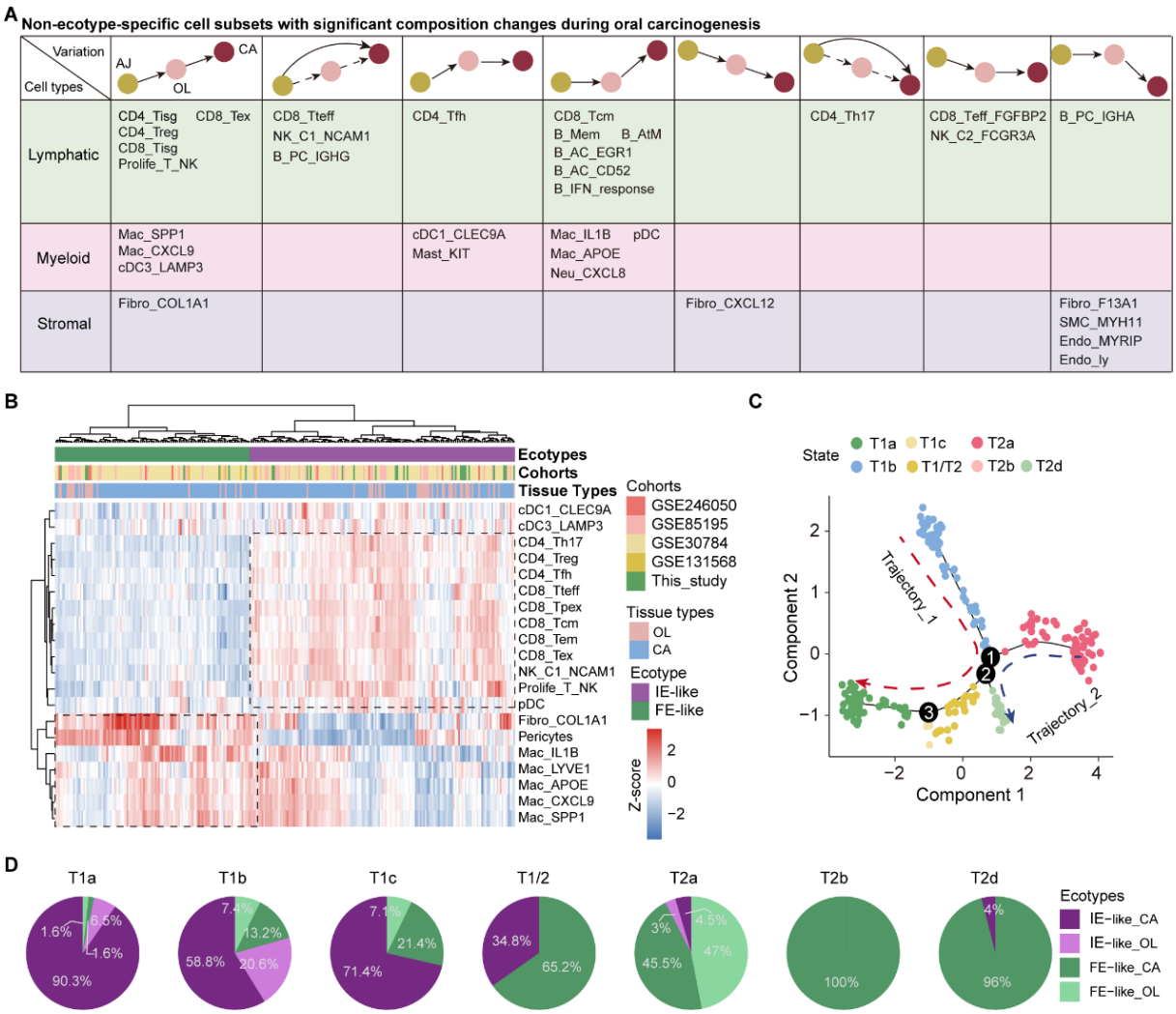

Figure S8. Immune and stromal cell dynamics during oral carcinogenesis. Related to Figure 4.

(A) Summary of non-ecotype-specific cell population shifts across the AJ-OL-CA continuum.

(B) Hierarchical clustering of immune/stromal cell relative abundance in OL and CA samples, revealing that OL lesions already cluster into IE-like and FE-like groups.

(C) Pseudotime trajectory of OL and CA samples based on inferred TME composition. Each dot represents one sample, colored by inferred pseudotime state.

(D) Proportion of IE-like and FE-like samples marked by OL or CA type within each pseudotime state.

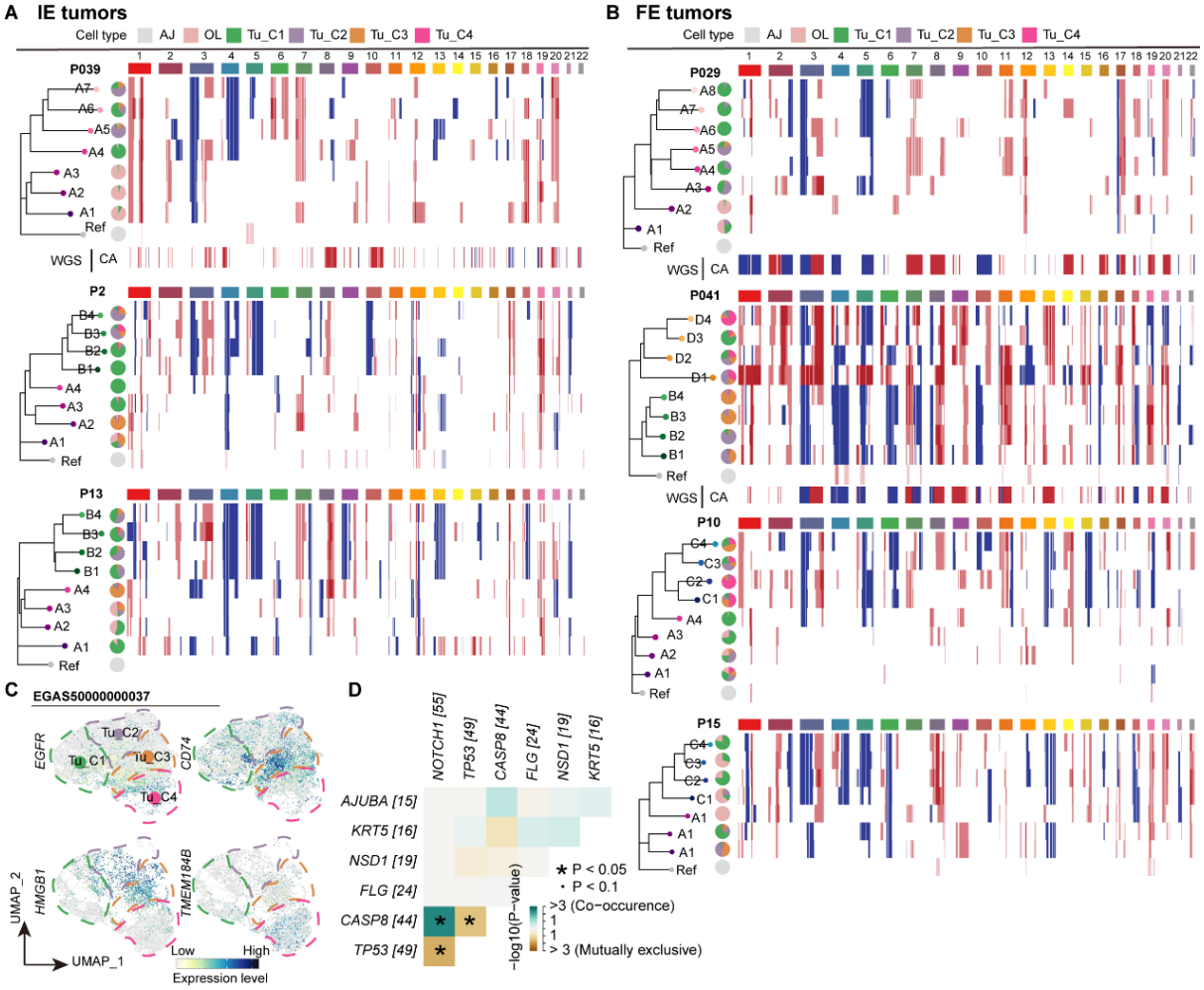

Figure S9. Clonal architecture and genomic instability across ecotypes during oral tumorigenesis. Related to Figure 5.

(A-B) Single-cell CNV profiles and clonal phylogenies for representative IE (A) and FE (B) tumors. Pie charts show clonal composition across tissue types. Matched WGS CNV profiles are shown below when available.

(C) UMAPs showing expression of ecotype-specific drivers (*EGFR*, *CD74*, *HMGB1*, *TMEM184B*) in tumor cells from an independent dataset.

(D) Co-occurrence analysis of somatic mutations.

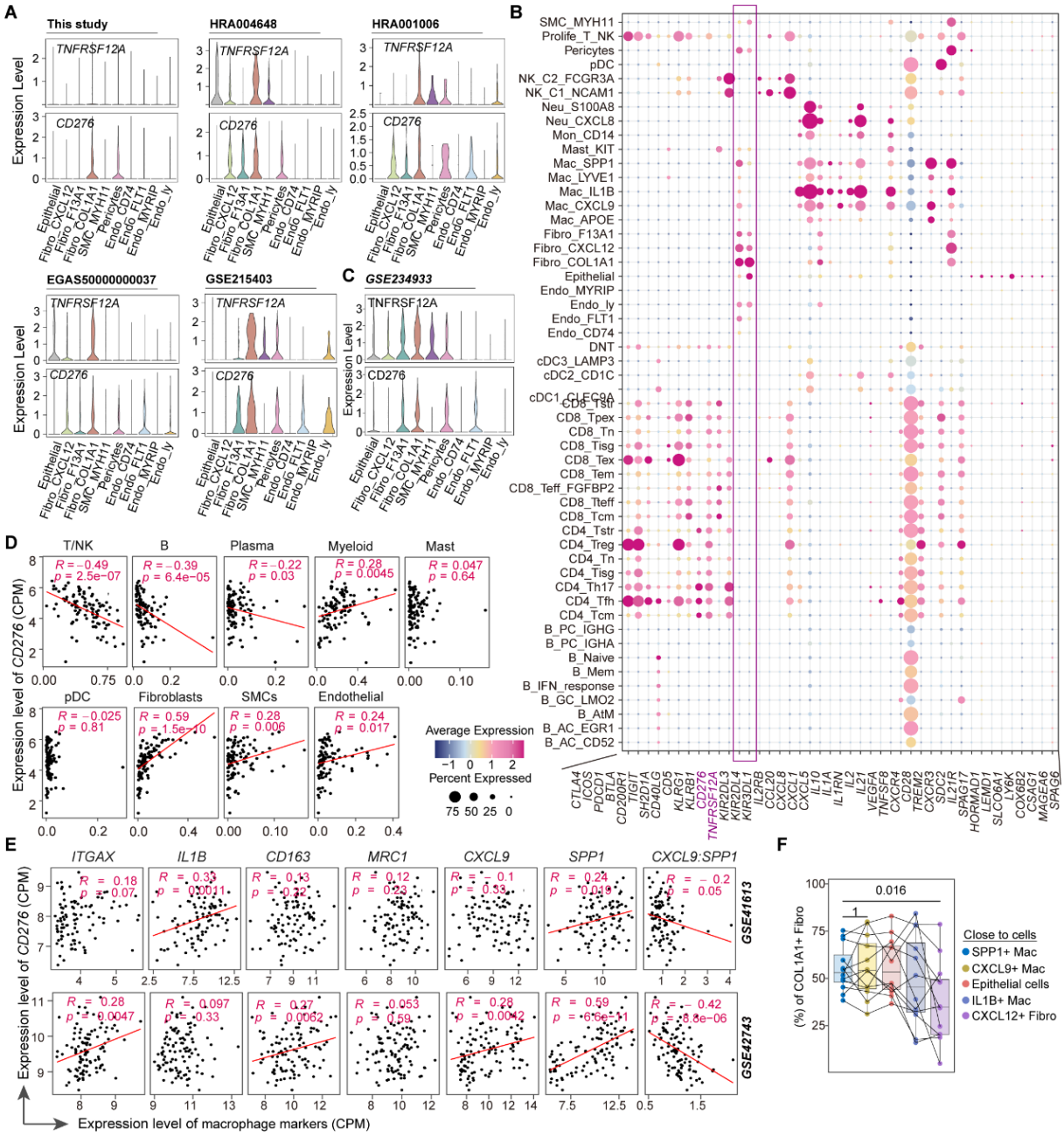

Figure S10. Immunoregulatory gene expression landscape. Related to Figure 6.

(A) Violin plots of *TNFRSF12A* and *CD276* expression across epithelial and stromal cell subpopulations in multiple datasets.

(B) Dot plot of selected immunomodulatory genes across cell subpopulations.

(C) Validation of *TNFRSF12A* and *CD276* expression patterns in an independent scRNA-seq dataset (GSE234933).

(D) Pearson correlation between *CD276* expression and immune/stromal cell abundance.

(E) Pearson correlation between *CD276* expression and macrophage markers.

(F) Spatial proximity analysis showing the proportion of *COL1A1*$^{+}$ fibroblasts in proximity to target cell subpopulations (n = 12). Significance: Wilcoxon test.

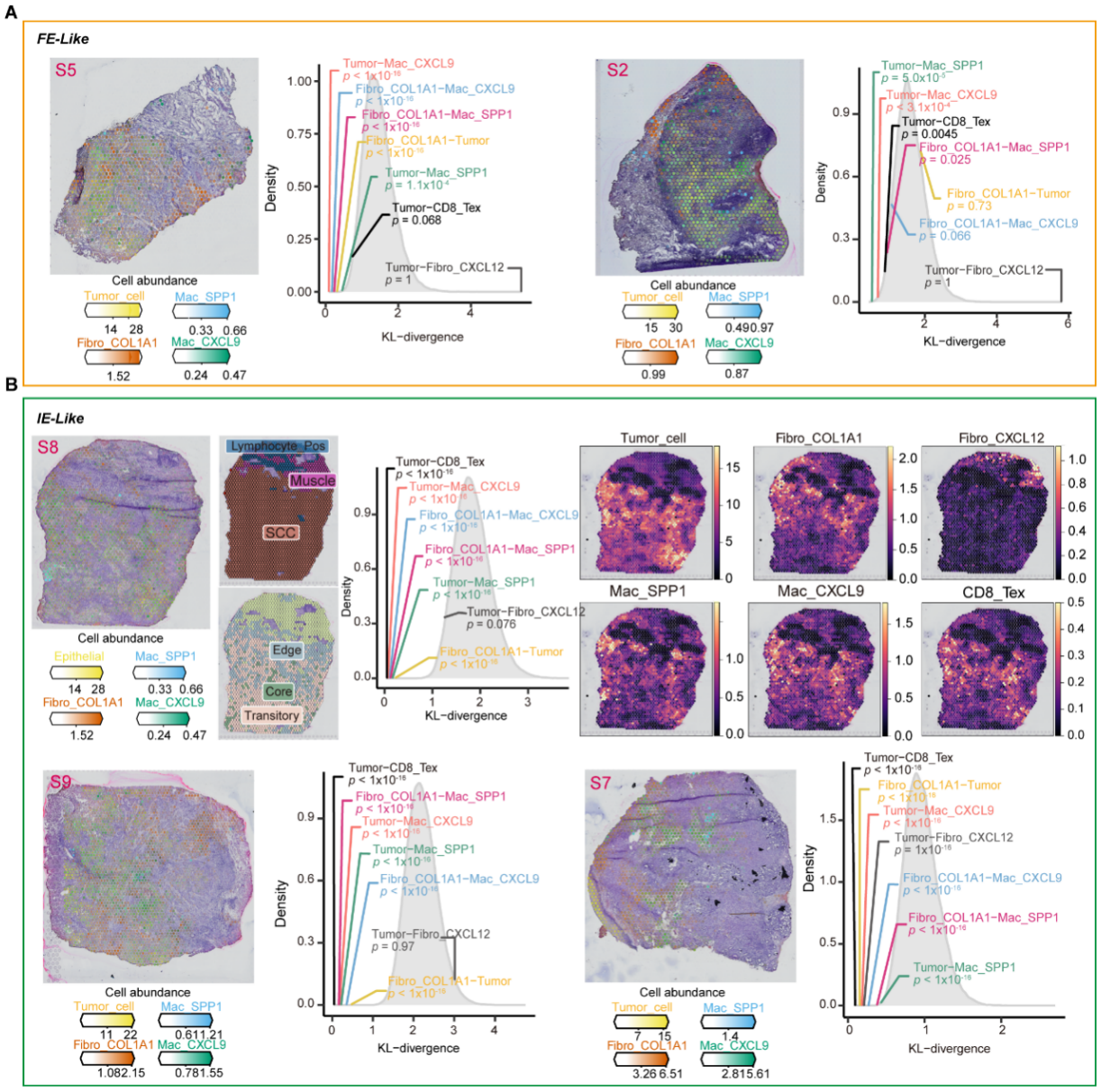

Figure S11. Spatial colocalization analysis. Related to Figure 6.

(A) Spatial co-localization analysis for FE-like samples. Left: Predicted spatial abundance (cell2location). Right: Density plot showing the KL divergence between cell subpopulations compared to a null distribution ($p < 0.05$ indicates significance).

(B) Same analysis for representative IE-like samples.

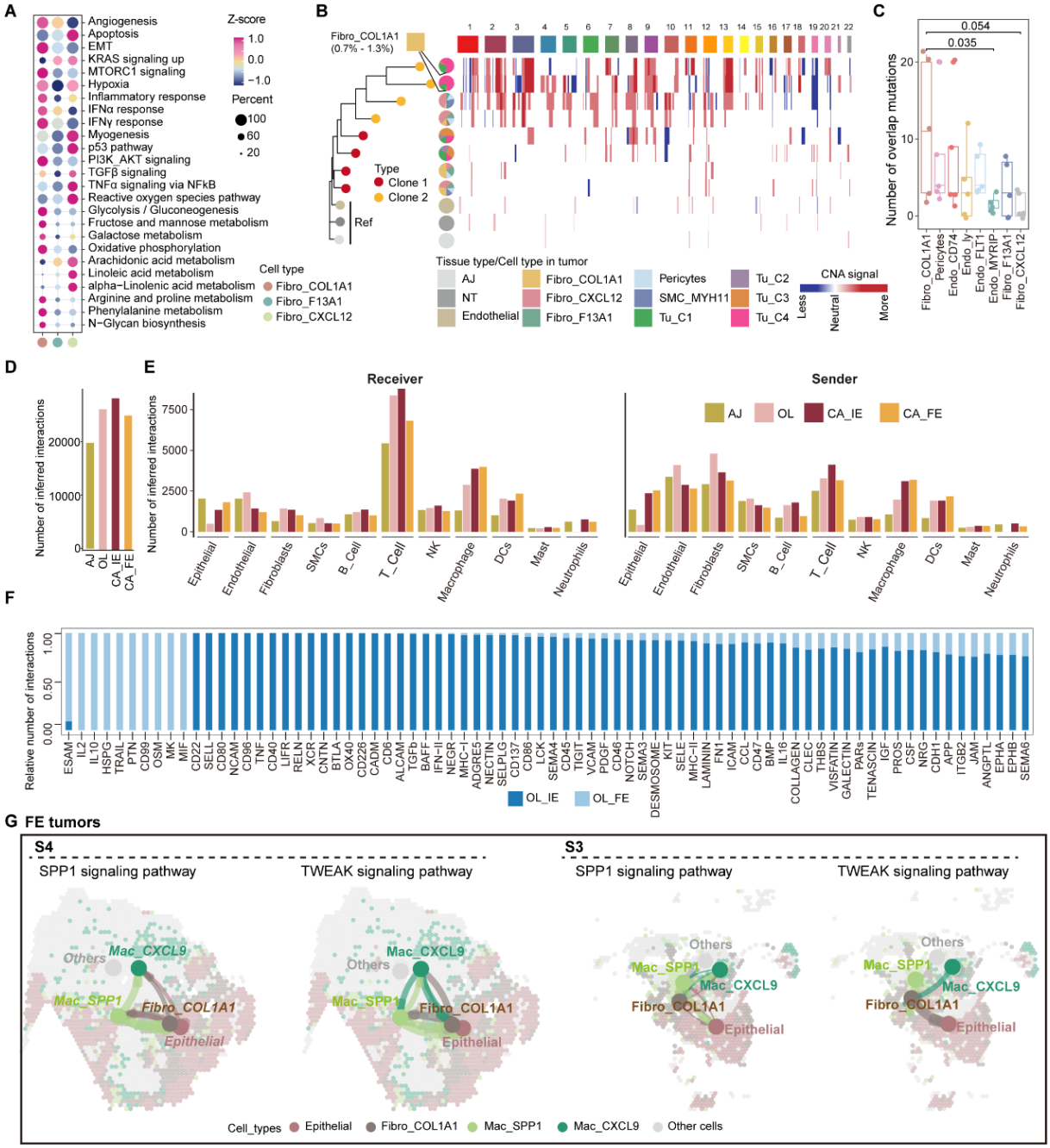

Figure S12. Functional features of stromal populations and signaling networks. Related to Figure 7.

(A) Dot plot of functional gene set activity across fibroblast subsets.

(B) Inferred single-cell CNV profiles and phylogenies for tumor and fibroblast cells. Pie charts depict the cellular composition of subclones across distinct tissue states.

(C) Boxplot showing the number of intersected somatic mutation between fibroblast subtypes and tumor cells. Statistical significance was determined using one-sided Wilcoxon rank-sum test.

(D) Quantification of total cell-cell communications across AJ, OL, IE-CA, and FE-CA samples.

(E) Quantification of communications among immune and stromal populations.

(F) Pathway-level comparison of ecotype-specific L-R signals for OL tissue.

(G) Spatial maps highlighting ecotype-enriched macrophage–fibroblast-tumor cell signaling pathways (SPP1 and TWEAK) in representative OSCC samples. The size of the line reflects the signal strength; the arrow indicates the predicted direction of interaction.

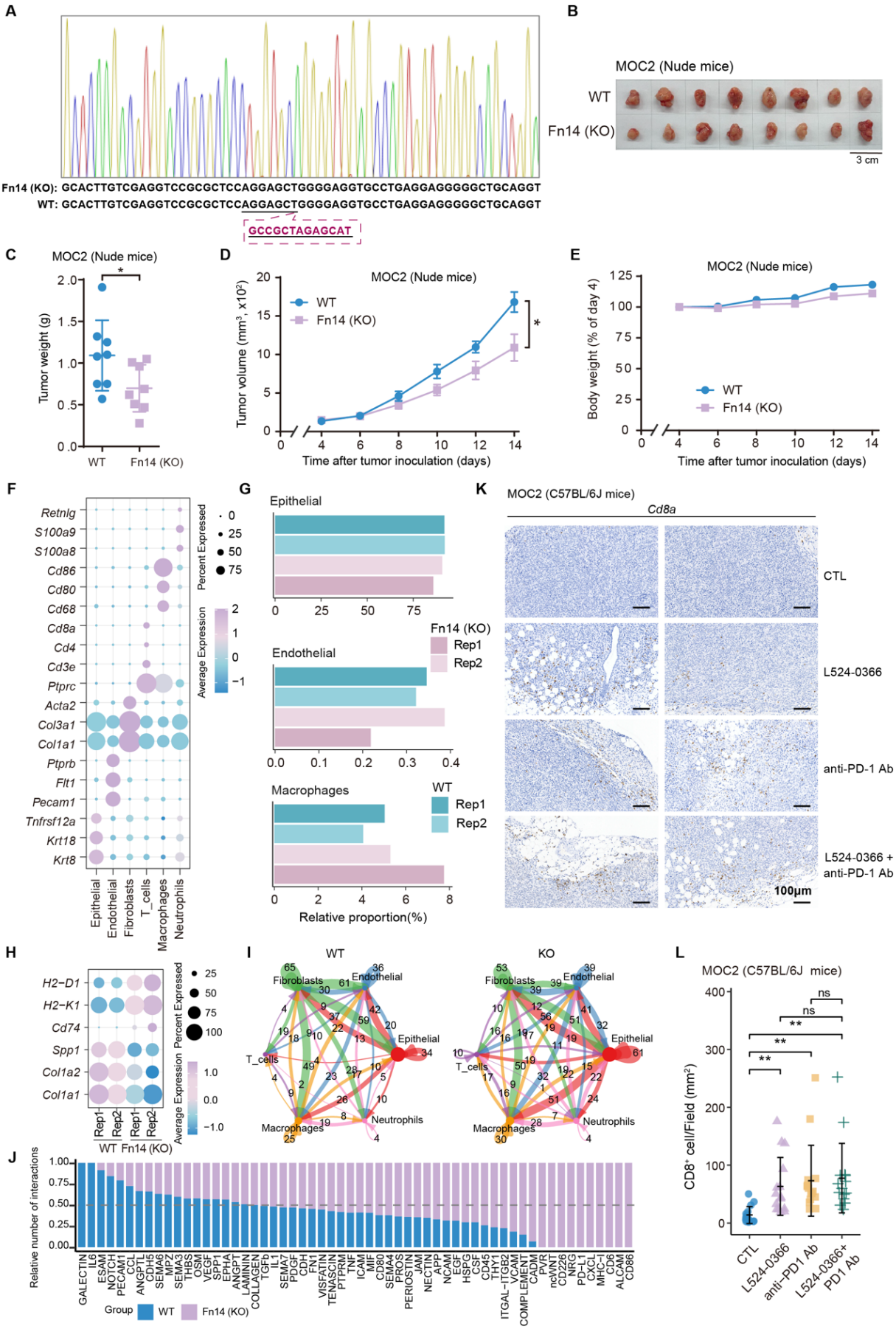
A
Fn14 (KO): GCACTTGTCGAGGTCCGCGCTCCAGGAGCTGGGGAGGTGCCTGAGGAGGGGGCTGCAGGT
WT: GCACTTGTCGAGGTCCGCGCTCCAGGAGCTGGGGAGGTGCCTGAGGAGGGGGCTGCAGGT
GCCGCTAGAGCAT
B
MOC2 (Nude mice)
WT
Fn14 (KO)
3 cm
C
MOC2 (Nude mice)
Tumor weight (g)
WT
Fn14 (KO)
D
MOC2 (Nude mice)
Tumor volume (mm3, x10²)
WT
Fn14 (KO)
Time after tumor inoculation (days)
E
MOC2 (Nude mice)
Body weight (% of day 4)
WT
Fn14 (KO)
Time after tumor inoculation (days)
F
Retnlg
S100a9
S100a8
Cd86
Cd80
Cd68
Cd8a
Cd4
Cd3e
Ptprc
Acta2
Col3a1
Col1a1
Ptprb
Flt1
Pecam1
Tnfrsf12a
Krt18
Krt8
Epithelial
Endothelial
Fibroblasts
T_cells
Macrophages
Neutrophils
Percent Expressed
Average Expression
G
Epithelial
Endothelial
Macrophages
Fn14 (KO)
Rep1
Rep2
WT
Rep1
Rep2
Relative proportion(%)
K
MOC2 (C57BL/6J mice)
Cd8a
CTL
L524-0366
anti-PD-1 Ab
L524-0366 +
anti-PD-1 Ab
100μm
H
H2−D1
H2−K1
Cd74
Spp1
Col1a2
Col1a1
Rep1
Rep2
Rep1
Rep2
WT
Fn14 (KO)
Percent Expressed
Average Expression
I
WT
KO
Fibroblasts
Endothelial
Epithelial
T_cells
Macrophages
Neutrophils
L
MOC2 (C57BL/6J mice)
CD8⁺ cell/Field (mm²)
CTL
L524-0366
anti−PD1 Ab
L524-0366+
PD1 Ab
ns
**
J
Relative number of interactions
Group
WT
Fn14 (KO)
GALECTIN
IL6
ESAM
NOTCH
PECAM1
CCL
ANGPTL
CDH5
SEMA6
MK
SEMA3
THBS
OSM
VEGF
SPP1
EPHA
ANGPT
LAMININ
COLLAGEN
TGFb
IL1
SEMA7
PDGF
FN1
VISFATIN
TENASCIN
PTPRM
TNF
ICAM
MIF
CD80
SEMA4
PROS
PERIOSTIN
JAM
NECTIN
APP
NCAM
EGF
HSPG
CSF
CD45
THY1
ITGAL−ITGB2
VCAM
COMPLEMENT
CADM
PVR
ncWNT
CD226
PD−L1
CXCL
MHC−I
CD6
ALCAM
CD86

Figure S13. Impact of Fn14 knockout *in vivo*. Related to Figure 8.

(A) Sanger sequencing confirmation of the Fn14 knockout in MOC2 cells. gRNA target underlined. Red indicates deleted sequences.

(B-C) Images (B) and weights (C) of tumors from nude mice inoculated with WT or Fn14-KO cells. *P < 0.05; Wilcoxon test.

(D-E) Tumor volume (D) and body weight (E) over time in the nude mouse model. *P < 0.05; Wilcoxon test. Data shown as mean ± s.e.m. N = 4 mice per group.

(F) Dot plot of marker gene expression in major cell types from murine tumor scRNA-seq (C57BL/6J model).

(G) Bar plots comparing the relative abundance of epithelial, endothelial, and macrophage populations in WT vs. Fn14-KO tumors.

(H) Dot plot of selected gene expression in fibroblasts.

(I) Cell-cell communication networks in WT (left) vs. Fn14-KO (right) tumors. Node size: cell abundance; edge width: number of significant ligand-receptor interactions.

(J) Pathway-level comparison of group-specific L–R signaling activity.

(K) IHC staining of $CD8^+$ T cells in tumors treated with control, L524-0366, anti-PD-1, or combination.

(L) Quantification of $CD8^+$ T cell density (cells/mm^2) across treatment groups. **P < 0.01, ns = not significant; Wilcoxon test. N = 15 fields per group.