Journal of Biomedical Research

Integrating bulk and single-cell sequencing reveals cellular heterogeneity between lung adenocarcinoma in smokers and never-smokers

Zihuan Zhao, Pan Yang, Yuhan Liu, Kai Wang, Xianfeng Xu, Yuzhuo Wang, Meng Zhu, Na Qin, Cheng Wang, Weimin Li, Hongxia Ma, Zhoufeng Wang, Hongbing Shen

Cite this article as:
Zihuan Zhao, Pan Yang, Yuhan Liu, Kai Wang, Xianfeng Xu, Yuzhuo Wang, Meng Zhu, Na Qin, Cheng Wang, Weimin Li, Hongxia Ma, Zhoufeng Wang, Hongbing Shen. Integrating bulk and single-cell sequencing reveals cellular heterogeneity between lung adenocarcinoma in smokers and never-smokers[J]. *Journal of Biomedical Research*, 2026, 40(3): 247-265. doi: 10.7555/JBR.39.20250160

View online: https://doi.org/10.7555/JBR.39.20250160

Articles you may be interested in

Association between soft drink consumption types and risk of lung cancer and all-cancer: A prospective study of PLCO data
The Journal of Biomedical Research. 2022, 36(6): 390 https://doi.org/10.7555/JBR.36.20220135

Characteristics of alveolar macrophages in bronchioalveolar lavage fluids from active tuberculosis patients identified by single-cell RNA sequencing
The Journal of Biomedical Research. 2022, 36(3): 167 https://doi.org/10.7555/JBR.36.20220007

A novel synthesized prodrug of gemcitabine based on oxygen-free radical sensitivity inhibited the growth of lung cancer cells
The Journal of Biomedical Research. 2023, 37(5): 355 https://doi.org/10.7555/JBR.37.20230022

GSH-responsive curcumin/doxorubicin encapsulated Bactrian camel serum albumin nanocomposites with synergistic effect against lung cancer cells
The Journal of Biomedical Research. 2020, 34(1): 54 https://doi.org/10.7555/JBR.33.20190036

A SARS-CoV-2 neutralizing antibody discovery by single cell sequencing and molecular modeling
The Journal of Biomedical Research. 2023, 37(3): 166 https://doi.org/10.7555/JBR.36.20220221

Single-cell omics analyses with single molecular detection: challenges and perspectives
The Journal of Biomedical Research. 2021, 35(4): 264 https://doi.org/10.7555/JBR.35.20210026

Available online at www.jbr-pub.org.cn

Open Access at PubMed Central

Journal of Biomedical Research, 2026 40(3): 247–265

Special Section On Cancer Research

JBR

Integrating bulk and single-cell sequencing reveals cellular heterogeneity between lung adenocarcinoma in smokers and never-smokers

Zihuan Zhao[1,△], Pan Yang[2,△], Yuhan Liu[1,△], Kai Wang[1,3], Xianfeng Xu[1], Yuzhuo Wang[1], Meng Zhu[2], Na Qin[1], Cheng Wang[1], Weimin Li[2], Hongxia Ma[1,✉], Zhoufeng Wang[2,✉], Hongbing Shen[1,4,✉]

[1]*Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 211166, China;*
[2]*Institute of Respiratory Health, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China;*
[3]*Department of Epidemiology, School of Public Health, Southeast University, Nanjing, Jiangsu 210096, China;*
[4]*Research Units of Cohort Study on Cardiovascular Diseases and Cancers, Chinese Academy of Medical Sciences, Beijing 100037, China.*

Abstract

Lung cancer in smokers (LCIS) and lung cancer in never-smokers (LCINS) are different entities with distinct molecular features. However, their cellular heterogeneity still requires further investigation. Through an integrated analysis of single-cell RNA sequencing and bulk sequencing data, we identified cell subpopulations associated with smoking and non-smoking patients. Subsequent transcriptomic analyses were performed to elucidate differences in cellular functions and the tumor microenvironment. We observed that smoking-associated cancer cells exhibited a higher degree of aggressiveness, which may correlate with an adverse prognosis in smoking patients. Additionally, immunosuppressive *CXCL10*$^{+}$ macrophages may contribute to tumorigenesis in smokers, and the immunoregulatory LGALS9-HAVCR2 axis could be a potential immunotherapeutic target. In non-smokers, the inflammatory microenvironment may be involved in tumor development. Moreover, the reduced anti-tumor cytotoxicity may be associated with their suboptimal immunotherapeutic response. Our study uncovered differences in oncogenic and immune escape mechanisms between LCIS and LCINS patients and suggests potential immunotherapeutic strategies.

Keywords: lung cancer in smokers, lung cancer in never-smokers, single-cell RNA sequencing, cellular heterogeneity

Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide[1]. Despite the strong association between lung cancer development and cigarette smoke, nearly 25% of lung cancer cases,

△These authors contributed equally to this work.

✉Corresponding authors: Hongxia Ma, E-mail: hongxiama@njmu.edu.cn, ORCID: 0000-0002-9821-6955; Zhoufeng Wang, E-mail: wangzhoufeng@scu.edu.cn, ORCID: 0000-0002-7881-1837; Hongbing Shen, E-mail: hbshen@njmu.edu.cn, ORCID: 0000-0002-2581-5906.

Received: 15 April 2025; Revised: 18 August 2025; Accepted: 25 August 2025; Available online: 25 August 2025; Published date: 21 May 2026

CLC number: R734, Document code: A

The corresponding author Hongbing Shen was a former editor-in-chief of this journal but had no involvement in the editorial handling or peer review process of this manuscript. The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

© 2026 by Journal of Biomedical Research.

https://doi.org/10.7555/JBR.39.20250160

predominantly lung adenocarcinoma (LUAD), occur in individuals who have never smoked[2]. With the decline in smoking rates, the proportion of lung cancer in never-smokers (LCINS) has increased and was estimated to be the fifth leading cause of cancer-related deaths in 2023[3–4]. LCINS preferentially affects females and individuals of East Asian descent[3], and it exhibits a comparatively better prognosis than lung cancer in smokers (LCIS)[5]. Although several risk factors have been found to be associated with lung cancer in lifelong non-smokers, including radon exposure, hormonal influences, and genetic factors[2], the exact pathogenesis of LCINS remains unclear.

Accumulating evidence has shown that LCIS and LCINS exhibit distinct genomic and transcriptomic features[3]. Previous studies have shown that non-smoking patients have a lower tumor mutational burden than smoking patients[6]. Moreover, 78%–92% of non-smoking patients carry driver gene alterations, while only 49.5% of smoking-related LUAD patients harbor driver gene mutations[6]. Specifically, *EGFR* is the most common driver gene mutation in LCINS, with a prevalence ranging from 60% to 74% among non-smoking female lung cancer patients of East Asian ethnicity[7]. *TP53* and *KRAS* driver gene mutations are more prevalent in smoking patients[8]. Transcriptomic studies have demonstrated distinct gene expression profiles and functional pathway activities between smokers and non-smokers. Smoking patients display increased expression of cell cycle-related genes and enrichment of the *p53* pathway, whereas non-smokers and long-term former smokers show lower stemness scores[9]. However, these findings are primarily derived from bulk sequencing, which only reflects averaged gene expression, thereby making it challenging to understand intratumoral heterogeneity[10].

The advent of single-cell RNA sequencing (scRNA-seq) technology has enabled the comparison of cellular composition across different phenotypes, the identification of cell subpopulations associated with specific phenotypes, and the exploration of cellular interaction differences[11–12]. For example, Martos *et al*[13] discovered a dysfunctional natural killer (NK)-like $CD8^+$ T cell subtype enriched in smokers, while Duclos *et al*[14] demonstrated that smoking might be linked to the loss of club cells and an increased number of goblet cells. A recent study comparing scRNA-seq data from smoking and non-smoking LUAD patients revealed that cancer cells in smoking patients exhibited greater aggressiveness, while the immunosuppressive tumor microenvironment might promote tumor growth in non-smoking patients[5].

However, these scRNA-seq studies typically involve small sample sizes, which limit the statistical power to correlate phenotypes with specific cell subtypes reliably. Recently, researchers have developed SCISSOR[11], a bioinformatics algorithm that leverages phenotype information from large bulk cohorts to identify cell subpopulations most relevant to disease phenotypes. By utilizing this algorithm, a previous study discovered that LUAD patients with the *EGFR* mutation were associated with dendritic cell (DC) infiltration, while those with *TP53* and *STK11* mutations exhibited higher $CD8^+$ T cell infiltration[15].

In the current study, to elucidate the cellular heterogeneity between smoking and non-smoking LUAD patients, we incorporated data from 140 LUAD patients who underwent bulk RNA sequencing from the Nanjing-Huaxi Lung Cancer Consortium (NHLCC) and 92 553 cells from 20 East Asian LUAD patients who underwent scRNA-seq. By integrating scRNA-seq and bulk sequencing data using the SCISSOR algorithm, we explored the cellular-level differences between smoking and non-smoking patients. Our results offer new insights into the distinct tumor microenvironment of smoking and non-smoking patients and provide potential treatment strategies.

Materials and methods

Study population

In our study, we used bulk RNA-sequencing data from 140 LUAD patients enrolled in the NHLCC cohort. Specifically, 78 patients were from the Nanjing Lung Cancer Cohort (NJLCC), and 62 were from West China Hospital. Each sample was assessed and examined by at least one pathologist to validate the histopathological attributes and tumor cell content, with only tumor samples containing a minimum of 70% tumor cells included. The study was approved by the Institutional Review Board of Nanjing Medical University (Nanjing, Jiangsu, China; project identification code: 2018.848) and West China Hospital (Chengdu, Sichuan, China; project identification code: 2017.114). The study was performed in accordance with the Declaration of Helsinki, and all participants provided written informed consent. The scRNA-seq data were downloaded from the Single-cell Lung Cancer Atlas (https://luca.icbi.at)[15]. A total of 92 553 primary LUAD tumor cells from 20 East Asian LUAD patients were extracted using the 10X Genomics platform. Baseline patient characteristics of the bulk sequencing

and scRNA-seq datasets are shown in ***Supplementary Tables 1*** and ***2***, respectively.

Bulk RNA sequencing data preprocessing

Total RNA was extracted using the Qiagen RNeasy kit, following the manufacturer's instructions. The standard Illumina RNA-seq protocol (http://support.illumina.com/training/online-courses/sequencing.html) was employed to produce transcription profiles of these samples, using the Illumina HiSeq 1500 platform. Subsequently, the raw RNA-seq reads were mapped to the GENCODE v19 genome assembly using STAR v2.4.1 with default parameters[16]. To mitigate technical variability, raw counts underwent ComBat-seq batch correction using the sva package (v3.38.0) with default parameters (group = NULL, covar_mod = NULL) to retain biological variation while adjusting for batch effects[17]. Before correction, counts were normalized to counts per million to account for sequencing depth differences. Bulk RNA-seq gene expression data and clinical annotations of The Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) cohort were retrieved from UCSC Xena (https://xena.ucsc.edu/), and the survival information was obtained from previous research[18]. A total of 498 patients with complete tumor staging and survival data were retained for downstream analysis.

Processing and annotation of scRNA data

The Seurat package (v4.0.1) was used for the scRNA-seq data processing[19]. All ribosomal genes were excluded from subsequent analyses. Raw counts were normalized using the 'NormalizeData' function. Because the scRNA-seq data came from two studies, the highly variable genes in each dataset were determined using the 'FindVariableFeatures' function, and the common highly variable genes were identified using the 'SelectIntegrationFeatures' function. Batch effects were addressed through Seurat's reciprocal principal component analysis-based integration workflow, which generated batch-corrected expression matrices by identifying integration anchors across batches using the 'FindIntegrationAnchors' function and then merging them using the 'IntegrateData' function. Batch correction was not applied to cancer cells to avoid masking biological variations. Principal component analysis, uniform manifold approximation and projection (UMAP) visualization, neighborhood analysis, and clustering analysis were performed using the 'RunPCA', 'RunUMAP', 'FindNeighbors', and 'FindClusters' functions from the Seurat package, respectively[19]. The identities of each cluster of major immune cell types were further annotated using marker genes identified by the 'FindAllMarkers' function.

Identification of smoking-associated and non-smoking-associated cell subgroups

The SCISSOR algorithm was applied to integrate bulk RNA-seq and scRNA-seq data to identify smoking-associated (Scissor_S) and non-smoking-associated (Scissor_NS) cell subpopulations in LUAD[11]. Briefly, a correlation matrix was first computed between each bulk sample and single-cell sample, based on common gene expression using Pearson correlation. A hybrid approach, combining L1-regularization (Lasso) and network-based penalization, was implemented on the correlation matrix, with the balance between sparsity and network coherence controlled by the parameter alpha. The parameter alpha was tuned within a default search range (0.05–0.90), terminating when the selected cells were less than 20% of the total. For each fixed alpha, the regularization strength λ of the L1 model was chosen using cross-validation with the minimum averaged error. The reliability of the Scissor_S/NS classification was validated by comparing the proportions of Scissor_S/NS cells in Scissor-selected cells, and we found that cells from smoking and non-smoking patients had higher proportions of Scissor_S cells and Scissor_NS cells, respectively (***Supplementary Fig. 1***).

Differential gene expression and functional enrichment analysis

Differential gene expression analysis of scRNA-seq data was performed using Seurat's 'FindMarkers' function (v4.0.1) with default parameters, including the Wilcoxon rank sum test, to compare batch-corrected normalized counts across groups, except for cancer cells, which retained uncorrected normalized counts as previously detailed. Genes with an absolute fold change (FC) greater than 1.25 and a false discovery rate (FDR) less than 0.05 were considered significantly differentially expressed genes. Gene set enrichment analysis (GSEA) was performed to assess pathway activity using the fgsea (v1.16.0) package[20]. Genes were ranked based on fold change, and gene sets were obtained from the Molecular Signatures Database (MsigDB)[21–22]. Gene sets with $P < 0.05$ and FDR < 0.25 were considered significantly enriched.

Single sample gene set enrichment analysis (ssGSEA)

To assess functional differences among different cell subgroups, the ssGSEA method was employed to

calculate the enrichment scores of known gene sets with specific biological significance[23]. The alveolar epithelial type Ⅱ cell (AT2) gene set was composed of known lung epithelial AT2 cell marker genes. The effector cell gene set consisted of cytotoxic T and NK cell marker genes, representing the level of T/NK cell infiltration[24]. The gene sets "immune suppression by myeloid cells", "checkpoint molecules", and "MHC-Ⅱ" comprised genes related to myeloid cell-mediated immune suppression, immune checkpoint molecules, and major histocompatibility complex (MHC)-Ⅱ antigen presentation, respectively, as described in the previous study[24].

Signature construction and survival analysis

The differentially expressed genes of the Scissor-selected cells were ranked based on FC values, and a gene signature was constructed using the top 50 genes with the highest FC values. Bulk sample signature scores were calculated using the ssGSEA algorithm[23]. Patients were separated into high-score and low-score groups based on the median value of the signature score. Kaplan-Meier survival analysis and the log-rank test were performed to assess the survival differences. Additionally, a multivariate Cox regression analysis, adjusting for tumor stage, was conducted to investigate whether the gene signature represented an independent risk factor.

Trajectory and cell-cell communication analysis

We used Monocle 3 for single-cell trajectory analysis to examine the differentiation differences between Scissor_S and Scissor_NS cancer cells[25]. To establish the differentiation trajectory, we included all cancer cells, along with normal AT2 and basal epithelial cells within the tumor samples, as these two cell types were recognized as the origins of lung cancer cells[26–28], and were designated as root cells. The normalization and clustering of all included cells were conducted as previously described, and normal epithelial cells were designated as root cells. The pseudotime values of Scissor_S and Scissor_NS cancer cells were compared using the two-sided Wilcoxon rank sum test.

Intercellular communication in the scRNA-seq data was computed and inferred using CellChat (v2.1.2)[12]. The manually curated CellChatDB database was used as the reference. Normalized counts were used for CellChat analysis, and the core functions 'computeCommunProb' and 'computeCommunProbPathway' were employed to calculate the communication probabilities and identify significant signaling pathways. Additionally, the 'aggregateNet' function was used to analyze the communication network by tallying and summarizing the interaction counts and communication probabilities.

Statistical analysis

For group comparisons, the Wilcoxon rank sum test was conducted for continuous variables, while the chi-square test and Fisher's exact test were employed for categorical variables. All reported *P* values were two-sided, with statistical significance defined as $P < 0.05$. *P* values were adjusted by FDR in multiple comparisons. Statistical analyses were performed using R software (v4.0.3), and all figures were generated using Seurat (v4.0.1), EnhancedVolcano (v1.12.0), ggplot2 (v3.3.5), and ggpubr (v0.4.0).

Results

Clinical characteristics and the study design

Among the 140 patients who underwent bulk sequencing, 47 were smokers (34%), and 93 were never-smokers (66%). The detailed clinical information is presented in ***Supplementary Table 1***. No significant differences were observed in clinical stage or age distribution between smoking and non-smoking patients. All smoking patients were male, while most non-smoking patients were female (70/93, 75%). Cox regression analysis revealed significantly poorer survival among smoking patients after adjusting for tumor stage (hazard ratio [HR] = 5.25, 95% confidence interval [CI]: 1.66–16.63, $P = 0.005$) (***Supplementary Fig. 2***). The clinical baseline information of the 20 LUAD patients undergoing scRNA-seq is detailed in ***Supplementary Table 2***. Fisher's exact test also revealed differences in sex distribution between smoking and non-smoking patients, with no significant differences observed in tumor stage or study origin.

The study workflow is shown in ***Fig. 1***. Initially, we defined smoking-associated (Scissor_S) cells and non-smoking-associated (Scissor_NS) cells using the Scissor algorithm. Next, we compared compositional and transcriptomic differences to uncover the underlying biological distinctions. Subsequently, we investigated the clinical relevance of Scissor-selected cells to identify potential prognostic markers and therapeutic strategies.

Smoking-associated and non-smoking-associated cell distribution

A total of 92 553 cells from 20 LUAD patients were

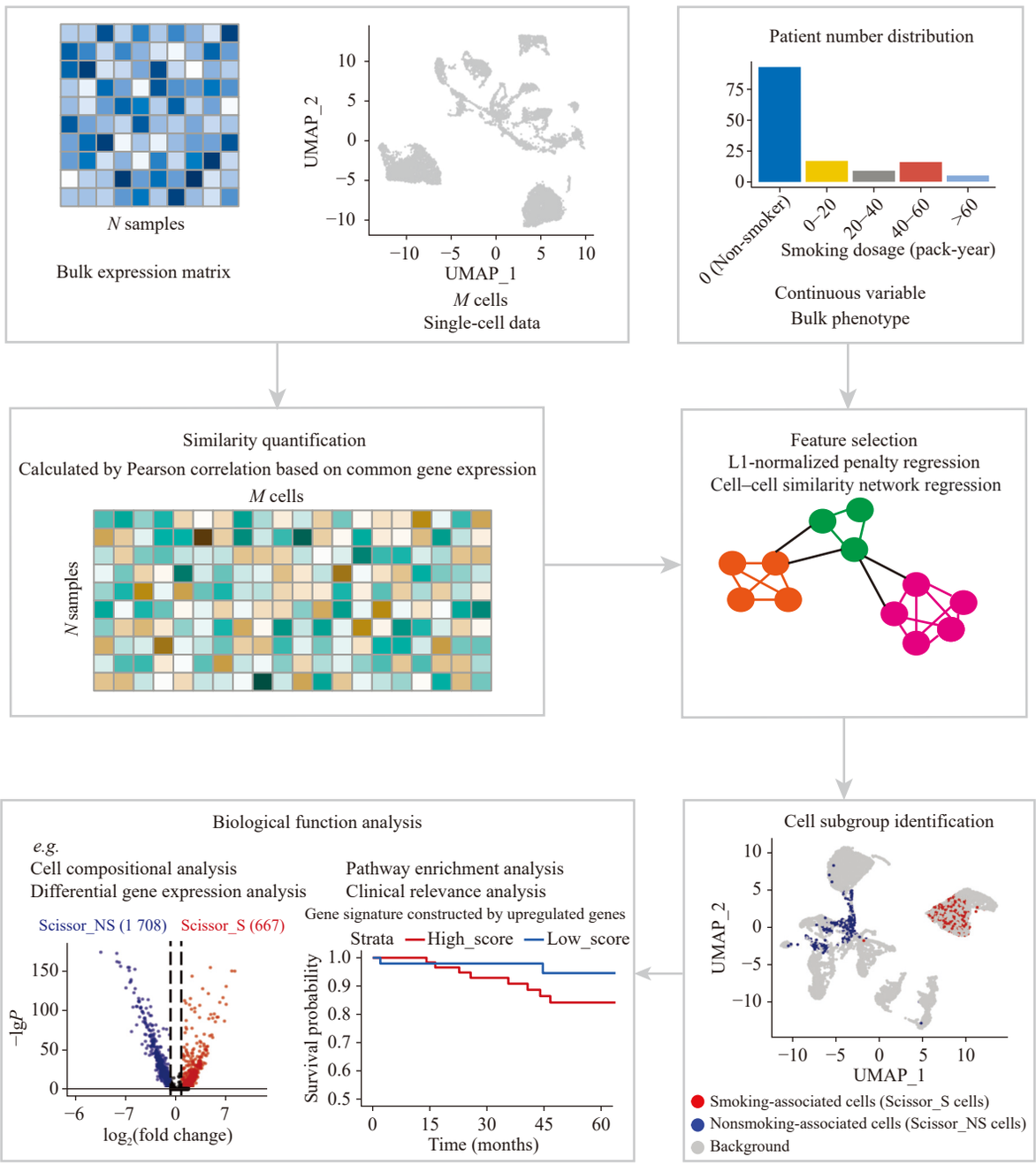


Fig. 1 **Overview of study workflow.** Step 1: Similarity quantification. A correlation matrix was generated using Pearson correlation based on shared gene expression between bulk and single-cell datasets. Step 2: Identification of smoking-associated (Scissor_S) and non-smoking-associated (Scissor_NS) cell subpopulations. The Scissor algorithm was employed to integrate bulk and single-cell datasets. Using 140 bulk RNA-seq samples with smoking dosage information, LASSO regression and network-constrained models were applied to identify Scissor_S and Scissor_NS cell subpopulations within the scRNA-seq dataset. Step 3: Biological function analysis. Compositional and transcriptomic analyses were conducted on Scissor_S and Scissor_NS cell subpopulations. In addition, we explored the clinical relevance of these Scissor-selected cell subgroups. Abbreviations: Scissor_NS, non-smoking-associated subpopulations (by the Scissor algorithm); Scissor_S, smoking-associated subpopulations (by the Scissor algorithm); UMAP, uniform manifold approximation and projection.

included in the current study, with 51 797 cells derived from smoking patients (56%) and 40 756 from non-smoking patients (44%). Epithelial cells, $CD4^+$ T cells, $CD8^+$ T/NK cells, macrophages, and DCs were the predominant cell populations, together accounting for over 80% of cells in all LUAD patients, as well as in both LCIS and LCINS subgroups (***Supplementary Fig. 3A*** and ***3B***). A direct comparison of cell proportions between 13 smoking and 7 non-smoking patients was performed, using linear regression with the dataset as the covariate. No significant difference was observed (***Supplementary Fig. 3C***), possibly due to the limited sample size. Therefore, we used the Scissor algorithm, guided by 140 bulk RNA-

sequencing samples, to identify smoking-associated and non-smoking-associated cell subpopulations.

A total of 5 266 Scissor_S cells and 3 164 Scissor_NS cells were identified (***Fig. 2A***). $CD4^+$ T cells (37% *vs.* 4%, FDR < 0.001) and $CD8^+$ T/NK cells (29% *vs.* 2%, FDR < 0.001) were predominantly present in Scissor_S cells, while DCs constituted the primary cell type in Scissor_NS cells (0% *vs.* 39%, FDR < 0.001) (***Fig. 2B***). Differential gene expression analysis (FC > 1.25, FDR < 0.05) revealed that 428 and 403 genes were upregulated in Scissor_S and Scissor_NS cells, respectively (***Fig. 2C***). IgG-related genes (*e.g.*, *IGHG3*) were highly expressed in Scissor_S cells, whereas the expression of IgA-related genes (*e.g.*, *IGHA1* and *IGHA2*) was elevated in Scissor_NS cells, suggesting potential differences in the infiltration of antibody-producing plasma cells between smoking and non-smoking patients (***Fig. 2C***

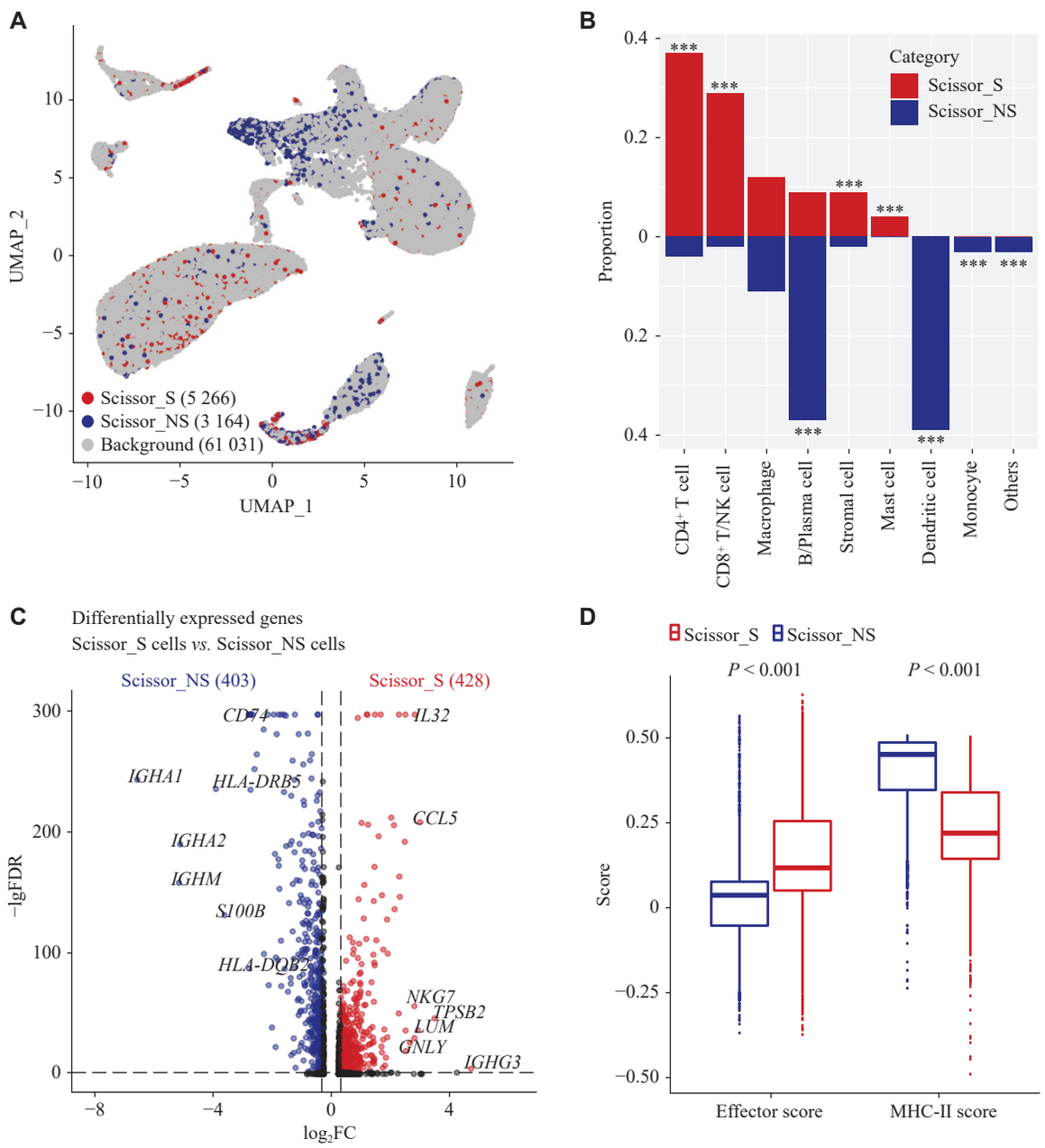


Fig. 2 **Smoking-associated and non-smoking-associated cell landscape.** A: UMAP plot of Scissor-selected smoking-associated and non-smoking-associated cells. B: Comparison of cell subtype proportions in Scissor_S and Scissor_NS cells. ***FDR < 0.001. C: Volcano plot of differentially expressed genes between Scissor_S and Scissor_NS cells (|FC| > 1.25 and FDR < 0.05). D: Boxplot showing the MHC-Ⅱ score and effector cell score in Scissor_S and Scissor_NS cells. *P* values were calculated by the two-tailed Wilcoxon rank sum test. Abbreviations: FC, fold change; FDR, false discovery rate; MHC, major histocompatibility complex; Scissor_NS, non-smoking-associated subpopulations (by the Scissor algorithm); Scissor_S, smoking-associated subpopulations (by the Scissor algorithm); UMAP, uniform manifold approximation and projection.

and ***Supplementary Table 3***). Additionally, the expression of several cell cytotoxicity-related genes (*e.g.*, *NKG7* and *GNLY*) was elevated in Scissor_S cells. Scissor_NS cells, on the other hand, expressed several HLA class Ⅱ molecules (*e.g.*, *HLA-DQB2* and *HLA-DRB5*) at high levels (***Fig. 2C*** and ***Supplementary Table 3***).

To further evaluate potential differences in the expression of cytotoxicity-related molecules and MHC class Ⅱ genes between Scissor_S and Scissor_NS cells, we computed the enrichment scores for an effector gene set, including $CD8^+$ T cell/NK cell marker genes, and an MHC class Ⅱ gene set in these two cell populations. The results revealed that the effector gene set score was elevated in Scissor_S cells (median, 0.12 *vs.* 0.04, $P < 0.001$), whereas the MHC-Ⅱ gene set score was increased in Scissor_NS cells (median, 0.22 *vs.* 0.45, $P < 0.001$) (***Fig. 2D***). These findings suggest that the tumor immune microenvironment of smokers may have a greater infiltration of cytotoxic cells, while non-smokers exhibit a more abundant presence of MHC class $Ⅱ^+$ antigen-presenting cells.

Higher aggressiveness of cancer cells in LCIS patients

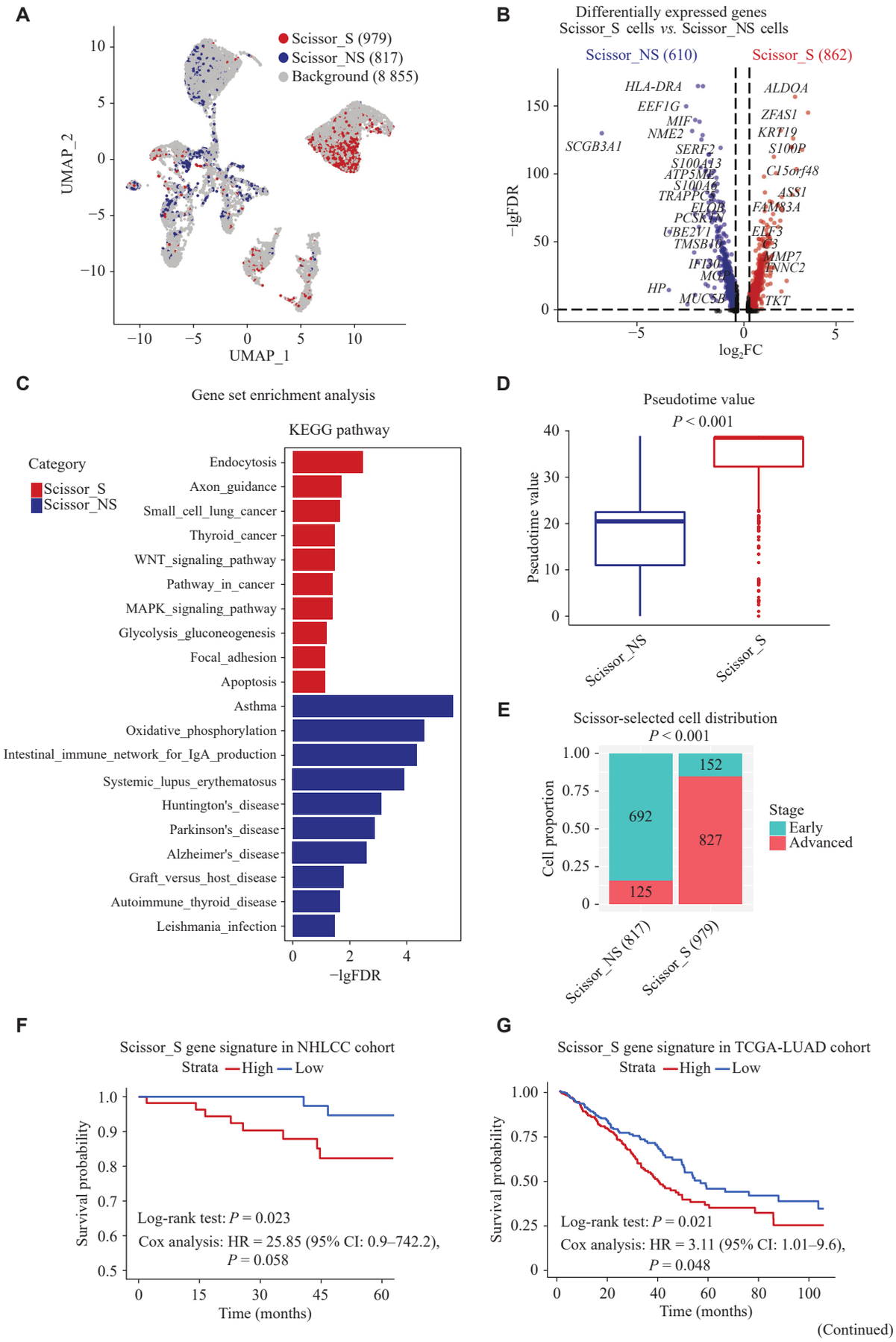
Cancer cells exhibit significant heterogeneity within each individual. Using the Scissor algorithm, we selected 1 796 cancer cells, including 979 Scissor_S cancer cells and 817 Scissor_NS cancer cells (***Fig. 3A***). Differential gene expression analysis revealed that 862 genes were upregulated in Scissor_S cells and 610 genes were upregulated in Scissor_NS cells (***Fig. 3B***). Several oncogenes, including *ZFAS1* and *ALDOA*, were found to be overexpressed in Scissor_S cancer cells[29–30]. GSEA revealed the activation of pathways related to endocytosis, axon guidance, and various cancer-related pathways, such as "Thyroid_cancer" and "Small_cell_lung_cancer", as well as oncogenic signaling pathways like WNT, MAPK, and p53, in Scissor_S cancer cells (***Fig. 3C*** and ***Supplementary Table 4***). Trajectory analysis was performed with AT2/basal cells designated as the root population to infer the pseudotime values of cancer cells (***Supplementary Fig. 4A***). Genes associated with lung epithelial cell differentiation, such as *SFTPC* and *SCGB3A1*[31], were inversely correlated with pseudotime values (***Supplementary Fig. 4B***), indicating that higher pseudotime corresponds to the loss of cellular differentiation and increased aggressiveness. We found that most Scissor_S cancer cells were positioned at the terminal end of the differentiation trajectory (***Fig. 3A*** and ***Supplementary Fig. 4A***), exhibiting significantly higher pseudotime values (median, 38.52 *vs.* 20.47, $P < 0.001$) than Scissor_NS cancer cells (***Fig. 3D***), indicating a more advanced stage of malignant progression. Correspondingly, a large proportion of Scissor_S cancer cells were derived from late-stage patients (84%), while most Scissor_NS cancer cells were sourced from early-stage tumor patients (85%) (***Fig. 3E***).

Subsequently, we selected the top 50 genes with the highest fold changes in Scissor_S cancer cells to construct the Scissor_S cancer cell signature and computed the signature score. Patients with higher expression levels of the Scissor_S gene signature had worse prognosis after adjusting for tumor stage in the NHLCC cohort (HR = 25.85, 95% CI: 0.9–742.2, $P = 0.058$) (***Fig. 3F***), and the association reached statistical significance in the TCGA-LUAD cohort (HR = 3.11, 95% CI: 1.01–9.60, $P = 0.048$) (***Fig. 3G***). Additionally, the normalized expression counts of MHC class Ⅱ molecules were elevated in Scissor_NS cancer cells, whereas MHC class Ⅰ molecules were found to be highly expressed in Scissor_S cancer cells (***Fig. 3H***). Furthermore, we observed a significantly higher expression of multiple AT2 cell markers (*e.g.*, *SFTPB* and *SFTPC*) in Scissor_NS cancer cells (***Supplementary Table 4***), along with a higher AT2 score (median, 0.11 *vs.* 0.31, $P < 0.001$) (***Fig. 3I***). AT2 cells play a key role in injury repair, and studies have shown that LUAD patients with higher AT2 scores have better prognoses[5]. Considering the established association between smoking and damage to the alveolar epithelium[14], smoking-induced injury to AT2 cells may be linked to adverse outcomes in smokers.

Non-smoking-associated cancer cells showed elevated levels of inflammation-related molecules, such as *HP* and *MIF* (***Fig. 3B*** and ***Supplementary Table 4***). GSEA indicated that Scissor_NS cancer cells were correlated with inflammation-related diseases (*e.g.*, asthma) (***Fig. 3C***), suggesting a potential role of chronic inflammation in lung cancer development in never-smokers. Additionally, the oxidative phosphorylation pathway was elevated in Scissor_NS cancer cells, which may also be linked to the inflammatory tumor microenvironment[32].

Immunosuppressive property of smoking-associated macrophages

Macrophages play an intrinsic role in the tumor immune microenvironment and were classified into six subgroups (***Fig. 4A*** and ***Supplementary Fig. 5***). A total of 1 437 Scissor_S macrophages and 1 275 Scissor_NS macrophages were identified (***Fig. 4B***).


A
Scissor_S (979)
Scissor_NS (817)
Background (8 855)
UMAP_2
UMAP_1
B
Differentially expressed genes
Scissor_S cells vs. Scissor_NS cells
Scissor_NS (610)
Scissor_S (862)
HLA-DRA
EEF1G
MIF
NME2
SCGB3A1
SERF2
S100A13
ATP5ME
S100A6
TRAPPC3
ELOB
PCSK1N
UBE2V1
TMSB10
IFI30
MGP
HP
MUC5B
ALDOA
ZFAS1
KRT19
S100P
C15orf48
ASS1
FAM83A
ELF3
C3
MMP7
TNNC2
TKT
−lgFDR
log2FC
C
Gene set enrichment analysis
KEGG pathway
Category
Scissor_S
Scissor_NS
Endocytosis
Axon_guidance
Small_cell_lung_cancer
Thyroid_cancer
WNT_signaling_pathway
Pathway_in_cancer
MAPK_signaling_pathway
Glycolysis_gluconeogenesis
Focal_adhesion
Apoptosis
Asthma
Oxidative_phosphorylation
Intestinal_immune_network_for_IgA_production
Systemic_lupus_erythematosus
Huntington's_disease
Parkinson's_disease
Alzheimer's_disease
Graft_versus_host_disease
Autoimmune_thyroid_disease
Leishmania_infection
−lgFDR
D
Pseudotime value
P < 0.001
Pseudotime value
Scissor_NS
Scissor_S
E
Scissor-selected cell distribution
P < 0.001
Cell proportion
692
125
152
827
Stage
Early
Advanced
Scissor_NS (817)
Scissor_S (979)
F
Scissor_S gene signature in NHLCC cohort
Strata High Low
Survival probability
Log-rank test: P = 0.023
Cox analysis: HR = 25.85 (95% CI: 0.9–742.2),
P = 0.058
Time (months)
G
Scissor_S gene signature in TCGA-LUAD cohort
Strata High Low
Survival probability
Log-rank test: P = 0.021
Cox analysis: HR = 3.11 (95% CI: 1.01–9.6),
P = 0.048
Time (months)

(Continued)

(Continued)

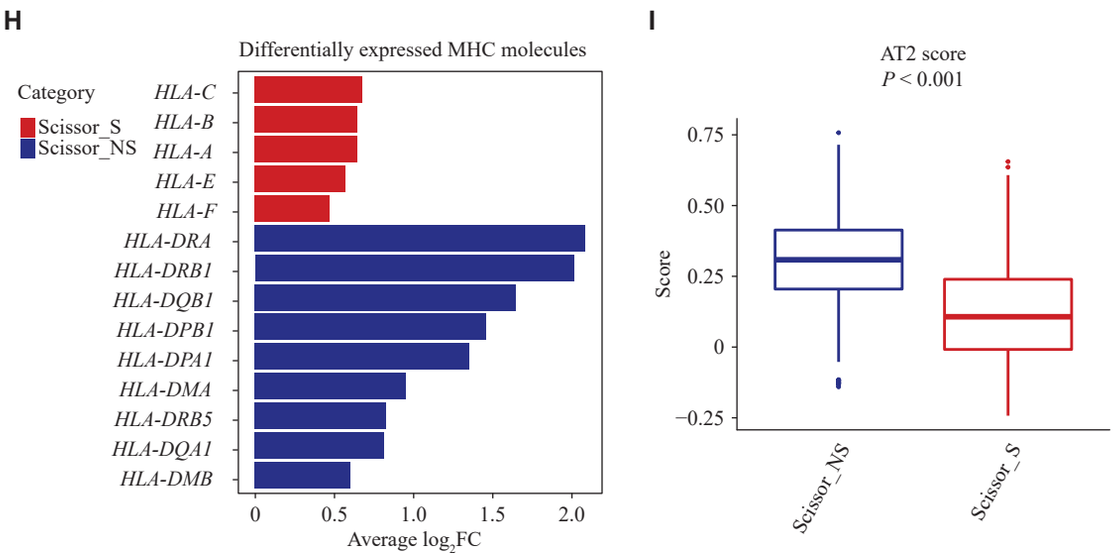
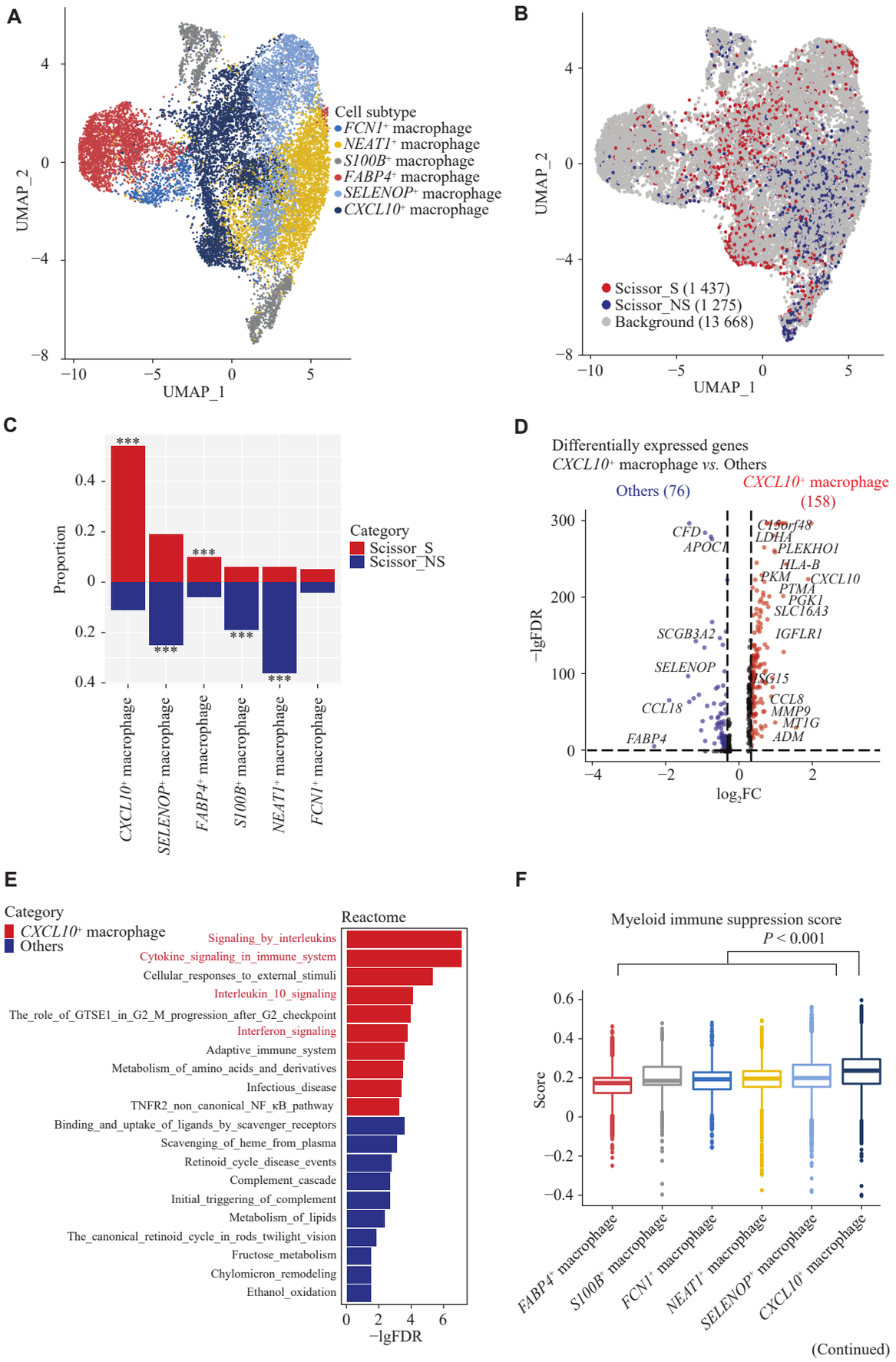


Fig. 3 **Characterization of smoking-associated and non-smoking-associated cancer cell subpopulations.** A: UMAP plot of Scissor-selected smoking-associated and non-smoking-associated cancer cells. B: Volcano plot of differentially expressed genes between Scissor_S and Scissor_NS cancer cells (|FC| > 1.25 and FDR < 0.05). C: Bar plot of gene set enrichment analysis showing the top 10 pathways with the lowest FDR values in Scissor_S and Scissor_NS cancer cells, respectively. D: Comparison of pseudotime values between Scissor_S and Scissor_NS cancer cells. The *P* value was derived from the two-sided Wilcoxon test. E: Proportions of Scissor_S and Scissor_NS cancer cells from early or late clinical stages. The *P* value was derived from the two-tailed chi-squared test. F: Kaplan-Meier survival curves comparing high and low Scissor_S gene signature subgroups in the NHLCC cohort. The *P* value was derived from the two-tailed log-rank sum test and multivariate Cox regression adjusting for tumor stage. G: Kaplan-Meier survival curves comparing high and low Scissor_S gene signature subgroups in the TCGA-LUAD cohort. The *P* value was derived from the two-tailed log-rank sum test and multivariate Cox regression adjusting for tumor stage. H: Differentially expressed MHC molecules in Scissor_S and Scissor_NS cancer cells. I: Comparison of AT2 scores between Scissor_S and Scissor_NS cancer cells. The *P* value was derived from a two-sided Wilcoxon test. Abbreviations: AT2, alveolar epithelial type Ⅱ cell; CI: confidence interval; FC, fold change; FDR, false discovery rate; HR, hazard ratio; MHC, major histocompatibility complex; NHLCC, Nanjing-Huaxi Lung Cancer Consortium; TCGA-LUAD, The Cancer Genome Atlas Lung Adenocarcinoma; Scissor_S, smoking-associated subpopulations (by the Scissor algorithm); Scissor_NS, non-smoking-associated subpopulations (by the Scissor algorithm); UMAP, uniform manifold approximation and projection.

The *CXCL10*$^+$ macrophages were significantly overrepresented in Scissor_S cells (54% *vs.* 11%, FDR < 0.001), while the *NEAT1*$^+$ macrophages were more prevalent among Scissor_NS macrophages (6% *vs.* 36%, FDR < 0.001) (***Fig. 4C***). A total of 158 genes were found to be overexpressed in the *CXCL10*$^+$ macrophages (***Fig. 4D***). Among them, multiple types of cytokines were elevated in the *CXCL10*$^+$ macrophages, including interleukins (*e.g.*, *IL4I1* and *IL1B*), chemokines (*e.g.*, *CXCL9* and *CXCL10*), and interferons (*e.g.*, *IFITM3* and *IFITM1*) (***Fig. 4D*** and ***Supplementary Table 5***). Pathway analysis also revealed that *CXCL10*$^+$ macrophages were enriched in interleukin, cytokine, and interferon signaling pathways (***Fig. 4E***), indicating that they were involved in immune regulation in the tumor microenvironment. Within cytokine-related pathways, the *IL-10* signaling pathway was highly active in *CXCL10*$^+$ macrophages. IL-10 is an anti-inflammatory cytokine that inhibits the activities of multiple immune cells[33]. Moreover, compared with other macrophages, *CXCL10*$^+$ macrophages had a significantly higher "myeloid cell immune suppression" score (0.24 *vs.* 0.19, *P* < 0.001) (***Fig. 4F***). Additionally, the immune suppression score of Scissor_S *CXCL10*$^+$ macrophages was significantly higher than that of Scissor_NS *CXCL10*$^+$ macrophages (median, 0.30 *vs.* 0.26, *P* < 0.001) (***Fig. 4G***). Taken together, immunosuppressive, smoking-associated *CXCL10*$^+$ macrophages may be involved in the carcinogenesis of smokers.

CXCL10 plays a critical role in recruiting cytotoxic CD8$^+$ T cells to the tumor site[34]. Correlation analysis in bulk tissues also revealed a significant association between *CXCL10*$^+$ macrophage infiltration and the expression of cytotoxic and exhausted T cell marker genes (*CD8A*, *GZMB*, *PRF1*, and *PDCD1*) (***Supplementary Fig. 6***). Additionally, although not reaching a significant threshold (FC > 1.25), several immune checkpoint molecules, including *PD-L1* (*CD274*), were overexpressed in *CXCL10*$^+$ macrophages (***Fig. 4H***). They also exhibited a significantly higher immune checkpoint score (0.08 *vs.* 0.02, *P* < 0.001) (***Fig. 4I***). Furthermore, the checkpoint score of Scissor_S *CXCL10*$^+$ macrophages was significantly higher than that of Scissor_NS


A
UMAP_2
UMAP_1
Cell subtype
FCN1+ macrophage
NEAT1+ macrophage
S100B+ macrophage
FABP4+ macrophage
SELENOP+ macrophage
CXCL10+ macrophage
B
Scissor_S (1 437)
Scissor_NS (1 275)
Background (13 668)
C
Proportion
Category
Scissor_S
Scissor_NS
CXCL10+ macrophage
SELENOP+ macrophage
FABP4+ macrophage
S100B+ macrophage
NEAT1+ macrophage
FCN1+ macrophage
D
Differentially expressed genes
CXCL10+ macrophage vs. Others
Others (76)
CXCL10+ macrophage (158)
−lgFDR
log2FC
CFD
APOC1
C15orf48
LDHA
PLEKHO1
HLA-B
PKM
CXCL10
PTMA
PGK1
SLC16A3
IGFLR1
SCGB3A2
SELENOP
ISG15
CCL18
CCL8
MMP9
MT1G
ADM
FABP4
E
Category
CXCL10+ macrophage
Others
Reactome
Signaling_by_interleukins
Cytokine_signaling_in_immune_system
Cellular_responses_to_external_stimuli
Interleukin_10_signaling
The_role_of_GTSE1_in_G2_M_progression_after_G2_checkpoint
Interferon_signaling
Adaptive_immune_system
Metabolism_of_amino_acids_and_derivatives
Infectious_disease
TNFR2_non_canonical_NF_κB_pathway
Binding_and_uptake_of_ligands_by_scavenger_receptors
Scavenging_of_heme_from_plasma
Retinoid_cycle_disease_events
Complement_cascade
Initial_triggering_of_complement
Metabolism_of_lipids
The_canonical_retinoid_cycle_in_rods_twilight_vision
Fructose_metabolism
Chylomicron_remodeling
Ethanol_oxidation
−lgFDR
F
Myeloid immune suppression score
P < 0.001
Score
FABP4+ macrophage
S100B+ macrophage
FCN1+ macrophage
NEAT1+ macrophage
SELENOP+ macrophage
CXCL10+ macrophage

(Continued)

(Continued)

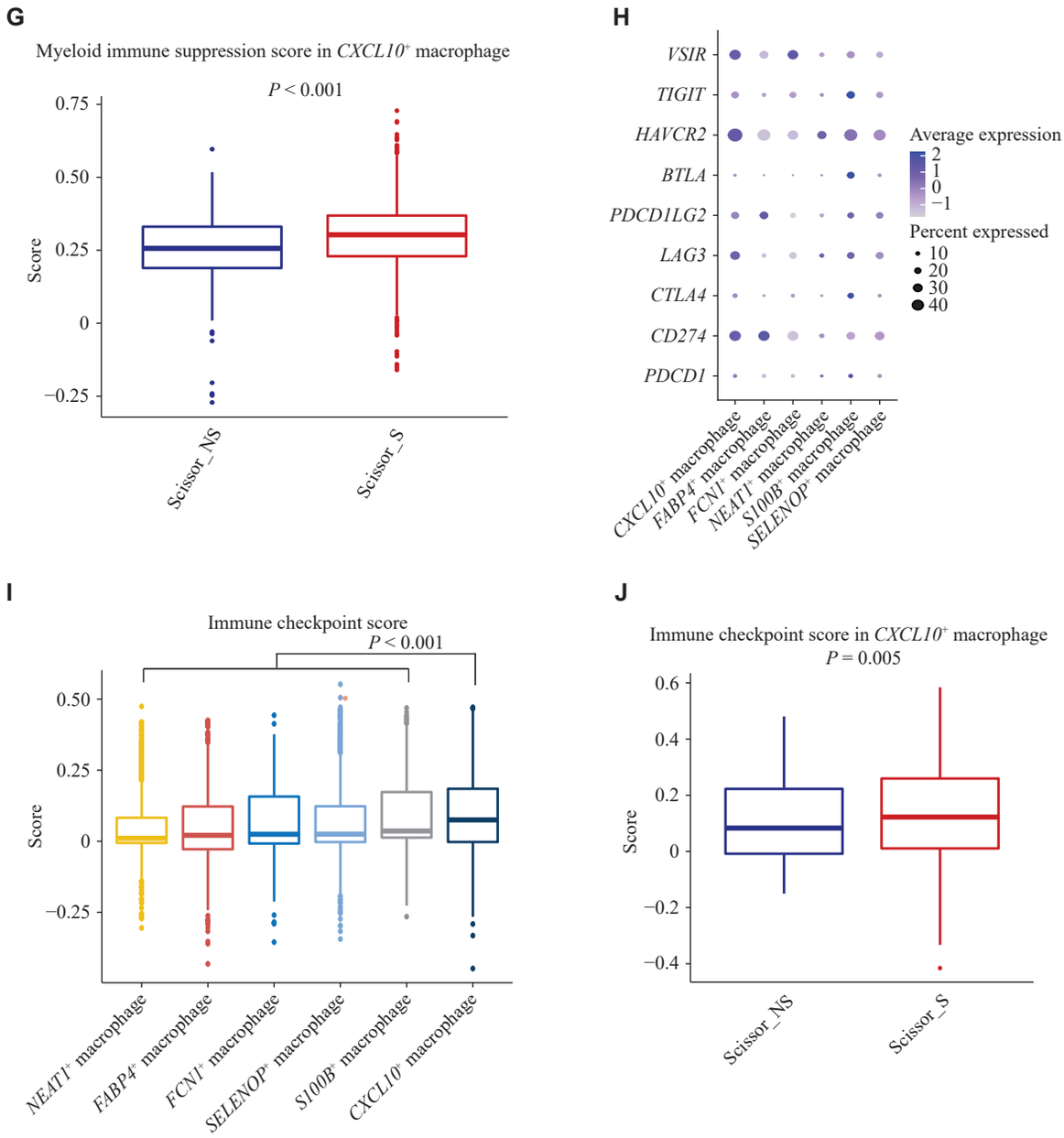


Fig. 4 **Characterization of smoking-associated and non-smoking-associated macrophage subpopulations.** A: UMAP plot of macrophages, colored by cell subtypes. B: UMAP plot of Scissor-selected smoking-associated and non-smoking-associated macrophages. C: Comparison of cell subtype proportions between Scissor_S and Scissor_NS macrophages. ***FDR < 0.001. D: Volcano plot of differentially expressed genes between *CXCL10*+ macrophages and other macrophages (|FC| > 1.25 and FDR < 0.05). E: Bar plot of gene set enrichment analysis showing the top 10 pathways with the lowest FDR values in *CXCL10*+ macrophages and other macrophages, respectively. F: Myeloid immune suppression scores among different macrophage subtypes. The *P* value was derived from the two-sided Wilcoxon test comparison between *CXCL10*+ macrophages and other macrophages. G: Comparison of myeloid immune suppression scores between Scissor_S and Scissor_NS *CXCL10*+ macrophages. The *P* value was derived from the two-sided Wilcoxon test. H: Dot plot showing the expression of checkpoint genes in each macrophage subtype. I: Immune checkpoint score among different macrophage subtypes. The *P* value was derived from the two-sided Wilcoxon test comparison between *CXCL10*+ macrophages and other macrophages. J: Comparison of immune checkpoint scores between Scissor_S and Scissor_NS *CXCL10*+ macrophages. The *P* value was derived from the two-sided Wilcoxon test. Abbreviations: FC, fold change; FDR, false discovery rate; Scissor_S, smoking-associated subpopulations (by the Scissor algorithm); Scissor_NS, non-smoking-associated subpopulations (by the Scissor algorithm); UMAP, uniform manifold approximation and projection.

macrophages (median, 0.12 *vs.* 0.08, $P = 0.005$) (***Fig. 4J***). Given that both cytotoxic cell infiltration and immune checkpoint molecule expression are predictive of immunotherapeutic efficacy, the enrichment of *CXCL10*+ macrophages may be associated with a favorable immunotherapy response in smoking patients.

NEAT1 has been reported to be involved in the

activation of the inflammasome in macrophages[35]. Additionally, multiple inflammation-related genes, such as *APOC1* and *IFI27*, were highly expressed in *NEAT1*$^+$ macrophages. Pathway analysis revealed that *NEAT1*$^+$ macrophages were enriched in the lysosome pathway (***Supplementary Table 5***). Therefore, the aggregation of *NEAT1*$^+$ macrophages in the Scissor_NS macrophage population may contribute to the inflammatory microenvironment observed in non-smoking patients.

Enrichment of cytotoxic plasmacytoid DCs in smoking-related DCs

DCs were classified into eight subgroups according to gene expression profiles (***Supplementary Figs. 7*** and ***8A***). We identified 358 Scissor_S DCs and 415 Scissor_NS DCs by the Scissor algorithm (***Supplementary Fig. 8B***). *CXCL10*$^+$ DCs (38% *vs.* 7%, FDR < 0.001) and plasmacytoid DCs (23% *vs.* 0%, FDR < 0.001) were found to be enriched in Scissor_S DCs (***Supplementary Fig. 8C***). A total of 300 genes were found to be highly expressed in *CXCL10*$^+$ DCs. Apart from *CXCL10*, multiple marker genes of monocytes, including *S100A9*, *S100A8*, and *CD14*, displayed the highest fold change values (***Supplementary Table 6***), suggesting that *CXCL10*$^+$ DCs may originate from circulating monocytes. Compared with other types of DCs, plasmacytoid DCs expressed cytotoxic molecules such as *GZMB*, indicating their tumor elimination capability (***Supplementary Fig. 8D*** and ***Supplementary Table 6***). The most prevalent cell subtype of Scissor_NS DCs was the *CD1A*$^+$ DC (4% *vs.* 32%, FDR < 0.001) (***Supplementary Fig. 8C***). *CD1A* plays a role in mediating antigen presentation to T cells[36], and our results suggest that DCs in non-smoking patients exhibit a higher antigen presentation capability than those in smoking patients. Pathway enrichment analysis revealed distinct pathway activities for *CXCL10*$^+$ DCs, plasmacytoid DCs, and *CD1A*$^+$ DCs. For example, the extracellular matrix-related pathways were highly active in *CXCL10*$^+$ DCs, while plasmacytoid DCs exhibited lower activity of the innate immune system and cytokine signaling pathways (***Supplementary Table 6***).

Distinct lymphocyte anti-tumor activities in smoking and non-smoking patients

CD8$^+$ T cells and NK cells are major cytotoxic effector cells, which were categorized into eight subgroups (***Fig. 5A*** and ***Supplementary Fig. 9***). Among them, NK cells and cytotoxic T lymphocytes (CTLs), collectively known as NK/CTL cells, expressed the highest levels of cytotoxic molecules (***Supplementary Fig. 9***). Using Scissor, we extracted 1 763 CD8$^+$ T/NK Scissor_S cells and 1 162 CD8$^+$ T/NK Scissor_NS cells (***Fig. 5B***). We observed that Scissor_S CD8$^+$ T/NK cells had a higher fraction of cytotoxic T/NK cells (30% *vs.* 12%, FDR < 0.001), whereas the proportion of *GZMK*$^+$CD8$^+$ T cells was elevated in Scissor_NS cells (14% *vs.* 37%, FDR < 0.001) (***Fig. 5C***). *GZMK*$^+$CD8$^+$ T cells were reported to be less cytotoxic[5], and compared with NK/CTL cells, *GZMK*$^+$CD8$^+$ T cells possessed lower levels of cytotoxic molecules (*e.g.*, *GNLY* and *PRF1*) (***Supplementary Fig. 9***). Pathway enrichment analysis revealed that multiple cytotoxicity-related pathways (*e.g.*, natural killer cell-mediated cytotoxicity) were activated in NK/CTL cells, but their activities were significantly decreased in *GZMK*$^+$CD8$^+$ T cells (***Supplementary Table 7***). Additionally, when comparing the transcriptomic differences between Scissor_S and Scissor_NS NK/CTL cells, we observed that the expression of cytotoxic molecules (*e.g.*, *FGFBP2*, *GZMB*, and *NKG7*) was downregulated in Scissor_NS NK/CTL cells-(***Fig. 5D***). Our results indicated that non-smoking patients exhibited decreased cytotoxic T/NK cell infiltration, along with reduced activity.

B/plasma cells represent a distinct type of lymphocyte that produces a significant amount of antibodies and were classified into CD20$^+$ B cells, dividing B cells, IgA$^+$ plasma cells, and IgG$^+$ plasma cells (***Fig. 5E*** and ***Supplementary Fig. 10***). After applying Scissor, 518 Scissor_S cells and 468 Scissor_NS cells were identified (***Fig. 5F***). Scissor_S cells had a higher proportion of CD20$^+$ B cells (59% *vs.* 47%, FDR < 0.001) and IgG$^+$ plasma cells (37% *vs.* 19%, FDR < 0.001), whereas IgA$^+$ plasma cells were predominantly found in Scissor_NS cells (2% *vs.* 27%, FDR < 0.001) (***Fig. 5G***). Concordantly, bulk tissue analysis revealed a significantly elevated expression of IgA in non-smokers (median value of mean IgA-related gene expression, 8.83 *vs.* 9.38, P = 0.012) (***Fig. 5H***). Our findings demonstrated variations in the infiltration of antibody-producing plasma cells between smoking and non-smoking patients.

Cell-cell communication analysis

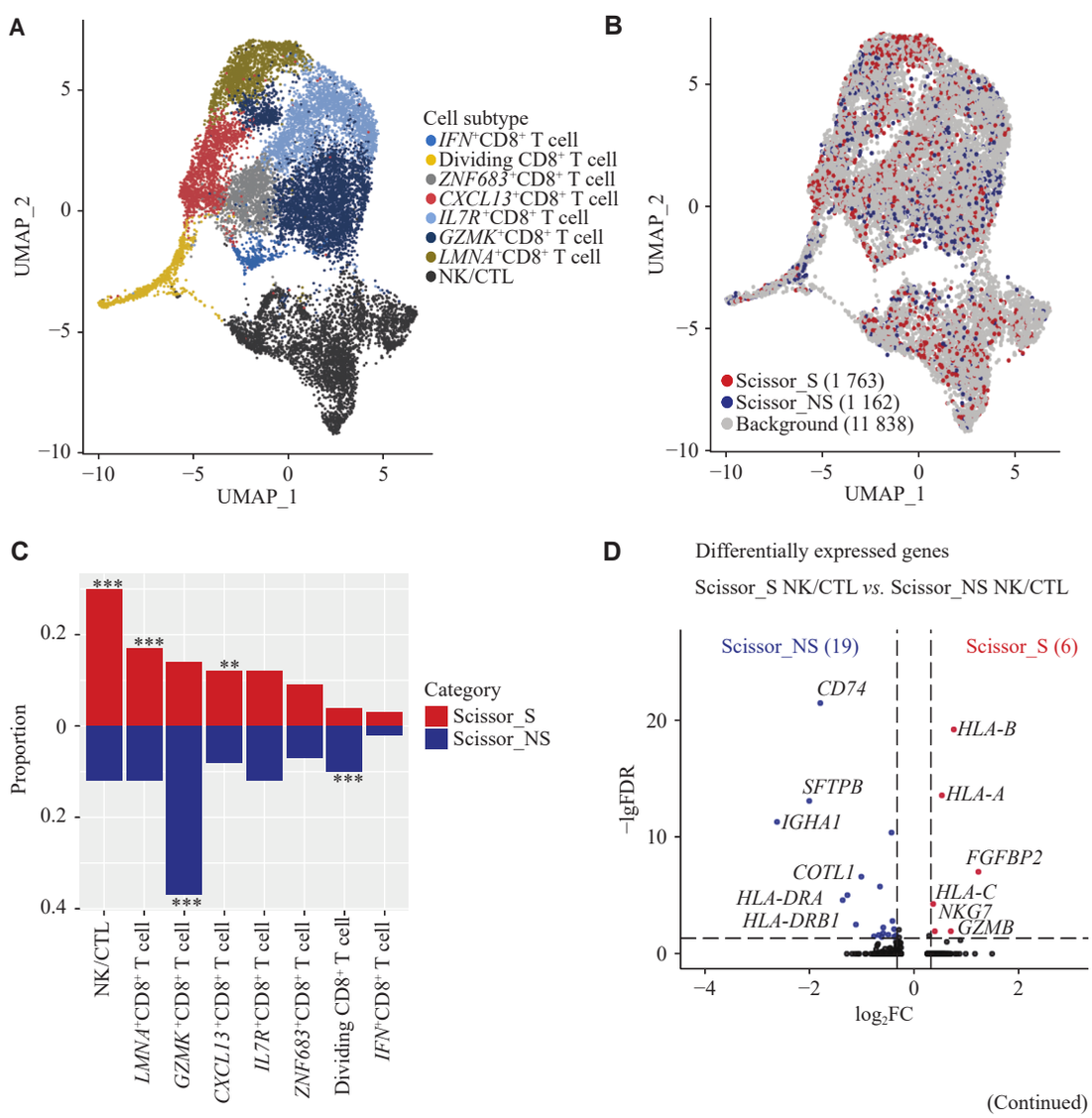
To clarify the difference in cell-cell communication between LCIS and LCINS, we first compared the interactions of Scissor_S and Scissor_NS cancer cells with other major cell types. The incoming signals to Scissor_S and Scissor_NS cancer cells did not show much difference (***Fig. 6A***), but Scissor_NS cancer cells sent more signals to other cells (***Fig. 6B***).

Despite interacting less frequently with other cells, Scissor_S cancer cells still exhibited strong communication signals with CD8+ T/NK cells (***Fig. 6B***). This phenomenon may be attributed to the active MHC- I –CD8A axis, given the high expression of MHC- I molecules on Scissor_S cancer cells (***Figs. 6C*** and ***3H***, and ***Supplementary Fig. 11***). The Scissor_NS cells, on the other hand, were activated in the macrophage migration inhibitory factor (MIF) signaling pathway (***Fig. 6C***), with *MIF* being predominantly expressed in Scissor_NS cancer cells (***Fig. 6D***). Although MIF is a pro-inflammatory molecule, in the context of cancer, it could create a favorable immune microenvironment to facilitate tumor growth[37].

CXCL10+ macrophages were found to be immunosuppressive and expressed several immune checkpoint molecules, including *PD-L1*, *VSIR*, and *HAVCR2* (***Fig. 4G***). Among these checkpoint molecules, we observed that *CXCL10*+ macrophages interacted with several other cell types through the immunoregulatory HAVCR2-LGALS9 axis (***Fig. 6E***). HAVCR2 activation is primarily induced by LGALS9[38], and we noted that *LGALS9* expression was higher in smoking patients (***Fig. 6F***). Therefore, our results suggest that the HAVCR2-LGALS9 axis might contribute to the immunosuppressive properties of *CXCL10*+ macrophages in smoking patients.

Discussion

Our study reveals distinct molecular and immunological landscapes between LUAD in smokers and never-smokers, highlighting distinct oncogenic mechanisms and immune microenvironments that shape different clinical outcomes and therapeutic responses[3]. Smoking patients exhibited enhanced tumor aggressiveness, with upregulated pathways



(Continued)

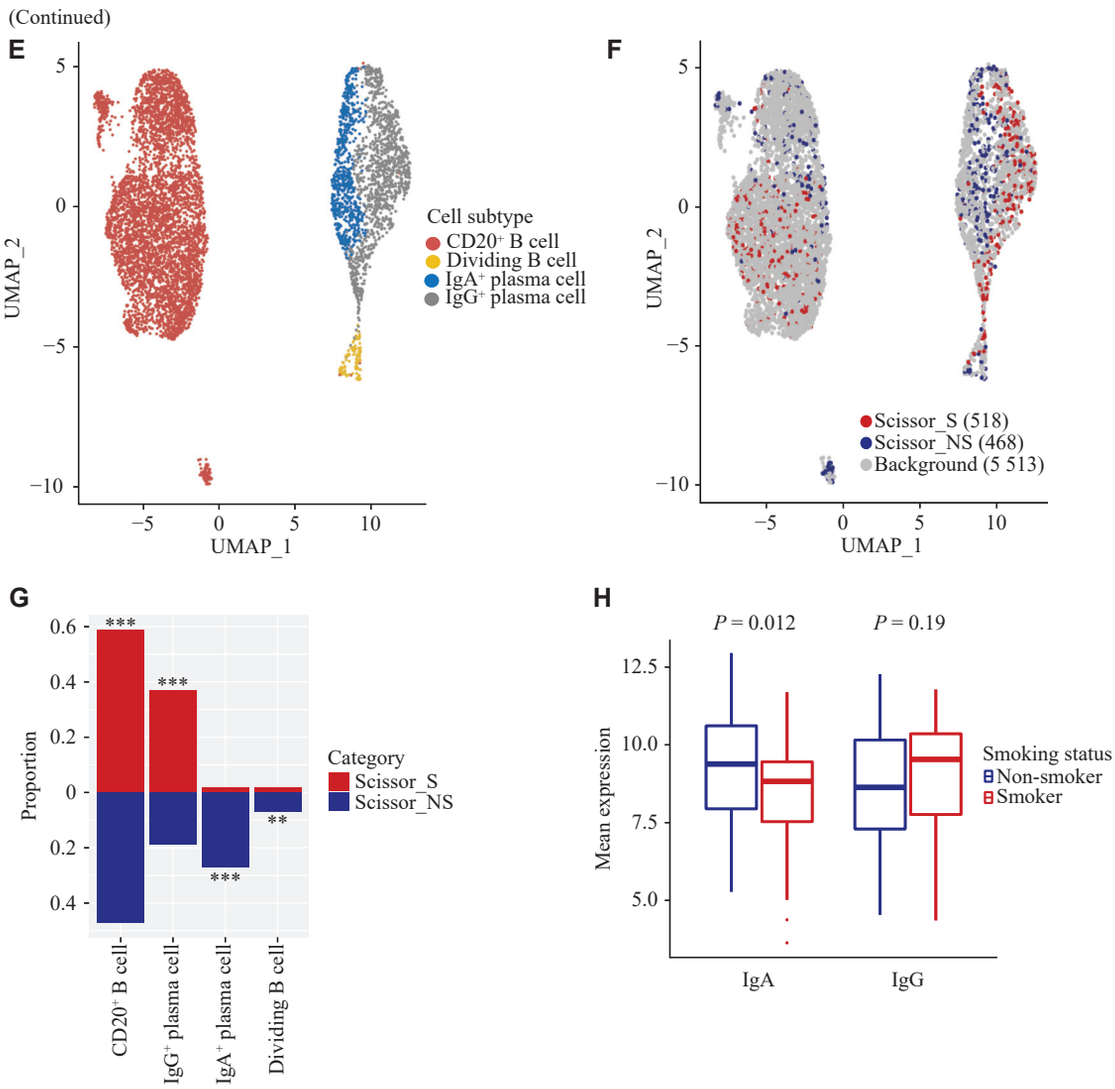


Fig. 5 **Characterization of smoking-associated and non-smoking-associated lymphocyte cell subpopulations.** A: UMAP plot of CD8⁺ T/NK cells, colored by cell subtypes. B: UMAP plot of Scissor-selected smoking-associated and non-smoking-associated CD8⁺ T/NK cell subpopulations. C: Comparison of cell subtype proportions between Scissor_S and Scissor_NS CD8⁺ T/NK cells. **FDR < 0.01 and ***FDR < 0.001. D: Volcano plot of differentially expressed genes between Scissor_S and Scissor_NS NK/CTL cells (|FC| > 1.25 and FDR < 0.05). E: UMAP plot of B/plasma cells, colored by cell subtypes. F: UMAP plot of Scissor-selected smoking-related and non-smoking-related B/plasma cell subpopulations. G: Comparison of cell subtype proportions between Scissor_S and Scissor_NS B/plasma cells. **FDR < 0.01 and ***FDR < 0.001. H: Comparison of IgA- and IgG-related gene expression between smoking and non-smoking bulk RNA-seq patients. Abbreviations: CTL, cytotoxic T lymphocyte; FC, fold change; FDR, false discovery rate; NK, natural killer; Scissor_NS, non-smoking-associated subpopulations (by the Scissor algorithm); Scissor_S, smoking-associated subpopulations (by the Scissor algorithm); UMAP, uniform manifold approximation and projection.

linked to cell proliferation and oncogenesis, such as the WNT and MAPK pathways[39–40]. These findings align with prior proteomic evidence and are consistent with the established link between smoking and WNT/MAPK pathway activation[41–42]. Additionally, smoking-associated LUAD displayed reduced MHC-Ⅱ expression on cancer cells and diminished infiltration of MHC-Ⅱ⁺ immune cells, correlating with reduced CD4⁺ T-cell infiltration, impaired CD4⁺ T-cell activation, and poor prognosis[5,43–48]. Smoking-induced AT2 cell disruption —a key population in lung repair —further suggests a mechanism for accelerated tumorigenesis, as AT2 cell loss was associated with poor prognosis[5].

For never-smokers, reduced antitumor cytotoxic activity may contribute to tumorigenesis. Non-smoking patients demonstrated lower expression of MHC-Ⅰ on tumor cells, reduced infiltration of cytotoxic CD8⁺ T and NK cells, and diminished CD8⁺ T-cell cytotoxicity. These characteristics correspond to their lower tumor mutational burden, which correlates with fewer neoantigens and a weaker cytotoxic response, explaining their suboptimal response to immune checkpoint inhibitors[6,8,49].

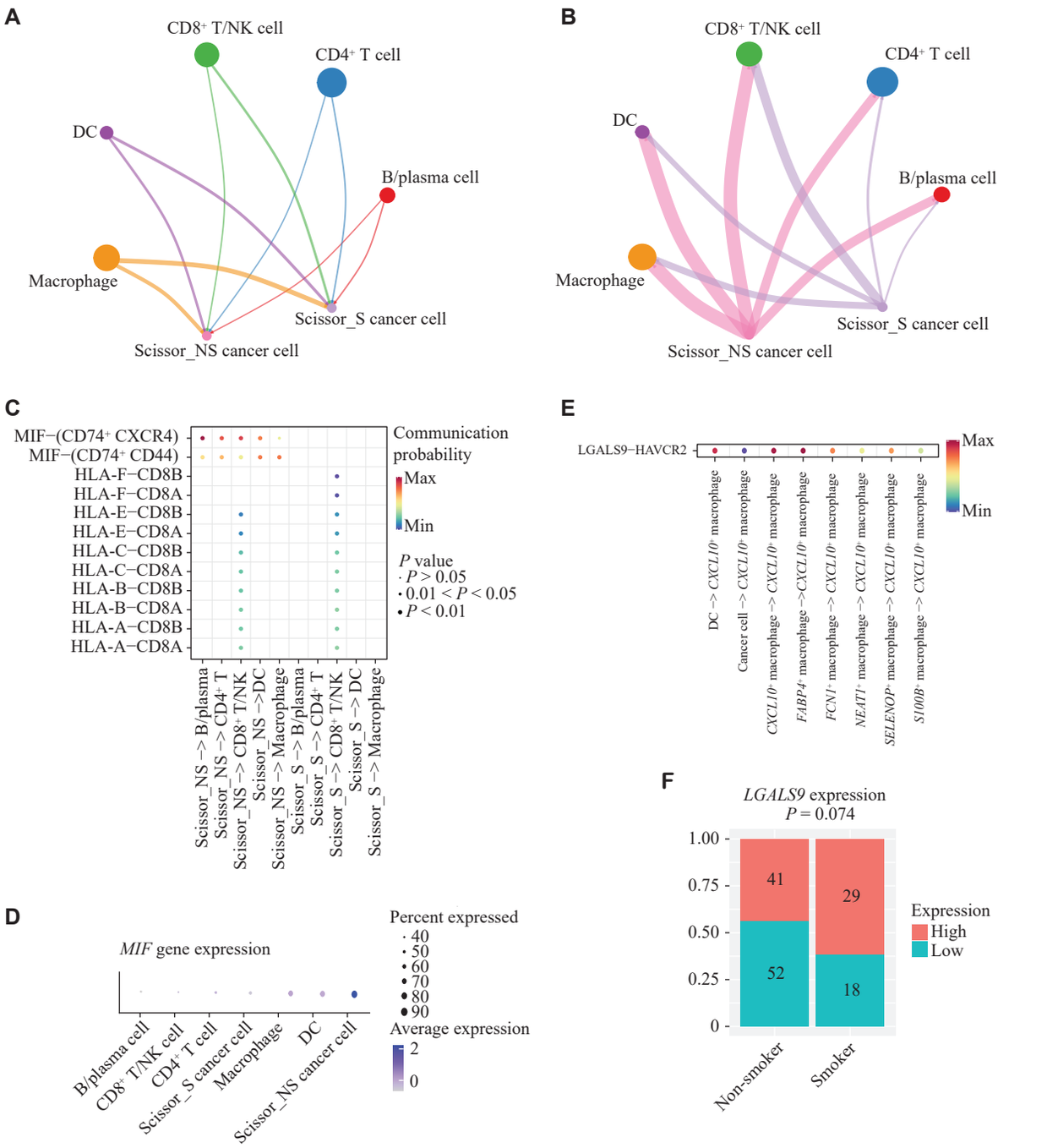


Fig. 6 **Cell-cell communication analysis.** A: Incoming signals to Scissor_S and Scissor_NS cancer cells. A thicker edge line indicates a stronger signal. B: Outgoing signals from Scissor_S and Scissor_NS cancer cells. A thicker edge line indicates a stronger signal. C: Significant interaction associated with MIF and MHC- I signaling sent by Scissor_S and Scissor_NS cells. D: *MIF* expression in various cell subtypes. E: Significant LGALS9-HAVCR2 interaction targeting *CXCL10*$^+$ macrophages. F: Comparison of bulk *LGALS9* expression in smoking and non-smoking patients. Abbreviations: MHC, major histocompatibility complex; Scissor_NS, non-smoking-associated subpopulations (by the Scissor algorithm); Scissor_S, smoking-associated subpopulations (by the Scissor algorithm).

Furthermore, non-smokers exhibited pronounced MIF signaling, which is a proinflammatory cytokine associated with chronic inflammation[37]. Previous research has also indicated that prolonged inflammation may reshape the tissue microenvironment and promote lung cancer tumor growth, particularly in non-smoking patients[50–51].

Modulating the tumor microenvironment has emerged as a pivotal focus in contemporary cancer therapy[52–54]. As critical orchestrators of the tumor microenvironment, macrophages have become a key area of investigation in both tumor pathogenesis and therapeutic responses[55–56]. In the current study, we discovered that *CXCL10*$^+$ macrophages were

significantly enriched in smoking-associated macrophages. Previous studies have suggested that the infiltration of *CXCL10*$^+$ macrophages is associated with a more favorable response to atezolizumab, a PD-L1 inhibitor[34]. In our investigation, we also observed the expression of several immune checkpoint molecules in this macrophage subtype, including *PD-L1* and *HAVCR2*. Therefore, the infiltration of *CXCL10*$^+$ macrophages in smoking patients may also be associated with a favorable response to immunotherapy[49]. Currently, the primary immunotherapeutic approach targets the PD-1–PD-L1 axis. However, patients undergoing anti-PD-1/anti-PD-L1 immunotherapy may develop drug resistance, potentially due to the upregulation of other co-inhibitory molecules, such as HAVCR2[57]. We observed significant interaction between *CXCL10*$^+$ macrophages and other cells through the HAVCR2-LGALS9 axis, with *LGALS9* found to be upregulated in smoking patients. *In vitro* experiments demonstrated that anti-HAVCR2 antibodies could reverse anti-PD-1 resistance[58]. To date, multiple clinical trials evaluating anti-HAVCR2 and anti-LGALS9 therapies have been initiated, and some of them have shown promising results[58–63]. Our evidence suggests that targeting the HAVCR2-LGALS9 axis could represent a promising immunotherapy strategy for smoking patients.

Nevertheless, our study has several limitations. Firstly, we employed the Scissor algorithm, which identified smoking-associated and non-smoking-associated cell subpopulations by leveraging bulk phenotype information and the cell-sample correlation matrix. We did not apply Scissor to small cell subgroups, as these subgroups contributed minimally to bulk gene expression, and the results calculated by the Scissor algorithm could be inaccurate. However, these cell subtypes, such as fibroblasts, also play a crucial role in cancer development[64]. Secondly, we found increased infiltration of immunosuppressive *CXCL10*$^+$ macrophages in LCIS patients through bioinformatic analysis. However, the immunosuppressive properties of *CXCL10*$^+$ macrophages require experimental confirmation. Thirdly, future investigations are needed into anti-HAVCR2 and anti-LGALS9 immunotherapy strategies for LUAD patients in both experimental and clinical studies.

In conclusion, we observed that smoking-associated cancer cells displayed increased aggressiveness, which potentially correlates with the diminished survival rates observed in smoking patients. For non-smoking patients, reduced anti-tumor cytotoxicity and the inflammatory microenvironment were identified as potential contributors to their oncogenesis. Additionally, immunosuppressive *CXCL10*$^+$ macrophages were found to facilitate tumor progression in smoking patients, and the immunoregulatory LGALS9-HAVCR2 axis emerged as a promising immunotherapeutic target. Our findings illustrate the distinct oncogenic and immune suppression mechanisms in smoking and non-smoking LUAD patients, and present promising immunotherapeutic strategies.

Funding

This study was supported by the National Natural Science Foundation of China (Grant No. 82388102 to H.S.) and the Research Unit of Prospective Cohort of Cardiovascular Diseases and Cancer, Chinese Academy of Medical Sciences (Grant No. 2019RU038 to H.S.).

Acknowledgments

We sincerely acknowledge all enrolled patients for their invaluable participation in this study.

Author contributions

Zihuan Zhao: Conceptualization, Data curation, Formal analysis, Writing —original draft, Writing —review & editing. **Pan Yang**: Conceptualization, Data curation, Formal analysis, Writing—original draft, Writing—review & editing. **Yuhan Liu**: Data curation, Formal analysis, Writing —review & editing. **Kai Wang**: Formal analysis, Writing—original draft, Writing—review & editing. **Xianfeng Xu**: Formal analysis, Writing —review & editing. **Yuzhuo Wang**: Supervision, Writing—review & editing. **Meng Zhu**: Supervision, Writing —review & editing. **Na Qin**: Writing —original draft, Supervision, Writing —review & editing. **Cheng Wang**: Supervision, Writing —review & editing. **Weimin Li**: Project administration, Writing—review & editing. **Hongxia Ma**: Conceptualization, Writing —original draft. **Zhoufeng Wang**: Conceptualization, Project administration, Supervision, Writing —review & editing. **Hongbing Shen**: Conceptualization, Project administration, Funding acquisition, Supervision, Writing—review & editing.

Additional information

The online version contains supplementary

materials available at http://www.jbr-pub.org.cn/article/doi/10.7555/JBR.39.20250160?pageType=en.

References

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2024, 74(3): 229–263.

[2] Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers —a different disease[J]. *Nat Rev Cancer*, 2007, 7(10): 778–790.

[3] LoPiccolo J, Gusev A, Christiani DC, et al. Lung cancer in patients who have never smoked—an emerging disease[J]. *Nat Rev Clin Oncol*, 2024, 21(2): 121–146.

[4] Pelosof L, Ahn C, Gao A, et al. Proportion of never-smoker non-small cell lung cancer patients at three diverse institutions[J]. *J Natl Cancer Inst*, 2017, 109(7): djw295.

[5] Luo W, Zeng Z, Jin Y, et al. Distinct immune microenvironment of lung adenocarcinoma in never-smokers from smokers[J]. *Cell Rep Med*, 2023, 4(6): 101078.

[6] Devarakonda S, Li Y, Martins Rodrigues F, et al. Genomic profiling of lung adenocarcinoma in never-smokers[J]. *J Clin Oncol*, 2021, 39(33): 3747–3758.

[7] Zhou F, Zhou C. Lung cancer in never smokers—the East Asian experience[J]. *Transl Lung Cancer Res*, 2018, 7(4): 450–463.

[8] Chen J, Yang H, Teo ASM, et al. Genomic landscape of lung adenocarcinoma in East Asians[J]. *Nat Genet*, 2020, 52(2): 177–186.

[9] Hu Y, Chen G. Pathogenic mechanisms of lung adenocarcinoma in smokers and non-smokers determined by gene expression interrogation[J]. *Oncol Lett*, 2015, 10(3): 1350–1370.

[10] Li Q, Wang R, Yang Z, et al. Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq[J]. *Genome Med*, 2022, 14(1): 87.

[11] Sun D, Guan X, Moran AE, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data[J]. *Nat Biotechnol*, 2022, 40(4): 527–538.

[12] Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat[J]. *Nat Commun*, 2021, 12(1): 1088.

[13] Martos SN, Campbell MR, Lozoya OA, et al. Single-cell analyses identify dysfunctional $CD16^+$ CD8 T cells in smokers[J]. *Cell Rep Med*, 2020, 1(4): 100054.

[14] Duclos GE, Teixeira VH, Autissier P, et al. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution[J]. *Sci Adv*, 2019, 5(12): eaaw3413.

[15] Salcher S, Sturm G, Horvath L, et al. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer[J]. *Cancer Cell*, 2022, 40(12): 1503–1520.e8.

[16] Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner[J]. *Bioinformatics*, 2013, 29(1): 15–21.

[17] Zhang Y, Parmigiani G, Johnson WE. *ComBat-seq*: Batch effect adjustment for RNA-seq count data[J]. *NAR Genom Bioinform*, 2020, 2(3): lqaa078.

[18] Liu J, Lichtenberg T, Hoadley KA, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics[J]. *Cell*, 2018, 173(2): 400–416.e11.

[19] Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data[J]. *Cell*, 2021, 184(13): 3573–3587.e29.

[20] Korotkevich G, Sukhov V, Budin N, et al. Fast gene set enrichment analysis[PP/OL]. BioRxiv (2021-02-01) [2025-04-01]. https://www.biorxiv.org/content/10.1101/060012v3.

[21] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles[J]. *Proc Natl Acad Sci U S A*, 2005, 102(43): 15545–15550.

[22] Liberzon A, Birger C, Thorvaldsdóttir H, et al. The molecular signatures database hallmark gene set collection[J]. *Cell Syst*, 2015, 1(6): 417–425.

[23] Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-seq data[J]. *BMC Bioinformatics*, 2013, 14: 7.

[24] Bagaev A, Kotlov N, Nomie K, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy[J]. *Cancer Cell*, 2021, 39(6): 845–865.e7.

[25] Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis[J]. *Nature*, 2019, 566(7745): 496–502.

[26] Wang Z, Li Z, Zhou K, et al. Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing[J]. *Nat Commun*, 2021, 12(1): 6500.

[27] Fukui T, Shaykhiev R, Agosto-Perez F, et al. Lung adenocarcinoma subtypes based on expression of human airway basal cell genes[J]. *Eur Respir J*, 2013, 42(5): 1332–1344.

[28] Zhang L, Zhang Y, Wang C, et al. Integrated single-cell RNA sequencing analysis reveals distinct cellular and transcriptional modules associated with survival in lung cancer[J]. *Signal Transduct Target Ther*, 2022, 7(1): 9.

[29] O'Brien SJ, Fiechter C, Burton J, et al. Long non-coding RNA ZFAS1 is a major regulator of epithelial-mesenchymal transition through miR-200/ZEB1/E-cadherin, vimentin signaling in colon adenocarcinoma[J]. *Cell Death Discov*, 2021, 7(1): 61.

[30] Ji S, Zhang B, Liu J, et al. ALDOA functions as an oncogene in the highly metastatic pancreatic cancer[J]. *Cancer Lett*, 2016, 374(1): 127–135.

[31] Martins LR, Sieverling L, Michelhans M, et al. Single-cell division tracing and transcriptomics reveal cell types and differentiation paths in the regenerating lung[J]. *Nat*

Commun, 2024, 15(1): 2246.

[32] Zuo L, Prather ER, Stetskiv M, et al. Inflammaging and oxidative stress in human diseases: From molecular mechanisms to novel treatments[J]. *Int J Mol Sci*, 2019, 20(18): 4472.

[33] Couper KN, Blount DG, Riley EM. IL-10: The master regulator of immunity to infection[J]. *J Immunol*, 2008, 180(9): 5771–5777.

[34] Cappuyns S, Philips G, Vandecaveye V, et al. PD-1^{-} CD45RA^{+} effector-memory CD8 T cells and CXCL10^{+} macrophages are associated with response to atezolizumab plus bevacizumab in advanced hepatocellular carcinoma[J]. *Nat Commun*, 2023, 14(1): 7825.

[35] Zhang P, Cao L, Zhou R, et al. The lncRNA Neat1 promotes activation of inflammasomes in macrophages[J]. *Nat Commun*, 2019, 10(1): 1495.

[36] Mizumoto N, Takashima A. CD1a and langerin: Acting as more than Langerhans cell markers[J]. *J Clin Invest*, 2004, 113(5): 658–660.

[37] Sumaiya K, Langford D, Natarajaseenivasan K, et al. Macrophage migration inhibitory factor (MIF): A multifaceted cytokine regulated by genetic and physiological strategies[J]. *Pharmacol Ther*, 2022, 233: 108024.

[38] Wada J, Kanwar YS. Identification and characterization of galectin-9, a novel beta-galactoside-binding mammalian lectin[J]. *J Biol Chem*, 1997, 272(9): 6078–6086.

[39] Bugter JM, Fenderico N, Maurice MM. Mutations and mechanisms of WNT pathway tumour suppressors in cancer[J]. *Nat Rev Cancer*, 2021, 21(1): 5–21.

[40] Pradhan R, Singhvi G, Dubey SK, et al. MAPK pathway: A potential target for the treatment of non-small-cell lung carcinoma[J]. *Future Med Chem*, 2019, 11(8): 793–795.

[41] Malyla V, Paudel KR, De Rubis G, et al. Cigarette smoking induces lung cancer tumorigenesis *via* upregulation of the WNT/β-catenin signaling pathway[J]. *Life Sci*, 2023, 326: 121787.

[42] Zhang W, Zhang Y, Zhu Q. Cigarette smoke extract-mediated FABP4 upregulation suppresses viability and induces apoptosis, inflammation and oxidative stress of bronchial epithelial cells by activating p38 MAPK/MK2 signaling pathway[J]. *J Inflamm (Lond)*, 2022, 19(1): 7.

[43] Axelrod ML, Cook RS, Johnson DB, et al. Biological consequences of MHC-Ⅱ expression by tumor cells in cancer[J]. *Clin Cancer Res*, 2019, 25(8): 2392–2402.

[44] Holling TM, Schooten E, van Den Elsen PJ. Function and regulation of MHC class Ⅱ molecules in T-lymphocytes: Of mice and men[J]. *Hum Immunol*, 2004, 65(4): 282–290.

[45] Johnson AM, Bullock BL, Neuwelt AJ, et al. Cancer cell-intrinsic expression of MHC class Ⅱ regulates the immune microenvironment and response to anti-PD-1 therapy in lung adenocarcinoma[J]. *J Immunol*, 2020, 204(8): 2295–2307.

[46] Park IA, Hwang SH, Song IH, et al. Expression of the MHC class Ⅱ in triple-negative breast cancer is associated with tumor-infiltrating lymphocytes and interferon signaling[J]. *PLoS One*, 2017, 12(8): e0182786.

[47] Bawden EG, Wagner T, Schröder J, et al. CD4^{+} T cell immunity against cutaneous melanoma encompasses multifaceted MHC Ⅱ-dependent responses[J]. *Sci Immunol*, 2024, 9(91): eadi9517.

[48] Callahan MJ, Nagymanyoki Z, Bonome T, et al. Increased HLA-DMB expression in the tumor epithelium is associated with increased CTL infiltration and improved prognosis in advanced-stage serous ovarian cancer[J]. *Clin Cancer Res*, 2008, 14(23): 7667–7673.

[49] El-Osta H, Jafri S. Predictors for clinical benefit of immune checkpoint inhibitors in advanced non-small-cell lung cancer: A meta-analysis[J]. *Immunotherapy*, 2019, 11(3): 189–199.

[50] Wang C, Yin R, Dai J, et al. Whole-genome sequencing reveals genomic signatures associated with the inflammatory microenvironments in Chinese NSCLC patients[J]. *Nat Commun*, 2018, 9(1): 2054.

[51] Lahiri A, Maji A, Potdar PD, et al. Lung cancer immunotherapy: Progress, pitfalls, and promises[J]. *Mol Cancer*, 2023, 22(1): 40.

[52] Bejarano L, Jordão MJC, Joyce JA. Therapeutic targeting of the tumor microenvironment[J]. *Cancer Discov*, 2021, 11(4): 933–959.

[53] Guo Z, Saw PE, Jon S. Non-invasive physical stimulation to modulate the tumor microenvironment: Unveiling a new frontier in cancer therapy[J]. *BIO Integr*, 2024, 5(1): 986.

[54] Luo Z, Mei J, Wang X, et al. Voluntary exercise sensitizes cancer immunotherapy *via* the collagen inhibition-orchestrated inflammatory tumor immune microenvironment[J]. *Cell Rep*, 2024, 43(9): 114697.

[55] Xu Y, Zeng Y, Xiao X, et al. Targeted imaging of tumor associated macrophages in breast cancer[J]. *BIO Integr*, 2023, 4(3): 114–124.

[56] Tang J, He J, Guo H, et al. PTBP2-mediated alternative splicing of IRF9 controls tumor-associated monocyte/macrophage chemotaxis and repolarization in neuroblastoma progression[J]. *Research (Wash D C)*, 2023, 6: 0033.

[57] Koyama S, Akbay EA, Li YY, et al. Adaptive resistance to therapeutic PD-1 blockade is associated with upregulation of alternative immune checkpoints[J]. *Nat Commun*, 2016, 7: 10501.

[58] Limagne E, Richard C, Thibaudin M, et al. Tim-3/galectin-9 pathway and mMDSC control primary and secondary resistances to PD-1 blockade in lung cancer patients[J]. *Oncoimmunology*, 2019, 8(4): e1564505.

[59] Lv Y, Ma X, Ma Y, et al. A new emerging target in cancer immunotherapy: Galectin-9 (LGALS9)[J]. *Genes Dis*, 2023, 10(6): 2366–2382.

[60] A phase 1, open-label, multicenter trial investigating the safety, tolerability, and preliminary antineoplastic activity of Sym023 (anti-TIM-3) in patients with advanced solid tumor malignancies or lymphomas[EB/OL]. [2018-04-05]. https://clinicaltrials.gov/study/NCT03489343.

[61] TIM3 inhibition with MBG453 for patients with lower risk MDS: An adaptive two-stage phase Ⅱ clinical trial[EB/OL].

[2025-04-01]. https://www.dana-farber.org/clinical-trials/20-637.

[62] Curigliano G, Gelderblom H, Mach N, et al. Phase I / I b clinical trial of sabatolimab, an anti-TIM-3 antibody, alone and in combination with spartalizumab, an anti-PD-1 antibody, in advanced solid tumors[J]. *Clin Cancer Res*, 2021, 27(13): 3620–3629.

[63] Davar D, Boasberg P, Eroglu Z, et al. A phase 1 study of TSR-022, an anti-TIM-3 monoclonal antibody, in combination with TSR-042 (anti-PD-1) in patients with colorectal cancer and post-PD-1 NSCLC and melanoma[J]. *J Immunother Cancer*, 2018, 6(S1): 115O21.

[64] Hanley CJ, Waise S, Ellis MJ, et al. Single-cell analysis reveals prognostic fibroblast subpopulations linked to molecular and immunological subtypes of lung cancer[J]. *Nat Commun*, 2023, 14(1): 387.

CLINICAL TRIAL REGISTRATION

The *Journal* requires investigators to register their clinical trials in a public trials registry for publication of reports of clinical trials in the *Journal*. Information on requirements and acceptable registries is available at **https: //clinicaltrials.gov/.**