Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ECM-CAFs Induce Myeloid M2 Polarization via the HLA- DRA-CD4 Axis and Contribute to Poor Prognosis in Esophageal Squamous Cell Carcinoma

YUJIE WENG
Inner Mongolia Medical University
RONG JIA
Inner Mongolia Medical University
HUIRU ZHAO
Inner Mongolia Medical University
PENGFEI NING
ningpengfei@immu.edu.cn

Inner Mongolia Medical University

Research Article

Keywords: esophageal squamous cell carcinoma, single-cell RNA sequencing, cancer-associated fibroblasts, myeloid cells, prognosis

Posted Date: October 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10979345/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Esophageal squamous cell carcinoma (ESCC) has a poor prognosis. The heterogeneity of cancer-associated fibroblasts (CAFs) within the tumor microenvironment and their interactions with immune cells warrant further investigation. This study employed single-cell RNA sequencing to analyze the subtypes and differentiation trajectories of CAFs in ESCC, as well as to explore the mechanisms underlying CAF–myeloid cell interactions and their prognostic significance.

Methods

Single-cell transcriptomic data for ESCC were downloaded from the GEO database, followed by quality control, dimensionality reduction, and clustering. Major cell types were identified using marker genes. CAF subtypes were characterized, differentiation trajectories were constructed, and Scissor was employed to identify cell subpopulations associated with prognosis. CellChat was used to examine receptor-ligand interactions between CAFs and prognosis-associated cell populations, while the M1/M2 polarization score assessed the polarization direction of myeloid cells. For prognostic validation, CIBERSORT estimated the abundance of M2 macrophages and evaluated their association with overall survival. Cross-validation was performed using an M2 marker gene module score independent of CIBERSORT. Finally, the prognostic relevance and cross-cohort consistency of key molecules were validated in the TCGA-ESCC dataset and two independent GEO cohorts.

Results

A total of 11 major cell types were annotated. Among the 2,749 cancer-associated fibroblast (CAF) cells, extracellular matrix (ECM)-CAFs accounted for 79% of tumor CAFs. Prognostic analysis revealed that 82.3% of the prognosis-associated cells identified by Scissor were myeloid cells. A subpopulation with high HLA expression was present within ECM-CAFs, and it was found that ECM-CAFs communicate with myeloid cells via the HLA-DRA-CD4 axis. Notably, 65.5% of myeloid cells expressed CD4 and exhibited M2 polarization. Batch transcriptomic validation yielded consistent results: HLA-DRA expression was positively correlated with M2 macrophage abundance, CD4 expression was associated with the M2 module score, and both the HLA-DRA and CD4 high-expression groups were linked to poorer prognosis. M2 macrophage abundance was identified as an independent prognostic factor, with consistent risk ratio directions observed across two GEO cohorts.

Conclusion

ECM-CAFs are the predominant CAF subtype in ESCC tumors. Their internal subpopulation, characterized by high HLA expression, interacts with myeloid cells through the HLA-DRA-CD4 axis, promoting the polarization of CD4-positive myeloid cells toward the M2 phenotype. The abundance of M2 macrophages, confirmed by two independent methods—CIBERSORT deconvolution and signature gene module scoring—is associated with poor prognosis.

1. Background

Esophageal cancer is a common malignant tumor of the digestive system worldwide, with esophageal squamous cell carcinoma (ESCC) as its primary subtype, which is exhibiting a rising incidence. The majority of ESCC patients are diagnosed at intermediate or advanced stages, with a five-year survival rate of less than 30 percent [1]. Although treatment methods continue to advance, improvements in patient prognosis remain limited, highlighting the need for a deeper understanding of the molecular mechanisms underlying its etiology and progression.

The tumor microenvironment plays a crucial role in the development and progression of ESCC. Fibroblasts, a major component of the tumor stroma, promote tumor growth, invasion, and immune evasion through mechanisms such as extracellular matrix (ECM) deposition and the secretion of growth factors and cytokines [2]. Single-cell transcriptomic studies have revealed that cancer-associated fibroblasts (CAFs) exhibit significant transcriptomic heterogeneity across various diseases [3].

Myeloid cells represent a significant population of immune cells within the tumor microenvironment. In various tumors, the M2 polarization of tumor-associated macrophages is closely linked to tumor progression and poor prognosis [4]. However, the mechanisms underlying the interactions between CAFs and myeloid cells, as well as their impact on the prognosis of ESCC, remain to be elucidated.

Esophageal squamous cell carcinoma is characterized by significant prognostic heterogeneity, with the tumor microenvironment playing a key role. This study begins by identifying CAFs subtypes using single-cell transcriptomic data and progressively analyzes their differentiation trajectories, prognostic associations, intercellular communication, and influence on myeloid cell polarization. These findings are validated using bulk transcriptomic data, aiming to provide new insights for prognostic assessment and immunotherapeutic interventions in esophageal squamous cell carcinoma.

2. Methods

2.1 Samples and Data Sources

The single-cell transcriptomic data used in this study were obtained from the GSE196756 dataset in the GEO database. This dataset comprises six esophageal tissue samples: three non-cancerous tissues adjacent to the tumor and three samples of esophageal squamous cell carcinoma. For prognostic validation, transcriptomic and clinical data for TCGA-ESCC were downloaded from the UCSC Xena

database [5, 6] and used to assess the association between key molecules in the HLA-DRA-CD4-M2 axis, as well as the relationship between M2 macrophage abundance and prognosis. Additionally, microarray expression profiles and survival data for two independent ESCC cohorts, GSE53624 and GSE53622, were obtained from the GEO database [7] to serve as external validation sets for the prognostic significance of M2 macrophage abundance.

2.2 Single-Cell Data Quality Control and Preprocessing

Data analysis was performed using Seurat v5 [8]. The quality control criteria were as follows: the number of genes had to be greater than 500 and less than 6,000; the UMI count had to be greater than 1,000 and less than 50,000; the proportion of mitochondrial genes had to be less than 25%; and the proportion of hemoglobin genes had to be less than 5%. Normalization was performed using SCTransform, with the influence of mitochondrial gene proportion regressed out [9]. UMAP dimensionality reduction and neighbor-based graph construction were conducted based on the top twenty principal components [10], and clustering was performed using the Louvain algorithm [11]. The resolution was determined to be 0.6 following a grid search ranging from 0.2 to 0.8.

2.3 Cell Type Annotation

Automated cell type annotation was conducted using SingleR alongside the BlueprintEncode and DatabaseImmuneCellExpressionData reference datasets [12, 13]. This process was supplemented by manual verification and correction based on classic marker genes for each cell type. For clusters in the annotation results where marker gene expression patterns were unclear or inconsistent with the clustering distribution, cluster-by-cluster corrections were applied in combination with differential expression analysis to definitively identify each cell population.

2.4 Analysis of CAF Subtypes

All fibroblasts were isolated for subclustering analysis. Principal component analysis, UMAP dimensionality reduction, and Louvain clustering were re-performed on 2,749 CAF cells. Gene set scoring was conducted based on the functional programs of CAFs reported in the literature [14, 15], and CAF subtypes were annotated by integrating the expression profiles of marker genes for each subgroup.

2.5 Pseudotemporal Analysis

Trajectory inference was performed on CAF cells using Monocle3 [16], with resting CAFs designated as the starting point of the trajectory. Differentiation trajectories were constructed, and their branching structures were analyzed. The pseudo-temporal distribution of each CAF subtype within the branches was calculated. Proliferation scores and the average number of detected genes per cell were used as surrogate indicators of cellular maturity to compare the maturity states of ECM-CAF and apCAF.

2.6 Identification of Prognostically Relevant Cell Subpopulations

The Scissor method was employed to integrate all CAF cells with 8,000 randomly selected non-CAF cells. The Cox proportional hazards model was then used to identify cell subpopulations associated with prognosis [17], applying a screening threshold of cutoff = 0.2. Enrichment analysis was conducted on Scissor-positive cells to evaluate both the proportion of these cells within each cell type and the strength of their association with prognosis.

2.7 Analysis of Intercellular Communication

CellChat was utilized to infer ligand-receptor interactions between different cell types [18], with a focus on signaling from ECM-CAFs to myeloid cells. Significant ligand-receptor pairs transmitted from ECM-CAFs to myeloid cells were identified, and the signal strength of the HLA-DRA-CD4 axis was compared with that of the classical ECM-associated signaling axis.

2.8 Analysis of M1 and M2 Polarization in Myeloid Cells

In tumor myeloid cells, a linear mixed-effects model was constructed with the sample as a random effect and continuous CD4 expression as a fixed effect [19] to assess the association between CD4 expression and M1 and M2 scores. The M1 and M2 module scores were calculated using the average expression levels of the corresponding gene sets [20], while Spearman correlation coefficients between CD4 expression and the M1 and M2 scores were also computed [21]. Descriptive comparisons between the CD4-positive and CD4-negative groups were performed using the Wilcoxon test.

2.9 Batch Transcriptomic Immune Deconvolution and Survival Analysis

Immune cell abundance deconvolution was performed using the local implementation of CIBERSORT, employing the LM22 signature matrix with 1,000 permutations and quantile normalization [22]. The analysis cohorts included TCGA-ESCC, GSE53624, and GSE53622. Univariate and multivariate Cox proportional hazards models were used to assess the association between M2 macrophage abundance and overall survival [23]. The multivariate model was adjusted for clinical stage, and the hazard ratio was calculated for each interquartile range increment.

To cross-validate the CIBERSORT results, an M2 module score was constructed using M2 marker genes compiled from 39 literature sources. The average expression values were calculated after standardizing each gene based on its mean and standard deviation within the discovery set GSE53624. Nine of these genes, including CD4, are members of the LM22 signature matrix; however, using CIBERSORT M2 abundance to directly assess the association between CD4 and M2 polarization would constitute circular reasoning. Therefore, these nine LM22 genes were excluded, resulting in a total of 29 genes used to construct an independent M2 module score (SELENOP was excluded because it was not detected in the TCGA dataset). Furthermore, using the union of ECM-CAF characteristic genes and M2 marker genes as candidates, Cox-LASSO regression was applied in the discovery set GSE53624 to

identify genes associated with prognosis [24]. The cross-cohort generalizability of these genes was then validated in the TCGA-ESCC dataset.

2.10 Statistical Analysis

Differences in cell type proportions and Scissor-positive cell enrichment were analyzed using Fisher's exact test [25]. The association between M1 and M2 scores and CD4 expression was examined using a linear mixed-effects model, with samples treated as random effects. Survival analysis was conducted using the Cox proportional hazards model and the Kaplan–Meier method [26], with intergroup comparisons of Kaplan–Meier curves performed using the log-rank test [27]. Correlation analysis was carried out using Spearman's rank correlation coefficient. A p-value less than 0.05 was considered statistically significant.

3. Results

3.1 Single-Cell Transcriptomic Profiles of ESCC

In this study, single-cell transcriptomic sequencing was conducted on tumor tissues from three cases of ESCC and three matched adjacent non-tumor tissues. The raw dataset comprised a total of 36,515 cells; after quality control filtering, 28,284 high-quality cells were retained for downstream analysis. The filtering criteria included selecting cells with gene counts between 500 and 6,000, UMI counts between 1,000 and 50,000, a mitochondrial gene proportion below 25%, and a hemoglobin gene proportion below 5% (Fig. 1A). The quality control scatter plot demonstrated that the vast majority of cells fell within these filtering thresholds (Fig. 1B).

Based on UMAP dimensionality reduction and unsupervised clustering, combined with the SingleR reference dataset and manual correction using signature genes, a total of 11 major cell types were annotated (Fig. 1C). The UMAP plot, colored by sample origin, shows that tumor cells and normal cells are regionally separated within the transcriptomic space (Fig. 1D).

Dot plots of classical signature genes validated the accuracy of the annotations for each cell type (Fig. 1E). Tumor epithelial cells express squamous epithelial marker genes such as KRT5, KRT14, and TP63. Notably, the expression levels of KRT14 and TP63 are higher in tumor epithelium than in normal epithelium, consistent with the known squamous differentiation characteristics of ESCC. These results indicate that the quality of this dataset and the reliability of the cell annotations are sufficient for subsequent analyses.

3.2 Three Functional Subtypes of CAFs

A total of 2,749 fibroblasts were extracted from the full dataset and subjected to an independent subclustering analysis, which identified three subtypes: ECM-CAF, resting CAF, and apCAF (Fig. 2A).

The expression of marker genes for each subtype is illustrated in the dot plot (Fig. 2B). ECM-CAFs exhibit high expression of extracellular matrix (ECM) remodeling-related genes, including MMP1, POSTN, and COL10A1. Resting-CAFs predominantly express resting fibroblast marker genes such as DCN, PI16, and C7. Additionally, apCAFs show elevated expression of antigen-presentation-related genes, including HLA-DRA and CD74.

ECM-CAFs and apCAFs are enriched in tumor tissue, whereas resting-CAFs are primarily found in normal tissue. ECM-CAFs constitute 79 percent of tumor CAFs and represent the dominant CAF subtype within tumors (Figs. 2C–D). These results suggest that fibroblasts in the ESCC tumor microenvironment transition from a resting state to an activated state associated with extracellular matrix remodeling.

3.3 CAF Differentiation Trajectory

Pseudo-temporal analysis was performed using Monocle3, with resting CAF set as the starting point of the trajectory, to reconstruct the CAF differentiation pathway. The analysis revealed that CAFs exhibit a bifurcated structure with a common main branch (Fig. 3A).

Purity analysis of terminal nodes showed that ECM-CAF accounted for 99 percent of the nodes at Y_108 (Fig. 3B). Pseudo-timeline position maps for each CAF subtype revealed that ECM-CAF is located at the very end of the branch, whereas apCAF only reaches the midpoint, indicating that ECM-CAF represents a true terminal differentiation state, while apCAF does not (Fig. 3C).

A comparison of maturity proxy indicators showed that ECM-CAF had a proliferation score of 0.11 and an average of 4,435 detected genes per cell. In contrast, apCAF had a proliferation score of 0.89 and an average of 1,836 genes detected per cell (Fig. 3D). These findings further demonstrate that ECM-CAF represents a mature, stable terminal state, whereas apCAF represents a proliferatively active transitional state, providing additional support for ECM-CAF as the endpoint of differentiation.

3.4 Identification of Cell Subsets Associated with Prognosis Using Scissor

Using the Scissor method, a combined analysis was conducted on all CAF cells and a subsample of 8,000 non-CAF cells, integrated with external survival data. A total of 468 Scissor-positive cells associated with poor prognosis were identified. Mapping these cells onto the whole-cell UMAP space revealed that the majority were located within the myeloid cell region (Fig. 4A).

Among the Scissor-positive cells, 82.3% were myeloid cells (Fig. 4B). Enrichment analysis by cell type revealed that Scissor-positive cells comprised 41.2% of myeloid cells, the only cell type exhibiting significant enrichment. Plasma cells showed a slight enrichment, while the remaining cell types did not reach statistical significance (Fig. 4C). Kaplan–Meier survival curves indicated that patients in the high-expression group for the Scissor-positive signature had significantly shorter overall survival compared to those in the low-expression group (Fig. 4D).

Gene enrichment analysis revealed that the apCAF signature was significantly enriched among differentially expressed genes in Scissor-positive cells, whereas the ECM-CAF signature was not enriched, and the resting signature showed negative enrichment (Fig. 4E). HLA module scores across cell types indicated that myeloid cells had the highest scores, followed by B cells, with fibroblasts exhibiting the lowest scores (Fig. 4F). This suggests that the apCAF signature enriched in Scissor-positive cells primarily originates from myeloid cells rather than from CAFs themselves.

These results indicate that the cell population associated with poor prognosis in ESCC consists of myeloid cells, and that CAFs may indirectly influence patient outcomes by regulating the HLA-related functions of myeloid cells, rather than directly driving poor prognosis.

3.5 The HLA-DRA–CD4 Axis Mediates Communication between ECM-CAFs and Myeloid Cells

Analysis of cell-cell communication revealed that, among the signals transmitted by ECM-CAFs to various immune cells, T cells exhibited the highest number of ligand-receptor pairs, followed by myeloid cells (Fig. 5A). CellChat analysis identified a total of 41 significant ligand-receptor pairs transmitted by ECM-CAFs to myeloid cells, with the strongest signal intensity observed for the HLA-DRA-CD4 axis (Fig. 5B). This finding indicates that ECM-CAFs primarily communicate with myeloid cells through this axis.

A further comparison of the communication probabilities between the HLA-DRA-CD4 axis and classical ECM communication axes revealed that the communication probability of the HLA-DRA-CD4 axis was significantly higher than that of classical ECM receptor-ligand pairs, such as COL1A1-CD44 (Figs. 5C–5D). These results indicate that communication between ECM-CAFs and myeloid cells does not depend on classical ECM pathways but is instead dominated by the HLA-DRA-CD4 axis.

3.6 Identification of the HLA-High-Expressing ECM-CAF Subpopulation

HLA-DRA exhibited a bimodal distribution in ECM-CAFs, with 14.6 percent comprising the HLA-high-expressing subpopulation (Fig. 6A). This subpopulation formed a distinct cluster in the UMAP space (Fig. 6B). After excluding interference from dimers (Fig. 6C), the HLA-high-expressing cells specifically overexpressed antigen-presentation-related genes (HLA-DRA, HLA-DRB1, HLA-DPB1, CD74), while there were no significant changes in the expression of immune cell marker genes, indicating that this was not due to immune cell contamination (Fig. 6D). These results suggest that HLA-high-expressing ECM-CAFs constitute a distinct subpopulation within ECM-CAFs with immunoregulatory potential.

3.7 CD4-Positive Myeloid Cells Exhibit M2 Polarization

Sixty-five point five percent of tumor myeloid cells expressed CD4, indicating that myeloid cells possess the receptor necessary to receive HLA-DRA signals (Fig. 7A). A linear mixed-effects model was constructed, with the sample as a random effect and continuous CD4 expression as a fixed effect. This analysis revealed that the M2 score of CD4-positive myeloid cells was significantly higher than that of CD4-negative cells (Wilcoxon $p = 3.3 \times 10^{-27}$) (Fig. 7B). Spearman's correlation analysis further demonstrated that CD4 expression levels were positively correlated with the M2 score, while showing a weaker correlation with the M1 score (Fig. 7C). These results indicate that CD4-positive myeloid cells exhibit M2 polarization.

3.8 Batch Transcriptomic Validation of the Mechanistic Pathway

To verify whether the HLA-DRA-CD4-M2 polarization axis identified at the single-cell level is reproducible in independent datasets, this study conducted correlation analyses and survival validation using batch transcriptomic data from TCGA-ESCC.

First, the association between HLA-DRA expression and M2 macrophage abundance was evaluated. HLA-DRA expression demonstrated a significant positive correlation with M2 macrophage abundance, as estimated by CIBERSORT (Fig. 8A). Additionally, the MHC-II module score—calculated as the standardized mean expression of four genes: HLA-DRA, HLA-DRB1, HLA-DPB1, and HLA-DQB1—also exhibited a significant positive correlation with M2 macrophage abundance (Fig. 8B). These findings indicate a consistent association between HLA-DRA, its related MHC-II molecular module, and the degree of M2 macrophage infiltration.

Secondly, the relationship between CD4 expression and M2 polarization was assessed. In this study, an independent M2 module score—excluding all LM22 members—was used as an indicator of M2 polarization. The results demonstrated a strong positive correlation between CD4 expression and this independent M2 module score (Fig. 8C). A comparative analysis using the complete module, comprising all 39 M2 marker genes, yielded consistent findings (Fig. 8D).

To validate the reliability of the data, positive and negative controls were established in this study. CD163, a classic marker of M2 macrophages, exhibited a significant positive correlation between its expression and the M2 abundance estimated by CIBERSORT, whereas the gene GAPDH showed no correlation with M2 abundance (Fig. 8E), confirming the validity of the analytical method.

Finally, survival analysis was conducted for validation. Patients with high HLA-DRA expression had significantly shorter overall survival than those with low expression (Fig. 8F). Similarly, high CD4 expression was associated with poor prognosis (Fig. 8G). Univariate Cox regression analysis revealed that for every increase of one interquartile range in CD4 expression, the risk of death increased by 49%. Additionally, for every increase of one interquartile range in the M2 module score, the risk of death increased by 67% (Fig. 8H).

Validation using TCGA-ESCC batch transcriptomics independently confirmed the HLA-DRA–CD4–M2 polarization axis identified at the single-cell level. HLA-DRA was positively correlated with M2 abundance, while CD4 was similarly positively correlated with the M2 module. Furthermore, both CD4 and the M2 module were significantly associated with poor prognosis. These findings support the mechanistic hypothesis that ECM-CAFs drive myeloid M2 polarization via the HLA-DRA–CD4 axis.

3.9 Association Between M2 Macrophage Abundance and ESCC Prognosis

To independently validate the association between M2 polarization and prognosis, CIBERSORT was used to perform immune cell abundance deconvolution on batch transcriptomic data. The coverage of the LM22 gene signature was 94% in TCGA and 85% in GSE53624 and GSE53622 datasets.

In the TCGA-ESCC cohort, a 50% increase in overall survival risk was observed for each interquartile range increase in M2 macrophage abundance (Fig. 9A). After adjusting for clinical stage, this association was further strengthened, indicating that M2 abundance is an independent prognostic factor. In two independent GEO cohorts, GSE53624 and GSE53622, the direction of the hazard ratio for M2 abundance was consistent but did not reach statistical significance (Fig. 9B). A pooled forest plot across the three cohorts demonstrated a consistent hazard ratio direction, supporting the cross-cohort consistency of the association between M2 macrophage abundance and prognosis (Fig. 9C).

To further confirm that the above association was not an artifact of a specific deconvolution method, cross-validation was performed using an M2 signature module score independent of CIBERSORT. The results showed that, in the TCGA-ESCC cohort, a one-interquartile range increase in the M2 module score was associated with a 66% increase in the risk of death, which was highly consistent with the CIBERSORT deconvolution results (Fig. 9D). The fact that two independent methods simultaneously validated the prognostic association of M2 further supports the robustness of this finding.

4. Discussion

This study elucidated the heterogeneity of fibroblasts in ESCC through single-cell transcriptomic analysis, revealing that ECM-CAFs are the predominant CAF subtype in the tumor stroma and are enriched in tumor tissue. Trajectory analysis indicated that these cells are in a state of terminal differentiation. However, Scissor analysis suggested that the core cell population associated with prognosis consists of myeloid cells rather than fibroblasts themselves, prompting further investigation into the mechanisms underlying the interaction between CAFs and myeloid cells. An HLA-high-expressing subpopulation was identified within ECM-CAFs, and CellChat analysis indicated communication with myeloid cells via the HLA-DRA-CD4 axis. Functional analysis further revealed that CD4-positive myeloid cells exhibit significant M2 polarization characteristics.

The association of key molecules in this mechanistic pathway was independently validated using the TCGA-ESCC batch transcriptomic data. HLA-DRA expression was positively correlated with M2 macrophage abundance, while CD4 expression was positively correlated with an independent M2 module score. Furthermore, high expression levels of both HLA-DRA and CD4, as well as the M2 module score, were associated with poor prognosis. This cross-method validation confirmed the communication inferred at the single-cell level within the batch data, thereby enhancing the credibility of the mechanistic hypothesis.

Regarding prognostic validation, CIBERSORT deconvolution confirmed that M2 macrophage abundance is an independent predictor of poor prognosis in TCGA-ESCC. This finding was consistent when cross-validated using an independent M2 module score. In two independent GEO cohorts, GSE53624 and GSE53622, the direction of the hazard ratio for M2 abundance was consistent but did not reach statistical significance. This may be due to differences in sample size and gene coverage across microarray platforms.

Although ECM-CAFs are the predominant CAF subtype in the ESCC tumor stroma and interact with myeloid cells via the HLA-DRA-CD4 axis, the ECM-CAF signature genes themselves do not exhibit independent prognostic value. These findings suggest that the impact of ECM-CAFs on prognosis is primarily mediated through their indirect role in promoting the polarization of myeloid cells within the tumor microenvironment, rather than through the direct influence of ECM-CAF gene expression. This also indicates that, in prognostic modeling of the tumor microenvironment, genetic signatures reflecting the state of immune cells may have greater prognostic significance than those reflecting stromal remodeling.

The causal relationship between the HLA-DRA-CD4 signaling axis and myeloid M2 polarization is primarily inferred from correlation analyses of single-cell and bulk transcriptomic data; it has not yet been directly validated through in vitro experiments and requires further experimental confirmation.

4.1 Limitations and Future Directions

Although this study provides valuable insights into the heterogeneity of CAFs in ESCC and their interactions with myeloid cells, it has several limitations. First, the sample size of the single-cell transcriptomic data is small and may not fully capture the heterogeneity of the ESCC tumor microenvironment. Second, although the HLA-DRA-CD4-M2 axis was validated in the TCGA-ESCC cohort, the prognostic association of M2 macrophage abundance did not reach statistical significance in two independent GEO cohorts; this finding requires validation in larger sample sets. Third, all analyses in this study were based on retrospective public data and lacked in vitro and in vivo experimental validation; therefore, the causal relationship between the HLA-DRA-CD4 axis and M2 polarization requires further experimental confirmation. Future research should focus on elucidating the specific molecular mechanisms underlying this signaling axis and evaluating its translational potential as a therapeutic target for ESCC.

5. Conclusions

ECM-CAF is the predominant CAF subtype in ESCC tumors; its HLA-high-expressing subpopulation promotes myeloid M2 polarization through the HLA-DRA-CD4 axis. The abundance of M2 macrophages is independently associated with poor prognosis in ESCC, whereas the ECM-CAF gene itself has no independent prognostic value.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

GSE196756 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE196756)

TCGA-ESCC (https://xenabrowser.net/datapages/)

GSE53624(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53624)
GSE53622(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53622)

Competing interests

Not applicable

Funding

Natural Science Foundation of the Inner Mongolia Autonomous Region

Authors' contributions

P.N. was responsible for the research concept and design, as well as overall supervision of the project. Y.W. was responsible for establishing the methodology, conducting the formal analysis, implementing the software and organising the data. R.J. and R.H. were responsible for the literature review, data interpretation and drafting the manuscript. All authors reviewed the manuscript.

Acknowledgements

Not applicable

Clinical trial number

Not applicable

References

1. Chamil C, Kenneth K. Squamous Cell Carcinoma of the Esophagus. Gastroenterol Clin North Am 2022, 51(3):457–84.
2. Chang J, Lu J, Liu Q et al. Single-cell multi-stage spatial evolutional map of esophageal carcinogenesis.Cancer Cell,2025, 43(3):380–97.
3. Zhan J, Li M, Li L et al. Targeting LTBP2 Derived from Cancer-Associated Fibroblasts Sensitizes Esophageal Squamous Cell Carcinoma to Chemotherapy.Cancer Res,2025, 85(13):2412–28.
4. Wang Y, Lyu Z, Qin Y et al. FOXO1 promotes tumor progression by increased M2 macrophage infiltration in esophageal squamous cell carcinoma.Theranostics,2020, 10(25):11535–48.
5. Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017;541(7636):169–75.
6. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675–8.
7. Li J, Chen Z, Tian L, et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut. 2014;63(11):1700–10.
8. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304.
9. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.
10. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. J Open Source Softw. 2018;3(29):861.
11. Blondel VD, Guillaume JL, Lambiotte R, et al. Fast unfolding of communities in large networks. J Stat Mech: Theory Exp. 2008;2008(10):P10008.
12. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163–72.
13. Monaco G, Lee B, Xu W, et al. RNA-seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. Cell Rep. 2019;26(6):1627–40.
14. Öhlund D, Handly-Santana A, Biffi G, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med. 2017;214(3):579–96.
15. Elyada E, Bolisetty M, Laise P, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov. 2019;9(8):1102–23.
16. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566(7745):496–502.

17. Sun D, Guan X, Moran AE, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. Nat Biotechnol. 2022;40(4):527–38.
18. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088.
19. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48.
20. Murray PJ, Allen JE, Biswas SK, et al. Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity. 2014;41(1):14–20.
21. Spearman C. The proof and measurement of association between two things. Am J Psychol. 1904;15(1):72–101.
22. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7.
23. Cox DR. Regression models and life-tables. J Roy Stat Soc: Ser B (Methodol). 1972;34(2):187–220.
24. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.
25. Fisher RA. The Design of Experiments. Edinburgh: Oliver & Boyd; 1935.
26. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457–81.
27. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966;50(3):163–70.

Figures

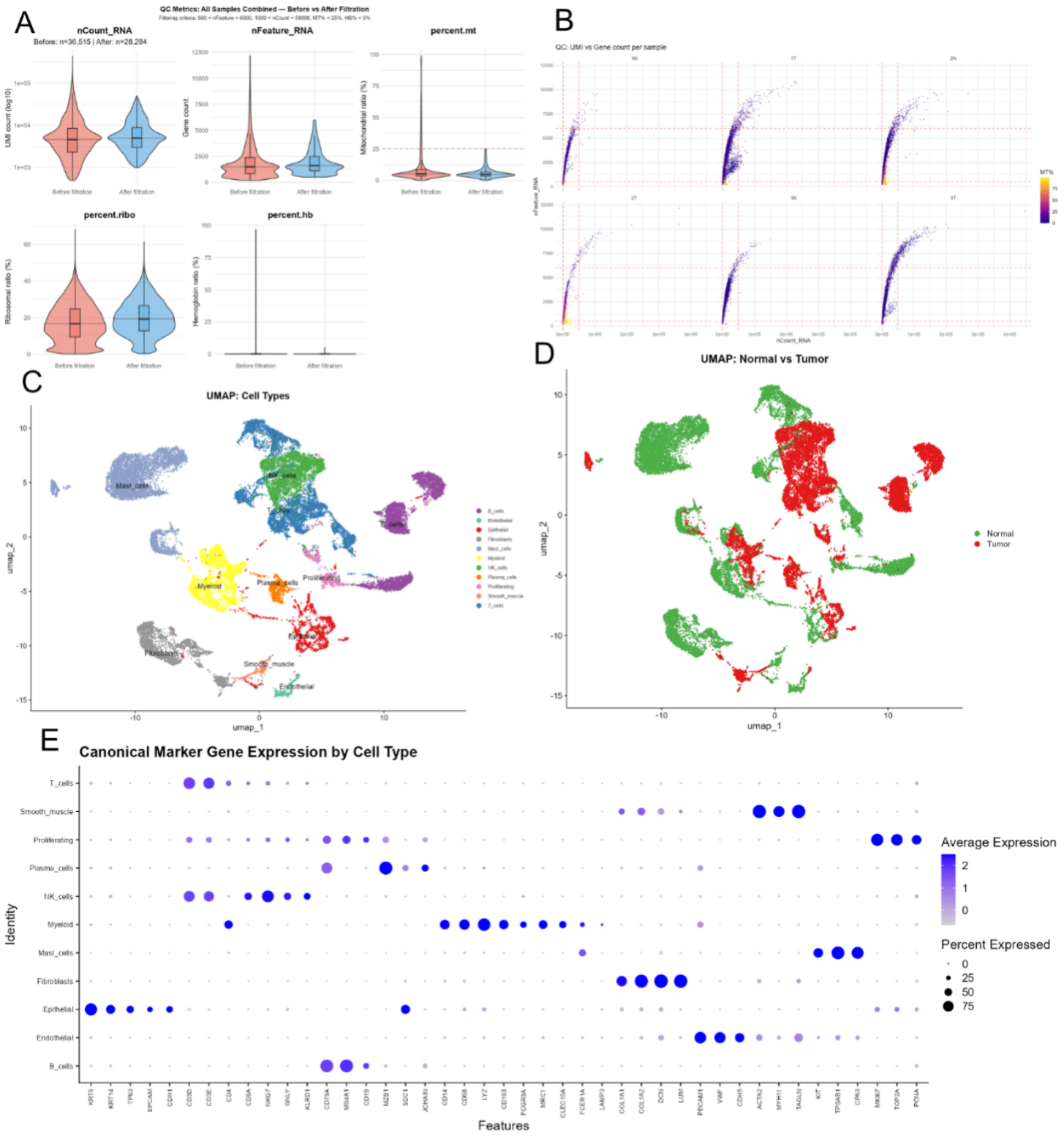


Figure 1

Single-cell transcriptomic profile of ESCC. (A) Bar charts showing cell counts before and after quality control for each sample. (B) Quality control scatter plot colored according to the proportion of mitochondrial genes, with the red dotted line indicating the filtering threshold boundary. (C) UMAP dimension-reduction plot showing the clustering distribution of 11 cell types in transcriptomic space. (D) UMAP dimension-reduction plot colored according to sample origin, with green representing cells from

normal tissue and red representing cells from tumor tissue. (E) Dot plots of classic marker genes for each cell type, where dot size represents the proportion of cells expressing the gene, and color represents the average expression level.

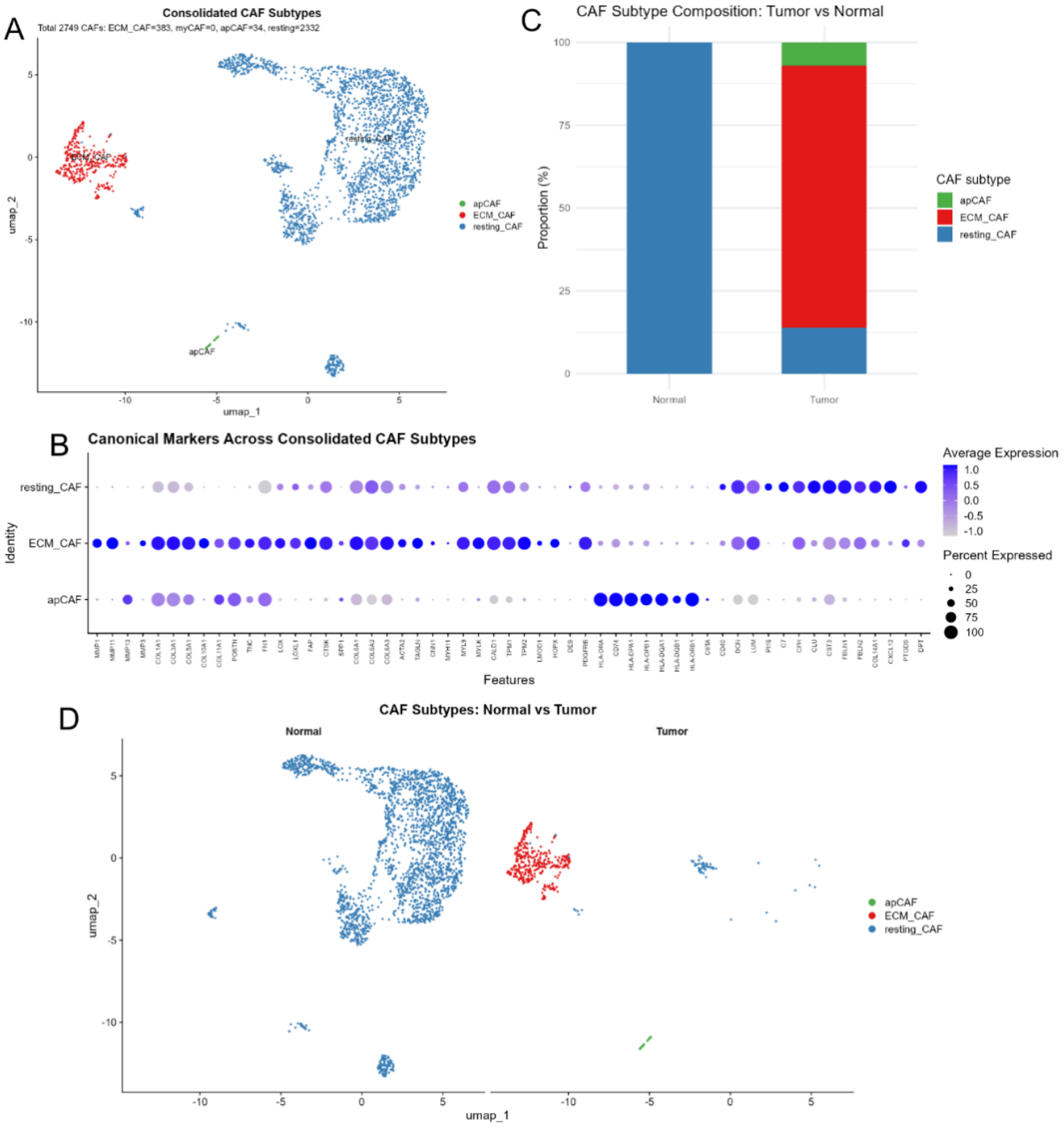


Figure 2

Three functional subtypes of CAFs. (A) UMAP plot of CAF subtypes, colored according to the three functional subtypes: ECM-CAF, resting-CAF, and apCAF. (B) Dot plots of marker genes for each CAF subtype, showing the expression levels and proportions of marker genes for ECM-CAF, resting-CAF, and apCAF. (C) Bar charts depicting the proportion of CAF subtypes in normal and tumor tissues, illustrating differences in the distribution of each subtype between the two groups. (D) Grouped UMAP plot demonstrating that ECM-CAF is the predominant CAF subtype in tumors.

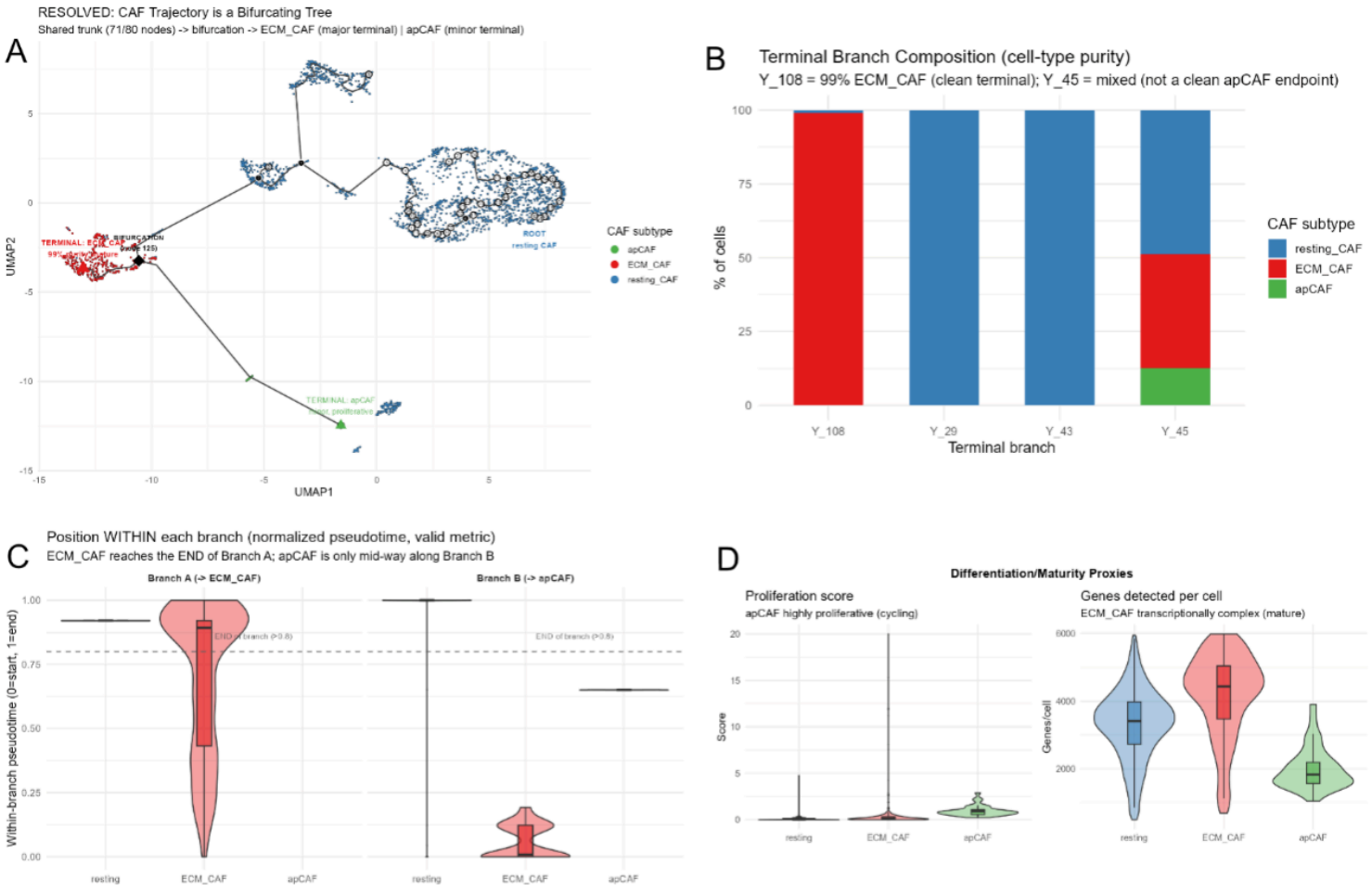


Figure 3

CAF Differentiation Trajectories. (A) Pseudotime analysis plot, colored according to CAF subtype, illustrating the bifurcating branch structure of CAF differentiation trajectories. (B) Analysis of cellular composition purity at terminal nodes, showing that ECM-CAFs constitute 99 percent of the cells at the Y_108 terminal node. (C) Distribution of pseudotemporal positions of each CAF subtype along the differentiation trajectory. (D) Comparison of maturity proxy indicators between ECM-CAFs and apCAFs.

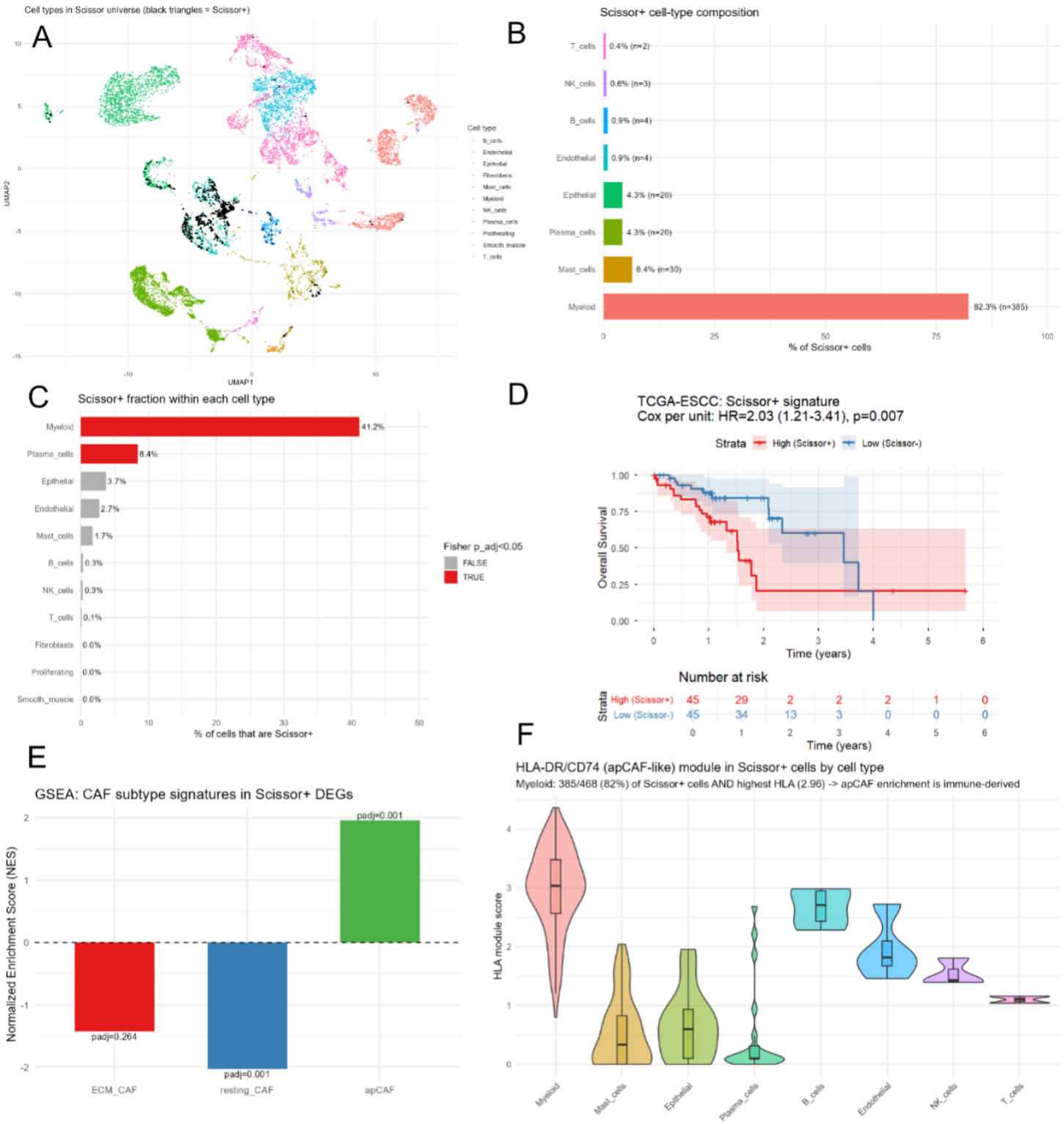


Figure 4

Scissor identifies cell subsets associated with prognosis. (A) Whole-cell UMAP plot colored by cell type; black triangles indicate the locations of Scissor-positive cells, the majority of which are situated in the myeloid cell region. (B) Composition analysis of Scissor-positive cells, showing the proportion of each cell type within the Scissor-positive population. (C) Enrichment analysis of the proportion of Scissor-positive cells within each cell type, demonstrating that myeloid cells are the only significantly enriched

cell type. (D) Kaplan–Meier curves for overall survival in the high- and low-expression groups of the Scissor-positive phenotype. (E) Enrichment analysis of CAF subtypes among differentially expressed genes in Scissor-positive cells. (F) Distribution of HLA module scores across cell types, indicating that myeloid cells have the highest scores.

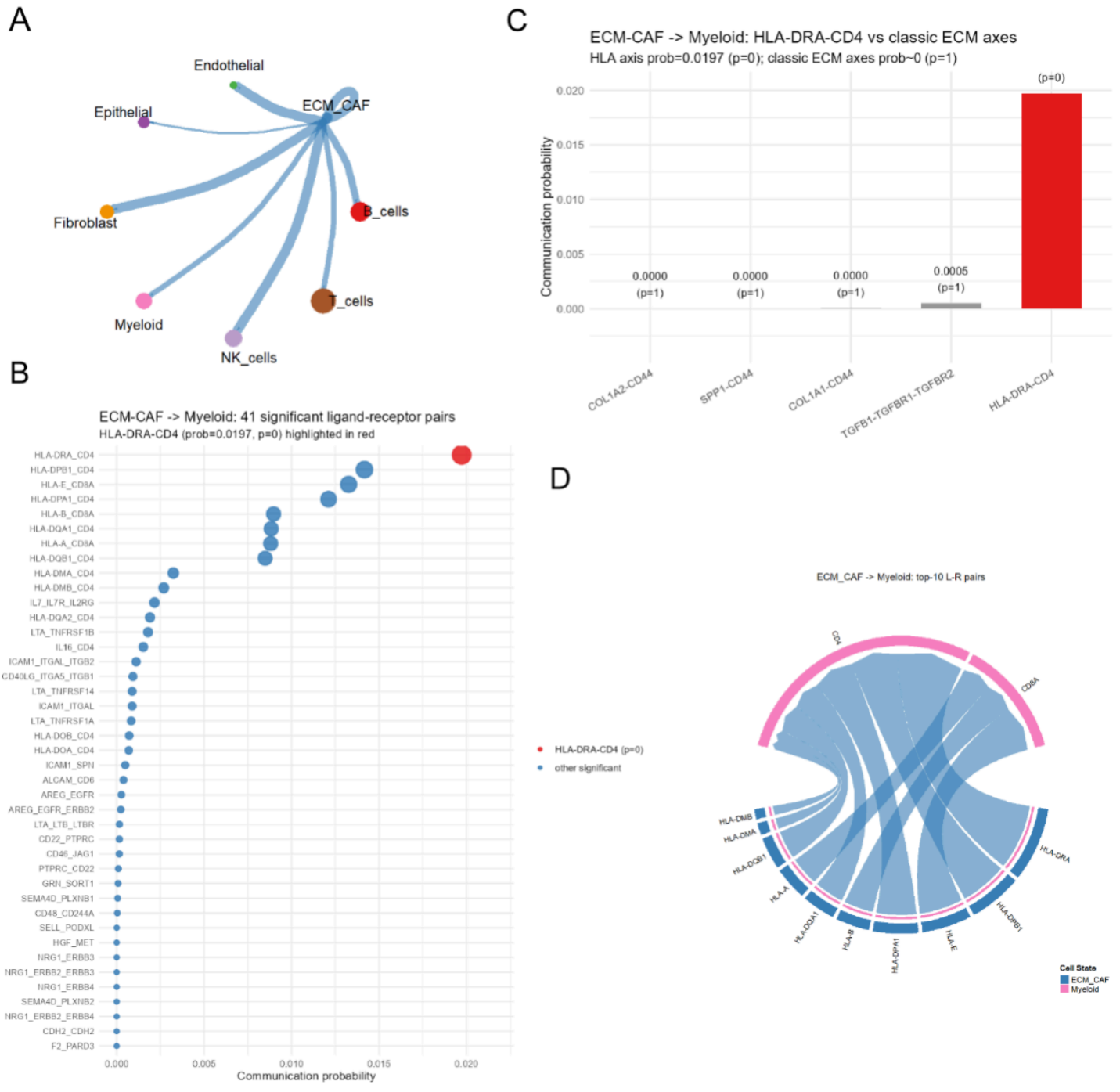


Figure 5

The HLA-DRA-CD4 axis mediates communication between ECM-CAFs and myeloid cells. (A) Network diagram depicting outgoing communication from ECM-CAFs; arrow thickness represents the strength of communication, while node size corresponds to the number of cells of each type. (B) Bubble plot of

ligand-receptor pairs transmitted from ECM-CAFs to myeloid cells, highlighting the HLA-DRA-CD4 signal as the strongest. (C) Comparison of communication probabilities between the HLA-DRA-CD4 axis and the classical ECM communication axis. (D) Chord diagram illustrating communication signals transmitted from ECM-CAFs to myeloid cells; the width of the bands connecting the two ends indicates communication probability. The HLA-DRA-CD4 axis forms the widest chord, signifying it as the primary channel through which ECM-CAFs transmit signals to myeloid cells.

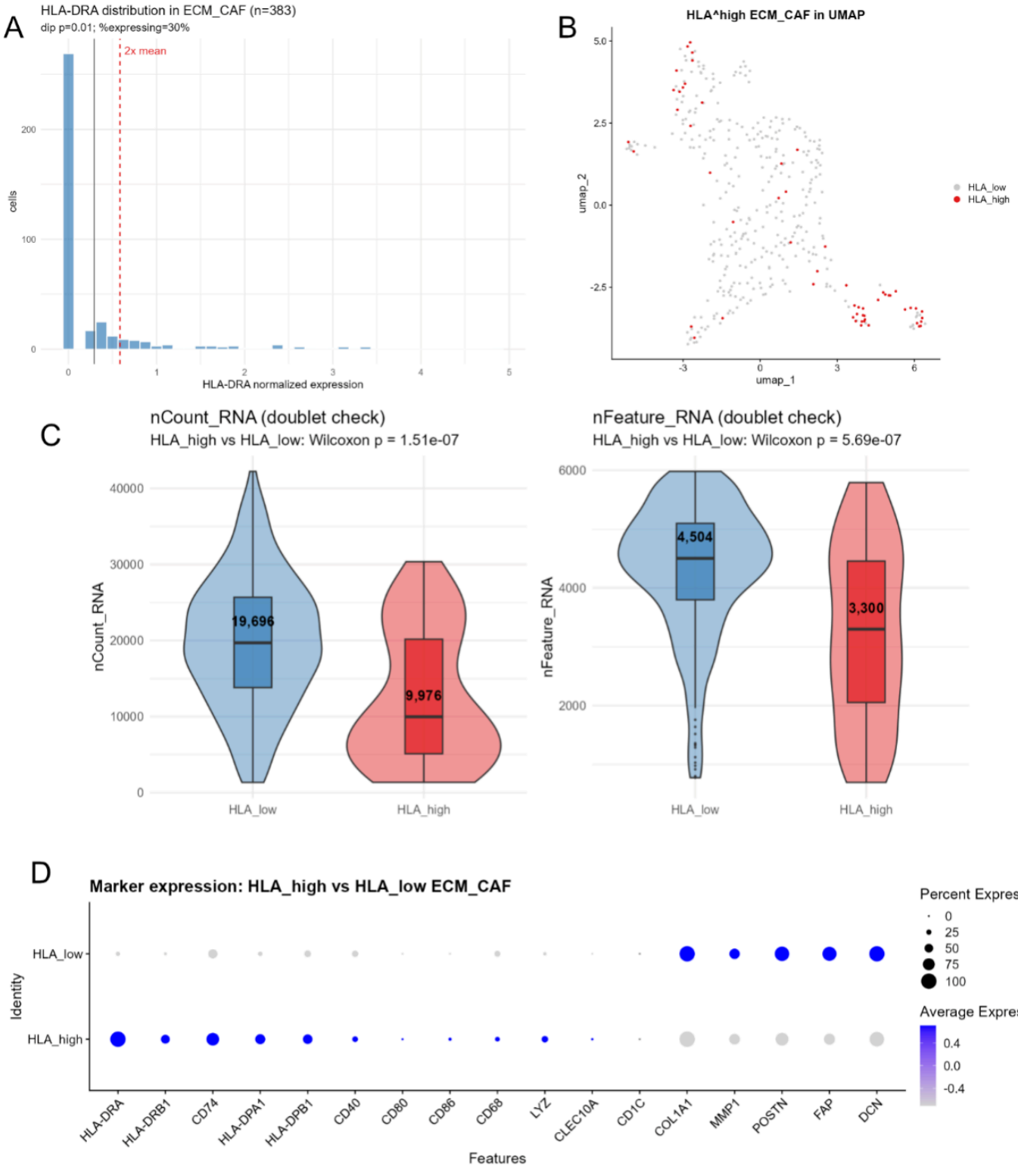


Figure 6

Identification of the HLA-high-expressing ECM-CAF subpopulation. (A) Histogram showing the distribution of HLA-DRA expression in ECM-CAFs, with the red dotted line indicating the threshold at twice the mean expression level. (B) UMAP plot of ECM-CAFs, with HLA-high-expressing cells shown in red and HLA-low-expressing cells in gray. (C) Violin plots of total molecular count and gene count for HLA-high and HLA-low cells, demonstrating that both metrics are lower in HLA-high cells than in HLA-low cells. (D) Scatter plots of HLA-high and HLA-low cells, illustrating the expression levels of antigen-presentation-related genes and immune cell marker genes.

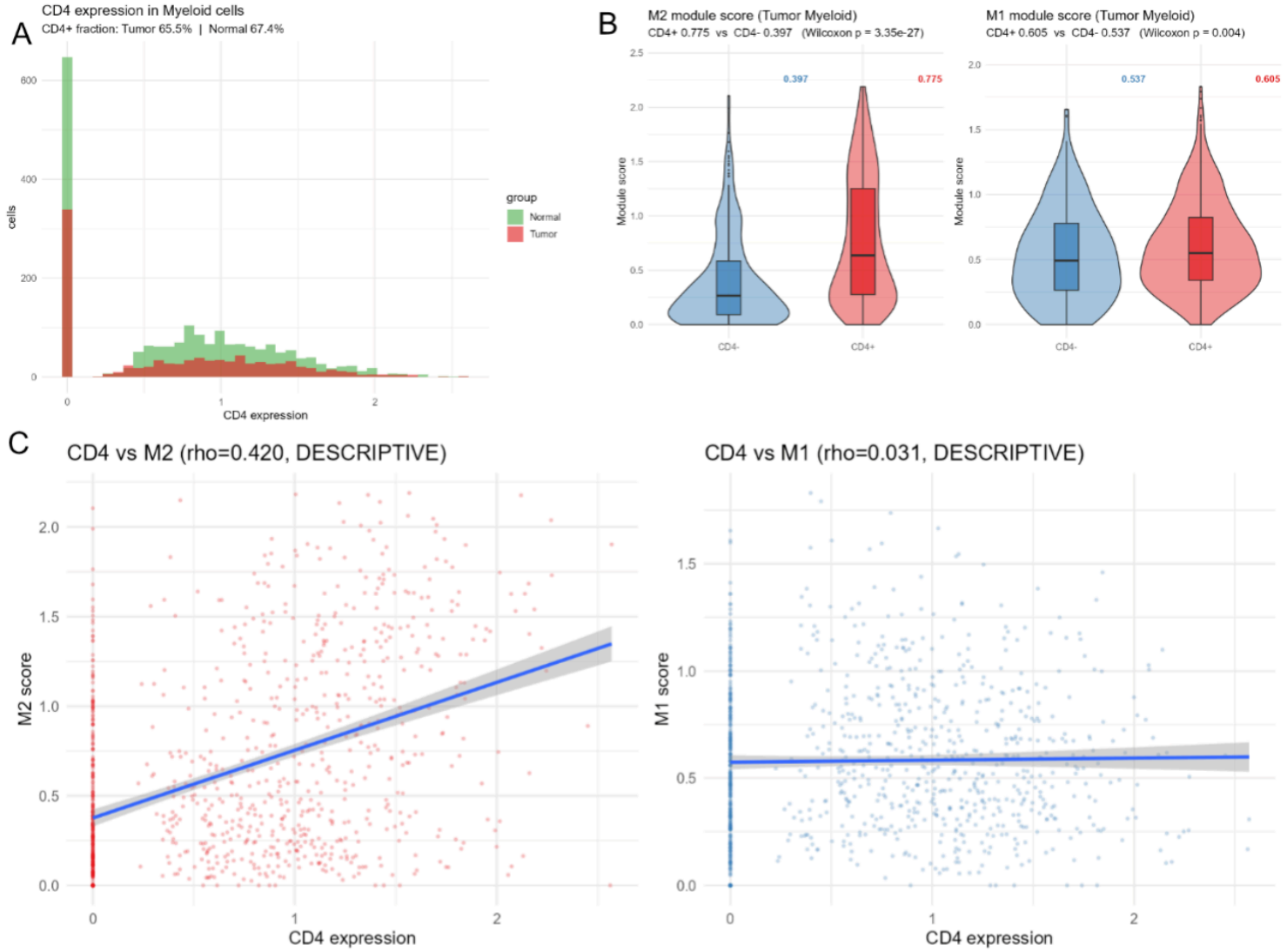


Figure 7

CD4-positive myeloid cells exhibit M2 polarization. (A) Histogram showing the distribution of CD4 expression in tumor myeloid cells, indicating that 65.5% of myeloid cells express CD4. (B) Comparison of M1 and M2 scores in CD4-positive and CD4-negative myeloid cells, demonstrating a significantly higher M2 score in CD4-positive cells. (C) Scatter plot showing the correlation between CD4 expression and M2

and M1 scores, indicating that CD4 expression is positively correlated with the M2 score but not with the M1 score.

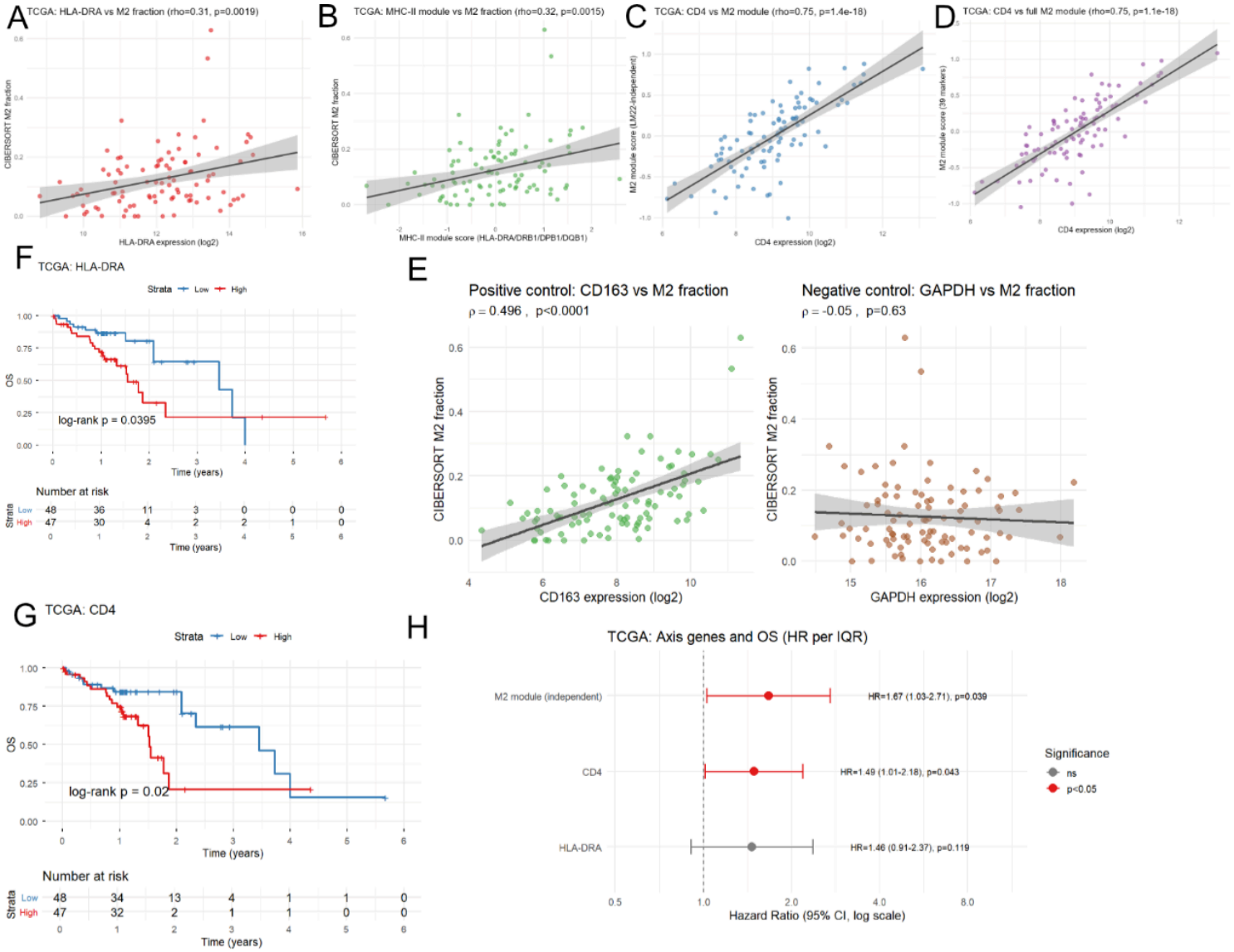


Figure 8

Validation of the HLA-DRA–CD4–M2 axis in the TCGA-ESCC batch transcriptome. (A) Scatter plot showing a significant positive correlation between HLA-DRA expression and M2 macrophage abundance, as estimated by CIBERSORT. (B) Scatter plot showing a significant positive correlation between the MHC-II module score and M2 macrophage abundance. (C) Scatter plot demonstrating a strong positive correlation between CD4 expression and the independent M2 module score. (D) Scatter plot showing a positive correlation between CD4 expression and the complete M2 module score. (E) Scatter plot illustrating that the positive control CD163 is significantly positively correlated with M2 abundance, whereas the negative control GAPDH shows no correlation with M2 abundance. (F) Kaplan–Meier curves depicting overall survival in high- and low-expression groups of HLA-DRA. (G) Kaplan–Meier curves depicting overall survival in high- and low-expression groups of CD4. (H) Forest plot showing the hazard ratio for death per interquartile range increase in CD4 and M2 module scores.

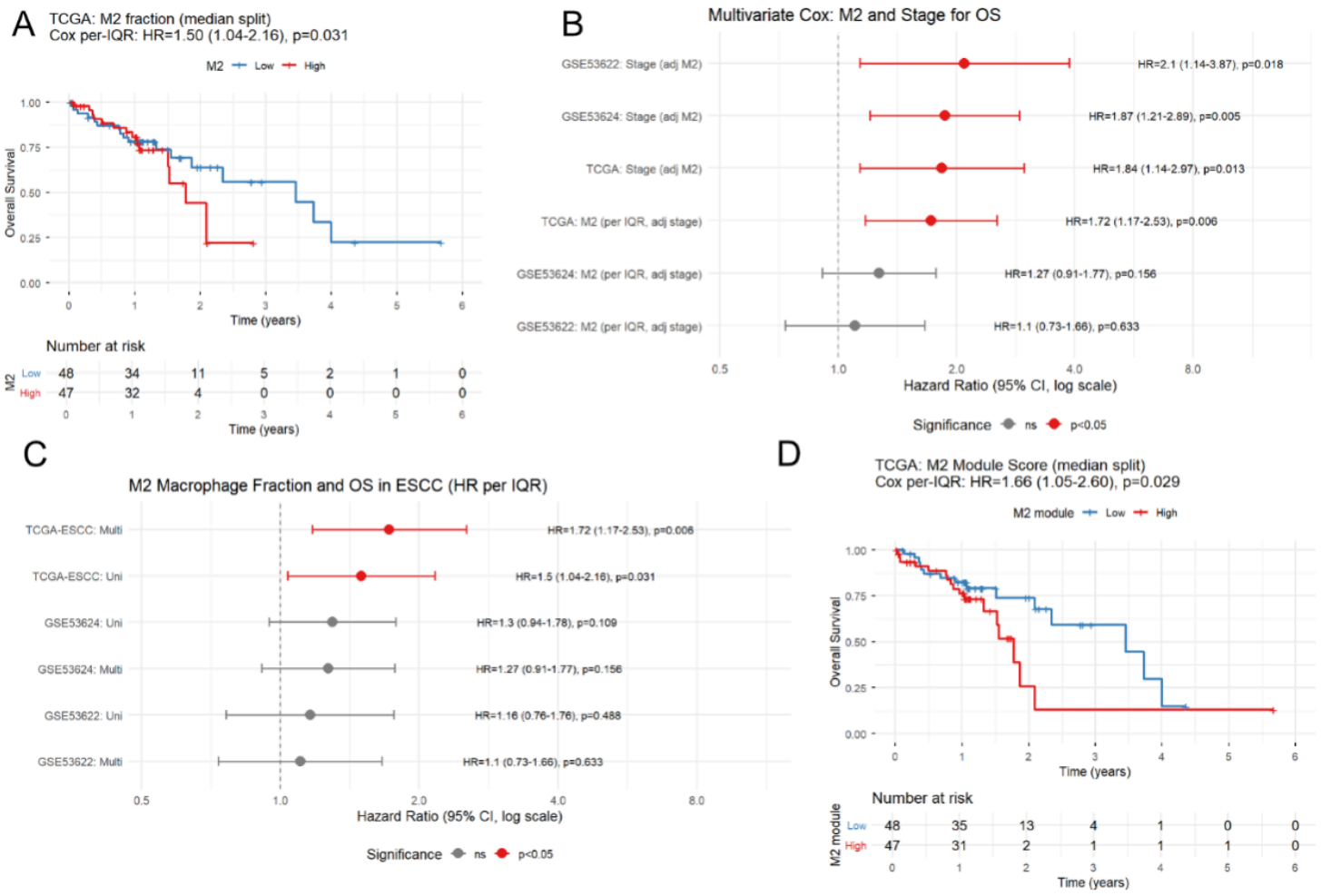


Figure 9

Association between M2 macrophage abundance and prognosis in esophageal squamous cell carcinoma. (A) Kaplan–Meier curves for overall survival according to M2 macrophage abundance in the TCGA-ESCC cohort; (B) Forest plot illustrating the association between M2 macrophage abundance and overall survival across the TCGA-ESCC, GSE53624, and GSE53622 cohorts; (C) Summary of univariate and multivariate Cox regression forest plots across the three cohorts; (D) Kaplan–Meier curve showing the association between the M2 signature gene module score and overall survival in the TCGA-ESCC cohort.