Translational Oncology 69 (2026) 102801

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Integrating spatial and single-cell transcriptomics via machine learning to characterize efferocytosis in hepatocellular carcinoma prognosis and immunotherapy

Qiuyun Yi [a,b,1], Bin Zhou [c,1], Xiaohui Gu [d,1], Zihui Ma [c], Bochen Pan [e], Zhou Yuan [d,*], Wen Wen [b,*], Chongde Lu [c,*]

[a] National Center for Liver Cancer, The Third Affiliated Hospital of Naval Medical University, Shanghai, China
[b] Department of Laboratory Diagnosis, The Third Affiliated Hospital of Naval Medical University (Second Military Medical University), Shanghai 200438, China
[c] Department of Hepatic Surgery VI, The Third Affiliated Hospital of Naval Medical University (Second Military Medical University), 225 Changhai Road, Shanghai 200433, China
[d] Department of General Surgery, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, 222 Huanhu Xisan Road, Shanghai 201306, China
[e] Department of Urology, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China

ARTICLE INFO

Keywords:
Hepatocellular carcinoma
Efferocytosis
SPP1+ macrophages
Spatial transcriptomics
Machine learning
TPI1

ABSTRACT

Background: The intratumoral heterogeneity and immunosuppressive microenvironment of hepatocellular carcinoma (HCC) significantly limit therapeutic efficacy. Although efferocytosis is fundamental for tissue homeostasis, its impact on the spatial architecture of the tumor microenvironment (TME) and patient prognosis in HCC remains uncharacterized.

Methods: We integrated single-cell RNA sequencing, spatial transcriptomics, and large-scale bulk datasets to construct an Efferocytosis-Related Scoring System (ERG score). We identified molecular subtypes and traced key cell subsets driving high-efferocytosis features. A robust prognostic model was established via large-scale machine learning screening (101 algorithms) and validated by in vitro experiments (qRT-PCR/Western Blot).

Results: The ERG score indicates a specific immunosuppressive state driven by the High_ERGs_SPP1_Mac subpopulation. Spatially, this subpopulation physically co-localizes with malignant cells, potentially mediated by the MDK-SDC2 axis. The machine learning-derived risk score serves as a reliable quantitative indicator of High_ERGs_SPP1_Mac abundance, effectively predicting "cold tumor" phenotypes and immune checkpoint upregulation (e.g., PD-1, CTLA4). Additionally, experimental validation confirmed the upregulation of the core gene *TPI1* (triosephosphate isomerase 1), aligning with the high glycolytic features of the high-risk subtype.

Conclusion: The ERG scoring system offers a novel perspective for deconvoluting HCC heterogeneity. Our findings suggest that *SPP1+* macrophages spatially reshape the TME, providing a theoretical basis for optimizing prognostic stratification and personalized immunotherapy.

Introduction

Hepatocellular carcinoma (HCC) ranks among the leading causes of cancer-related mortality worldwide, characterized by high incidence, insidious onset, and significant heterogeneity [1]. Although surgical resection, liver transplantation, and ablation are effective interventions for early-stage HCC, the majority of patients are diagnosed at advanced stages, thereby missing the optimal window for curative surgery [2,3]. In recent years, systemic therapies represented by immune checkpoint inhibitors (ICBs) have achieved notable progress in advanced HCC. However, constrained by the complex tumor microenvironment (TME), only a fraction of patients derive durable clinical benefit, while resistance and recurrence remain severe challenges [4,5]. Given the current lack of specific molecular biomarkers to accurately predict prognosis and immunotherapy response, the overall clinical outcome for HCC patients remains suboptimal. Therefore, deeply characterizing the

* Corresponding authors.
E-mail addresses: zhouyuan851@163.com (Z. Yuan), wenwen_smmu@163.com (W. Wen), 86246337@qq.com (C. Lu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.tranon.2026.102801
Received 22 February 2026; Received in revised form 15 April 2026; Accepted 27 April 2026

molecular features of HCC and identifying novel prognostic indicators and therapeutic targets are of great significance for improving patient stratification and optimizing treatment strategies.

Efferocytosis orchestrates a pivotal homeostatic framework within the tumor microenvironment (TME). As an evolutionarily conserved physiological phenomenon, it entails the systematic recognition, internalisation, and subsequent catabolism of apoptotic debris by professional phagocytes, most notably macrophages [6]. Within a steady-state physiological milieu, this cascade is rigorously governed by an interplay between "find-me" molecules (e.g., ATP and CX3CL1) and "eat-me" ligands, particularly externalized phosphatidylserine (PtdSer) [7,8]. The ligation of these cues to cognate phagocytic surface receptors, including the TAM kinase family (Tyro3, AXL, and MERTK), precipitates an immunologically quiescent clearance pathway [9]. Divergent from the inflammatory phagocytosis triggered by exogenous pathogens, efferocytosis is intrinsically non-immunogenic. It actively attenuates pro-inflammatory signaling (e.g., NF-κB) while augmenting the secretion of pro-resolving factors like IL-10 and TGF-β, thereby safeguarding against autoimmunity [10,11]. Nevertheless, in malignant settings, this homeostatic machinery is frequently sequestered by the tumor, especially when extensive cell death is driven by hypoxia, metabolic deprivation, or iatrogenic anti-tumor interventions [12].

Efficient efferocytosis within the TME acts as a "double-edged sword": while it prevents secondary necrosis and the release of autoantigens, the continuous engulfment of apoptotic tumor cells induces profound functional reprogramming of tumor-associated macrophages (TAMs) [13]. Mechanistically, efferocytic signaling not only drives macrophage polarization toward a pro-tumorigenic M2-like phenotype but also induces metabolic adaptations (e.g., upregulation of glycolysis and lipid metabolism) to handle the massive influx of cellular cargo [14,15]. Recent milestone studies indicate that metabolites derived from efferocytosis (such as lactate) can further suppress anti-tumor immunity and promote angiogenesis [16]. Accumulating evidence highlights the clinical significance of this process across various malignancies. In non-small cell lung cancer (NSCLC), high expression of efferocytosis receptors (e.g., AXL, MERTK) has been confirmed as a potent predictor of shortened survival and resistance to EGFR-targeted therapy [17,18]. Similarly, in breast cancer, efferocytosis-driven signaling has been found to promote tissue remodeling and metastasis by upregulating wound-healing-associated cytokines [19]. Furthermore, pan-cancer analyses suggest a significant negative correlation between efferocytosis-related gene signatures and cytotoxic T-cell infiltration, underscoring its potential as a universal marker of immune exclusion [20]. Although aberrant expression of efferocytosis-related genes (ERGs) has been linked to poor prognosis in multiple cancers, their specific expression landscape in HCC, their role in reshaping the spatial architecture of the TME, and their clinical value as prognostic biomarkers remain systematically underexplored.

To address these gaps, we established and validated an efferocytosis-related gene (ERG) score for HCC. Here, “ERG” refers specifically to efferocytosis-related genes and should be distinguished from the ETS-related gene ERG oncogene. By integrating single-cell RNA sequencing (scRNA-seq), spatial transcriptomics (ST), and bulk transcriptomic datasets, we characterized the HCC immune landscape and assessed its prognostic relevance. Unsupervised clustering identified molecular subtypes with distinct efferocytic activity. Integrating single-cell and spatial data further enabled the identification of key cell subsets driving high-efferocytosis features and their interactions with malignant cells. We then developed a prognostic model and nomogram based on signature genes. Overall, this study clarifies the role of efferocytosis in HCC progression and provides a basis for risk stratification, microenvironment profiling, and personalized treatment.

Materials and methods

Data acquisition

This study comprehensively utilized multi-omics data for analysis. First, three scRNA-seq datasets (GSE149614, GSE151530, and GSE156625) were retrieved and downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). From these datasets, HCC tumor tissues and adjacent non-tumor tissues were filtered for downstream analysis. Second, large-scale bulk transcriptome sequencing data were included, comprising the TCGA-LIHC cohort (containing 374 HCC tissues and 50 normal tissues) obtained from the UCSC Xena platform (https://xena.ucsc.edu), as well as two external validation cohorts from GEO (GSE14520 [Platform: GPL3921] and GSE116174). Furthermore, to elucidate the spatial structural characteristics of the tumor, ST data of HCC (Sample IDs: VISDS000512, VISDS000516) were acquired from the CROST database (https://ngdc.cncb.ac.cn/crost/). Genes associated with efferocytosis were based on previously published studies (Table S1) [6,11].

Consensus clustering

In the present work, we determined a set of 29 phagocytosis-associated genes possessing significant prognostic value. To stratify HCC patients into two distinct molecular clusters, an unsupervised consensus clustering analysis was implemented utilizing the "ConsensusClusterPlus" R Bioconductor package [21]. Analysis across two independent cohorts revealed that these identified subgroups exhibited substantial disparities in terms of clinical manifestations and prognostic outcomes.

Assessment of immune infiltration and checkpoints in HCC

To quantify the immunological landscape, the ESTIMATE algorithm was applied to the TCGA transcriptional profiles for calculating immune cell infiltration across various cohorts. We utilized the "IOBR" R package [22] to further refine the assessment of immune infiltration within the HCC patient population from TCGA. Concurrently, the Single-Sample Gene Set Enrichment Analysis (ssGSEA) [23] was implemented to characterize the enrichment levels of immune cell subsets within the HCC TME. Following this, we explored the expression heterogeneity of immune checkpoints across distinct HCC molecular subgroups, with visualization and statistical analysis facilitated by the "ggplot2", "immuneeconv", and "pheatmap" R environments.

Somatic mutation analysis

Genomic alteration profiles for HCC patients were retrieved from the TCGA database, encompassing Single Nucleotide Variations (SNVs) and Copy Number Variations (CNVs) generated from Whole Exome Sequencing (WES) and high-density SNP arrays (Affymetrix Genome-Wide Human SNP Array 6.0). Integrated clinical annotations corresponding to these samples were also acquired. For the subsequent analysis and visualization of the somatic mutational landscape, we implemented the "maftools" (v2.14.0) [24] and "GISTIC2.0" [25] R packages to generate comprehensive summary statistics and identify recurrently altered regions.

scRNA-seq data processing

Single-cell transcriptomic data were processed and analyzed leveraging the Seurat R library (v4.4.0) [26]. To ensure data integrity, a stringent quality control pipeline was implemented, exclusively retaining cells that harbored over 200 detected features (nFeature) and a mitochondrial gene fraction (percent.mt) below 20%. For downstream analysis, the filtered expression matrix underwent normalization

through the LogNormalize technique (scaling factor = 10,000). We subsequently identified the top 2000 highly variable genes (HVGs) using the FindVariableFeatures function, followed by centered scaling via ScaleData to minimize technical artifacts. Dimensionality reduction was executed through Principal Component Analysis (PCA), with the leading 30 principal components selected as the basis for further computation. To rectify potential batch effects inherent in multi-sample datasets (refer to Fig. S3A, B for sample distribution), the Harmony algorithm was integrated into the workflow for batch correction [26]. Seurat was used for downstream single-cell processing and visualization [27]. Cellular connectivity was modeled by establishing a Shared Nearest Neighbor (SNN) graph. Subsequently, unsupervised clustering was performed at a resolution of 1 to delineate distinct cell populations (Fig. S3C). For low-dimensional representation and visual exploration, Uniform Manifold Approximation and Projection (UMAP) was utilized to project the high-dimensional clusters onto a 2D coordinate system.

Identification of CNVs in scRNA-seq data

To estimate chromosomal copy number variations (CNVs) within the epithelial compartments of each HCC specimen, the InferCNV R library (v1.4.0) was utilized [28]. The analytical workflow initiated with the extraction of the raw count matrix from the processed Seurat object as the input dataset. To establish a reliable baseline for normal genomic profiles, T cells and NK cells isolated from adjacent non-malignant tissues were designated as the reference control. The inferential process was executed with a defined cut-off threshold of 0.1, whereas all supplementary parameters remained consistent with the manufacturer's default configurations.

Construction of cell-cell communication networks

To systematically analyze signaling communication and metabolic characteristics between different cell types, we employed the CellChat tool [29]. Intercellular communication between different cell clusters was explored to identify potential ligand-receptor pairs, which were then projected onto a protein-protein interaction (PPI) network to calculate intercellular communication probabilities. The structure of the communication network and signal strength between cell clusters were visualized using functions such as netVisual_circle and netVisual_bubble.

Pseudotime trajectory and cell stemness analysis

To elucidate the dynamic transitions between cellular states, we conducted pseudotime trajectory reconstruction based on the scRNA-seq profiles employing the Monocle2 algorithm (v2.16.0) [30]. Furthermore, the differentiation capacity and developmental stemness of individual cells were computationally predicted via the implementation of the CytoTRACE framework (v0.3.3) [31].

ST data processing

The processing and visual exploration of ST datasets were conducted within the Seurat R environment. For data normalization, the SCTransform (SCT) framework was implemented. To harmonize disparate spatial datasets, a systematic integration pipeline was executed, encompassing the SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData functions. Subsequently, an unsupervised clustering approach was adopted to group spatial spots with shared transcriptomic profiles. The taxonomic annotation of these clusters was facilitated by a comparative analysis of Hematoxylin and Eosin (H&E) histological images alongside cluster-specific highly variable genes. Finally, the spatial distribution and expression patterns across the tissue sections were visualized leveraging the SpatialDimPlot and SpatialFeaturePlot utilities.

Cell type deconvolution of ST data

To project the transcriptomic identities from the scRNA-seq reference onto the ST landscape, we implemented the Robust Cell Type Decomposition (RCTD) algorithm [32]. Prior to decomposition, lineage-specific marker genes were characterized via the Seurat FindAllMarkers utility, specifically filtering for genes with positive log2-transformed fold changes. Adhering to the established RCTD computational pipeline, the deconvolution was executed by integrating the reference profiles with the Visium ST datasets, utilizing the 'full' parameter configuration (doublet mode) to ensure comprehensive cell-type assignment.

Construction of spatial interaction patterns

MISTy (Multiview Intercellular SpaTial modeling) functions as an interpretable analytical framework designed to decipher intercellular crosstalk by integrating diverse spatial contexts, termed "views." This methodology facilitates a holistic exploration of the spatial architecture and communicative networks within the tissue microenvironment. In this study, we implemented the MISTyR library (v1.6.1) [33] to characterize distinct spatial microenvironments within the ST datasets, utilizing intercellular spatial proximity as the fundamental metric for modeling.

Neighborhood network analysis

To quantify cellular organization patterns and interaction characteristics in space, we analyzed spatial neighborhood structures using functions from the STUtility package. The GetSpatNet function was used to calculate homotypic neighborhood scores for each spatial spot. Observed spatial aggregation patterns were then compared with random distributions within the sample to determine whether specific cell types exhibited aggregation trends higher than random expectation. A higher homotypic score indicates stronger spatial organization and aggregation of that cell type in local regions. Furthermore, to quantify heterotypic interactions between different cell types, the RegionNeighbors function was used to calculate the frequency with which spatial spots of one cell type were adjacent to spots of another cell type.

Identification of differentially expressed genes (DEGs)

To determine differentially expressed genes (DEGs) between HCC specimens and normal controls within the TCGA-LIHC cohort, we applied linear modeling via the limma R package. To ensure statistical stringency and minimize both false positive and false discovery rates, the significance thresholds were established at an absolute log2-transformed fold change $|\log2FC| > 1$ along with an adjusted P-value < 0.05.

Enrichment analysis and gene set scoring

To investigate biological functions, gene set enrichment analysis (GSEA) [33] and subsequent functional characterization were executed leveraging the clusterProfiler library [34] (v4.6.2). For the quantification of specific biological signatures, gene set scores were calculated via the AddModuleScore utility within the Seurat framework. The requisite KEGG gene sets utilized for these analyses were retrieved from the Molecular Signatures Database (MSigDB).

Construction of the prognostic model

Gene expression matrices and complete clinical survival data for HCC patients were systematically obtained from three cohorts: TCGA-LIHC, GSE14520, and GSE116174. To ensure data quality, samples with high missing value rates or incomplete clinical information were

excluded, and data were standardized. The TCGA-LIHC cohort was designated as the training set, while GSE14520 and GSE116174 served as external validation sets. To identify candidate genes with prognostic value, the scope was narrowed to key feature genes of the High_ERGs_SPP1_Mac subpopulation. Univariate Cox regression analysis was performed to screen for genes significantly associated with Overall Survival (OS) ($p < 0.05$). Based on this, an integrated machine-learning strategy was adopted, constructing 101 combination models using 10 classical machine-learning algorithms. The 10 algorithms included Lasso, Ridge, Elastic Net, Stepwise Cox, CoxBoost, Random Survival Forest (RSF), Gradient Boosting Machine (GBM), survival-SVM, SuperPC, and plsRcox, which were selected to represent complementary survival-modeling strategies for comprehensive comparison. To evaluate model prediction performance and generalizability, the Concordance Index (C-index) was calculated for all models across the training set and both validation sets. Through systematic screening and comparison, Lasso + GBM was determined as the optimal predictive model, identifying 10 core feature genes: SPP1, FABP5, TPI1, LDHA, TREM1, S100A9, IL1RN, SLC2A1, ANGPTL4, and MIF.

Prediction of immunotherapy response

To evaluate potential immune escape mechanisms, the Tumor Immune Dysfunction and Exclusion (TIDE) metric, alongside scores for T-cell dysfunction and exclusion, were quantified through the TIDE computational framework (http://tide.dfci.harvard.edu/). This algorithmic approach facilitated the assessment of therapeutic response and immune evasion patterns across the patient cohorts.

Cell lines and cell culture

The human HCC cell lines (Huh7 and SMMC-7721) and the immortalized human hepatocyte line MIHA were procured from Procell Life Science & Technology Co., Ltd. (Wuhan, China). Standard culture conditions involved the use of Dulbecco's Modified Eagle Medium (DMEM; Procell) enriched with 10% fetal bovine serum (FBS; Procell) and 1% penicillin-streptomycin. All experimental cell populations were sustained within a humidified atmosphere containing 5% CO_2 at a constant temperature of 37 °C.

Western blotting analysis

To evaluate protein expression, total cellular protein was harvested from MIHA, Huh7, and SMMC-7721 lines utilizing RIPA lysis buffer enriched with protease and phosphatase inhibitors. Subsequent to quantification via a BCA Protein Assay Kit, equivalent protein masses were resolved on 10% SDS-PAGE gels and electroblotted onto polyvinylidene fluoride (PVDF) membranes. Nonspecific binding sites were occluded by incubating membranes in 5% non-fat milk for 1 h at ambient temperature, followed by a primary antibody incubation at 4 °C overnight. After thorough rinsing with TBST, the blots were exposed to horseradish peroxidase (HRP)-conjugated secondary antibodies for 1 h at room temperature. Protein signals were developed using an Enhanced Chemiluminescence (ECL) detection system and digitized via a gel imaging platform. Densitometric analysis was performed using ImageJ software (NIH, Bethesda, MD, USA), with the target protein (TPI1) expression normalized against the internal loading control (GAPDH). The specific primary antibodies implemented were anti-TPI1 (1:1000, 10,713-1-AP; Proteintech, China) and anti-GAPDH (1:5000, 60,004-1-Ig; Proteintech, China). Correspondingly, HRP-linked goat anti-rabbit (1:5000, DY60202; Diyibio, China) and goat anti-mouse IgG (1:5000, DY60203; Diyibio, China) were utilized as secondary detection reagents.

RNA extraction and quantitative real-time PCR (RT-qPCR)

To evaluate transcript levels, total RNA was isolated from MIHA, Huh7, and SMMC-7721 cell lines utilizing TRIzol reagent (Invitrogen, USA), with its concentration and purity determined via a NanoDrop spectrophotometer. Subsequently, 1 μg of the extracted RNA was reverse-transcribed into cDNA using the PrimeScript RT Master Mix (Takara, Japan). The quantitative real-time PCR (RT-qPCR) assays were executed on an ABI 7500 Real-Time PCR System, employing TB Green Premix Ex Taq II (Takara, Japan) for fluorescence detection. The amplification utilized specific primer sets (5′-3′): TPI1 forward (TGGCATGATCAAAGACTGC) and reverse (CCCAATCAGCTCATCTGAC). The thermal cycling profile consisted of an initial denaturation phase at 95 °C for 30 s, followed by 40 amplification cycles (95 °C for 5 s and 60 °C for 34 s). All biological assays were conducted in triplicate, and the $2^{-\Delta\Delta Ct}$ method was implemented to calculate the relative abundance of TPI1 mRNA.

Statistical analysis

Computational processing and statistical evaluations were executed within the R environment (version 4.3.1). To assess prognostic divergences between identified patient clusters, Kaplan-Meier survival curves alongside the log-rank test were generated for each dataset. For continuous variables that deviated from a normal distribution, the Wilcoxon rank-sum test or Kruskal-Wallis H-test was implemented to ensure analytical accuracy. All statistical assessments were two-tailed, and a threshold of $P < 0.05$ was defined as the criterion for statistical significance.

Results

Identification of efferocytosis-related prognostic genes in HCC

Through a systematic literature review, a total of 167 ERGs were curated, among which univariate Cox regression analysis identified 37 candidates significantly correlated with overall survival (OS) (Fig. 1A). Genomic analysis revealed prevalent DNA copy number variations (CNVs) across these prognostic ERGs. Notably, widespread CNV amplifications were observed in genes such as FCER1G, S100A9, NCF2, FABP4, and FABP5, whereas other ERGs predominantly exhibited CNV deletions (Fig. 1B); their specific chromosomal distribution is visualized in Fig. 1C. In parallel, somatic mutation profiling indicated that ERGs (including PER1, CYP27A1, and ODC1) were altered in 7.07% (26/368) of the HCC cohort (Fig. 1D). Furthermore, an investigation into the expression profiles within the TCGA cohort demonstrated that the majority of these 37 prognostic genes were significantly dysregulated in tumor tissues relative to normal controls (Fig. S1A).

Consensus clustering identifies two molecular subtypes in HCC

To establish a resilient molecular classification, we prioritized 29 ERGs shared between the TCGA-LIHC and GSE14520 cohorts. Utilizing the expression landscapes of these conserved genes, unsupervised consensus clustering was implemented to delineate novel endocytosis-associated subtypes in HCC. An optimal cluster stability at k = 2 facilitated the stratification of the TCGA cohort into Cluster 1 (n = 194) and Cluster 2 (n = 180) (Fig. 2A, B). Assessment of cumulative gene expression revealed that Cluster 2 possessed a significantly elevated ERG score relative to Cluster 1 (Fig. 2C). In alignment with this transcriptomic profile, Kaplan-Meier analysis demonstrated that Cluster 2 patients exhibited a substantially inferior survival outcome compared to their Cluster 1 counterparts ($P < 0.001$) (Fig. 2D). These prognostic discrepancies were successfully recapitulated in the GSE14520 validation set (Fig. S1B–E). At the single-gene level, Cluster 2 was characterized by the robust upregulation of pivotal ERGs, including SPP1, MMP9, MMP14, and S100P (Fig. 2E). To decipher the biological underpinnings of these subtypes, GSEA was conducted across Hallmark and KEGG frameworks. We observed that Cluster 2 was predominantly

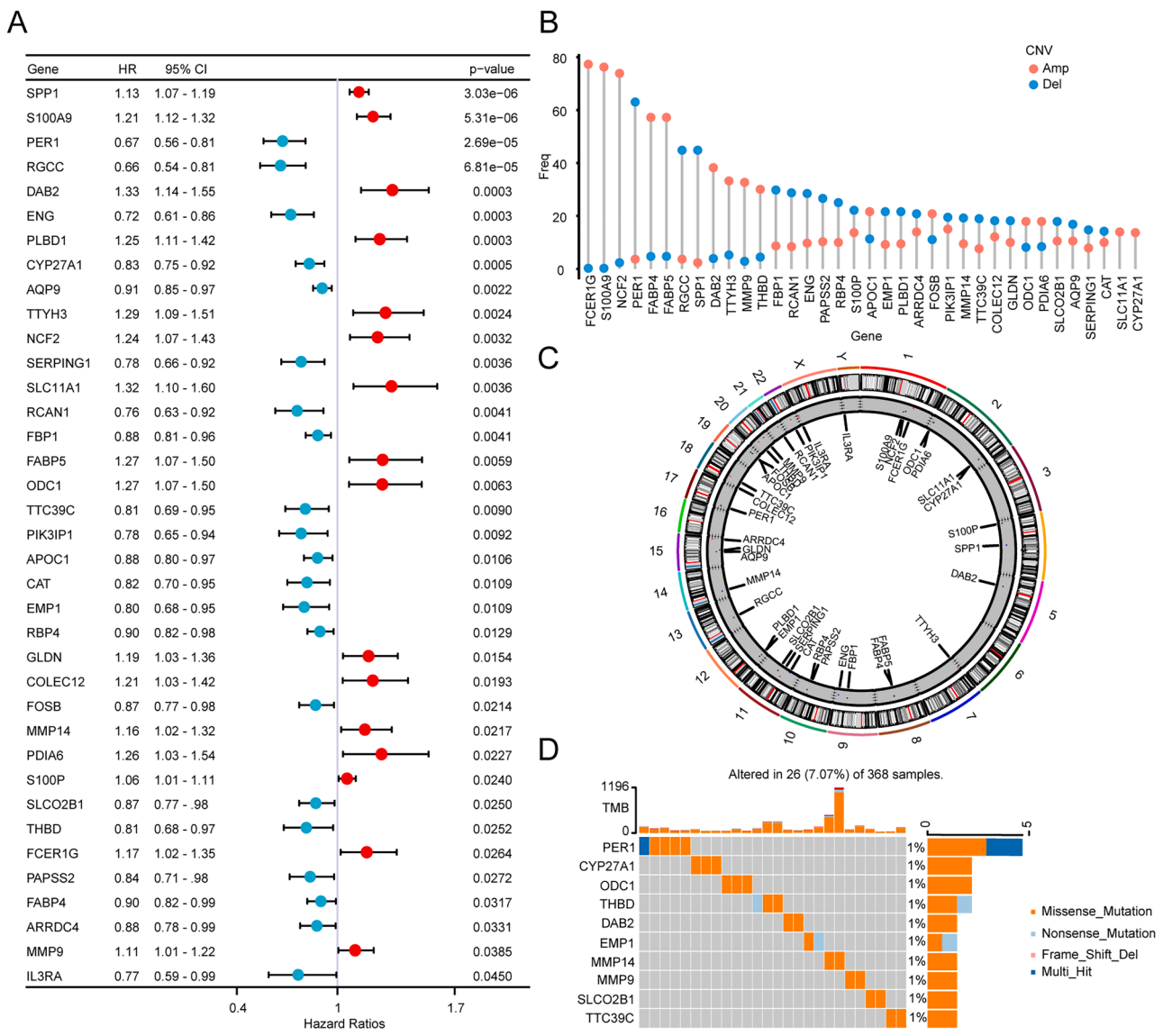

Fig. 1. Identification of prognosis-associated endocytosis genes within the TCGA dataset. (A) Forest plot of the 37 prognostic efferocytosis-related genes (ERGs) with $P < 0.05$. (B) Recurrence of CNVs among ERGs in HCC; column height correlates with the specific alteration frequency. (C) Genomic topography illustrating the chromosomal distribution of CNV events in ERGs. (D) Mutational landscape and prevalence of ERG alterations across the TCGA-LIHC cohort.

characterized by the activation of oncogenic and immunoinflammatory cascades, such as the Epithelial-Mesenchymal Transition (EMT), glycolysis, and inflammatory signaling. Conversely, metabolic signatures intrinsic to mature hepatocytes, including bile acid, fatty acid, and xenobiotic metabolism, alongside the PPAR signaling pathway, were markedly attenuated in Cluster 2 (Fig. 2F). This divergent landscape implies that while Cluster 1 maintains a more differentiated hepatocyte-like phenotype, Cluster 2 is defined by aggressive malignant traits and metabolic reprogramming. Ultimately, univariate and multivariate Cox proportional hazards models identified the cluster assignment as a robust, independent prognostic indicator for HCC (Fig. 2G, H).

Somatic variation analysis and immune infiltration landscape of the two subtypes

To investigate the genomic alterations associated with the two subtypes, we analyzed the somatic mutation profiles of HCC patients. Analysis of SNVs revealed that missense mutations were the predominant variant type in both groups, primarily consisting of cytosine to thymine (C > T) transitions (Fig. 3A, B). We displayed the mutation types of the top 25 most frequently mutated genes. Genes such as TTN, MUC16, and ALB were common mutation targets in both groups. Notably, CTNNB1 mutations were observed in 31% of patients in Cluster 1 (compared to 26% in Cluster 2). Conversely, TP53 mutations were more prevalent in Cluster 2, with a frequency of 35% (compared to 22% in Cluster 1) (Fig. 3C, D). This suggests that distinct driver gene mutations may govern the biological behaviors of the different subtypes. To further explore genomic instability, we evaluated somatic CNVs using GISTIC2 software and calculated G-scores for each genomic locus based on the frequency and amplitude of copy number changes. By selecting regions with a false discovery rate (FDR) q-value < 0.05, we obtained the global CNV profiles for each subtype. Although both subtypes exhibited widespread chromosomal instability, they displayed significant specificity in focal amplifications: the dominant event unique to Cluster 1 was the specific amplification of 11q13.3, whereas the high-risk Cluster 2 was synergistically driven by dual high-frequency amplifications at 8q24.21 and 1q21.3. Regarding deletions, Cluster 1 alterations were concentrated at 13q14.2, while those in Cluster 2 shifted to 13q13.1 (Fig. S2). These results indicate that specific amplifications at 8q and 1q may serve as the critical genetic basis maintaining the

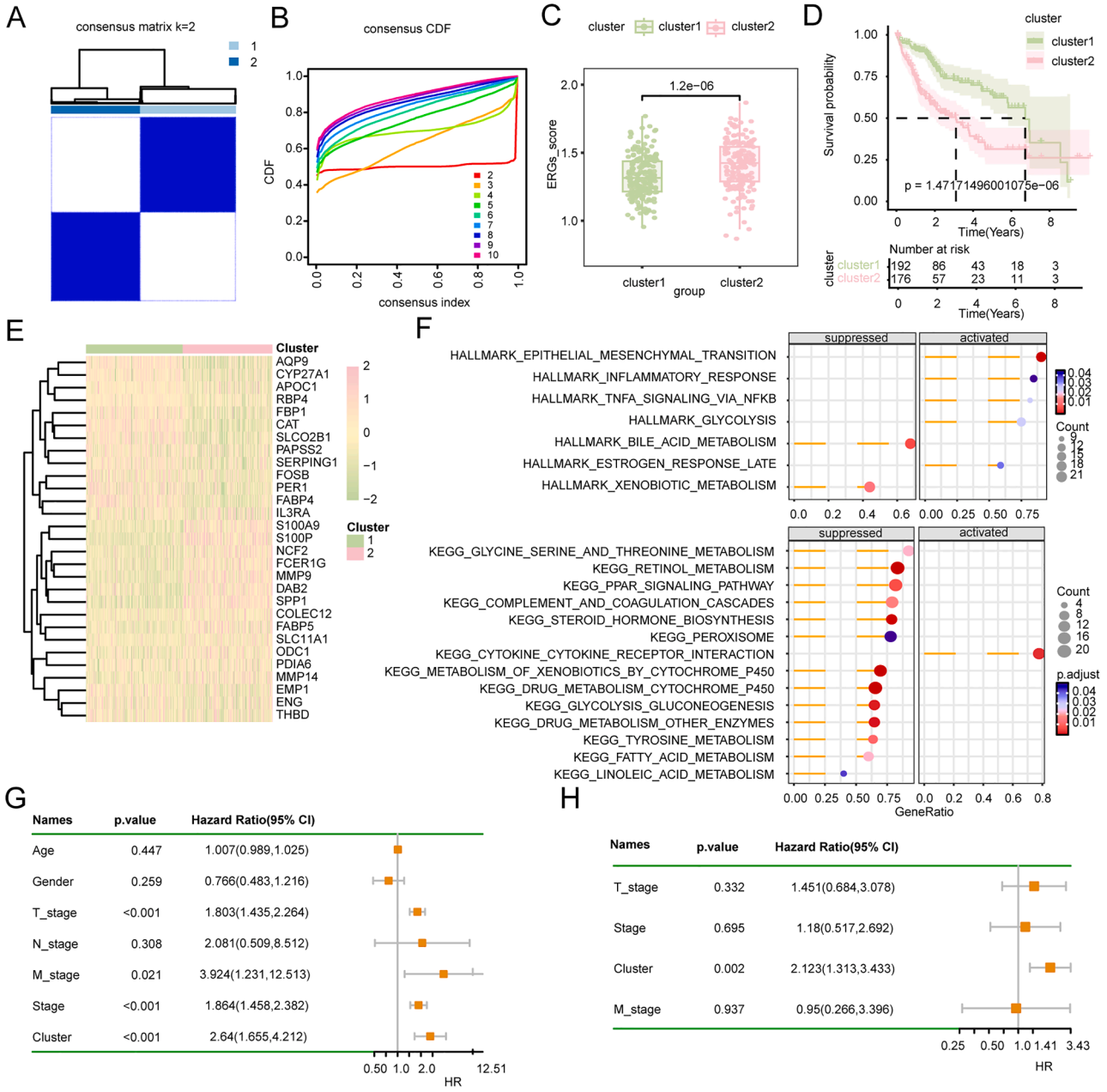

Fig. 2. Molecular stratification of HCC patients into divergent subtypes via consensus clustering. (A) Consensus clustering matrix of HCC patients based on 29 ERGs with k = 2. (B) Indication that k = 2 is the optimal cluster number. (C) Discrepancies in the global expression profiles of ERGs across the identified molecular subtypes. (D) Kaplan-Meier curves demonstrating superior overall survival outcomes for patients in Cluster 1. (E) Hierarchical clustering heatmap illustrating the divergent expression patterns of ERGs between the two subgroups. (F) Gene Set Enrichment Analysis (GSEA) utilizing HALLMARK and KEGG libraries, highlighting signaling cascades significantly upregulated in Cluster 2 relative to Cluster 1. (G) Univariate Cox regression assessing the prognostic impact of clinicopathological parameters and ERG-based subtypes. (H) Multivariate Cox regression identifying independent prognostic indicators among clinicopathological variables and ERG subtypes.

malignant phenotype of Cluster 2.

Subsequently, to explore the roles of these subtypes within the HCC TME, we first utilized the ESTIMATE algorithm to evaluate overall immune infiltration between the two subgroups, including stromal, estimate, and immunity scores (Fig. 3E). Using the CIBERSORT algorithm, we found that the Cluster 1 subgroup exhibited immune-active features, being significantly enriched in cytotoxic activated NK cells and monocytes. In sharp contrast, although the Cluster 2 subgroup presented a high immune score, it was dominated primarily by resting or undifferentiated cell populations. Cluster 2 was significantly enriched in abundant M0 macrophages and regulatory T cells (Tregs), suggesting that this subtype is in a state of immune stagnation or suppression (Fig. S1F, Fig. 3F). Furthermore, immune checkpoint analysis revealed that key molecules such as CD47, CTLA4, and PDCD1 were significantly upregulated in Cluster 2 (Fig. 3G). TIDE analysis further demonstrated that Cluster 2 had significantly higher TIDE scores, T cell dysfunction scores, and T cell exclusion scores (Fig. S1G, Fig. 3H), suggesting that the Cluster 2 subgroup is more prone to immune escape, whereas the Cluster 1 subgroup may derive greater benefit from immunotherapy.

Single-cell atlas landscape of HCC

To further investigate the specific cellular localization and activity of ERGs within the HCC tumor TME, we analyzed scRNA-seq data derived

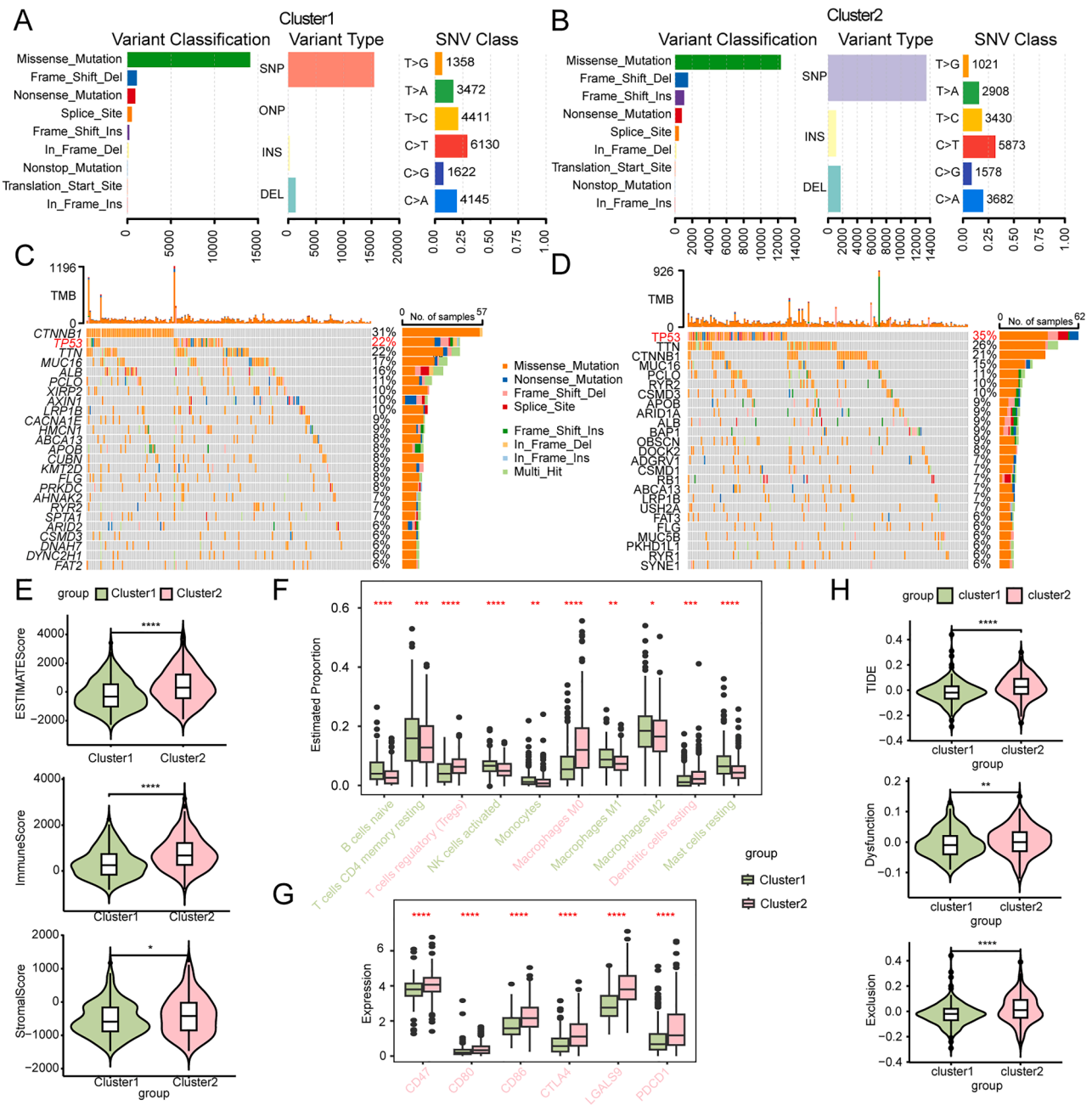

Fig. 3. Somatic mutational landscape and immune infiltration profiles across distinct molecular subtypes. (A-B) Overview of variant classifications, counts, types, and SNV categorization for Cluster 1 (n = 194) (A) and Cluster 2 (n = 180) (B) cohorts. (C-D) Comprehensive mutational architecture of the top 25 recurrently altered genes in Cluster 1 (C) and Cluster 2 (D). Corresponding mutational frequencies and specific variant types for each gene are delineated on the right panel. (E) Comparative assessment of Stromal, Immune, and ESTIMATE scores between the identified molecular subtypes. (F) Heterogeneity in immune cell infiltration levels among patients within Cluster 1 and Cluster 2. (G) Differential expression of immune checkpoints between Cluster 1 and Cluster 2 patients. (H) Comparison of TIDE, T-cell dysfunction, and T-cell exclusion scores between Cluster 1 and Cluster 2 patients.

from tumor and adjacent non-tumor tissues of HCC patients. After rigorous quality control and doublet removal, a total of 133,239 high-quality single cells were obtained. Based on the expression profiles of canonical marker genes, we identified these cells into nine major clusters: T cells (CD3D, CD3E), B cells (MS4A1, CD79A), NK cells (NKG7, GNLY), myeloid cells (LYZ, CD14), endothelial cells (PECAM1, VWF), fibroblasts (COL1A1, DCN), pericytes (RGS5, HIGD1B), plasma cells (JCHAIN, IGHG2), and epithelial cells (ALB, KRT8) (Fig. 4A, B). Cell proportion analysis revealed distinct cellular compositional differences between tissues; notably, compared to normal tissues, the proportions of myeloid cell and T cell populations were significantly expanded in HCC tissues, whereas endothelial cells were reduced (Fig. 4C). We further utilized CellChat analysis to explore intercellular interactions. Results showed that myeloid cells served as the primary communication hubs in HCC tissues. Compared to the normal group, myeloid cells in HCC exhibited the highest strength in both incoming and outgoing signaling interactions (Fig. 4D), highlighting their pivotal role in orchestrating TME signaling networks. Subsequently, we employed inferCNV to re-cluster epithelial cells into normal epithelial cells and malignant cells (Fig. 4E, F, Fig. S3D-F) and utilized CytoTRACE to infer their differentiation states. As expected, the CytoTRACE scores of malignant cells were significantly higher than those of normal epithelial cells, indicating

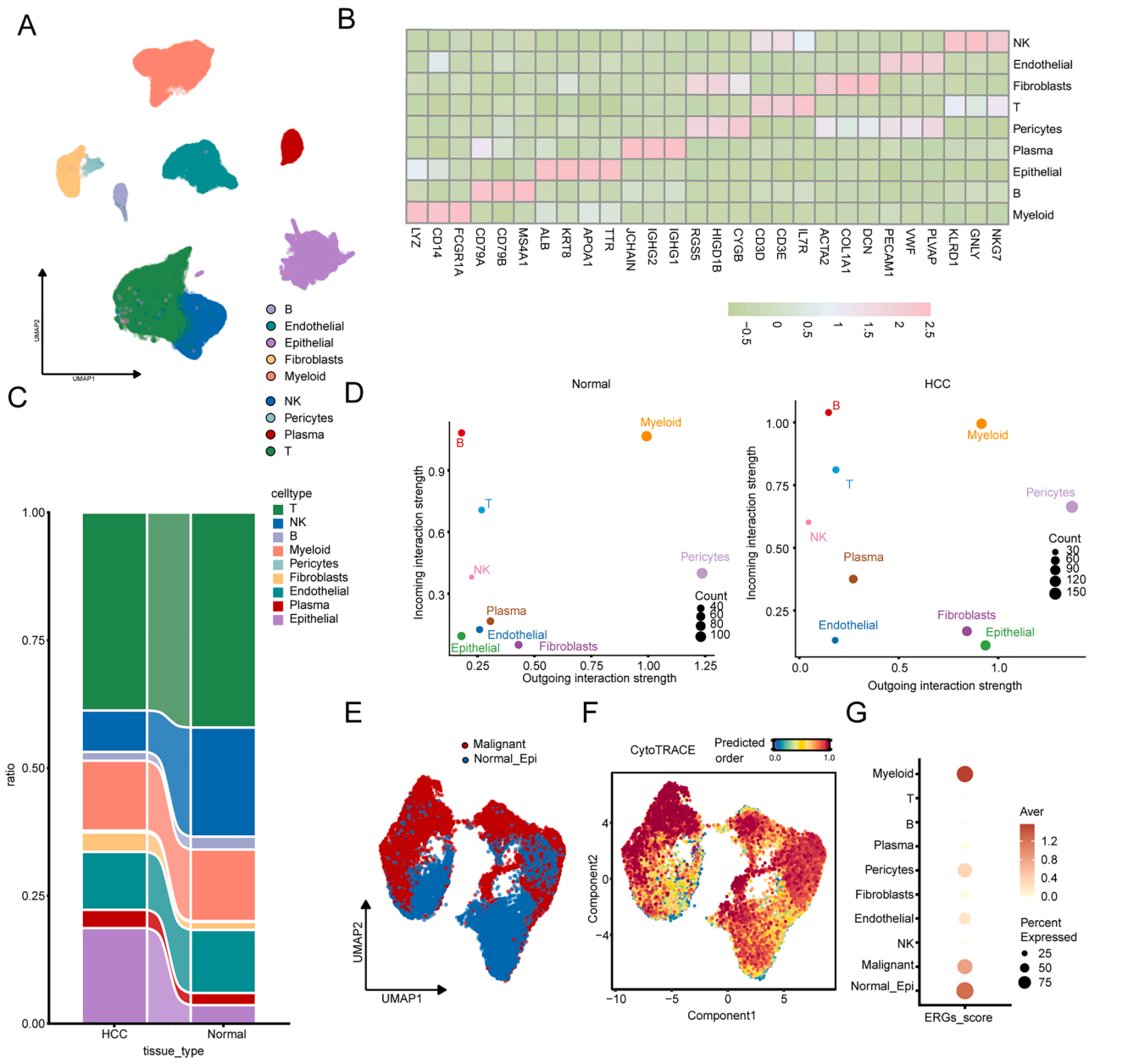

Fig. 4. Single-cell landscape of liver samples from HCC patients. (A) UMAP projection visualizing the spatial distribution of nine principal cellular lineages. (B) Heatmap illustrating the transcriptional signatures of canonical markers across the identified cell populations. (C) Fractional composition of sample origins within each of the nine major cell clusters. (D) Dot plot illustrating the differences in cellular communication activity of each cell type across different tissue types. (E) UMAP plot delineating the segregation of malignant cells from normal epithelial counterparts. (F) Differences in stemness indices between malignant cells and normal epithelial cells. (G) Bubble plot demonstrating the preferential enrichment of ERGs within the myeloid compartment.

that malignant cells reside in a less differentiated, stem-like state associated with aggressive proliferative characteristics. Next, we evaluated the expression profiles of the identified ERGs across all cell types. Notably, ERG scores were preferentially enriched in myeloid cells (Fig. 4G). This specific localization suggests that ERGs are not ubiquitously expressed but are intrinsically linked to the actively interacting myeloid cells.

Elevated efferocytic activity characterizes the SPP1+ macrophage subset

Given the observed enrichment of ERGs in myeloid cells within the global landscape, we re-clustered the myeloid cell population to precisely pinpoint the specific cell subpopulations driving this phenotype. Utilizing UMAP dimensionality reduction and clustering, and based on the top 10 specific marker genes identified via the FindAllMarkers function, we identified eight distinct myeloid subtypes: SPP1+ macrophages, TREM2+ macrophages, CXCL10+ macrophages, CCL3+ macrophages, MARCO+ macrophages, monocyte-derived DC-like cells, dendritic cells (DCs), and monocytes. Tissue distribution analysis revealed significant heterogeneity: SPP1_Mac and TREM2_Mac were primarily enriched in HCC tumor tissues, whereas monocytes were more abundant in normal tissues (Fig. 5A, Table S2). Each subtype exhibited unique marker gene expression patterns, as illustrated in the heatmap (Fig. 5B).

To investigate the dynamic changes of ERGs during myeloid differentiation, we mapped ERG scores onto the pseudotime trajectory and performed stratified analysis based on tissue origin (Fig. 5C). The results showed that ERG scores were significantly upregulated in the HCC

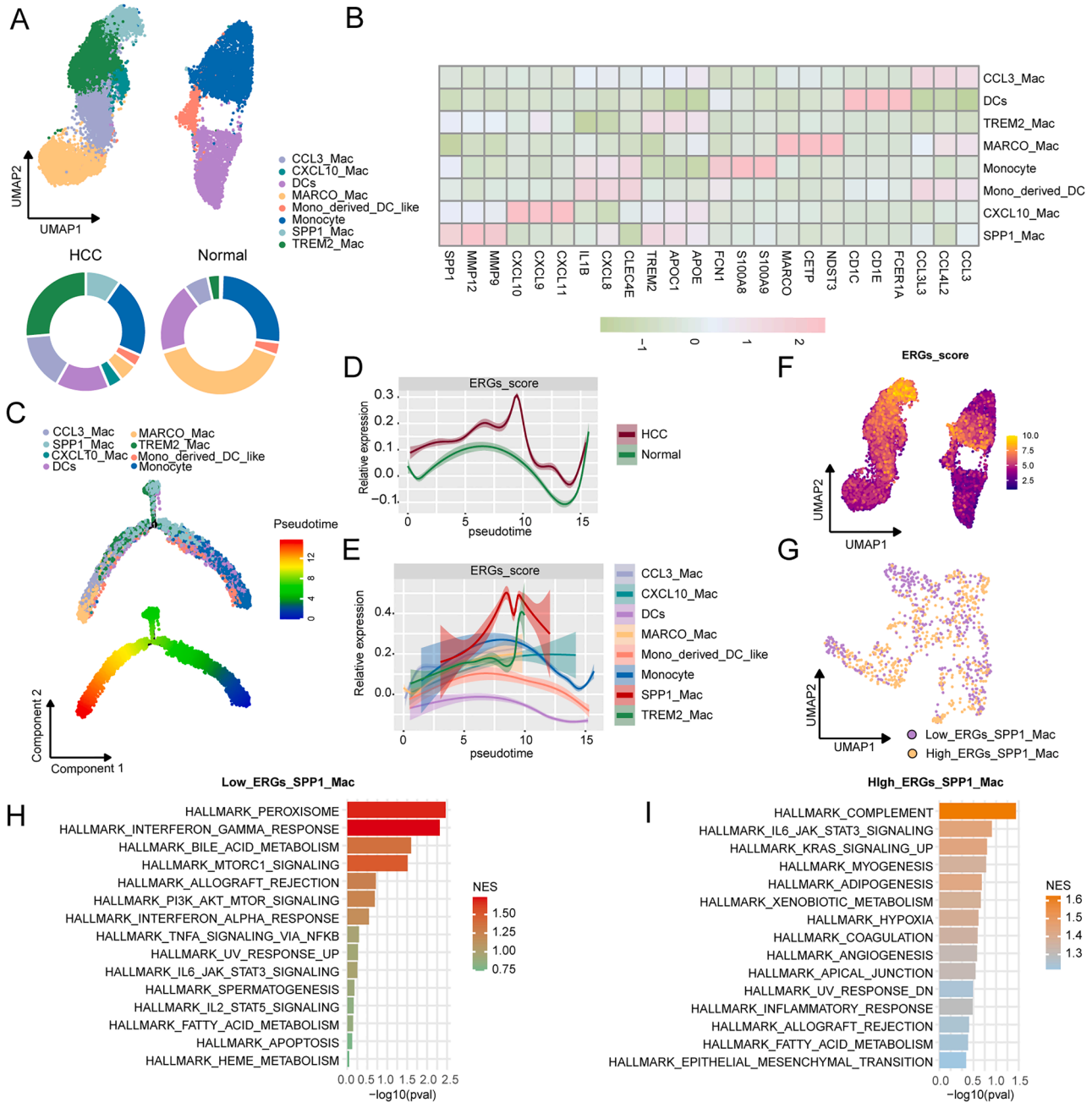

Fig. 5. Evaluation of macrophage involvement in efferocytic activity. (A) Subclustering of myeloid cells into 8 subpopulations and their proportional distribution differences across samples. (B) Expression patterns of the top 3 characteristic marker genes across each myeloid subset. (C) Pseudotime developmental trajectory analysis of myeloid subtypes. (D-E) Dynamic expression changes of ERGs across different samples (D) and different myeloid subtypes (E) along the trajectory. (F) UMAP projection illustrating the transcriptomic distribution of ERGs within the myeloid population. (G) UMAP plot delineating the segregation and spatial distribution of High_ERGs_SPP1_Mac and Low_ERGs_SPP1_Mac clusters. (H-I) GSEA functional enrichment analysis revealing differential pathway activities in High_ERGs_SPP1_Mac (H) and Low_ERGs_SPP1_Mac (I).

tumor group: throughout the differentiation trajectory, myeloid cells derived from HCC tissues exhibited consistently higher ERG expression levels compared to those from normal tissues (Fig. 5D). To determine the specific cell subpopulations responsible for this upregulation, we further analyzed the expression dynamics of different subtypes along the pseudotime. The results indicated that this tumor-specific high ERG score was primarily driven by SPP1+ macrophages, with partial involvement of TREM2+ macrophages. With the progression of pseudotime, SPP1+ macrophages displayed the most significant and robust increasing trend in ERG scores; although TREM2+ macrophages also showed a certain degree of score upregulation, the magnitude of induction in SPP1+ macrophages was more prominent at the end of the trajectory (Fig. 5E, F). This suggests that the activation of the ERG signaling pathway is a key feature of macrophage maturation and function, particularly for the SPP1+ subset, within the TME.

To further dissect the functional heterogeneity inherent in this pivotal subpopulation, SPP1_Mac cells were stratified into High_ERGs_SPP1_Mac and Low_ERGs_SPP1_Mac cohorts, utilizing the median ERG score as the classification threshold (Fig. 5G). Gene Set Enrichment Analysis (GSEA) delineated divergent biological landscapes between these two subgroups. Specifically, the Low_ERGs_SPP1_Mac cluster exhibited an enrichment of pathways associated with immune activation

and metabolic homeostasis, notably the interferon response and bile acid metabolism (Fig. 5H). Conversely, the High_ERGs_SPP1_Mac subset manifested a significant predilection for aggressive, pro-tumorigenic signatures, including Hypoxia, Angiogenesis, Epithelial-Mesenchymal Transition (EMT), and the IL6-JAK-STAT3 signaling axis (Fig. 5I). Collectively, these results suggest that elevated ERG expression identifies a specialized niche of SPP1+ macrophages; their pronounced hypoxic adaptation and pro-angiogenic orientation indicate a functional synergy with malignant cells within the tumor microenvironment.

Integrated spatial and single-cell analysis reveals close interaction between High_ERGs_SPP1_Mac and malignant cells

To investigate the functional role of High-ERG SPP1+ macrophages in the TME, we first utilized CellChat to analyze the intercellular communication network at the single-cell level. The analysis revealed that High_ERGs_SPP1_Mac acts as a core "communication hub" within the TME, exhibiting extremely high incoming and outgoing interaction strengths across all cell types. Notably, interaction count heatmaps and network plots consistently confirmed that, compared to other cell types, High_ERGs_SPP1_Mac established robust signaling crosstalk with

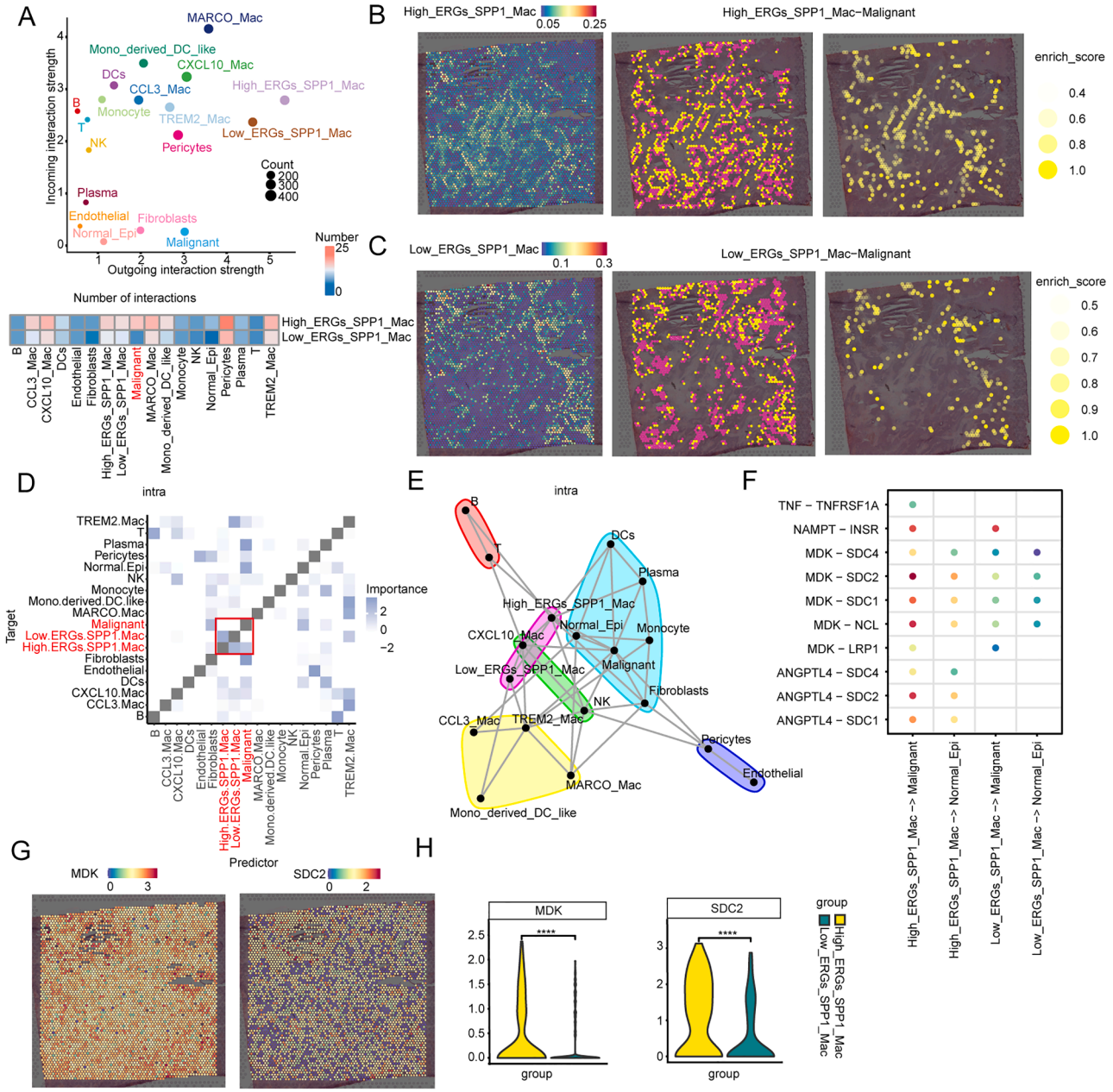

Fig. 6. Intercellular communication network between High_ERGs_SPP1_Mac and malignant cells. (A) Cellular signaling networks reconstructed via CellChat. The scatter plot delineates the incoming and outgoing signaling intensities across various lineages, while the heatmap provides a quantitative assessment of interaction frequencies between SPP1_Mac subsets and other cellular components. (B, C) Visualization of the in situ spatial distribution of High_ERGs_SPP1_Mac (B) and Low_ERGs_SPP1_Mac (C) on tissue sections, along with neighborhood analysis indicating the intensity of their spatial co-localization with malignant cells. (D, E) Spatial microenvironments and intercellular interactions identified by MISTy analysis based on sample VISDS000512, revealing that High_ERGs_SPP1_Mac and malignant cells form a specific spatial niche. (F) Dot plot showing specific ligand-receptor pairs mediating the crosstalk between SPP1_Mac subpopulations and malignant/normal epithelial cells. (G) Spatial expression distribution of MDK and SDC2 molecules on tissue sections. (H) Violin plots depicting the divergent expression of MDK and SDC2 within identified SPP1_Mac subpopulations.

malignant cells, a feature significantly more pronounced than in the Low_ERGs_SPP1_Mac group (Fig. 6A). To validate the findings from single-cell analysis at the in situ tissue level, we performed ST analysis on HCC tumor tissues (samples VISDS000512 and VISDS000516). First, we applied rigorous filtering and normalization to the spatial sequencing data using standard quality control metrics (Fig. S4A, B). Combined with unsupervised clustering analysis, we identified spatial regions with distinct transcriptional signatures (Fig. S4C, D). To systematically map the single-cell-defined cell identities to spatial locations, we applied the RCTD algorithm to deconvolute the ST data (Fig. S5A–D). Subsequently, to explore the spatial neighborhood relationships between macrophage subpopulations and tumor cells, we conducted cellular neighborhood network analysis. The results revealed that High_ERGs_SPP1_Mac formed a tighter heterotypic cellular network with malignant cells and exhibited significant co-localization signals (Fig. 6B, C). MistyR spatial dependency analysis further confirmed that, compared to Low_ERGs_SPP1_Mac, High_ERGs_SPP1_Mac displayed stronger spatial interaction potential, with particularly prominent interaction strength with malignant cells (Fig. 6D, E; Fig. S5E, F). To uncover the molecular mechanisms maintaining this specific spatial distribution, we focused on comparing the signaling characteristics of interactions with malignant cells between High_ERGs_SPP1_Mac and

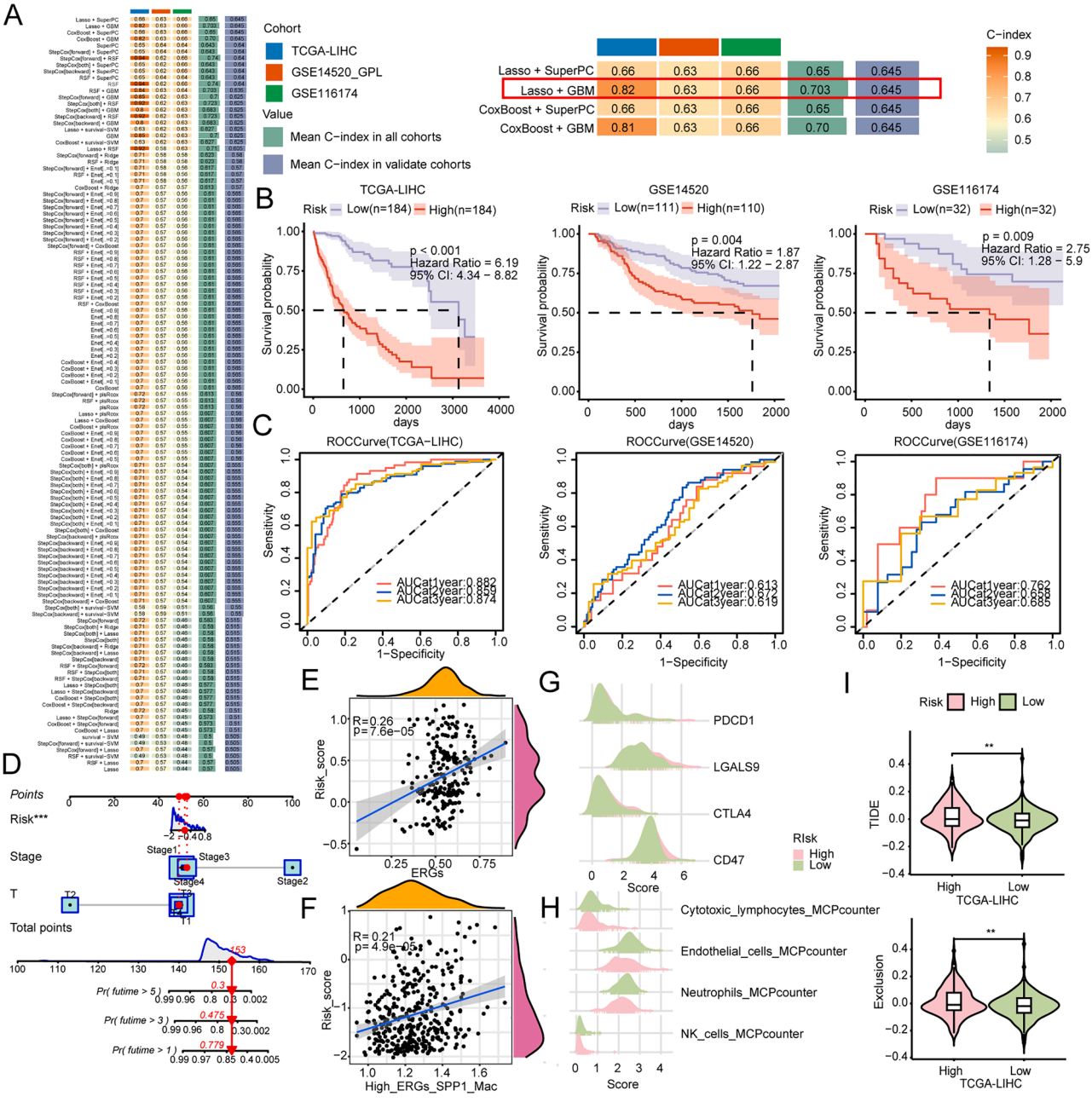

Fig. 7. Construction of an HCC prognostic model based on High_ERGs_SPP1_Mac. (A) Heatmap illustrating the C-indices of 101 integrative models across the TCGA-LIHC, GSE14520, and GSE116174 datasets. The color spectrum from red to green denotes a gradient from high to low predictive accuracy. The Lasso-GBM ensemble was identified as the optimal configuration for model assembly. (B) Kaplan-Meier survival curves depicting overall survival (OS) probabilities across all cohorts. (C) Time-dependent ROC curves visualizing AUC values for 1-, 2-, and 3-year survival forecasting across datasets. (D) Construction of a clinical nomogram based on the risk score, T stage, and clinical stage. (E, F) Correlations between the risk score and ERGs (E), as well as High_ERGs_SPP1_Mac infiltration (F). (G, H) Differences in immune checkpoint expression (G) and immune cell infiltration (H) between the high- and low-risk score groups. (I) Variations in TIDE and T-cell exclusion metrics across the high- and low-risk score categories.

Low_ERGs_SPP1_Mac. The analysis results suggested a distinct difference in the activity of the MDK signaling pathway between the two subgroups: compared to the relatively weak signaling activity in the Low_ERGs_SPP1_Mac group, this pathway showed significantly greater enrichment in the communication directed from High_ERGs_SPP1_Mac to malignant cells (Fig. 6F; Fig. S5G). Network centrality analysis further delineated this functional divergence, indicating that High_ERGs_SPP1_Mac appeared to assume a more dominant role as an MDK signal "sender" compared to the Low group, suggesting a stronger potential for this subpopulation to modulate malignant cells (Fig. S5H).

To further investigate the molecular details behind this difference, we evaluated the relative contributions of specific ligand-receptor pairs. The results showed that the MDK–SDC2 interaction pair exhibited a high interaction probability in the communication between High_ERGs_SPP1_Mac and malignant cells (Fig. S5I). Subsequently, we visualized the in situ distribution of the ligand and receptor. Spatial feature plots revealed extensive overlapping expression of MDK and SDC2 within the tumor tissue, confirming the physical basis for their interaction within the spatial neighborhood (Fig. 6G). To clarify the specific cellular source of this spatial signal, subsequent gene expression analysis demonstrated that the expression level of the ligand MDK was significantly higher in High_ERGs_SPP1_Mac than in Low_ERGs_SPP1_Mac (Fig. 6H). Coupled with the specific upregulation of MDK in the High group, this spatial gene co-expression pattern implies that the High-ERG subpopulation may more effectively establish connections with malignant cells via the MDK–SDC2 axis, whereas the Low group fails to form similar interactions due to the lack of key ligand expression.

Construction of a prognostic model based on High_ERGs_SPP1_Mac related genes

To develop a robust prognostic framework, we first identified 148 characteristic genes of High_ERGs_SPP1_Mac through the FindMarkers utility (Table S3). An expansive machine learning integration strategy, encompassing 10 distinct algorithms, was subsequently deployed to construct 101 predictive candidates. These models were rigorously benchmarked based on their mean concordance index (C-index) across the TCGA-LIHC discovery cohort and two external validation sets (GSE14520, GSE116174). Ultimately, the Lasso and GBM combination emerged as the optimal signature due to its superlative predictive stability (Fig. 7A). Kaplan-Meier survival analysis demonstrated that the resulting risk score effectively stratified patients into divergent prognostic groups across all cohorts ($P < 0.01$) (Fig. 7B). Time-dependent ROC analysis further validated this efficacy; in the training cohort, AUC values reached 0.882, 0.859, and 0.874 at 1, 2, and 3 years, respectively. The model exhibited consistent performance in GSE14520 (AUC: 0.613–0.672) and GSE116174 (AUC: 0.658–0.762) (Fig. 7C). Clinical utility was further substantiated by a meta-analysis of univariate Cox regression, identifying the risk score as a potent independent prognosticator (Combined HR > 3, $P < 0.003$) (Fig. S6A). Subsequent screening of clinicopathological features confirmed that only T stage and Pathological Stage maintained significant prognostic relevance ($P < 0.001$), while demographic factors such as age and gender showed no association (Fig. S6B). Consequently, a visual Nomogram was constructed by integrating the risk score with these validated clinical parameters, achieving high-precision multivariate survival forecasting (Fig. 7D). To decipher the biological underpinnings of this model, we explored the intrinsic relationship between the risk score and High_ERGs_SPP1_Mac attributes. Correlation analysis within the TCGA cohort revealed positive associations between the risk score and both ERGs scores ($R = 0.26$) and High_ERGs_SPP1_Mac density ($R = 0.21$) (Fig. 7E, F). These findings were corroborated by external validation, where high-risk patients consistently displayed elevated ERGs signatures and High_ERGs_SPP1_Mac infiltration ($P < 0.05$) (Fig. S6C, D). Comprehensive immunoprofiling unveiled a profound immunosuppressive landscape in the high-risk group. Specifically, inhibitory checkpoints including PDCD1 (PD-1), CTLA4, CD47, and LGALS9 were markedly upregulated, signaling potential immune exhaustion (Fig. 7G). Conversely, MCP-counter analysis indicated that cytotoxic lymphocytes and NK cells were significantly sequestered in the low-risk group but depleted in the high-risk group (Fig. 7H). This "cold" immune phenotype was further supported by elevated TIDE and immune exclusion scores in the high-risk cohort (Fig. 7I, Fig. S6E). Collectively, the prognostic utility of our model reflects the immunosuppressive TME driven by the High_ERGs_SPP1_Mac subpopulation, characterized by checkpoint-mediated evasion and effector cell exclusion.

Screening and validation of the key core gene TPI1 from prognostic models

To discern the primary molecular drivers of dismal prognosis and identify viable clinical biomarkers, we evaluated the selection frequency of candidate genes across the 101 integrated machine learning frameworks. Data revealed that TPI1, together with SPP1, S100A9, and LDHA, emerged as the most frequently selected features (Frequency > 15), underscoring their pivotal roles in maintaining model predictive stability (Fig. 8A). Subsequent univariate Cox regression analysis was implemented to prioritize the clinical relevance of these high-frequency candidates. TPI1 exhibited the most substantial Hazard Ratio (HR $=$ 1.672, 95% CI: 1.278–2.188, $P < 0.001$), establishing it as a preeminent risk factor for mortality (Fig. 8B). Based on these findings, TPI1 was selected as the focal point for downstream investigation. Comparative expression profiling within the TCGA-LIHC and GSE14520 cohorts confirmed a significant upregulation of TPI1 in HCC malignancies relative to adjacent normal liver tissues (Fig. 8C). To further ascertain its independent prognostic significance, patients were stratified by TPI1 expression levels. Kaplan-Meier analysis demonstrated that elevated TPI1 expression was associated with markedly curtailed overall survival (OS) across both discovery and validation cohorts ($P < 0.01$) (Fig. 8D). To complement these in silico findings with cellular evidence, we quantified TPI1 expression in the immortalized hepatocyte line MIHA alongside HCC-derived Huh7 and SMMC-7721 lines. RT-qPCR assays demonstrated that TPI1 mRNA abundance was significantly higher in HCC cell lines compared to MIHA ($P < 0.01$) (Fig. 8E). This trend was corroborated by Western blotting, which revealed a distinct enrichment of TPI1 protein in HCC cells relative to the normal hepatocyte control ($P < 0.01$) (Fig. 8F). These in vitro observations are highly congruent with our bioinformatic insights, collectively validating the overexpression and oncogenic potential of TPI1 in HCC.

Discussion

The pronounced intratumoral heterogeneity of HCC and its intricate immunosuppressive microenvironment constitute primary obstacles limiting clinical therapeutic efficacy. Although the advent of ICIs has revolutionized HCC treatment, the highly heterogeneous nature of the TME results in significantly variable clinical benefits for patients [20]. Traditional tumor staging systems often fail to capture these complex microenvironmental features, thereby hindering the implementation of precision medicine. To address this gap, we constructed a robust ERG scoring system. Unlike traditional single-gene biomarkers, this system integrates multi-omics features to quantify the biological heterogeneity of HCC. Our findings suggest that the ERG score is not merely a statistical prognostic marker, but also a functional indicator of a specific immunosuppressive state.

The ERG scoring system demonstrates unique biological utility in resolving the TME landscape. A high ERG score defines a molecular subtype with distinct malignant features. Metabolically, this subtype exhibits significant upregulation of glycolysis and hypoxia pathways. Immunologically, it is characterized by an immunosuppressive microenvironment dominated by M0 macrophages and Tregs, together with high expression of checkpoint molecules such as PD-1 and CTLA4. This finding is consistent with the immunometabolic framework of

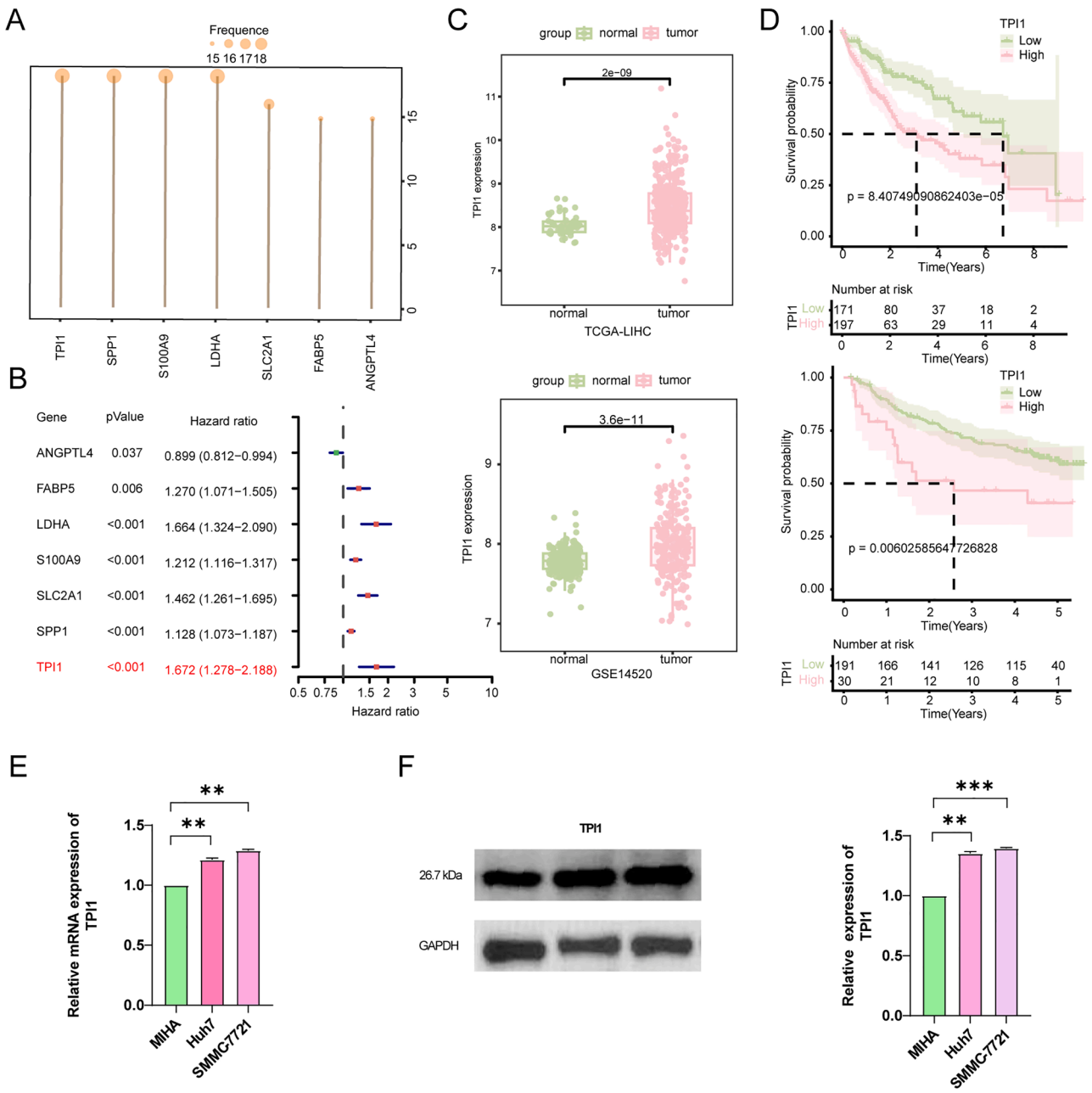

Fig. 8. Identification and multi-level validation of TPI1 as a pivotal prognostic driver. (A) Lollipop chart showing the selection frequency of feature genes across 101 machine learning models, identifying TPI1 as a high-frequency core gene. (B) Univariate Cox regression analysis of candidate genes; TPI1 exhibited the most substantial Hazard Ratio (HR), characterizing it as a preeminent risk factor. (C) Expression profiling of TPI1 across malignant versus paracancerous tissues within the TCGA-LIHC discovery cohort (upper) and GSE14520 validation cohort (lower). (D) Kaplan-Meier overall survival curves comparing patients with high and low TPI1 expression in the TCGA-LIHC (top) and GSE14520 (bottom) cohorts. (E) Relative mRNA expression levels of *TPI1* in the immortalized human hepatocyte line (MIHA) and HCC cell lines (Huh7, SMMC-7721) determined by RT-qPCR. (F) Representative Western blot images (left) and densitometric quantification (right) of TPI1 protein levels in MIHA, Huh7, and SMMC-7721 cells. GAPDH served as the internal loading control. Data are expressed as mean ± SD. $^{**}P < 0.01$, $^{***}P < 0.001$.

efferocytosis. Previous studies have shown that the clearance of apoptotic cells in the TME promotes the release of anti-inflammatory mediators by phagocytes and is accompanied by the accumulation of metabolites such as lactate [35,36]. We therefore speculate that this metabolic reprogramming may create a nutrient-restricted microenvironment that suppresses effector T-cell activity and contributes to checkpoint upregulation and immune exclusion.

At cellular resolution, the biological basis of the ERG score is further anchored to a specific cell subpopulation—High_ERGs_SPP1_Mac. Our analysis identifies this subpopulation as the core driver of high ERG scores. Recent single-cell sequencing studies have defined SPP1+ macrophages as a major pro-tumorigenic population in HCC, closely associated with poor prognosis [37]. While existing literature has linked them to hypoxia or angiogenesis, the specific contribution of endocytic activity in shaping their functional heterogeneity remains underexplored. Our results suggest that this pro-tumorigenic phenotype may be linked to high efferocytic activity. This is consistent with the view that the metabolic adaptation of macrophages within specific microenvironments dictates their functional polarization [38]. We speculate that High_ERGs_SPP1_Mac may represent a terminally differentiated

macrophage state shaped by sustained efferocytic activity, with enhanced capacity to produce pro-angiogenic and immunosuppressive signals.

Evidence from ST further elucidates the spatial mechanisms underlying this scoring system. The risk reflected by a high ERG score spatially corresponds to a significant physical co-localization between High_ERGs_SPP1_Mac and malignant cells. The MDK-SDC2 axis was identified as a potential mechanism mediating this cellular neighborhood interaction. MDK is a heparin-binding growth factor that has been widely reported to promote tumor growth, metastasis, and chemotherapy resistance in multiple cancers [39]. SDC2, in turn, has been implicated in cell adhesion and signal transduction [40]. This tight spatial dependency suggests that high-endocytic macrophages may directly support tumor cells via paracrine MDK signaling, echoing findings regarding macrophage-derived factors supporting the tumor survival niche [41]. However, this interaction remains computationally inferred and spatially supported, and future co-culture, blocking, or knockdown experiments will be needed for direct functional validation.

To bridge the gap between single-cell biological insights and clinical utility, the present study orchestrated a large-scale integrative machine-learning framework. By systematically evaluating 10 distinct algorithms and their 101 combinatorial iterations, the Lasso-GBM ensemble was identified as the superlative predictive architecture. This rigorous computational selection underscores the robust generalizability of the ERG-related risk score across diverse, multicenter cohorts. Furthermore, meta-analytical evidence established the score as a potent independent prognosticator, which effectively complements traditional T-staging and pathological frameworks to facilitate the construction of a high-precision nomogram.

Crucially, the prognostic fidelity of this model is intrinsically tethered to its underlying biological connotation. Our correlation analysis reveals that the ERG-related risk score essentially serves as a "digital surrogate" for the infiltration density of High_ERGs_SPP1_Mac within the TME. High-risk patients not only manifested a pronounced enrichment of these specific macrophages but also exhibited a systemic upregulation of inhibitory checkpoints (PDCD1, CTLA4, CD47, and LGALS9) alongside a profound depletion of cytotoxic lymphocytes. Consequently, this "cold tumor" architecture, characterized by elevated TIDE and immune exclusion metrics, concurs with our hypothesis that immune evasion orchestrated by high-efferocytic macrophages constitutes the fundamental driver of dismal survival. To further support this interpretation, we validated the overexpression of the hub gene TPI1 in HCC by RT-qPCR and confirmed increased TPI1 protein levels by Western blotting. As a key glycolytic enzyme, TPI1 may reflect tumor-cell metabolic adaptation within an immunosuppressive microenvironment and may contribute to poor clinical outcome. Our findings concur with previous reports identifying TPI1 as a catalyst for malignant progression in gastric [42] and lung carcinomas [43]. Collectively, the integration of computational and experimental data strongly supports the role of TPI1 as a critical metabolic-immune node, linking glycolytic reprogramming to the establishment of an immunosuppressive landscape.

Despite these findings, several limitations should be acknowledged. First, the present study was primarily based on retrospective public cohorts, and prospective validation will be needed to further support its clinical applicability. Second, although our transcriptomic and spatial analyses strongly supported an efferocytosis-related phenotype of High_ERGs_SPP1_Mac, direct functional assays confirming its efferocytic capacity were not performed in the present study. Third, the MDK-SDC2 interaction was inferred computationally and supported by spatial co-localization, but was not directly validated experimentally. Finally, although TPI1 was validated in HCC cell lines, additional validation in clinical tissue samples would further strengthen the translational relevance of our findings.

Ethics approval and consent to participate

Not applicable.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This study is funded by Project of Naval Medical University and Eastern Hepatobiliary Surgery Hospital (2022QN098/2022GZR007) and Project of Shanghai Municipal Science and Technology Commission (22Y11908900).

CRediT authorship contribution statement

Qiuyun Yi: Software, Methodology, Investigation, Formal analysis, Data curation. **Bin Zhou:** Writing – original draft, Methodology, Investigation, Data curation. **Xiaohui Gu:** Writing – original draft, Visualization, Formal analysis, Data curation. **Zihui Ma:** Writing – original draft, Methodology, Investigation. **Bochen Pan:** Writing – original draft, Software, Methodology, Data curation. **Zhou Yuan:** Writing – review & editing, Supervision, Resources, Project administration. **Wen Wen:** Writing – review & editing, Software, Resources, Project administration, Conceptualization. **Chongde Lu:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2026.102801.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA *Cancer J. Clin.* 71 (3) (2021) 209–249.
[2] J.A. Marrero, L.M. Kulik, C.B. Sirlin, A.X. Zhu, R.S. Finn, M.M. Abecassis, L. R. Roberts, J.K. Heimbach, Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American association for the study of liver diseases, Hepatol. (Baltim., Md) 68 (2) (2018) 723–750.
[3] EASL clinical practice guidelines on the management of hepatocellular carcinoma, J. Hepatol. 82 (2) (2025) 315–374.
[4] J.M. Llovet, F. Castet, M. Heikenwalder, M.K. Maini, V. Mazzaferro, D.J. Pinato, E. Pikarsky, A.X. Zhu, R.S. Finn, Immunotherapies for hepatocellular carcinoma, Nat. Rev. Clin. Oncol. 19 (3) (2022) 151–172.
[5] R.S. Finn, S. Qin, M. Ikeda, P.R. Galle, M. Ducreux, T.Y. Kim, M. Kudo, V. Breder, P. Merle, A.O. Kaseb, et al., Atezolizumab plus Bevacizumab in unresectable hepatocellular carcinoma, N. Engl. J. Med. 382 (20) (2020) 1894–1905.
[6] E. Boada-Romero, J. Martinez, B.L. Heckmann, D.R. Green, The clearance of dead cells by efferocytosis, Nat. Rev. Mol. Cell Biol. 21 (7) (2020) 398–414.
[7] S. Nagata, R. Hanayama, K. Kawane, Autoimmunity and the clearance of dead cells, Cell 140 (5) (2010) 619–630.
[8] M.R. Elliott, F.B. Chekeni, P.C. Trampont, E.R. Lazarowski, A. Kadl, S.F. Walk, D. Park, R.I. Woodson, M. Ostankovich, P. Sharma, et al., Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance, Nature 461 (7261) (2009) 282–286.
[9] C.V. Rothlin, S. Ghosh, E.I. Zuniga, M.B. Oldstone, G. Lemke, TAM receptors are pleiotropic inhibitors of the innate immune response, Cell 131 (6) (2007) 1124–1136.
[10] I. Kourtzelis, G. Hajishengallis, T. Chavakis, Phagocytosis of apoptotic cells in resolution of inflammation, Front. Immunol. 11 (2020) 553.

[11] A.C. Doran, A. Yurdagul Jr., I. Tabas, Efferocytosis in health and disease, Nat. Rev. Immunol. 20 (4) (2020) 254–267.
[12] L. Xiao, L. Zhang, C. Guo, Q. Xin, X. Gu, C. Jiang, J. Wu, "Find me" and "eat me" signals: tools to drive phagocytic processes for modulating antitumor immunity, Cancer commun. (Lond., Engl.) 44 (7) (2024) 791–832.
[13] A. Yurdagul Jr., M. Subramanian, X. Wang, S.B. Crown, O.R. Ilkayeva, L. Darville, G.K. Kolluru, C.C. Rymond, B.D. Gerlach, Z. Zheng, et al., Macrophage metabolism of apoptotic cell-derived arginine promotes continual efferocytosis and resolution of injury, Cell Metab. 31 (3) (2020) 518–533.e10.
[14] Q. Yan, K. Li, L. Chen, A. Wang, Y. Xi, H. Xiao, L. Yuan, Metab. Reprogramming Efferocytosis 13 (2025), 2025.
[15] B. Cai, C. Kasikara, A.C. Doran, R. Ramakrishnan, R.B. Birge, I. Tabas, MerTK signaling in macrophages promotes the synthesis of inflammation resolution mediators by suppressing CaMKII activity, Sci. Signal. 11 (549) (2018).
[16] D. Ngai, M. Schilperoort, I. Tabas, Efferocytosis-induced lactate enables the proliferation of pro-resolving macrophages to mediate tissue repair, Nat. Metabol. 5 (12) (2023) 2206–2219.
[17] Y. Tang, H. Zang, Q. Wen, S. Fan, AXL in cancer: a modulator of drug resistance and therapeutic target, J. Exp. Clin. Cancer Res. 42 (1) (2023) 148.
[18] A. Rayford, F. Gärtner, M.P. Ramnefjell, J.B. Lorens, D.R. Micklem, M. Aanerud, A. S.T. Engelsen, AXL expression reflects tumor-immune cell dynamics impacting outcome in non-small cell lung cancer patients treated with immune checkpoint inhibitor monotherapy, Front. Immunol. 15 (2024) 1444007.
[19] J.C. Stanford, C. Young, D. Hicks, P. Owens, A. Williams, D.B. Vaught, M. M. Morrison, J. Lim, M. Williams, D.M. Brantley-Sieders, et al., Efferocytosis produces a prometastatic landscape during postpartum mammary gland involution, J. Clin. Invest. 124 (11) (2014) 4737–4752.
[20] H. Li, J. Zhang, Y. Zhang, Z. Li, Y. Han, Y. Lun, A pan-cancer study of tumour-associated efferocytosis core genes and preliminary exploration of TIMD4 in renal cell carcinoma, J. Cell. Mol. Med. 29 (12) (2025) e70671.
[21] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinform.(Oxf.Engl.) 26 (12) (2010) 1572–1573.
[22] D. Zeng, Z. Ye, R. Shen, G. Yu, J. Wu, Y. Xiong, R. Zhou, W. Qiu, N. Huang, L. Sun, et al., IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures, Front. Immunol. 12 (2021) 687975.
[23] M.S. Rooney, S.A. Shukla, C.J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity, Cell 160 (1–2) (2015) 48–61.
[24] A. Mayakonda, D.C. Lin, Y. Assenov, C. Plass, H.P. Koeffler, Maftools: efficient and comprehensive analysis of somatic variants in cancer, Genome Res. 28 (11) (2018) 1747–1756.
[25] C.H. Mermel, S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (4) (2011) R41.
[26] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.R. Loh, Raychaudhuri S: fast, sensitive and accurate integration of single-cell data with harmony, Nat. Methods 16 (12) (2019) 1289–1296.
[27] Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck III, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, et al., Integrated analysis of multimodal single-cell data, Cell 184 (13) (2021) e3529.
[28] I. Tirosh, A.S. Venteicher, C. Hebert, L.E. Escalante, A.P. Patel, K. Yizhak, J. M. Fisher, C. Rodman, C. Mount, M.G. Filbin, et al., Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma, Nature 539 (7628) (2016) 309–313.
[29] S. Jin, M.V. Plikus, Q. Nie, CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics, Nat. Protoc. 20 (1) (2025) 180–219.
[30] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories, Nat. Methods 14 (10) (2017) 979–982.
[31] G.S. Gulati, S.S. Sikandar, D.J. Wesche, A. Manjunath, A. Bharadwaj, M.J. Berger, F. Ilagan, A.H. Kuo, R.W. Hsieh, S. Cai, et al., Single-cell transcriptional diversity is a hallmark of developmental potential, Sci. (N. Y., NY) 367 (6476) (2020) 405–411.
[32] D.M. Cable, E. Murray, L.S. Zou, A. Goeva, E.Z. Macosko, F. Chen, R.A. Irizarry, Robust decomposition of cell type mixtures in spatial transcriptomics, Nat. Biotechnol. 40 (4) (2022) 517–526.
[33] J. Tanevski, R.O.R. Flores, A. Gabor, D. Schapiro, J. Saez-Rodriguez, Explainable multiview framework for dissecting spatial relationships from highly multiplexed data, Genome Biol. 23 (1) (2022) 97.
[34] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, et al., ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innov. (Cam. (Mass)) 2 (3) (2021) 100141.
[35] S. Morioka, C. Maueröder, K.S. Ravichandran, Living on the edge: efferocytosis at the interface of homeostasis and pathology, Immunity 50 (5) (2019) 1149–1162.
[36] H. Rostamian, M. Khakpoor-Koosheh, L. Jafarzadeh, E. Masoumi, K. Fallah-Mehrjardi, M.J. Tavassolifar, J. MP, H.R. Mirzaei, J. Hadjati, Restricting tumor lactic acid metabolism using dichloroacetate improves T cell functions, BMC Cancer 22 (1) (2022) 39.
[37] A. Sharma, J.J.W. Seow, C.A. Dutertre, R. Pai, C. Blériot, A. Mishra, R.M.M. Wong, G.S.N. Singh, S. Sudhagar, S. Khalilnezhad, et al., Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma, Cell 183 (2) (2020) e321.
[38] S. Wang, G. Liu, Y. Li, Y. Pan, Metabolic reprogramming induces macrophage polarization in the tumor microenvironment, Front. Immunol. 13 (2022) 840029.
[39] L. Yan, J. Lv, M. Xu, H. Jia, S. Li, High midkine expression correlates with poor prognosis and immune cell infiltration in hepatocellular carcinoma, Int. J. Gen. Med. Volume 18 (2025) 2567–2579.
[40] edn. Edited by E.-S. Oh, J.R. Couchman, Syndecan-2 biology and its role in colorectal carcinoma, in: I. Kovalszky, M. Franchi, LD. Alaniz (Eds.), The Extracellular Matrix and the Tumor Microenvironment, Springer International Publishing, Cham, 2022, pp. 75–92. edn. Edited by.
[41] F. Verona, S. Di Bella, R. Schirano, C. Manfredi, F. Angeloro, G. Bozzari, M. Todaro, G. Giannini, G. Stassi, V. Veschi, Cancer stem cells and tumor-associated macrophages as mates in tumor progression: mechanisms of crosstalk and advanced bioinformatic tools to dissect their phenotypes and interaction, Front. Immunol. 16 (2025) 1529847.
[42] T. Chen, Z. Huang, Y. Tian, B. Lin, R. He, H. Wang, P. Ouyang, H. Chen, L. Wu, Clinical significance and prognostic value of triosephosphate isomerase expression in gastric cancer, Med. (Baltim.) 96 (19) (2017) e6865.
[43] P. Liu, S.-J. Sun, Y.-J. Ai, X. Feng, Y.-M. Zheng, Y. Gao, J.-Y. Zhang, L. Zhang, Y.-P. Sun, Y. Xiong, et al., Elevated nuclear localization of glycolytic enzyme TPI1 promotes lung adenocarcinoma and enhances chemoresistance, Cell Death Dis. 13 (3) (2022) 205.