Multi-omics Atlas of Tumor Microenvironment Heterogeneity in Gastrointestinal Cancers and Novel Therapeutic Strategies

Dissertation

Klinik und Poliklinik für Allgemein-, Viszeral und Thoraxchirurgie

zur Erlangung des Grades eines Doktors der Medizin

an der

Medizinischen Fakultät der Universität Hamburg

vorgelegt von:

Heming Ge

aus

Jiangxi (China)

2025

Betreuer:in / Gutachter:in der Dissertation: Prof. Dr. Matthias Reeh

Gutachter:in der Dissertation: Prof. Dr. Stefan Bonn

Vorsitz der Prüfungskommission: Prof. Dr. Stefan Bonn

Mitglied der Prüfungskommission: PD Dr. Tarik Ghadban

Mitglied der Prüfungskommission: PD Dr. Marianne Sinn

Datum der mündlichen Prüfung: 10.02.2026

Contents

1. Presentation of the publications

According to data from the International Agency for Research on Cancer (IARC), approximately 20 million new cancer cases and 9.7 million cancer deaths occurred globally in 2022. It is estimated that approximately one in five people will develop cancer during their lifetime, with new cancer cases projected to reach 35 million by 2050 (Bray et al. 2024). Cancer has emerged as a significant public health and economic challenge this century. Among all types of cancer, malignant tumors of the digestive system (including colorectal, gastric, esophageal, pancreatic, liver, and gallbladder cancers) exhibit the highest incidence and mortality rates (accounting for 24.56% and 34.13%, respectively) (Bray et al. 2024). Colorectal cancer demonstrates the highest incidence rate among digestive system cancers, while pancreatic cancer has the highest mortality rate within this group (Bray et al. 2024). Despite ongoing advancements in treatment strategies such as surgery, chemotherapy, radiotherapy, immunotherapy, and targeted therapy, the overall five-year survival rate for gastrointestinal cancers remains suboptimal (Wang et al. 2024). Furthermore, due to their anatomical location often leading to subtle and nonspecific symptoms, the majority of patients are diagnosed at an advanced stage, contributing further to the high mortality rates.

The intricate tumor microenvironment (TME) heterogeneity inherent to gastrointestinal cancers is a key factor driving resistance to conventional therapies. For instance, in pancreatic cancer, the densely desmoplastic tumor-associated stroma not only impedes drug delivery but also fosters a hypoxic niche (Ho et al. 2020, Kwon et al. 2025). This hypoxic state significantly suppresses immune cell responses, drives tumor metabolic reprogramming, and mediates therapeutic resistance (Tao et al. 2021). Conversely, in colorectal cancer, the unique intestinal microbiota profoundly impacts tumorigenesis, progression, and the balance of immune responses by modulating specific immune cell functions, such as those of $CD4^+$ T cells and specific macrophage subsets (Chen et al. 2024). Therefore, dissection of TME heterogeneity and the development of novel, highly effective microenvironment-based therapeutic targets and predictive biomarkers remain a critical clinical need.

To address these challenges, my research integrates multi-omics technologies, including single-cell sequencing, spatial transcriptomics, and Mendelian randomization to perform an in-depth dissection of the cellular and molecular heterogeneity within the TME of gastrointestinal cancers, specifically focusing on colorectal cancer and pancreatic ductal adenocarcinoma (PDAC). This research aims to identify critical cell types (e.g., hypoxia-responsive macrophage subpopulations) and key regulatory genes. Furthermore, it seeks to construct TME feature-based predictive models (such as a hypoxia-responsive macrophage score and tumor thrombus-related gene signatures) with the goal of guiding individualized treatment strategies and optimizing chemotherapeutic response in patients. Concurrently, we validate the differential expression of pivotal genes in chemoresistant pancreatic cancer cell lines using RNA-Seq and Western blot analysis *in vitro*. This experimental validation provides crucial evidence for identifying novel therapeutic targets to overcome chemotherapy resistance and enhance the efficacy of immunotherapy.

1.1 Spatial transcriptomics deciphers the immunosuppressive microenvironment in colorectal cancer with tumour thrombus

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy globally and stands as the second leading cause of cancer-related deaths (Bray et al. 2024). Microscopic examinations often reveal clusters of tumor cells within the blood vessels, indicative of vascular tumour thrombus, which is recognized as an early stage to metastasis and is associated with a higher risk of postoperative metastasis and poorer prognoses. The presence of tumour thrombus categorizes CRC patients, especially at stage II, as high-risk, necessitating adjuvant chemotherapy, as guided by NCCN and ESMO (Argiles et al. 2020) clinical guidelines. Previous research reveal that tumour thrombus is associated with a higher survival hazard ratio than lymph node metastasis in CRC (Song et al. 2021). Moreover, 21.48% of CRC cases presented with tumour thrombus, and these patients suffered a more dire prognosis comparing to those without tumour thrombus in our data. Regretfully, little is known about the factors and mechanism promoting the formation of tumour thrombus, leading to a lack of targeted treatment plans for CRC cases with tumour thrombus. Previous research suggested

that the formation of tumour thrombus is related to the malignant characteristics of tumor cells in hepatoma (Liu et al. 2012), which cannot fully explain the directional invasion and migration of tumor cells to vascular areas. Shi et al. believed that not only cancer cells but also tumor-infiltrating cells may participate in the process of tumour thrombus (Shi et al. 2022). Those tumor-promoting cells are able to provide a protective microenvironment for the survival of tumor cells and susceptible to the invasion and migration of tumor cells into specific immunosuppressive areas (Li et al. 2021, Xiang et al. 2021), providing a favorable microenvironment for tumor cells to invade blood vessels and form tumour thrombus. Therefore, we sought to investigate the multicellular heterogeneity of tumor cell communities of tumour thrombus to provide valuable biological and clinical insights into CRC tumour thrombus. While the constraints in previous technology elucidating the cell-cell crosstalk in tissues impeded a thorough understanding of the immune microenvironment of tumour thrombus in CRC, recent advancements in spatial transcriptomics now make such studies feasible (de Vries et al. 2020).

This study employed spatial transcriptomics to analyze primary tumor samples from four colorectal cancer (CRC) patients with vascular tumour thrombus (TT), comparing them against existing datasets from four non-tumour thrombus (NTT) CRC samples (Park et al. 2024). Analysis revealed that while malignant epithelial cells from TT and NTT samples exhibited no significant differences in gene expression profiles, consensus molecular subtypes (CMS/iCMS), or stem cell/ epithelial-mesenchymal transition features, critical distinctions resided within the tumor microenvironment (TME). Spatial profiling demonstrated that TT regions were enriched in $CD8^+$ T cells and macrophages, with a concomitant reduction in $CD4^+$ T cells. Paradoxically, this immune cell enrichment did not confer a better prognosis; functional analysis revealed that these abundant immune cells exhibited a state of dysfunction. Further investigation delineated a pronounced immunosuppressive landscape within the TT microenvironment: characterized by enriched immunosuppressive cell populations - including regulatory T cells (Tregs), M2 macrophages, monocytic and polymorphonuclear myeloid-derived suppressor cells (M-MDSCs, PMN-MDSCs), and tolerogenic dendritic cells (tDCs) -

concomitant with reduced conventional dendritic cells (cDCs), diminished expression of antigen-presenting molecules, and significant upregulation of immune checkpoint molecules (e.g., PDCD1/PD-1, CD274/PD-L1) and immunosuppressive cytokines/genes. Cell-cell communication analysis identified frequent, direct cell-contact-dependent inhibitory signaling between $CD4^+$ T cells and $CD8^+$ T cells within TTs, mediated by ligand-receptor pairs such as CD80/86-CTLA4, LGALS9-CD45 and CD274-PDCD1, which suppressed $CD8^+$ T cell function - a core mechanism underlying immune cell dysfunction in TT. Based on differential gene expression analysis, a 145-gene TT signature was constructed. This signature robustly predicted poor prognosis (reduced survival in the high-score group) in independent cohorts like TCGA-CRC. The high-score group also exhibited elevated immune/stromal scores, increased infiltration of immunosuppressive cells (e.g., Tregs), reduced levels of antigen-presenting cells, and predicted higher resistance to key chemotherapeutic agents. Pan-cancer analysis further demonstrated that this TT signature and its associated immunosuppressive state (positive correlation with Tregs, negative correlation with activated $CD4^+$ memory T cells and activated DCs) were significantly correlated with adverse outcomes across multiple epithelial-derived solid tumors.

This study elucidates the heterogeneous cellular composition and distribution within CRC tumour thrombus microenvironment through spatial transcriptomics. These findings successfully unlocked that immune cells exhibited an impaired anti-tumor functionality through the upregulation of immune checkpoints, immunosuppressive genes and attenuated antigen presentation processes, even though they were abundant in thrombus microenvironment, emphasizing the predominance of an immunosuppressive microenvironment. Given the reliance of current immunotherapies on the activation and efficacy of pre-existing immune cells, our findings suggest that CRC patients with tumour thrombus may derive enhanced benefit from immunotherapeutical strategies.

1.2 Gut microbiota influences colorectal cancer through immune cell interactions: a Mendelian randomization study

The diversity of human gut microbiota constitutes a complex and unique ecosystem whose stability is critical for maintaining host health, while disruption of this balance may trigger multiple pathological conditions (Al-Hujaily et al. 2022). Colorectal cancer (CRC) development is significantly associated with gut dysbiosis, where microenvironmental homeostasis exerts protective effects through maintaining epithelial barrier integrity, modulating inflammatory responses, and regulating immune surveillance (Chen and Fu 2022, Chen et al. 2024). 16S rRNA gene sequencing analyses revealed reduced bacterial diversity in fecal microbiota of CRC patients, accompanied by significant enrichment of pathogenic bacteria such as *Catabacter*, *Mogibacterium*, and *Fusobacteria* (Du et al. 2020). Specific gut microbial (GM) communities, along with their virulence factors or microbially derived metabolites, can directly promote oncogenic transformation of epithelial cells or facilitate CRC progression through interactions with the host immune system (El Tekle et al. 2024, Hu et al. 2024). Mendelian Randomization (MR) serves as a robust causal inference framework that leverages common genetic variants as instrumental variables (IVs). Capitalizing on the random allocation of genotypes during conception, MR effectively mitigates confounding biases inherent in conventional epidemiological studies (Davies et al. 2018). This approach recapitulates randomization procedures in clinical trials, thereby strengthening causal inference (Smith and Ebrahim 2003), and characterizes lifelong effects of exposures—overcoming limitations in assessing long-term exposure periods faced by traditional observational and interventional studies (Cornish et al. 2019).

This study systematically analyzed the associations of 473 gut microbial species and 731 immune cell signatures with CRC using data derived from large-scale genome-wide association studies (GWAS). The results revealed significant causal associations between five specific gut microbes (*Alloprevotella*, *Holdemania*, *Megamonas*, *Psychroserpens*, *Succinivibrionaceae*) and nine key immune phenotypes (including $CD4^+$ T cells, CD3 on CD28- CD8br T cells, etc.) and CRC. Specifically, *Alloprevotella* was identified as a risk factor for CRC, whereas microbes such

as *Holdemania* exhibited protective effects. Mediation analysis further revealed that *Alloprevotella* mediates CRC progression through the suppression of $CD4^+$ T cell activity (mediation effect: 6.48%), while the protective effect of *Holdemania* operates via the CD3 on CD28- CD8br immune pathway (mediation effect: 9.29%). Reverse Mendelian randomization analyses ruled out reverse causality (i.e., the effect of CRC on microbial abundance or immune traits). Sensitivity analyses confirmed the absence of significant heterogeneity or horizontal pleiotropy.

This study comprehensively assessed the association between the GM, immune cells, and CRC. These findings indicate that specific gut microbial species can exert either positive or negative effects on colorectal cancer development and progression by modulating specific immune cell populations. This suggests a novel therapeutic approach for CRC: targeting specific gut microbes (such as *Alloprevotella* or *Holdemania*) to influence corresponding immune cells and thereby confer protection against colorectal cancer.

1.3 Development of a hypoxia-responsive macrophage prognostic model using single-cell and bulk RNA sequencing in pancreatic cancer

Pancreatic ductal adenocarcinoma (PDAC), the most prevalent histological subtype of pancreatic cancer (Vincent et al. 2011, Güngör et al. 2014), is globally acknowledged as one of the most lethal malignancies, characterized by a dismal five-year survival rate approximating 10% despite significant research advances (Siegel et al. 2024). Current therapeutic modalities are routinely employed; however, patient response rates remain suboptimal (Chen et al. 2022, Sonkin et al. 2024). Notably, while immune checkpoint inhibitors have significantly improved outcomes in diverse tumor types, their efficacy in enhancing PDAC prognosis remains limited (Yoon et al. 2023). Thus, there is an urgent need to identify novel molecular biomarkers and develop effective therapeutic targets for this disease. Hypoxia constitutes a fundamental feature of the solid tumor microenvironment, including PDAC, and contributes critically to multiple pathological processes driving tumor progression (Hielscher and Gerecht 2015). These include suppression of anti-tumor immune responses, metabolic reprogramming,

promotion of epithelial-mesenchymal transition, and the development of therapy resistance (Choudhry and Harris 2018, Qiao et al. 2023). Crucially, the presence of hypoxic regions within PDAC tumors is strongly associated with adverse patient outcomes and serves as an independent prognostic indicator (Qin et al. 2014). Furthermore, tumor-associated macrophages (TAMs) exhibit a pronounced accumulation specifically within these hypoxic niches of the tumor stroma (Bai et al. 2022).

This study integrates single-cell RNA sequencing (scRNA-seq) data and bulk RNA sequencing data from the TCGA-PAAD database. We initially identified a macrophage subcluster exhibiting significant hypoxia responsiveness ("macrophage cluster1"), demonstrating a distinctive gene expression pattern within the hypoxic tumor microenvironment. Subsequently, we identified 13 core hypoxia- and prognosis-associated genes (*LYZ*, *SCN1B*, *PLAU*, *INSIG2*, *DSC2*, *MICAL1*, *U2AF1*, *KRTCAP2*, *DDX60L*, *SATB1*, *SAMD9*, *LTC4S*, *IGLL5*) and constructed a novel hypoxia-derived prognostic model. This model demonstrated robust independent predictive capacity within the TCGA-PAAD cohort: patients categorized into the high-hypoxia-score group exhibited significantly shorter overall survival (Hazard Ratio > 1). The model's predictive accuracy, as quantified by area under the curve (AUC) values for 1-year, 2-year, and 3-year survival (0.774, 0.727, and 0.711, respectively), significantly surpassed that of traditional clinicopathological features. Further validation revealed that the high-hypoxia-score group demonstrated significantly reduced sensitivity to first-line chemotherapeutic agents such as gemcitabine and oxaliplatin (characterized by increased IC_{50} values) and exhibited concomitant deterioration of the immune microenvironment. This deterioration manifested as reduced infiltration of antitumor naïve B cells, enrichment of protumoral M0 macrophages, along with elevated tumor mutational burden (TMB) and increased frequencies of copy number variations (CNV). Of particular importance, pan-cancer analysis identified *KRTCAP2* as a key cross-cancer biomarker. High *KRTCAP2* expression was positively correlated with adverse prognosis, advanced tumor stage, and infiltration of immunosuppressive regulatory T cells (Tregs), highlighting its potential therapeutic target value. Cumulatively, this research not only provides a prognostication tool with clinical translation potential for PDAC, but also elucidates

mechanistic insights whereby the hypoxic TME drives therapeutic resistance through its capacity to reshape the immune landscape and promote genomic instability.

Through integrative analysis of single-cell and bulk transcriptomic data, our study established a thirteen-gene hypoxia-based prognostic model for pancreatic cancer. Demonstrating strong independent predictive value for outcomes, this signature further elucidates the tumor immune microenvironment and chemotherapeutic response. Strikingly, it also holds significant potential for identifying novel biomarkers and actionable targets.

Experimental work: RARRES1 drives chemoresistance via an immune-cold malignant epithelial niche in pancreatic ductal adenocarcinoma

Pancreatic ductal adenocarcinoma (PDAC) is one of the most aggressive solid malignancies. Its treatment efficacy is frequently limited by the common development of chemoresistance (Ding et al. 2025). The mechanisms driving this resistance are highly complex, involving the accumulation and combined effects of multiple gene mutations within tumor cells (Won et al. 2021). Together, these create a substantial obstacle to effective treatment. Specifically, activating mutations in the *KRAS* (G12D) gene are the most common genetic alteration in PDAC. They occur in over 90% of cases and are recognized as a key driver of the disease (Dilly et al. 2024). Despite KRAS's central role in cancer-related signaling pathways, developing treatments that directly target mutant KRAS has proven difficult. Clinical success with such approaches remains limited (Singhal et al. 2024). Therefore, further research is needed to study other co-mutated genes and their related signaling pathways in PDAC. This research aims to understand how these genes contribute to chemoresistance and identify new, targetable treatment options.

We analyzed single-cell RNA sequencing data from public databases (Werba et al. 2023), covering 17 primary PDAC samples and 10 liver metastasis samples. Eight patients had received neoadjuvant chemotherapy with the following responses: 2 Partial Response (PR), 2 Stable Disease (SD), 3 Progressive Disease (PD), and 1 Nonevaluable. Following initial

clustering, we identified 86,286 epithelial cells (Fig. 1A). Tissue distribution analysis showed a significantly higher proportion of epithelial cells in the PD group compared to the Naive, PR, and SD groups (Naive: 1.05; PR: 0.53; SD: 1.09; PD: 1.57) (Fig. 1B). Additionally, liver metastases contained a higher proportion of epithelial cells than primary pancreatic tissues (Pancreas: 0.88; Liver: 1.25) (Fig. 1C).

Subcluster analysis of epithelial cells identified 20 distinct subclusters (c01-c20) (Fig. 1D). To exclude non-malignant cells, we performed copy number variation (CNV) analysis, selecting c01-c14 subclusters with high CNV scores (Fig. 1E). Expression patterns of marker genes confirmed these subclusters as malignant epithelial cells. Using established PDAC subtype classifications (Moffitt et al. 2015), c01-c03 were classified as classical subtype, c06-c08 as basal-like subtype (associated with poor prognosis), while c04 and c05 showed mixed expression patterns of both subtypes. Subclusters c09-c14 lacked defined subtype markers (Fig. 1F). Importantly, c05 demonstrated significant enrichment in liver metastases compared to primary tumors (Fig. 1G), higher abundance in chemotherapy-exposed versus treatment-naive samples (Fig. 1H), and elevated proportions in PD versus Naive, PR, and SD groups (Fig. 1I). Patients with high c05 signature gene expression in TCGA-PAAD showed significantly shorter overall survival ($P = 0.037$), a finding consistently observed in independent PACA-CA ($P = 0.0069$) and PACA-AU ($P = 0.00035$) cohorts (Fig. 1J-L).

Cell-cell communication analysis indicated that the c05 subcluster had significantly weaker signaling compared to other malignant epithelial cells, particularly classical and basal-like subtypes (Fig. 2A). c05 received minimal signals through ligand-receptor interactions and showed very limited input from immune cells including mast cells, myeloid cells, T/NK/ILCs, and B/Plasma cells (Fig. 2B). Bulk RNA analysis using CIBERSORT (Newman et al. 2015) further supported this phenotype: samples with high c05 signature expression showed increased immunosuppressive regulatory T cells (Tregs) but lacked anti-tumor effector immune cell infiltration (Fig. 2C). Non-negative matrix factorization (NMF) analysis (Chen et al. 2024) revealed that the c05-enriched NMF5 program corresponded to an immune-cold phenotype

(no immune cells present), while immune cells were concentrated in the NMF1 program (Fig. 2D). Importantly, NMF5-enriched cells primarily came from PD samples, whereas NMF1-enriched cells mainly originated from PR and SD samples (Fig. 2E). These patterns - immune exclusion in the microenvironment and spatial distribution differences - suggest how c05 drives treatment resistance and poor outcomes.

Pseudotime analysis (Trapnell et al. 2014) showed c05 subcluster and PD-derived cells were significantly enriched at differentiation trajectory endpoints (Fig. 3A). This indicates they collectively represent a terminal, treatment-resistant state in pancreatic cancer progression.

We assessed the treatment-resistant features associated with the c05 subcluster using drug sensitivity data from the GDSC2 database. Compared to other malignant epithelial subclusters, samples with high c05 signature gene expression showed significantly reduced sensitivity to first-line pancreatic cancer chemotherapies, including Gemcitabine, 5-Fluorouracil, Oxaliplatin, Irinotecan, and Paclitaxel (Fig. 3B). These findings demonstrate that the c05 signature predicts multi-drug resistance in pancreatic cancer patients.

Analysis of single nucleotide variants (SNVs) revealed that while high and low c05-expression groups shared similar mutation frequencies in the top five mutated genes (*KRAS*, *TP53*, *CDKN2A*, *SMAD4*, *TTN*), the high-expression group showed significantly higher *KRAS* mutation rates (81% vs. 45%) (Fig. 3C) and elevated tumor mutational burden (TMB; P = 0.0223) (Fig. 3D).

Differential gene expression analysis across four comparison groups: (1) c05 vs. other malignant epithelial subclusters, (2) PD vs. Naive/PR/SD samples, (3) liver metastases vs. primary pancreatic tumors, and (4) chemotherapy-treated vs. treatment-naive samples. We identified the top 50 most differentially expressed genes in each group. Intersection of these gene sets revealed four consistently upregulated genes: *SNCG*, *CST6*, *CAPS*, and *RARRES1* (Fig. 3E). To test their link to drug resistance, we generated gemcitabine-resistant L3.6pl cells (GEM-

R) using gradual dose escalation (0.1-2 µmol/L over 6 months). Both RNA sequencing and Western blotting confirmed significantly higher RARRES1 RNA- and protein expression in GEM-R cells than in the parental wild type chemosensitive lines (Fig. 3F, G), aligning with single-cell data. Importantly, our prior study employing microfluidic circulating tumor cell (CTC) analysis in resected PDAC patients demonstrated that RARRES1 protein expression on CTCs strongly correlated with early tumor relapse, further substantiating the adverse prognostic role of RARRES1-overexpressing tumor cells (Nitschke et al. 2022). These multi-method validation results establish RARRES1 as a key regulator of drug resistance development in pancreatic cancer.

Pan-cancer survival analysis identified high RARRES1 expression as a biomarker of poor prognosis across multiple cancer types. Patients with elevated RARRES1 levels had significantly worse overall survival in adrenocortical carcinoma, esophageal carcinoma, kidney renal clear cell carcinoma, brain lower grade glioma, lung adenocarcinoma, ovarian serous cystadenocarcinoma, thymoma, and uterine corpus endometrial carcinoma (Fig. 3H).

This study defines key mechanisms of chemotherapy resistance in PDAC through integrated single-cell RNA sequencing and functional validation. We identified a distinct malignant epithelial subcluster (c05) specifically enriched in liver metastases and progressive disease samples that correlates with poor patient outcomes. The c05 subcluster demonstrates: 1) an immune-cold phenotype with disrupted cell signaling and deficient anti-tumor immune infiltration, 2) a treatment-resistant state featuring frequent *KRAS* mutations and multi-drug resistance, and 3) RARRES1 as a central resistance driver validated across experimental models and clinical cohorts where its expression associates with treatment failure and reduced survival. These findings establish c05 as a cellular resistance niche and identify RARRES1 targeting as a promising strategy to overcome treatment resistance in PDAC.

2. Articles

2.1 Spatial transcriptomics deciphers the immunosuppressive microenvironment in colorectal cancer with tumour thrombus

Ge H, Pei Z, Zhou Z, Pei Q, Güngör C, Zheng L, Liu W, Li F, Zhou J, Xiang Y, Pei H, Li Y, Liu W (2024). Clin Transl Med. 14(12):e70112.

Received: 29 August 2024 | Revised: 9 November 2024 | Accepted: 15 November 2024

DOI: 10.1002/ctm2.70112

CLINICAL AND TRANSLATIONAL MEDICINE Open Access WILEY

LETTER TO THE JOURNAL

Spatial transcriptomics deciphers the immunosuppressive microenvironment in colorectal cancer with tumour thrombus

Dear Editor,

Vascular tumour thrombus (TT), a prominent indicator of early metastasis associated with poor prognosis, classifies colorectal cancer (CRC) patients as high-risk according to the National Comprehensive Cancer Network and the European Society for Medical Oncology[1] guidelines. Regretfully, the mechanisms by which TT occurs and develops remain unclear. In this study, we identified a distinctive transcriptomic profile of immune cells with a compromised phenotype in the TT microenvironment, where $CD4^+$ T cells as negative regulators directly signal to $CD8^+$ T cells.

Note that, 6150 CRC surgery patients at Xiangya Hospital from 2014 to 2019 were retrospectively analyzed, of which 1321 cases behaved with pathological vascular TT (Figure 1A,B and Table S1). Survival data available for 5109 patients demonstrated a significant separation in overall survival (OS) between those with and without TT (Figure 1C and Figure S1).

We performed spatial transcriptomics (ST) on four primary CRC samples with TT (TT 1–4; Figure S2A–D and Table S2) and compared them with four non-TT (NTT) samples from dataset GSE226997[2] (NTT 1–4; Figure S2E–H). Malignant epithelial cells were identified using CNA[3] and deconvolution methods (Figure 1D–F and Figure S3A,B), while non-malignant spots were classified into nine distinct clusters (Figure 1G,H and Figure S3C–K).[4]

The gene expression patterns of malignant epithelial cells were conserved in thrombus status (Figure S4A). Pathway enrichment related to angiogenesis and tumour growth was observed in malignant epithelial cells in TTs (Figure S4B,C and Tables S3 and S4). However, cancer-promoting transcription factors revealed comparable expression activated in both TTs and NTTs (Figure S4D and Table S5). When categorising malignant epithelial cells into consensus molecular subtypes CMS1-4 and iCMS2/3, we observed no significant differences in proportions or spatial distributions between TTs and NTTs (Figures S4E–J and S5A–E), nor in stem cell and epithelial-mesenchymal transition scores (Figure S5F–I), indicating that the prognostic differences were predominantly driven by the microenvironment.

Spatial profiling showed an increase of $CD8^+$ T cells and macrophages and a reduction of $CD4^+$ T cells in TTs (Figure 2A–F and Figure S6A–F), confirmed by immunofluorescence (Figure 2G–J), with a heightened co-localization of CD4+ and CD8+ T cells in TTs (Figure S7A–C).

Contrary to expectations, the increased immune cells in TTs did not contribute to a better prognosis. We performed a functional analysis of immune cells to explore this paradox,[5] indicating these immune cells with a state of dysfunction mostly within TTs albeit their abundance (Figure 2K).

Further analysis revealed a higher prevalence of immunosuppressive cells, including regulatory T cells (Tregs), M2 macrophages, monocytic myeloid-derived suppressor cells (M-MDSCs), polymorphonuclear MDSCs and tolerogenic dendritic cells in TTs (Figure 3A–E), suggesting a microenvironment dominated by immunosuppression. Immunostaining confirmed these findings (Figures 2I,J and 3G,H). Given that Tregs can inhibit antigen-presenting cells, we examined the spatial distribution of conventional dendritic cells, which was reduced in TTs (Figure 3F), confirmed by immunofluorescence (Figure 3I,J). Additionally, classical antigen-presenting molecules also exhibited lower expression levels in TTs (Figure S9). Chemokines, such as CCL5 and CXCL12, were notably elevated in TTs, which may partly explain the recruitment of both immune and immunosuppressive cells (Figures S8A,B and S9). Immune checkpoint genes like PDCD1 and CD274 were upregulated in TTs (Figures S8C,D and S9), validated by qPCR likewise (Figure 3K).

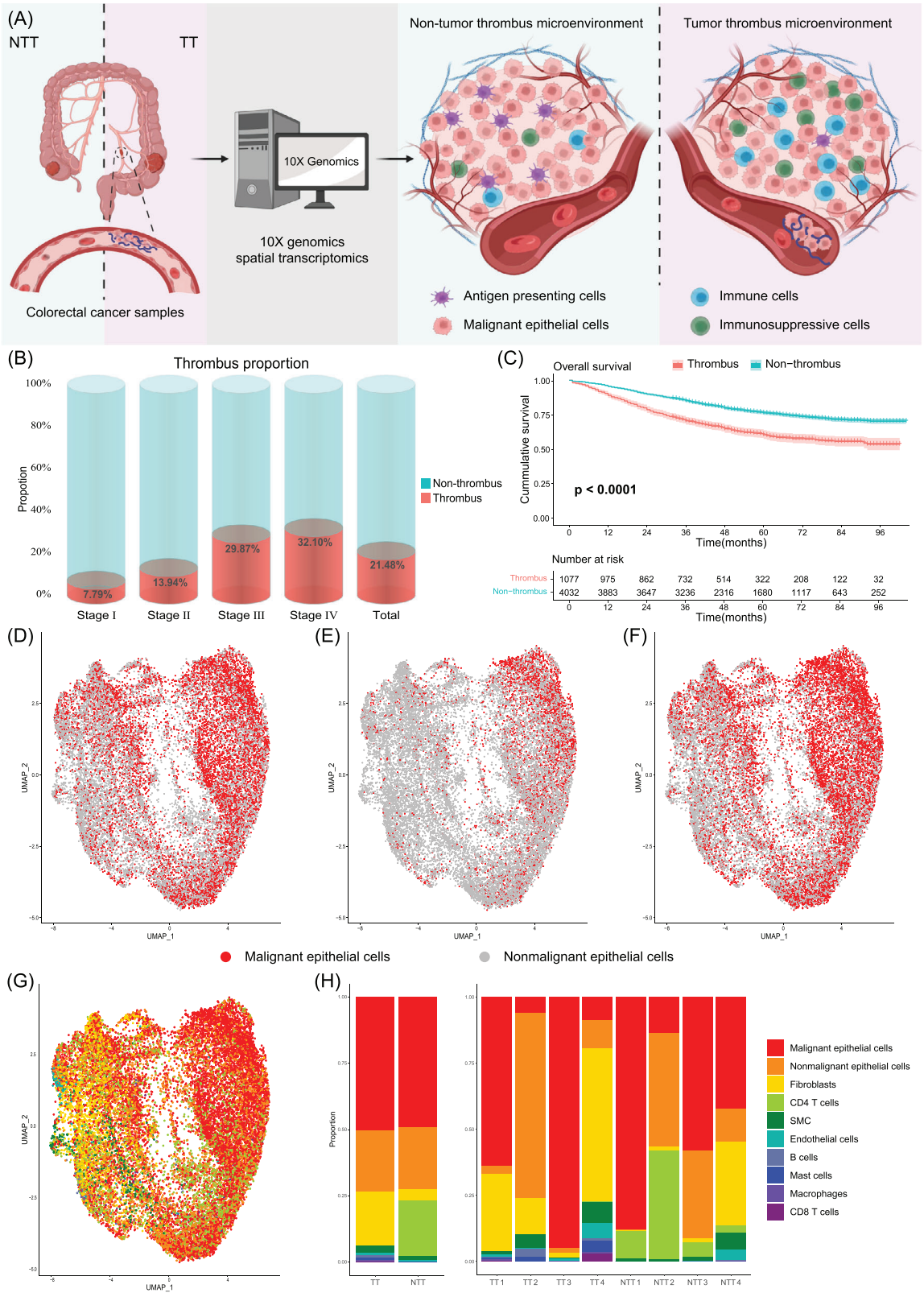

FIGURE 1 Overview of study design and spatial transcriptomics (ST) spots annotation. (A) Schematic representation of this study design. Tumour thrombus and non-tumour thrombus tissues from colorectal cancer (CRC) patients were used for spatial transcriptomics RNA sequencing. (B) The proportion of patients with tumour thrombus across different stages of CRC. (C) Survival curve for CRC patients with and without tumour thrombus. (D–F) UMAP plots of malignant spots identified by CNA analysis (D), deconvolution analysis (E), and the combination of CNA analysis and deconvolution results (F). (G) ST spots annotation based on CNA analysis and deconvolution in carcinoma regions. (H) Proportion of different cell type spots within the carcinoma regions of TT and NTT samples. TT, tumour thrombus; NTT, non-tumour thrombus; SMC, smooth muscle cells.

CLINICAL AND TRANSLATIONAL MEDICINE

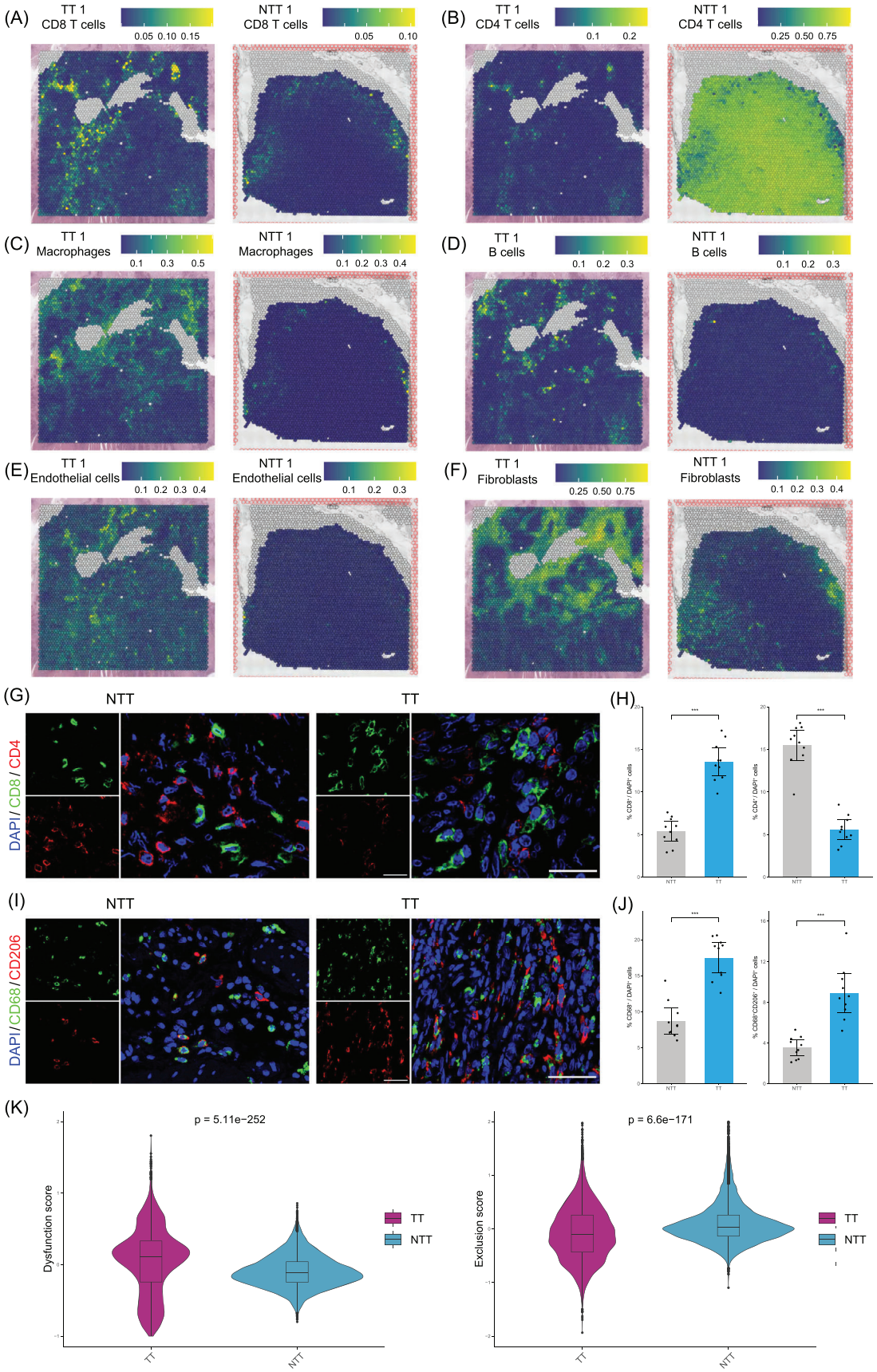

FIGURE 2 Spatial distribution of various cell types in TT and NTT samples. Spatial plots of $CD8^+$ T cells (A), $CD4^+$ T cells (B), macrophages (C), B cells (D), endothelial cells (E) and fibroblasts (F) in TT1 and NTT1 samples. (G) Immunofluorescence for $CD8^+$ and $CD4^+$ T cells on TT and NTT samples. Scale bar 25 μm. (H) Percentage of $CD8^+$ and $CD4^+$ T cells among total cells between TT and NTT groups. (I)

Immunosuppressive genes and cytokines were also elevated in TTs (Figures S8E,F and S9).

Intercellular communication[6] showed more signalling events in NTTs, whereas TTs exhibited frequent interactions between $CD8^+$ and $CD4^+$ T cells (Figure S10A–D and Table S6). Receptor-ligand pairs between $CD8^+$ and $CD4^+$ T cells, including CD80 - CTLA4, CD86 - CTLA4, LGALS9 - CD45 and CD274 - PDCD1, were identified as inhibitors of $CD8^+$ T cell activation (Figure S11). Interestingly, these interactions were cell-contact-dependent rather than mediated by secreted factors, aligning with the spatial proximity between $CD8^+$ and $CD4^+$ T cells. Enrichment analysis showed a significant inhibitory pathway in $CD8^+$ T cells within the TTs, whereas an obvious activation pathway was observed in NTTs (Figure S8G). This $CD4^+$ T cell-mediated inhibition was confirmed in thrombus samples from hepatocellular carcinoma, which demonstrated higher proportions of $CD8^+$ T cells with suppressed activity due to $CD4^+$ T cell interactions (Figure S12).[7] Collectively, our findings suggested that $CD4^+$ T cells exert an inhibitory effect on $CD8^+$ T cells within TTs, resulting in immune cell dysfunction.

Differential gene expression analysis between TTs and NTTs identified 145 TT-specific genes associated with poor prognosis (Table S7 and S8). TCGA-CRC patients were stratified into high- and low-score groups based on this gene signature, with the high-score group exhibiting a worse prognosis (Figure 4A and Figure S13A,B). Two distinct datasets validated these findings (Figure S13C,D).[8,9] Notably, the high-score group showed elevated immune scores (Figure 4B), accompanied by increased immunosuppressive Tregs and decreased activated/resting DCs and CD4+ T cells (Figure 4C and Figure S13E).[10] These findings were consistent with our results from ST samples, substantiating the reduction of $CD4^+$ T cells and antigen-presenting cells and the simultaneous augment of Tregs within TTs. Additionally, TT score correlated positively with higher half inhibitory concentrations of key chemotherapeutic agents (Figure 4D), reinforcing the association with poorer clinical outcomes in patients with TT.

Pan-cancer analysis exhibited a significant correlation between the aforementioned gene signature and poor prognosis across several epithelial-origin solid tumours (Figure 4E). Immunosuppressive Tregs showed a positive correlation with the TT gene signature, whereas both activated/resting CD4 memory T cells and activated DCs unveiled a negative correlation (Figure 4F). These findings aligned with findings from ST samples, indicating that an enhanced immunosuppressive state is a key feature of the TT microenvironment across diverse cancer types.

In conclusion, our study highlights that, despite their abundance in the TT microenvironment, immune cells demonstrate impaired anti-tumour functionality due to upregulated immune checkpoints, immunosuppressive genes, and attenuated antigen presentation, which emphasizes the predominance of an immunosuppressive microenvironment. Given the reliance of current immunotherapies on the activation and efficacy of pre-existing immune cells, our findings suggest that CRC patients with TT may benefit from enhanced immunotherapeutic strategies targeting immune cell reactivation.

AUTHOR CONTRIBUTIONS

Heming Ge, Yuqiang Li and Wenxue Liu conceived and supervised the study. Zhongyi Zhou, Qian Pei, Fengyuan Li, Jingxuan Zhou and Haiping Pei contributed to the clinical data collection and patients' follow-up. Heming Ge, Zhengda Pei, Zhongyi Zhou, Qian Pei, Cenap Güngör and Yao Xiang contributed to the data analysis. Heming Ge, Zhengda Pei, Linyi Zheng and Wei Liu prepared figures and tables. Heming Ge, Zhongyi Zhou and Qian Pei wrote the first version of the manuscript. Cenap Güngör, Yuqiang Li and Wenxue Liu reviewed and edited the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank all patients involved in this study. We thank the assistance of Hanhui Hong and Bihui Zhang from Accuramed (Guangzhou, China) Biotech Co. Ltd. in spatial-transcriptomic analysis. We thank BioRender tools (https://www.biorender.com/) for their assistance in the creation of the flowchart. We thank genescloud tools for providing a free online platform for data analysis (https://www.genescloud.cn). Heming Ge gratefully acknowledges the financial support from the China Scholarship Council program (Project ID: 202206370048).

Immunofluorescence for macrophages ($CD68^+$) and M2 macrophages ($CD68^+CD206^+$) on TT and NTT samples. Scale bar 50 μm. (J) Percentage of macrophages and M2 macrophages among total cells between TT and NTT groups. (K) Differences in the TIDE dysfunction and exclusion scores were observed between TT and NTT samples. In TT samples, the lower exclusion score indicated that immune cells successfully infiltrated the tumour. However, the higher dysfunction score in TT samples suggested that, despite this infiltration, the immune cells had their functions suppressed, preventing them from effectively killing tumour cells. TT, tumour thrombus; NTT, non-tumour thrombus; SMC, smooth muscle cells.

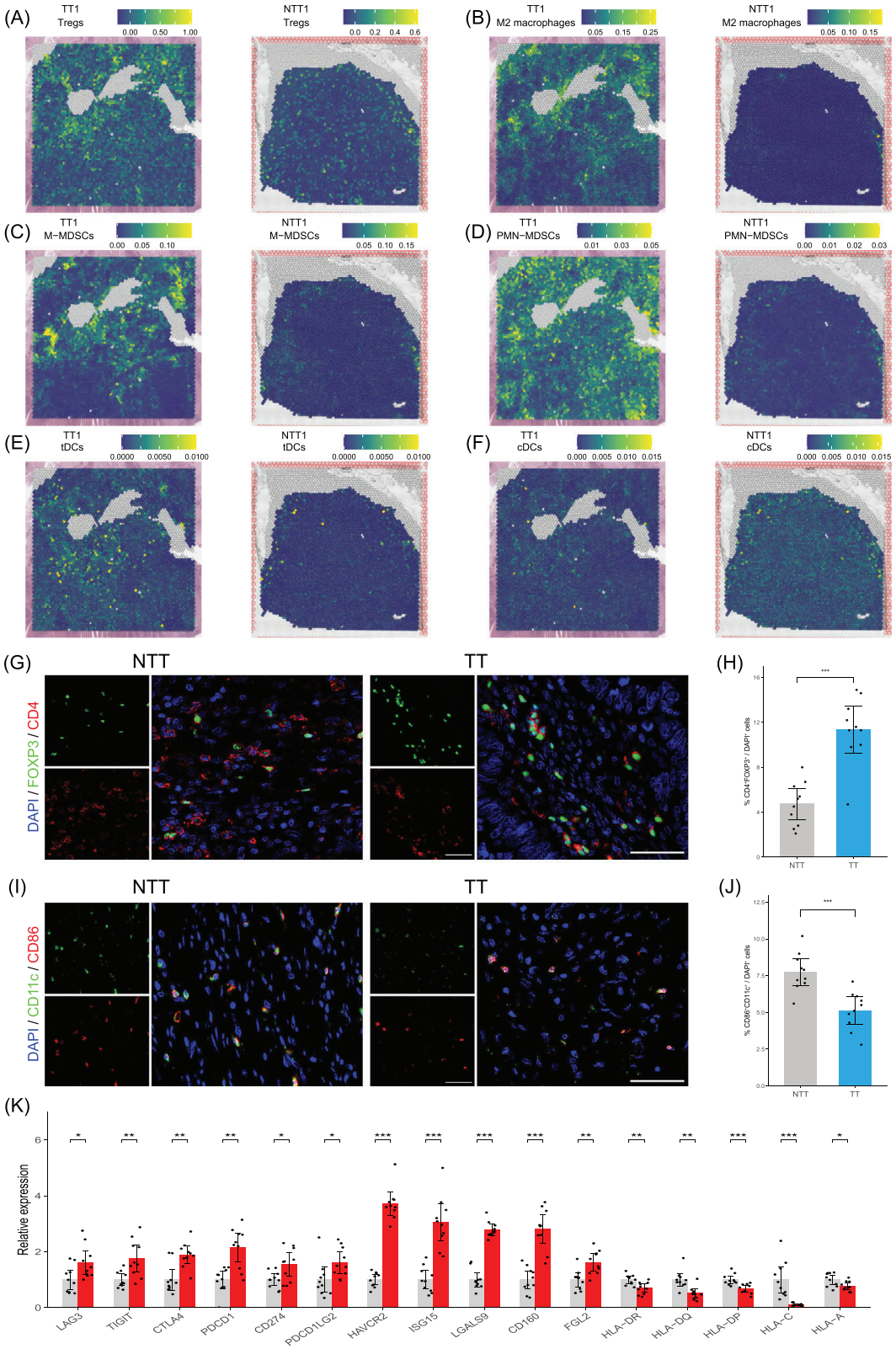

FIGURE 3 Immunosuppressive landscape of colorectal cancer (CRC) tumour thrombus. Spatial plots of regulatory T cells (Tregs) (A), M2 macrophages (B), monocytic myeloid-derived suppressor cells (M-MDSCs) (C), polymorphonuclear MDSCs (PMN-MDSCs) (D), tolerogenic dendritic cells (tDCs) (E) and conventional dendritic cells (cDCs) (F) in TT1 and NTT1 samples. (G) Immunofluorescence for Tregs ($CD4^+FOXP3^+$) on TT and NTT samples. Scale bar 50 μm. (H) Percentage of Tregs among total cells between TT and NTT groups. (I) Immunofluorescence for cDCs ($CD86^+CD11c^+$) on TT and NTT samples. Scale bar 50 μm. (J) Percentage of cDCs among total cells between TT and NTT groups. (K) qPCR analysis of key immune checkpoint genes, immunosuppressive genes, and antigen-presenting molecules in TT and NTT samples. TT, tumour thrombus; NTT, non-tumour thrombus.

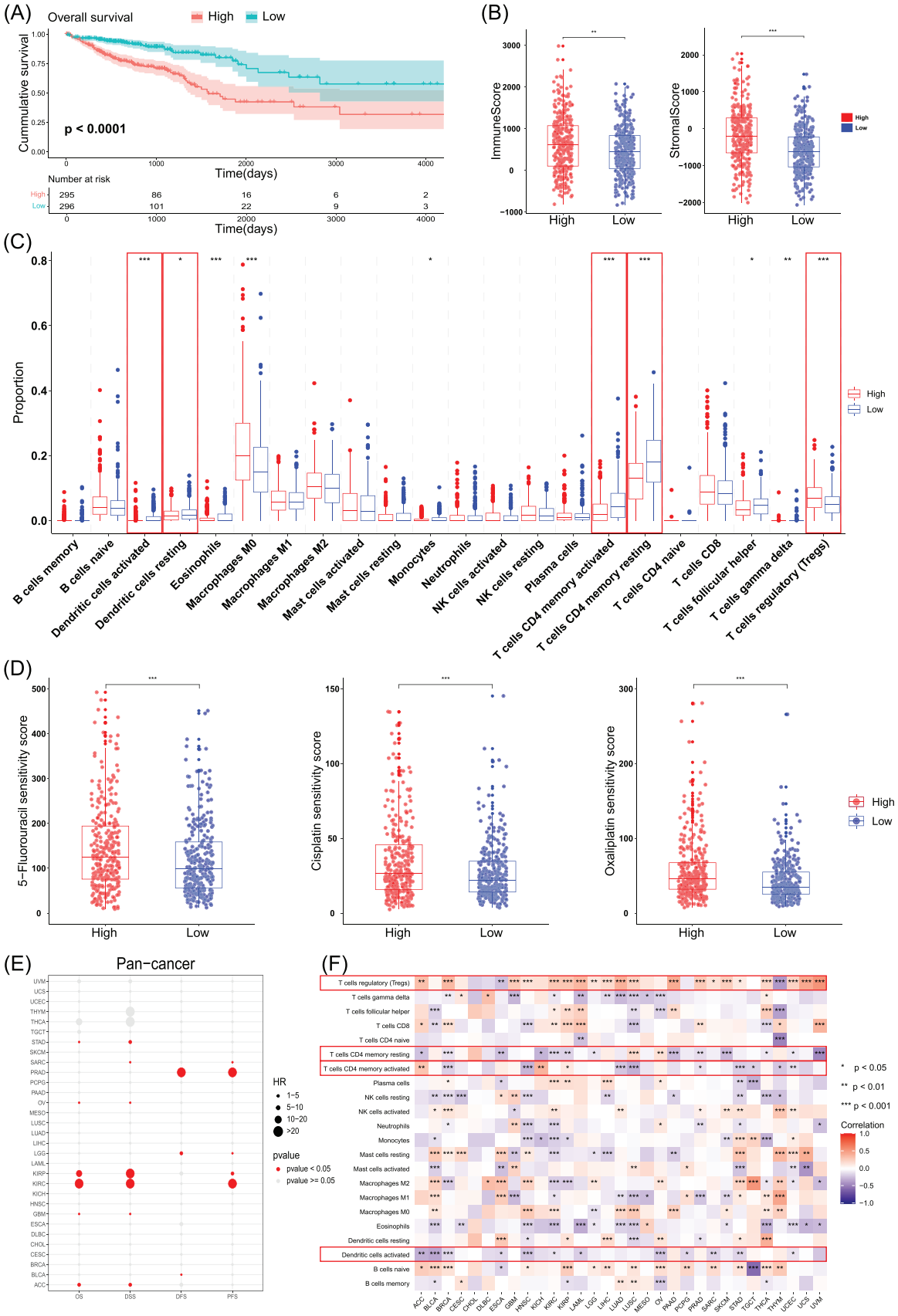

FIGURE 4 Construction of the tumour thrombus gene signature and pan-cancer analysis. (A) Survival curve of high and low tumour thrombus score groups in TCGA-CRC patients. (B) Comparison of immune and stromal scores between high and low tumour thrombus score groups. The high-score group exhibited significantly higher proportions of immune and stromal cells compared to the low-score group. (C)

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

FUNDING INFORMATION

This study was supported by the National Natural Science Foundation of China (82273077), the Natural Science Foundation of Hunan Province (2022JJ40799) and the Medical Research Project of Longhua District Medical Association (2023LHMA22).

DATA AVAILABILITY STATEMENT

The raw data of spatial transcriptomics RNA sequencing of four TT samples were uploaded to the Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa-human) of the National Genomics Data Center with accession number HRA010942. The raw data of spatial transcriptomics RNA sequencing of four NTT samples had been previously deposited in Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) repository with accession number GSE226997.[2] The scRNA-seq dataset reused in this study are available in the GEO database under accession number GSE132465.[4] The normalized gene expression data of colon adenocarcinoma and rectum adenocarcinoma were obtained from TCGA data portal (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). The bulk RNA-seq datasets of CRC patients were downloaded from GEO, including GSE39582[8] and GSE38832.[9] The bulk RNA-seq datasets for pan-cancer analysis were downloaded from the TCGA data portal.

ETHICS STATEMENT

This study was approved by the Ethics Committee of the Xiangya Hospital of Central South University (No. 2024020206).

PATIENT CONSENT STATEMENT

All human tissue samples were obtained from patients with written informed consent.

Heming Ge[1,2,3]
Zhengda Pei[4]
Zhongyi Zhou[1,2]
Qian Pei[1,2]
Cenap Güngör[3]
Linyi Zheng[1,2]
Wei Liu[1,2]
Fengyuan Li[1,2]
Jingxuan Zhou[5]
Yao Xiang[6]
Haiping Pei[1,2]
Yuqiang Li[1,2]
Wenxue Liu[2,7]

[1]*Department of General Surgery, Xiangya Hospital, Central South University, Changsha, China*
[2]*National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China*
[3]*Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[4]*Graduate Collaborative Training Base of Hunan Cancer Hospital, Hengyang Medical School, University of South China, Hengyang, China*
[5]*Department of Endocrinology, Endocrinology Research Center, Xiangya Hospital, Central South University, Changsha, China*
[6]*Department of Pathology, People's Hospital of Longhua District, Shenzhen, China*
[7]*Department of Geriatric Medicine, Xiangya Hospital, Central South University, Changsha, China*

Correspondence
Yuqiang Li, Department of General Surgery, Xiangya Hospital, Central South University, Changsha, China.
Email: liyuqiang22@csu.edu.cn

The proportions of 22 immune-related cells between the tumour thrombus score groups by the CIBERSORT method. (D) Correlation analysis between tumour thrombus scores and chemotherapeutic drug resistance. (E) Pan-cancer outcomes (OS, DSS, DFS and PFS) for tumour thrombus scores derived from 31 common cancer types from TCGA. (F) The relationship between tumour thrombus scores and immune-related cell infiltration across 31 common cancer types from TCGA. OS, overall survival; DSS, disease specific survival; DFS, disease free survival; PFS, progression free survival; ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; DLBC, diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma. *$p < .05$, **$p < .01$, ***$p < .001$.

Wenxue Liu, Department of Geriatric Medicine, Xiangya Hospital, Central South University, Changsha, China.
Email: liuwenxue@csu.edu.cn

Heming Ge, Zhengda Pei, Zhongyi Zhou and Qian Pei contributed equally to this work.

REFERENCES

1. Argiles G, Tabernero J, Labianca R, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(10):1291-1305. doi:10.1016/j.annonc.2020.06.022
2. Park SS, Lee YK, Choi YW, et al. Cellular senescence is associated with the spatial evolution toward a higher metastatic phenotype in colorectal cancer. *Cell Rep*. 2024;43(3):113912. doi:10.1016/j.celrep.2024.113912
3. Patel AP, Tirosh I, Trombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014;344(6190):1396-1401. doi:10.1126/science.1254257
4. Lee HO, Hong Y, Etlioglu HE, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet*. 2020;52(6):594-603. doi:10.1038/s41588-020-0636-z
5. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2018;24(10):1550-1558. doi:10.1038/s41591-018-0136-1
6. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;12(1):1088. doi:10.1038/s41467-021-21246-9
7. Li K, Zhang R, Wen F, et al. Single-cell dissection of the multicellular ecosystem and molecular features underlying microvascular invasion in HCC. *Hepatology*. 2024;79(6):1293-1309. doi:10.1097/HEP.0000000000000673
8. Marisa L, de Reynies A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*. 2013;10(5):e1001453. doi:10.1371/journal.pmed.1001453
9. Tripathi MK, Deane NG, Zhu J, et al. Nuclear factor of activated T-cell activity is associated with metastatic capacity in colon cancer. *Cancer Res*. 2014;74(23):6947-6957. doi:10.1158/0008-5472.CAN-14-1592
10. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457. doi:10.1038/nmeth.3337

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

2.2 Gut microbiota influences colorectal cancer through immune cell interactions: a Mendelian randomization study

Zheng L, Li Y, Güngör C, **Ge H** (2025). Discov Oncol. 16(1):747.

Discover Oncology

Analysis

Gut microbiota influences colorectal cancer through immune cell interactions: a Mendelian randomization study

Linyi Zheng[1] · Yuqiang Li[1] · Cenap Güngör[2] · Heming Ge[1,2]

Received: 22 October 2024 / Accepted: 24 April 2025
Published online: 13 May 2025

Abstract

Background Colorectal cancer (CRC) is the most prevalent malignant tumor of the digestive system globally, posing a significant threat to human health and quality of life. Recent studies have established associations between gut microbiota and immune cells with CRC; however, the mechanisms by which gut microbiota influence the development and progression of CRC through immune mediators remain poorly understood.

Methods We conducted a two-sample, bidirectional Mendelian randomization analysis. We utilized 731 immune cell types and 473 gut microbial species along with colorectal cancer statistics from published summary statistics from genome-wide association studies (GWAS).The analysis employed several methodologies, including inverse variance-weighted (IVW) analysis, MR-Egger regression, the weighted median method, and both weighted and simple model approaches.Sensitivity analyses were performed to confirm the reliability of the Mendelian randomization results, and reverse Mendelian randomization was used to assess the overall impact of CRC on gut microbiota and immune cells.

Results Our findings suggest a causal relationship involving nine immunophenotypes and five specific gut microbial taxa with CRC. Notably, the gut microbes *Alloprevotella* and *Holdemania,* along with immune cell types CD3 on CD28-CD8br and CD4 + T cells, demonstrated significant causal associations with CRC. Mediation analysis revealed that the association between *Alloprevotella* and CRC was mediated by CD4 + T cells, with a mediation effect of 6.48%. Additionally, *Holdemania* was found to mediate its association with CRC through CD3 on CD28- CD8br, exhibiting a mediation effect of 9.29%. Reverse Mendelian randomization did not indicate any causal effect of CRC on specific immune cells or gut microbiota. Two-sided sensitivity analyses revealed no evidence of heterogeneity or horizontal pleiotropy in our findings.

Conclusions This comprehensive Mendelian randomization study enhances our understanding of the mechanisms by which gut microbiota affects CRC through immune cell interactions. Further investigations are warranted to unravel the underlying mechanisms linking gut microbiota, immune cells, and colorectal cancer.

Keywords Colorectal cancer · Gut microbiota · Immune cells · Mendelian randomization

Abbreviations

GWAS	Genome-wide association analysis
GM	Gut microbiota
CRC	Colorectal cancer

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02486-3.

✉ Heming Ge, heming9502@163.com | [1]Department of Gastrointestinal Surgery, Xiangya Hospital, Central South University, Changsha 410013, China. [2]Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

AC	Absolute cell
MFI	Median fluorescence intensity
MP	Morphological parameters
RC	Relative cell
GTDB	Genome Taxonomy Database
IVW	Inverse variance weighted
MR	Mendelian randomization
IVs	Instrumental variables
SNPs	Single-nucleotide polymorphisms
LD	Linkage disequilibrium
MAF	Minimal allele frequency
UVMR	Univariate Mendelian randomisation
SCFA	Short-chain fatty acid
BCR	B-cell antigen receptor
MAPK	Mitogen-activated protein kinase
PRR	Pattern recognition receptor

1 Introduction

Colorectal cancer is a prevalent malignancy, ranking as the third most common cancer and the second leading cause of cancer-related mortality globally [1]. According to the most recent data on cancer burden from the World Health Organization's International Agency for Research on Cancer, it was estimated that in 2020, there were more than 1.9 million new cases of colorectal cancer and approximately 935,000 associated deaths. This accounts for approximately one in ten of all cancer cases and deaths [2]. Given the rising incidence of colorectal cancer, there is an urgent need for more sensitive biomarkers to enhance early detection and improve prognostic assessment of the disease.

Recent research has highlighted the extensive diversity of microorganisms inhabiting the human gut, forming a complex and unique microbial ecosystem [3]. The stability of the gut microbiota (GM) is crucial for maintaining human health, and disruptions to this balance can lead to a range of diseases. One risk factor for colorectal cancer (CRC) is intestinal dysbiosis, and the stability of the intestinal microenvironment is essential for preventing CRC development by maintaining intestinal barrier function, mediating intestinal inflammation, and regulating immune responses [4, 5] 0.16S rRNA gene sequencing has revealed decreased bacterial diversity in the fecal microbiota of CRC patients, accompanied by increased abundances of pathogenic bacteria such as *Catabacter, Mogibacterium, and Fusobacteria* [6]. Increasing evidence suggests that specific GM are directly associated with CRC development and progression. Particular microbial species and their virulence factors or associated small molecules can contribute to CRC by directly affecting the neoplastic transformation of epithelial cells or by interacting with the host immune system [7, 8].Recent studies have shown that *Lactococcus lactis HkyuLL 10* can inhibit colorectal tumorigenesis through the production of α-mannosidase, while *Fusobacterium nucleatum* can inhibit pyroptosis, leading to chemoresistance in CRC [9, 10].These results suggest that GM plays an important role in the development and progression of CRC.

The colorectum, as a digestive organ, also functions as a critical immune organ, participating in both innate and adaptive immune responses [11]. The immune system plays a central role in the pathogenesis and progression of CRC. Emerging evidence indicates that CRC is influenced by a confluence of factors, including interactions between innate and adaptive immune cells, cytokine signaling pathways, and GMs [12]. Histopathological analysis revealed that CRC is characterized by a significant infiltration of various immune cell types [13]. Immune cell populations within CRC encompass both innate and adaptive immune components. Specifically, the innate immune response includes macrophages, neutrophils, mast cells, and natural killer cells. In contrast, the adaptive immune response is represented by T lymphocytes and B lymphocytes. Innate immunity serves as the body's initial defense mechanism, or "gut response,"against cancer and operates independently of specific antigen recognition [14–16].

In addition to the roles of T and B lymphocytes, innate immune cells contribute to the inflammatory milieu that can either promote or inhibit tumor growth [17]. During CRC progression, adaptive immune cells are recruited and exhibit both tumor-promoting and anti-tumor effects. T lymphocytes are implicated in the processes of inflammation, tumor development, and progression, as well as in the generation of anti-cancer immunity. B lymphocytes also play a

complementary role in the host's response to tumors, highlighting the complex interplay between different immune system components in the context of CRC [18, 19].

Mendelian randomization (MR) is a robust analytical framework that employs common genetic variants as instrumental variables (IVs) to investigate causal relationships between exposures and outcomes. This method capitalizes on the fact that genotypes are randomly allocated at conception, thereby minimizing the confounding biases inherent in traditional epidemiological studies [20]. By leveraging the random assignment of genetic variants, MR studies approximate the randomization process found in clinical trials, thereby enhancing their capacity to infer causality [21]. Furthermore, MR studies capture the lifelong effects of exposure, providing a temporal advantage over conventional epidemiological studies and clinical trials that may be limited in their ability to account for long-term exposure periods [22].

Although extensive research has elucidated the roles of GM and immune cells in CRC development, there are still relatively few relevant studies on whether GM modulates the relationship between immunity and CRC.In this study, we employed MR analysis to systematically explore the causal effects of GM and 731 immune signatures on CRC (Fig. 1). This investigation utilized genome-wide association study (GWAS) data sourced from public databases to inform our analysis. Subsequently, we assessed the impact of GM on immune cell signatures and evaluated whether GM could influence CRC progression by modulating the immune system. This comprehensive approach aimed to elucidate the potential interactions between GM and immune responses in the context of CRC.

2 Materials and methods

2.1 Design

This study consisted of two steps of analysis. First, the association between 473 GMs and CRC was assessed using a two-sample MR approach, using SNPs as a instrumental variable (IV) for each factor. In the second step, mediators were selected from the 731 immune cell profiles to explore the role of immune cells in GM and CRC. The mediation effects of these immune traits were calculated using a two-step MR-mediated approach. MR analysis should be based on the following three assumptions: (1) Genetic IVs must be valid instruments, meaning that they are associated with exposure. (2) Genetic IVs must not be associated with known risk factors for the outcome of interest (i.e., confounding variables). (3) SNPs as instrumental variables are not directly associated with outcomes but are only associated with outcomes through exposure (Fig. 1) [20, 23, 24].

2.2 GWAS data sources for GM and immune cells

Summary statistics for immune signatures, derived from genome-wide association studies (GWAS), are publicly accessible in the GWAS Catalog, with accession numbers ranging from GCST0001391 to GCST0002121[25].This comprehensive dataset encompasses a total of 731 immunophenotypes, classified into distinct categories: absolute cell (AC) counts (n = 118), median fluorescence intensity (MFI) reflecting surface antigen levels (n = 389), morphological parameters (MP) (n = 32), and relative cell (RC) counts (n = 192). Specifically, the MFI, AC, and RC profiles include various immune cell types such as

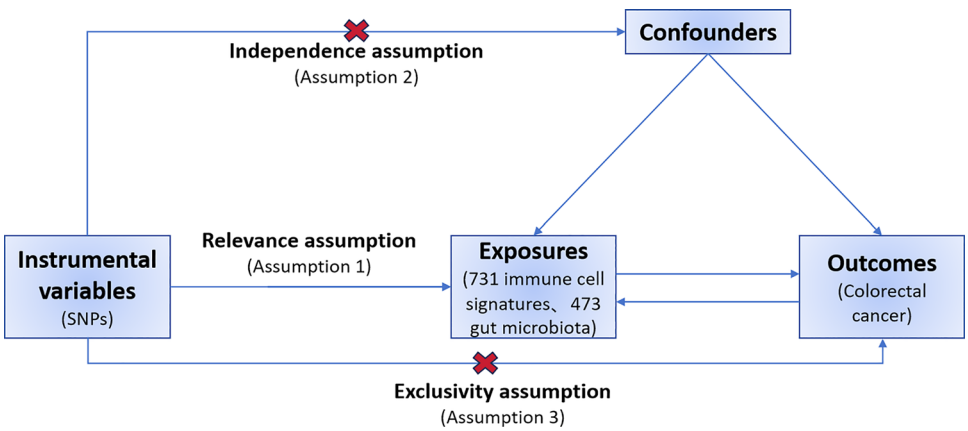

Fig. 1 Schematic diagram of the core assumptions of Mendelian randomization

B cells, circulating dendritic cells (CDCs), T cell maturation stages, monocytes, myeloid cells, T cells, B cells, natural killer (NK) cells, and regulatory T cells (Tregs), while the MP profiles focus on CDC and TBNK (T cells, B cells, NK cells) panels.

The human gut microbiota (GM) data utilized in this study originates from the largest GWAS published to date, representing an extensive multi-ethnic examination of human autosomal genetic variations in relation to GM composition,with accession numbers ranging from GCST90032172 to GCST90032644. Genome-wide association tests were performed on 2,801 microbial taxa alongside 7,967,866 human genetic variants obtained from a cohort of 5959 individuals enrolled in the FR02 study. Our analysis identified a total of 471 distinct taxa from the Genome Taxonomy Database (GTDB), representing 17% of all assessed taxa. This dataset encompasses 11 phyla, 19 classes, 24 orders, 62 families, 146 genera, and 209 species, providing a robust framework for investigating the intricate relationships between immune mechanisms and GM [26].

2.3 SNPs associated with colorectal cancer

The GWAS summary statistics of colorectal cancer were obtained from the FinnGen database (https://www.finngen.fi/en). The study performed a GWAS on European individuals (Ncase = 3022, Ncontrol = 174,006), with approximately 16380321 number of SNPs.

2.4 SNPs selection

To ensure the accuracy and validity of the causal relationship between GM, immune cells and CRC risk, we added the following restrictions to the IV inclusion criteria. First, only SNPs with $p < 1e\text{-}05$ were included as IV in the exposure and outcome analyses in the MR study. Second, SNPs with linkage disequilibrium (LD) $r2 < 0.001$) within a distance of 10,000 kb were removed using the TwoSampleMR R package [27].Third, SNPs that were significantly associated with the results were excluded (significance threshold of 1e-05). Fourth, palindromic SNPs were removed to ensure that SNP effects on exposure and outcome corresponded to the same allele. Finally, we calculated F-statistic values to measure the strength of IV,retaining SNPs with F-values greater than 10, and excluding SNPs with minimal allele frequency (MAF) less than 0.01 [28] (Table S1-3). IVs must not influence the outcome through pathways other than the exposure of interest. To avoid this violation, we screened for common SNPs associated with CRC traits using the LDtrait Tool (https://ldlink.nih.gov/) (Table S4-5).

2.5 Statistical analysis

Analyses were conducted using R version 4.3.0, employing the "Two-Sample MR" package to facilitate the formatting, harmonization, and semi-automated analysis of summary data from genetic association studies. Statistical significance was defined as a threshold of $p < 0.05$ [29]. Following the selection of valid single nucleotide polymorphisms (SNPs), we utilized inverse variance weighting (IVW) as the primary approach for estimating the parameters in our Mendelian randomization (MR) analysis [30, 31]. The IVW method is deemed the most accurate method for assessing the overall causal impact of exposure on outcomes, assuming that all selected SNPs are valid [32].

To bolster our findings, we incorporated complementary methods for causal inference, including the weighted median, MR Egger, weighted mode, and simple mode techniques [33]. The weighted median method offers robust estimates when more than half of the SNPs are valid, whereas the MR Egger is capable of providing reliable effect estimates even in cases where all SNPs are invalid [34]. Further evaluations included MR-Egger regression and MR pleiotropy residuals and outliers (MR-PRESSO) tests to ascertain potential horizontal pleiotropy among SNPs [35].

Additionally, we investigated reverse causality through reverse MR analysis involving the GM, immune cell phenotypes, and CRC, employing the same methodological framework. To mitigate the impact of multiple testing, we established that at least one of the MR-Egger and weighted median estimates must achieve significance ($p < 0.05$).

Subsequently, we performed a mediation analysis. Initially, we identified immune cells that were causally affected by the gut microbiota using univariate Mendelian randomization (UVMR) and calculated their effect values (β1). Next, we determined whether the mediator exhibited a causal relationship with the outcome independent of exposure, with its effect quantified asβ2. We implemented two-sample mediation analysis (TSMR) to decompose the direct effects (without mediators) and mediating effects (through mediators) of pathways linking exposure and outcomes. It is essential that both the direct and indirect effects of exposure on the outcome are aligned in the same direction. The mediating effect was derived using the formula$\beta 12 = \beta 1 \times \beta 2$. The proportion of the mediating effect relative to

the total effect was calculated as β12_p = β12/β_all × 100% [36]. Finally, the direct effect of exposure on the outcome was determined using the formula β_dir = β_all—β12 (Fig. 2).

3 Result

3.1 Exploration of the causal effect of immunophenotypes on colorectal cancer

We used two-sample MR analysis to detect the relationship between immunophenotypes and colorectal cancer. Preliminary results of IVW analysis showed that 35 suggestive immunophenotypes were identified,of which 5 were in the TBNK panel, 9 in the Treg panel,6 in the Myeloid cell panel,5 in the Maturation stages of T cell panel,7 in the B cell panel,2 in the cDC panel and 1 in the Monocyte panel (Fig S1). Combined with complementary and sensitivity analyses, nine immunophenotypes that met stringent screening criteria were identified as candidate immunophenotypes.The following immunophenotypes include CD25hi CD45RA + CD4 not Treg AC(OR 0.94, 95% CI: 0.90–0.98, $p = 0.002$),CD4 + AC(OR 0.95, 95% CI: 0.91–0.99, $p = 0.018$),CD19 on IgD + CD24-(OR 0.97, 95% CI: 0.95–1.00, $p = 0.022$),CD19 on IgD + CD38dim(OR 0.98, 95% CI: 0.95–1.00, $p = 0.032$),FSC-A on CD14 + monocyte(OR 0.95, 95% CI: 0.91–0.99, $p = 0.008$),CD4 on CD39 + secreting Treg (OR 0.97, 95% CI: 0.95–1.00, $p = 0.039$)suggesting a protective effect.Of these, CD3 on CD28- CD8br (OR 1.07, 95% CI: 1.01–1.13, $p = 0.016$),HLA DR on CD14 + CD16- monocyte(OR 1.04, 95% CI: 1.00–1.09, $p = 0.047$) and SSC-A on CD14 + monocyte(OR 1.06, 95% CI: 1.01–1.11, $p = 0.011$) are considered risk factors (Fig. 3, Table S6). Among the above nine immunophenotypes, each IVW resulted in a significant estimate ($p < 0.05$), and at least one of the MR-Egger and Weighted median estimates was significant ($p < 0.05$), and the direction and magnitude of the IVW, MR-Egger, and Weighted median estimates were consistent.To address potential pleiotropy issues, we scanned all SNPs used as IV in the study using the LDtrait tool. We identified one SNP (rs3184504) associated with CRC and reviewed the relevant literature to investigate its association with CRC (Table S4). After excluding this SNP, we performed MR analysis using the IVW method. The results showed that after removing this SNP, the p-value obtained using the IVW method remained greater than 0.05. After removing the outliers, the MR-PRESSO results did not support the presence of heterogeneous SNPs. Cochran's Q test ($p > 0.05$) and the MR-Egger's intercept test ($p > 0.05$) similarly demonstrated the absence of heterogeneous SNPs. The results of the MR leave-one-out sensitivity analysis indicate that individual SNPs do not cause bias in the MR estimation. Reverse MR analysis showed no forward causal association between the nine immunophenotypes and CRC, demonstrating the reliability of the results.Forest plots,scatter plots,funnel plots and leave-one-out sensitivity analysis plots associated with the MR analysis of the above positive results can be found in the supplementary files (Figs S2-S5).

3.2 The causal effect of GM on colorectal cancer

To further explore the causal effect of CRC on GM, we also chose two-sample MR analysis, which showed a causal relationship between a total of five GM and CRC: *Alloprevotella* (OR 1.17, 95% CI: 1.02–1.34, $p = 0.02$), factors.*Holdemania*

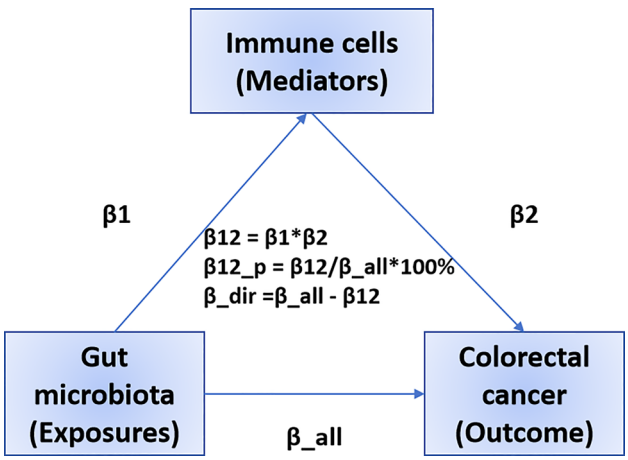

Fig. 2 The analytical methods of mediation analysis

Main.outcome	Method	nSNP	P.value	OR(95%CI)	Heterogeneity.Test.P	MR.Egger.Intercept	MR.PRESSO.Global.Test.P
CD25hi CD45RA+ CD4 not Treg AC	MR Egger	25	0.050	0.943(0.892 to 0.997)	0.066	0.818	0.142
	Weighted median	25	0.022	0.943(0.897 to 0.992)			
	Inverse variance weighted	25	0.002	0.939(0.902 to 0.977)			
	Simple mode	25	0.136	0.938(0.865 to 1.017)			
	Weighted mode	25	0.018	0.944(0.903 to 0.987)			
CD4+ AC	MR Egger	23	0.067	0.942(0.887 to 1.001)	0.095	0.795	0.166
	Weighted median	23	0.030	0.928(0.868 to 0.993)			
	Inverse variance weighted	23	0.018	0.947(0.906 to 0.991)			
	Simple mode	23	0.713	0.978(0.871 to 1.099)			
	Weighted mode	23	0.065	0.941(0.884 to 1.001)			
CD19 on IgD+ CD24-	MR Egger	25	0.034	0.970(0.944 to 0.996)	0.844	0.568	0.907
	Weighted median	25	0.088	0.974(0.945 to 1.004)			
	Inverse variance weighted	25	0.022	0.974(0.952 to 0.996)			
	Simple mode	25	0.310	0.964(0.900 to 1.033)			
	Weighted mode	25	0.065	0.977(0.953 to 1.000)			
CD19 on IgD+ CD38dim	MR Egger	24	0.035	0.969(0.943 to 0.996)	0.575	0.409	0.649
	Weighted median	24	0.283	0.984(0.954 to 1.014)			
	Inverse variance weighted	24	0.033	0.976(0.954 to 0.998)			
	Simple mode	24	0.344	0.971(0.914 to 1.031)			
	Weighted mode	24	0.064	0.973(0.947 to 1.000)			
CD3 on CD28- CD8br	MR Egger	20	0.222	1.086(0.956 to 1.233)	0.68	0.776	0.733
	Weighted median	20	0.003	1.114(1.037 to 1.196)			
	Inverse variance weighted	20	0.016	1.067(1.012 to 1.125)			
	Simple mode	20	0.078	1.136(0.994 to 1.300)			
	Weighted mode	20	0.035	1.128(1.016 to 1.252)			
FSC-A on CD14+ monocyte	MR Egger	23	0.137	0.956(0.903 to 1.012)	0.84	0.633	0.885
	Weighted median	23	0.021	0.930(0.875 to 0.989)			
	Inverse variance weighted	23	0.008	0.947(0.909 to 0.986)			
	Simple mode	23	0.928	0.996(0.908 to 1.092)			
	Weighted mode	23	0.099	0.952(0.900 to 1.007)			
HLA DR on CD14+ CD16- monocyte	MR Egger	22	0.065	1.084(1.000 to 1.174)	0.162	0.298	0.231
	Weighted median	22	0.040	1.055(1.002 to 1.110)			
	Inverse variance weighted	22	0.047	1.044(1.001 to 1.089)			
	Simple mode	22	0.051	1.106(1.005 to 1.218)			
	Weighted mode	22	0.048	1.049(1.003 to 1.096)			
CD4 on CD39+ secreting Treg	MR Egger	22	0.032	0.961(0.929 to 0.994)	0.803	0.291	0.798
	Weighted median	22	0.052	0.964(0.930 to 1.000)			
	Inverse variance weighted	22	0.039	0.972(0.946 to 0.999)			
	Simple mode	22	0.999	1.000(0.931 to 1.074)			
	Weighted mode	22	0.134	0.972(0.937 to 1.007)			
SSC-A on CD14+ monocyte	MR Egger	21	0.218	1.052(0.973 to 1.138)	0.483	0.746	0.518
	Weighted median	21	0.025	1.083(1.010 to 1.162)			
	Inverse variance weighted	21	0.011	1.063(1.014 to 1.115)			
	Simple mode	21	0.124	1.103(0.979 to 1.243)			
	Weighted mode	21	0.028	1.092(1.015 to 1.175)			

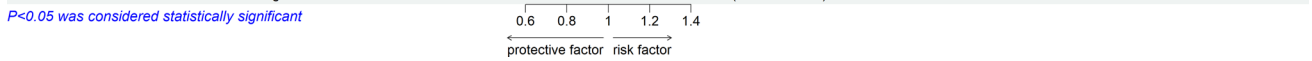

Fig. 3 Forest plot illustrating the causal effects between immune cell traits and CRC as determined by five MR analyses

(OR 0.82, 95% CI: 0.68–0.99, $p = 0.04$), *Megamonas* (OR 0.88, 95% CI: 0.79–0.97, $p = 0.01$), *Psychroserpens* (OR 0.64, 95% CI: 0.49–0.84, $p = 0.001$), and *Succinivibrionaceae* (OR 0.81, 95% CI: 0.72–0.92, $p = 0.001$) showed a negative association with CRC (Fig. 4, Table S7). Similarly, we scanned all SNPs used as IV in the study using the LDtrait tool and identified one SNP (rs 1,446,585) associated with CRC (Table S5). After excluding this SNP, we performed an MR analysis using the IVW method. The results showed that after removing this SNP, the p-value obtained using the IVW method remained greater than 0.05.All the forest plots,scatter plots,funnel plots and leave-one-out sensitivity analysis plots associated with the MR analysis of the above positive results can be found in the supplementary files (Fig S6-S9).

3.3 Mediation analyses of potential mediators

After identifying potential mediators, we performed a two-step MR analysis and revealed how GM impacts CRC through immune cells. We measured the effect of exposure (GM) on mediation (immune cells) by calculating the mediation effect, which showed a causal relationship between *Alloprevotella* and CD4 + AC (OR 0.829, 95% CI: 0.692, 0.993, $p = 0.042$), *Holdemania* and CD3 on CD28- CD8br (OR 1.327, 95% CI: 1.024–1.719, $p = 0.032$) (Fig. 5, Table S8). We found that CD4 + AC mediated the causal associations between *Alloprevotella* and CRC, with a mediation proportion of 6.48% (Fig. 6). CD3 on CD28- CD8br mediated the association between *Holdemania* and CRC, with a mediation proportion of 9.29% (Fig. 7).

Main.outcome	Method	nSNP	P.value	OR(95%CI)	Heterogeneity.Test.P	MR.Egger.Intercept	MR.PRESSO.Global.Test.P
Alloprevotella abundance	MR Egger	21	0.318	1.162(0.872 to 1.549)	0.822	0.961	0.881
	Weighted median	21	0.027	1.233(1.024 to 1.485)			
	Inverse variance weighted	21	0.024	1.170(1.020 to 1.341)			
	Simple mode	21	0.314	1.178(0.863 to 1.609)			
	Weighted mode	21	0.275	1.195(0.876 to 1.629)			
Holdemania abundance	MR Egger	19	0.481	0.798(0.432 to 1.475)	0.068	0.926	0.054
	Weighted median	19	0.030	0.790(0.638 to 0.978)			
	Inverse variance weighted	19	0.043	0.820(0.677 to 0.994)			
	Simple mode	19	0.120	0.680(0.427 to 1.081)			
	Weighted mode	19	0.114	0.696(0.454 to 1.068)			
Megamonas abundance	MR Egger	46	0.024	0.804(0.670 to 0.965)	0.84	0.277	0.837
	Weighted median	46	0.162	0.908(0.792 to 1.040)			
	Inverse variance weighted	46	0.009	0.877(0.794 to 0.968)			
	Simple mode	46	0.654	0.931(0.684 to 1.268)			
	Weighted mode	46	0.628	0.931(0.700 to 1.239)			
Psychroserpens abundance	MR Egger	17	0.228	0.656(0.340 to 1.265)	0.965	0.934	0.969
	Weighted median	17	0.021	0.651(0.453 to 0.937)			
	Inverse variance weighted	17	0.001	0.639(0.486 to 0.841)			
	Simple mode	17	0.107	0.588(0.320 to 1.081)			
	Weighted mode	17	0.118	0.591(0.317 to 1.103)			
Succinivibrionaceae abundance	MR Egger	47	0.038	0.756(0.586 to 0.977)	0.969	0.534	0.978
	Weighted median	47	0.001	0.748(0.629 to 0.890)			
	Inverse variance weighted	47	0.001	0.812(0.716 to 0.921)			
	Simple mode	47	0.036	0.635(0.420 to 0.959)			
	Weighted mode	47	0.028	0.646(0.443 to 0.942)			

P<0.05 was considered statistically significant

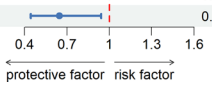

Fig. 4 Forest plot illustrating the causal effects between GMs and CRC as determined by five MR analyses

Expoursure	Outcome	Method	nSNP	P.value	OR(95%CI)	Heterogeneity.Test.P	MR.Egger.Intercept	MR.PRESSO.Global.Test.P
Alloprevotella abundance	CD4+ AC	MR Egger	21	0.147	0.741(0.502 to 1.092)	0.342	0.528	0.423
		Weighted median	21	0.047	0.787(0.621 to 0.996)			
		Inverse variance weighted	21	0.042	0.829(0.692 to 0.993)			
		Simple mode	21	0.135	0.763(0.543 to 1.072)			
		Weighted mode	21	0.053	0.754(0.576 to 0.987)			
Holdemania abundance	CD3 on CD28- CD8br	MR Egger	19	0.519	1.325(0.573 to 3.060)	0.2	0.996	0.228
		Weighted median	19	0.114	1.334(0.933 to 1.907)			
		Inverse variance weighted	19	0.032	1.327(1.024 to 1.719)			
		Simple mode	19	0.233	1.464(0.799 to 2.680)			
		Weighted mode	19	0.287	1.371(0.781 to 2.406)			

P<0.05 was considered statistically significant

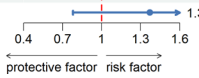

Fig. 5 Forest plot illustrating the causal effects between GMs and immune cell traits as determined by five MR analyses

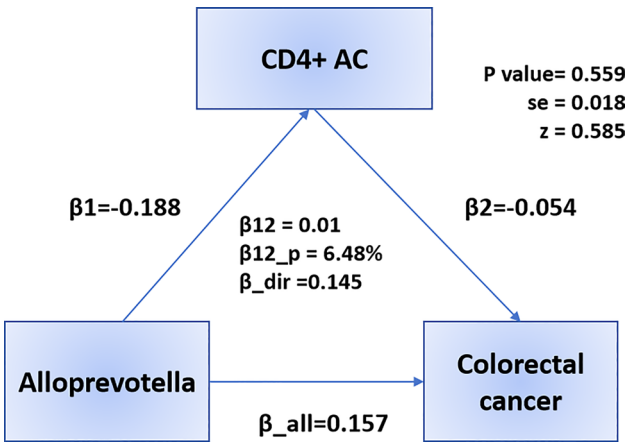

Fig. 6 Mediation effect of CD4 + AC in the association between *Alloprevotella* and colorectal cancer

Finally, we performed heterogeneity and pleiotropy analyses of the results of the aforementioned GM, immune cell, and mediation analyses. Our results showed p-values greater than 0.05, indicating the absence of heterogeneous and pleiotropic SNPs. In addition, we performed leave-one-out analysis, which further demonstrated the stability of our results.All the forest plots,scatter plots,funnel plots and leave-one-out sensitivity analysis plots associated with the MR analysis of the above positive results can be found in the supplementary files (Fig S10-S13).

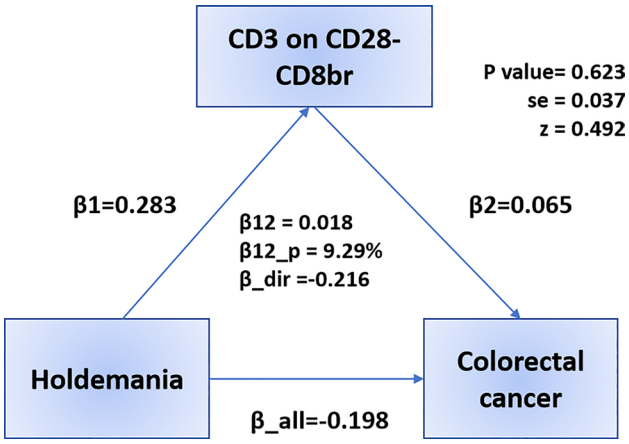

Fig. 7 Mediation effect of CD3 on CD28- CD8br in the association between *Holdemania* and colorectal cancer

4 Discussion

Colorectal cancer (CRC) has emerged as a global health concern, with half of all cases diagnosed at locally advanced stages or beyond [37]. To address this challenge, we conducted an intermediary Mendelian randomization (MR) study to elucidate the causal relationships among GM, immune cells, and CRC. This study represents a novel approach for exploring the interconnectedness of multiple GM, immune cells, and CRC through MR analysis.

Our Mendelian randomization findings revealed that CRC exhibited causal effects on nine distinct immunophenotypes, whereas five specific GM demonstrated an impact on CRC risk. Notably, our mediation MR analysis identified five CRC-associated GM and nine CRC-associated immune cell types. Among these, we discovered causal associations between *Alloprevotella* and CD4 + T cells, as well as between *Holdemania* and CD3 on CD28- CD8br, which helped us to gain more insight into the complex relationship between GM, immune cells, and CRC.

Numerous studies have highlighted the significant association between GM and CRC. Mechanistic insights suggest that dysbiosis, the presence of pathogenic GM, and their metabolites may contribute to CRC development and progression [38, 39]. In our investigation, we observed a notable positive correlation between CRC and *Alloprevotella* abundance of *Alloprevotella*. Recent research has identified *Alloprevotella* as a potential oral biomarker of the intestinal microbiota of patients with gastric cancer [40]. Additionally, evidence indicates that *Alloprevotella* is inversely associated with cytokines such as TNF-α, IFN-γ, and CXCR4 [41]. Li et al. reported elevated levels of *Alloprevotella* in CRC tissues compared to normal intestinal mucosa, consistent with our findings, suggesting a role for *Alloprevotella* in promoting CRC [42].

Furthermore, *Alloprevotella* has been implicated in the synthesis of short-chain fatty acids (SCFAs), which are key metabolites generated from the fermentation of insoluble dietary fibers by gut microbes. SCFAs are known to enhance gut health and have been shown to influence adoptive immunotherapy for cancer by modulating CD8 + T cells [43]. Based on these observations, we hypothesized that the metabolites produced by *Alloprevotella* may modulate immune cell function and consequently reduce CRC risk. This hypothesis underscores the need for further research to elucidate the precise mechanisms by which *Alloprevotella* and its metabolic products influence CRC pathogenesis and the immune response.

In contrast,our study identified four GM—*Holdemania, Megamonas, Psychroserpens, and Succinivibrionaceae*, which exhibit a negative association with CRC. *Holdemania* has been minimally investigated in the context of CRC; however, existing data suggest a higher abundance of *Holdemania* in younger CRC patients than in their older counterparts [9]. *Megamonas,* a member of the Firmicutes phylum, has been previously associated with CRC [44]. Notably, *Megamonas* was found to be significantly enriched in CRC patients, with an increased abundance observed in patients harboring KRAS mutations [45].*Psychroserpens*, a psychrophilic bacterium isolated from Antarctic marine sediments, is yet to be studied in relation to CRC [46]. Similarly, there is currently no available research on the association between *Succinivibrionaceae* and CRC, which underscores the need for further investigation into the roles of these microbiota in CRC pathogenesis, particularly given the limited existing evidence and potential implications for understanding the influence of microbes on CRC development.

Immune cells play a pivotal role in CRC development and progression. In our study, we observed a significant negative correlation between the presence of CD19 in IgD + CD24 – B cells and CD19 in IgD + CD38 dim B cells and CRC risk. CD19, a member of the immunoglobulin superfamily expressed exclusively on B lymphocytes, serves as

a critical co-receptor for B-cell antigen receptor (BCR) signaling. Co-ligation of CD19 with BCR has been shown to synergistically enhance calcium flux, mitogen-activated protein kinase (MAPK) activity, and cellular proliferation [47].

Prior research has indicated that depletion of CD19 + B lymphocytes can facilitate immune escape and adversely affect patient prognosis in CRC [48]. Our findings suggest that CD19 on IgD + CD24 – B cells and CD19 on IgD + CD38 dim B cells may act as protective factors against CRC. This observation aligns with recent studies that highlight the similar protective roles of these B-cell subsets. IgD + B cells, characterized by their surface expression of IgD, represent a subset of B cells involved in the immune response. IgD can coexist with other immunoglobulins in B cells and participate in the co-regulation of immune responses [49]. These results underscore the importance of CD19-expressing B-cell subsets in CRC and suggest potential avenues for further investigation of their roles as biomarkers or therapeutic targets in CRC management.

CD14, a prominent monocyte marker, functions as a pattern recognition receptor (PRR) and is integral in enhancing the innate immune response by mediating intracellular signaling upon encountering bacterial pathogens [50]. In our study, we found that HLA-DR expression in CD14 + CD16 – monocytes was positively associated with CRC risk. HLA-DR, an MHC class II cell surface receptor encoded by the human leukocyte antigen (HLA) complex on chromosome 6p21, along with CD14 and CD16, are crucial surface markers involved in signal recognition, signal transduction, and amplification of immune responses [51].

Previous investigations have demonstrated that HLA-DR expression in CD14 + CD16 – monocytes is notably elevated in CRC tissues and correlates with poor patient prognosis [52]. This subset of monocytes may contribute to tumor progression by producing angiogenic mediators or growth factors that promote tumor cell proliferation, thereby enhancing CRC migration and invasion [53, 54]. These findings suggest that HLA-DR + CD14 + CD16 – monocytes may play a significant role in the pathophysiology of CRC and may serve as potential biomarkers or therapeutic targets for improving disease outcomes.

Numerous studies underscore the pivotal role of CD4 + T cells in antitumor immunity, demonstrating their capacity to directly eliminate tumor cells and coordinate tumor destruction through cytokine production within the tumor microenvironment (TME) [55]. In CRC, GM has been identified as a key regulator of CD4 + T cell function. Recent research has highlighted a dichotomy within Bacteroides fragilis strains, where toxigenic strains promote tumorigenesis, whereas non-toxigenic strains confer protective effects by enhancing T follicular helper (Tfh) cell infiltration and the development of ectopic lymphoid structures [56]. In agreement with these observations, our findings revealed that CD4 + T cells may serve as a critical mediator between *Alloprevotella* and CRC. Specifically, while *Alloprevotella* has been identified as a potential risk factor for CRC, CD4 + T cells exhibit a protective role against the disease. We observed a negative correlation between *Alloprevotella* and CD4 + T cells, suggesting that *Alloprevotella* may facilitate CRC progression by suppressing CD4 + T cell activity, which may be attributable to *Alloprevotella*-induced inflammation, which could disrupt epithelial barrier integrity and enhance bacterial translocation. This cascade of events may lead to chronic inflammation and a sustained immune response that inhibits CD4 + T cell function, thereby contributing to the progression of CRC. These findings provide insights into the complex interplay between microbial components and immune cell dynamics in CRC progression, suggesting potential therapeutic strategies targeting microbial and immune interactions to mitigate CRC development [57].

Furthermore,we found a causal relationship between *Holdemania* and CD3 on CD28- CD8br.Our results indicate that CD3 on CD28- CD8br inhibits the protective effect of *Holdemania* against CRC.Previous studies have shown that *Holdemania* association with colorectal cancer suggests that it may play a role in immune function, inflammation, and hormone levels [58, 59].A study on irritable bowel syndrome showed that RPL9P33 and RP11-730G20.2 were positively correlated with the abundance of *Holdemania*, while RP11-730G20.2 was significantly enriched in CD8 + T cells, suggesting that *Holdemania* may have some link with CD8 + T cells [60]. In this study, ANXA2P2 was also found to be positively correlated with the abundance of *Holdemania*, and ANXA2P2 was significantly correlated with CD8 + T cells depletion in gliomas, which further indicated the association between *Holdemania* and CD8 + T cells [61, 62]. However, the specific mechanisms require further exploration.

Despite the insights provided by this study, several limitations must be acknowledged. First, the associations between genetic instrumental variables and phenotypes may be weakened, potentially resulting in a "weak instrument"phenomenon [63]. This limitation underscores the need for a cautious interpretation of our findings. Second, our primary analysis utilized inverse variance weighting (IVW) for Mendelian randomization (MR). While IVW was central to our analysis, it was essential to ensure that at least one of the MR-Egger or Weighted Median estimates was statistically significant ($p < 0.05$) to confirm robustness. However, these methods do not entirely mitigate the risk of false positives, and the mediation analysis yielded low effect sizes, suggesting a limited mediation impact. Additionally, the data used were predominantly derived from individuals

of European ancestry and were restricted to adults, with no stratification by sex or age. These factors may affect the generalizability and accuracy of the results.

Future studies should involve larger and more diverse cohorts to enhance the robustness and applicability of our conclusions. Validation across various populations will be crucial to affirm the generalizability of our findings and address the limitations of this study.

5 Conclusion

Overall, our study comprehensively assessed the association between GM, immune cells, and CRC.These findings suggest that specific gut microbiota can positively or negatively affect the development and progression of colorectal cancer by affecting specific immune cells. This provides new ideas for the treatment of colorectal cancer: targeting specific gut microbiota (such as *Alloprevotella* or *Holdemania*), by affecting the corresponding immune cells, and then play a protective role in colorectal cancer. Considering the current widespread use of immunotherapy in colorectal cancer, our subsequent studies would further explore the impact of targeting gut microbiota or immune cells on colorectal cancer immunotherapy.

Acknowledgements We extend our sincere gratitude to the consortia and investigators behind the GWAS for providing the summary data that made this research possible. We are also deeply appreciative of the participants whose contributions were instrumental in the success of these studies.

Author contributions LZ (Linyi Zheng): Designing the study, formal analysis, writing of original draft YL (Yuqiang Li): Data analysis, funding acquisition and formal analysis CG (Cenap Güngör): Formal analysis, guiding and editing HG(Heming Ge): Data curation, methodology, writing, reviewing, and editing All the authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Natural Science Foundationof Hunan Province (No. 2022JJ40799).

Data availability The Data of immune cells and GMs for the present study can be download in GWAS (GWAS ID included in the article), and the GWAS data of colorectal cancer were obtained from the FinnGen database (https://www.finngen.fi/en).Further inquiries can be directed to the corresponding author.

Declarations

Competing interests The authors declare no competing interests.

Ethics statement The GWAS summary statistics utilized for immune cells, GM, and CRC in this study were extracted from publicly available, ethically reviewed publications. These data were generated with the approval of their respective institutional review boards. Given that our research entailed no collection of fresh primary data and was conducted entirely online, we did not require any new ethical approval.

References

1. Siegel RL, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020;70:145–64. https://doi.org/10.3322/caac.21601.
2. Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49. https://doi.org/10.3322/caac.21660.
3. Al-Hujaily EM, Al-Sowayan BS, Alyousef Z, Uddin S, Alammari F. Recruiting immunity for the fight against colorectal cancer: current status and challenges. Int J Mol Sci. 2022. https://doi.org/10.3390/ijms232213696.

4. Chen J, et al. Multi-omics analysis of gut microbiota and host transcriptomics reveal dysregulated immune response and metabolism in young adults with irritable bowel syndrome. Int J Mol Sci. 2024;25:3514.
5. Chen L, Fu B. T cell exhaustion assessment algorism in tumor microenvironment predicted clinical outcomes and immunotherapy effects in glioma. Front Genet. 2022;13:1087434.
6. Du P, et al. ANXA2P2/miR-9/LDHA axis regulates Warburg effect and affects glioblastoma proliferation and apoptosis. Cell Signal. 2020;74: 109718.
7. El Tekle G, Andreeva N, Garrett WS. The role of the microbiome in the etiopathogenesis of colon cancer. Annu Rev Physiol. 2024;86:453–78.
8. Hu Y, Zhou P, Deng K, Zhou Y, Hu K. Targeting the gut microbiota: a new strategy for colorectal cancer treatment. J Transl Med. 2024;22:915.
9. Kharofa J, Apewokin S, Alenghat T, Ollberding NJ. Metagenomic analysis of the fecal microbiome in colorectal cancer patients compared to healthy controls as a function of age. Cancer Med. 2023;12:2945–57. https://doi.org/10.1002/cam4.5197.
10. Martin-Gallausiaux C, et al. Fusobacterium nucleatum promotes inflammatory and anti-apoptotic responses in colorectal cancer cells via ADP-heptose release and ALPK1/TIFA axis activation. Gut Microbes. 2024;16:2295384.
11. Wong CC, Yu J. Gut microbiota in colorectal cancer development and therapy. Nat Rev Clin Oncol. 2023;20:429–52. https://doi.org/10.1038/s41571-023-00766-x.
12. Xing C, et al. Interaction between microbiota and immunity and its implication in colorectal cancer. Front Immunol. 2022;13: 963819. https://doi.org/10.3389/fimmu.2022.963819.
13. McLean MH, et al. The inflammatory microenvironment in colorectal neoplasia. PLoS ONE. 2011;6: e15366. https://doi.org/10.1371/journal.pone.0015366.
14. Peddareddigari VG, Wang D, Dubois RN. The tumor microenvironment in colorectal carcinogenesis. Cancer Microenviron. 2010;3:149–66. https://doi.org/10.1007/s12307-010-0038-3.
15. Halama N, et al. The localization and density of immune cells in primary tumors of human metastatic colorectal cancer shows an association with response to chemotherapy. Cancer Immun. 2009;9:1.
16. Galon J, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science. 2006;313:1960–4. https://doi.org/10.1126/science.1129139.
17. Wissfeld J, Werner A, Yan X, Ten Bosch N, Cui G. Metabolic regulation of immune responses to cancer. Cancer Biol Med. 2022;19:1528–42. https://doi.org/10.20892/j.issn.2095-3941.2022.0381.
18. Mantovani A, Romero P, Palucka AK, Marincola FM. Tumour immunity: effector response to tumour and role of the microenvironment. Lancet. 2008;371:771–83. https://doi.org/10.1016/S0140-6736(08)60241-X.
19. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454:436–44. https://doi.org/10.1038/nature07205.
20. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362: k601. https://doi.org/10.1136/bmj.k601.
21. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32:1–22. https://doi.org/10.1093/ije/dyg070.
22. Cornish AJ, Tomlinson IPM, Houlston RS. Mendelian randomisation: a powerful and inexpensive method for identifying and excluding non-genetic risk factors for colorectal cancer. Mol Aspects Med. 2019;69:41–7. https://doi.org/10.1016/j.mam.2019.01.002.
23. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. Am J Epidemiol. 2012;175:332–9. https://doi.org/10.1093/aje/kwr323.
24. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol. 2011;40:755–64. https://doi.org/10.1093/ije/dyr036.
25. Orr V, et al. Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet. 2020;52(10):1036–45.
26. Qin Y, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat Genet. 2022;54:134–42. https://doi.org/10.1038/s41588-021-00991-z.
27. Genomes Project C, et al. A global reference for human genetic variation. Nature. 2015;526:68–74. https://doi.org/10.1038/nature15393.
28. Luo J, et al. Systemic inflammatory markers in relation to cognitive function and measures of brain atrophy: a Mendelian randomization study. Geroscience. 2022;44:2259–70. https://doi.org/10.1007/s11357-022-00602-7.
29. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
30. Kamat MA, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics. 2019;35:4851–3. https://doi.org/10.1093/bioinformatics/btz469.
31. Chen L, et al. Mendelian randomization rules out causation between inflammatory bowel disease and non-alcoholic fatty liver disease. Front Pharmacol. 2022;13: 891410. https://doi.org/10.3389/fphar.2022.891410.
32. de Lange KM, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49:256–61. https://doi.org/10.1038/ng.3760.
33. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693–8. https://doi.org/10.1038/s41588-018-0099-7.
34. Chen H, et al. The association between genetically predicted systemic inflammatory regulators and polycystic ovary syndrome: a Mendelian randomization study. Front Endocrinol (Lausanne). 2021;12:731569. https://doi.org/10.3389/fendo.2021.731569.
35. Burgess S, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res. 2019;4:186. https://doi.org/10.12688/wellcomeopenres.15555.3.
36. Zhu X, et al. Causal associations of BAFF-R on IgD+ CD24- B cell immune cell trait with hepatocellular carcinoma and the mediating role of phenylacetylglutamate levels: a Mendelian randomization study. J Cancer. 2024;15:4591–603. https://doi.org/10.7150/jca.96059.
37. Murphy CC, Zaki TA. Changing epidemiology of colorectal cancer - birth cohort effects and emerging risk factors [J]. Nat Rev Gastroenterol Hepatol. 2024;21(1):25–34.
38. Qu R, Zhang Y, Ma Y, Zhou X, Sun L, Jiang C, Zhang Z, Fu W. Role of the gut microbiota and its metabolites in tumorigenesis or development of colorectal cancer. Adv Sci (Weinh). 2023;10(23):e2205563. https://doi.org/10.1002/advs.202205563.

39. Wang Z, Dan W, Zhang N, Fang J, Yang Y. Colorectal cancer and gut microbiota studies in China. Gut Microbes. 2023;15(1):2236364. https://doi.org/10.1080/19490976.2023.2236364.
40. Liu Y, Wang H, Jiang H, Sun Z, Sun A. Alloprevotella can be considered as a potential oral biomarker in intestinal metaphase of gastric patients. Stud Health Technol Inform. 2023;308:155–67. https://doi.org/10.3233/SHTI230836.
41. Yu H, et al. Fecal microbiota transplantation inhibits colorectal cancer progression: reversing intestinal microbial dysbiosis to enhance anti-cancer immune responses. Front Microbiol. 2023;14:1126808. https://doi.org/10.3389/fmicb.2023.1126808.
42. Li Y, et al. Gut microbiota signatures in tumor, para-cancerous, normal mucosa, and feces in colorectal cancer patients. Front Cell Dev Biol. 2022;10: 916961. https://doi.org/10.3389/fcell.2022.916961.
43. Hou H, et al. Gut microbiota-derived short-chain fatty acids and colorectal cancer: ready for clinical translation? Cancer Lett. 2022;526:225–35. https://doi.org/10.1016/j.canlet.2021.11.027.
44. Morotomi M, Nagai F, Sakon H. Genus megamonas should be placed in the lineage of firmicutes; clostridia; clostridiales; 'acidaminococcaceae.' Megamonas Int J Syst Evol Microbiol. 2007;57:1673–4. https://doi.org/10.1099/ijs.0.65150-0.
45. Yuan D, et al. A comprehensive analysis of the microbiota composition and host driver gene mutations in colorectal cancer. Invest New Drugs. 2022;40:884–94. https://doi.org/10.1007/s10637-022-01263-1.
46. Kristyanto S, et al. *Psychroserpens ponticola* sp. nov. and *Marinomonas maritima* sp. nov., isolated from seawater. Int J Syst Evol Microbiol. 2023. https://doi.org/10.1099/ijsem.0.006090.
47. Sorrentino C, D'Antonio L, Fieni C, Ciummo SL, Di Carlo E. Colorectal cancer-associated immune exhaustion involves T and B lymphocytes and conventional NK cells and correlates with a shorter overall survival. Front Immunol. 2021;12: 778329. https://doi.org/10.3389/fimmu.2021.778329.
48. Li X, et al. CD19, from bench to bedside. Immunol Lett. 2017;183:86–95. https://doi.org/10.1016/j.imlet.2017.01.010.
49. Rampoldi F, et al. gammadelta T cells license immature B cells to produce a broad range of polyreactive antibodies. Cell Rep. 2022;39: 110854. https://doi.org/10.1016/j.celrep.2022.110854.
50. Sakakura K, et al. Immunological features of circulating monocyte subsets in patients with squamous cell carcinoma of the head and neck. Clin Immunol. 2021;225: 108677. https://doi.org/10.1016/j.clim.2021.108677.
51. Prat M, et al. Circulating CD14(high) CD16(low) intermediate blood monocytes as a biomarker of ascites immune status and ovarian cancer progression. J Immunother Cancer. 2020. https://doi.org/10.1136/jitc-2019-000472.
52. Vayrynen JP, et al. Prognostic significance of myeloid immune cells and their spatial distribution in the colorectal cancer microenvironment. J Immunother Cancer. 2021. https://doi.org/10.1136/jitc-2020-002297.
53. Jaillon S, et al. Neutrophil diversity and plasticity in tumour progression and therapy. Nat Rev Cancer. 2020;20:485–503. https://doi.org/10.1038/s41568-020-0281-y.
54. De Palma M, Biziato D, Petrova TV. Microenvironmental regulation of tumour angiogenesis. Nat Rev Cancer. 2017;17:457–74. https://doi.org/10.1038/nrc.2017.51.
55. Poncette L, Bluhm J, Blankenstein T. The role of CD4 T cells in rejection of solid tumors. Curr Opin Immunol. 2022;74:18–24. https://doi.org/10.1016/j.coi.2021.09.005.
56. Borst J, Ahrends T, Babala N, Melief CJM, Kastenmuller W. CD4(+) T cell help in cancer immunology and immunotherapy. Nat Rev Immunol. 2018;18:635–47. https://doi.org/10.1038/s41577-018-0044-0.
57. Kirkpatrick C, Lu YW. Deciphering CD4(+) T cell-mediated responses against cancer. Mol Carcinog. 2024;63:1209–20. https://doi.org/10.1002/mc.23730.
58. Saus E, Iraola-Guzman S, Willis JR, Brunet-Vega A, Gabaldon T. Microbiome and colorectal cancer: roles in carcinogenesis and clinical potential. Mol Aspects Med. 2019;69:93–106.
59. Sheng Q, et al. Characteristics of fecal gut microbiota in patients with colorectal cancer at different stages and different sites. Oncol Lett. 2019;18:4834–44.
60. Su ACY, et al. Lactococcus lactis HkyuLL 10 suppresses colorectal tumourigenesis and restores gut microbiota through its generated alpha-mannosidase. Gut. 2024;73:1478–88.
61. Wang N, et al. Fusobacterium nucleatum induces chemoresistance in colorectal cancer by inhibiting pyroptosis via the Hippo pathway. Gut Microbes. 2024;16:2333790.
62. Xu YJ, et al. Multiomics analysis revealed colorectal cancer pathogenesis. J Proteome Res. 2024;23:2100–11.
63. Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. Stat Med. 2011;30:1312–23. https://doi.org/10.1002/sim.4197.

2.3 Development of a hypoxia-responsive macrophage prognostic model using single-cell and bulk RNA sequencing in pancreatic cancer

Ge H, Wolters-Eisfeld G, Hackert T, Li Y, Güngör C (2025). PLoS One. 20(5):e0322618.

RESEARCH ARTICLE

Development of a hypoxia-responsive macrophage prognostic model using single-cell and bulk RNA sequencing in pancreatic cancer

Heming Ge[1], Gerrit Wolters-Eisfeld[1], Thilo Hackert[1], Yuqiang Li[2], Cenap Güngör[1]*

1 Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, **2** Department of General Surgery, Xiangya Hospital, Central South University, Changsha, China

* c.guengoer@uke.de

OPEN ACCESS

Citation: Ge H, Wolters-Eisfeld G, Hackert T, Li Y, Güngör C (2025) Development of a hypoxia-responsive macrophage prognostic model using single-cell and bulk RNA sequencing in pancreatic cancer. PLoS One 20(5): e0322618. https://doi.org/10.1371/journal.pone.0322618

Editor: Surinder K. Batra, University of Nebraska Medical Center, UNITED STATES OF AMERICA

Received: August 30, 2024

Accepted: March 24, 2025

Published: May 2, 2025

Data availability statement: All scRNA-seq files are available from the Gene Expression Omnibus database (accession number GSE155698).

Abstract

Objective

Pancreatic ductal adenocarcinoma (PDAC) is characterized by a low survival rate and limited responsiveness to current therapies. The role of hypoxia in the tumor microenvironment is critical, influencing tumor progression and therapy resistance. The aim of this study was to implement the complex dynamics of the hypoxic tumor microenvironment in PDAC in a hypoxia-related prognosis model.

Methods

We utilized single-cell RNA sequencing (scRNA-seq) data and integrated it with TCGA-PAAD database to identify hypoxia-responsive macrophage subsets and related genes. Kaplan-Meier survival analysis, Cox regression, and Lasso regression methods were employed to construct and validate a hypoxia-related prognostic model. The model's effectiveness was evaluated through its predictive capabilities regarding chemotherapy sensitivity and overall survival.

Results

Our research integrated data from scRNA-seq and the TCGA-PAAD database to construct a hypoxia-related prognostic model that encompassed 13 critical genes. This hypoxia model independently predicted chemotherapy response and poor outcomes, outperforming traditional clinicopathologic features. Additionally, a pan-cancer analysis affirmed the relevance of our hypoxia-related genes across multiple malignancies, particularly highlighting *KRTCAP2* as a pivotal biomarker associated with worse prognosis and reduced immune infiltration.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Conclusion

Our findings underscored the prognostic potential of hypoxia-related model and offered a novel avenue for therapeutic targeting, aiming to ameliorate outcomes in pancreatic cancer.

Introduction

Pancreatic cancer is globally recognized as one of the most lethal malignancies [1], with ductal adenocarcinoma (PDAC) representing the predominant pathological type [2]. Despite noteworthy advancements in the field of PDAC research, the five-year survival rate for patients still remains approximately 10% [1]. Current systemic therapies encompassing surgical intervention, chemotherapy, and immunotherapy are widely administered; however, the response rate among patients is still extremely low [3,4]. Notably, even though immune checkpoint inhibitors have markedly enhanced outcomes in a variety of tumor types recently, their efficacy in improving pancreatic cancer prognosis remains limited [5]. Consequently, there is a pressing need to identify novel molecular markers and develop effective therapeutic targets for pancreatic cancer.

Hypoxia is a hallmark of the tumor microenvironment in solid malignancies, including pancreatic cancer, where it contributes to several pathological processes such as impaired immune responses, metabolic reprogramming, epithelial-mesenchymal transition, and increased therapy resistance [6,7]. In pancreatic cancer, the presence of hypoxic areas is critically implicated in adverse patient outcomes and serve as independent prognostic markers [8]. Moreover, tumor-associated macrophages (TAMs) are notably accumulated in these hypoxic regions of the tumor. Previous studies have demonstrated that TAMs not only facilitate tumorigenesis but also correlate with poor survival outcomes due to their involvement in inflammation, further emphasizing their potential as therapeutic targets in pancreatic cancer [9,10]. The dynamic interactions between cancer cells and TAMs within these hypoxic environments are pivotal in driving tumorigenesis and therefore present promising targets for innovative therapeutic strategies in cancer management [11].

In this study, our findings indicated that hypoxia significantly affects macrophages more than any other investigated immune cell type within the PDAC microenvironment. We pinpointed a distinct subset of macrophages using single-cell RNA sequencing (scRNA-seq) data from pancreatic cancer, which displayed heightened susceptibility to hypoxic conditions. By integrating data from the TCGA-PAAD database, we identified a subset of hypoxia-related genes and subsequently developed a novel hypoxia-related prognostic model. This model is independent from current clinicopathologic features in pancreatic cancer. We further validated the prognostic efficacy of our model across multiple public databases. Our results confirmed that this model effectively predicts the sensitivity to common chemotherapeutic agents, including gemcitabine, oxaliplatin, cisplatin, 5-fluorouracil, and paclitaxel, within the pancreatic cancer context. Additionally, the model demonstrated robust prognostic prediction capabilities in a pan-cancer analysis.

Materials and methods

Data acquisition

The 200 hypoxia hallmark genes were retrieved from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). According to the MSigDB, this set consists of genes that are up-regulated in response to low oxygen levels (hypoxia).

The scRNA-seq data GSE155698 [12] were obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), which comprised 19 samples: 16 primary PDAC tissues and 3 non-malignant pancreas tissues.

Bulk RNA-seq data, along with copy number variation (CNV), single-nucleotide variants (SNV), and associated clinicopathologic information of PAAD were accessed from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/), which included 159 pancreatic cancer tissues, following the exclusion of samples lacking survival data and clinical details. For external validation, databases from two cohorts, PACA-CA and PACA-AU, were utilized, encompassing 142 and 76 samples respectively, after removing samples without follow-up. These were accessed from the International Cancer Genome Consortium (ICGC) database (https://dcc.icgc.org/).

Processing of scRNA-seq data

ScRNA-seq data were processed using the R package "Seurat" (v5.0.1). Initial quality control was performed with the following criteria:

1) Genes expressed in fewer than three single cells were excluded;

2) Cells with total gene counts <200 or > 5,000 were removed to filter low-quality cells and potential doublets;

3) Cells with mitochondrial gene content >20% were discarded to eliminate apoptotic cells.

Normalization and feature selection. Gene expression was normalized using the LogNormalize method (scale factor = 10,000) via the NormalizeData function. To identify highly variable genes for downstream analysis, 2,000 variable features were selected using the FindVariableFeatures function with the "vst" method.

Data scaling and dimensionality reduction. Expression values were standardized (z-score transformation) using ScaleData. Principal component analysis (PCA) was performed on scaled data (RunPCA), and the top 30 principal components (PCs) were selected based on the elbow plot for downstream clustering.

Cell clustering and annotation. Cell neighborhoods were constructed using FindNeighbors (dims = 1:30), followed by FindClusters (resolution = 0.4) to partition cells into distinct clusters. Uniform Manifold Approximation and Projection (UMAP) was applied for 2D visualization (RunUMAP, dims = 1:30). Cluster identity was determined by manually curating the top 10 marker genes from FindAllMarkers (Wilcoxon test, min.pct = 0.25, logfc.threshold = 0.25) against canonical cell-type markers (Fig 2D).

Calculation of hypoxic microenvironment scores

For bulk RNA-seq data (TCGA-PAAD). The tumor hypoxic microenvironment scores were computed via single-sample Gene Set Enrichment Analysis (ssGSEA) using the GSVA R package (v1.46.0). The 200 hypoxia hallmark genes were ranked per sample, and enrichment scores were calculated with 1,000 permutations. All 200 genes were included in the analysis. Subsequently, samples were classified into high and low hypoxic microenvironment groups based on the median score.

For scRNA-seq data. Hypoxia activity was quantified using two methods

AddModuleScore: The average expression of hypoxia genes was subtracted by the mean expression of 100 control gene sets to generate per-cell hypoxia scores.

AUCell: Gene rankings per cell were computed using the AUCell R package (v 1.20.2), and area under the curve (AUC) values were computed using the top 10% of ranked genes.

Gene set enrichment analysis

To investigate differential signaling pathways, we performed Gene set enrichment analysis (GSEA) using the package "clusterProfiler" (v4.6.0). This analysis compared macrophage cluster 1 with macrophage cluster 2 in scRNA-seq data, as well as high and low hypoxia groups within the TCGA-PAAD dataset. Pathway marker genes were downloaded from the MSigDB. The analysis was performed with the permutations parameter set to 1000. Results were considered significant if the normalized enrichment score (NES) had an absolute value greater than 1, with a FDR below 0.25 and p-values less than 0.05.

Construction and evaluation of the prognostic model

The FindMarkers function within the package "Seurat" was employed to identify DEGs between normal and tumor tissues in macrophage cluster 1 in scRNA-seq data, using Wilcoxon tests with a threshold value of 0.25. Genes were further filtered based on statistical significance and effect size, retaining only those with a p-value less than 0.05 and an absolute log2 fold change (|log2FC|) greater than 0.25. A univariate Cox regression analysis was performed to evaluate the prognostic value of these DEGs for OS in patients from the TCGA-PAAD dataset using the package "survival" (v3.4.0) [13]. Subsequent least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation (glmnet, v4.1.8) was used to penalize overfitting genes [14–16]. The optimal lambda ($\lambda = 0.0432$) was selected via minimum partial likelihood deviance.

Utilizing the 13 hypoxia-related genes, ssGSEA was performed to calculate the enrichment fraction of hypoxia in each sample from the TCGA-PAAD database. Patients were stratified into high and low hypoxia groups based on the median hypoxia score. The prognostic efficacy of the hypoxia model was validated through the construction of receiver operating characteristic (ROC) curves and Kaplan–Meier (K–M) survival curves, using the packages "timeROC" (v0.4) and "survminer" (v0.4.9), respectively. Decision curve analysis (ggDCA, v1.2) was used to assess clinical utility. Additionally, the validity of the model was corroborated using external databases from PACA-CA and PACA-AU.

Mutation landscape analysis

SNV and TMB analysis: Mutation annotation format files were processed using package "maftools" (v2.14.0). Tumor mutational burden (TMB) was calculated as mutations per megabase. CNV analysis: Segmented copy number data from TCGA were analyzed using GISTIC 2.0 (thresholds: amplification = 0.2, deletion = -0.2).

Immune landscape analysis

To investigate the characteristics of immune cells across different hypoxia groups, we employed the R package "estimate" (v1.0.13) to calculate immune and stromal scores. Furthermore, we adopted the CIBERSORT algorithm [17] to evaluate the infiltration of 22 immune cell types (LM22 signature matrix) in both high and low hypoxia groups using package "CIBERSORT" (v0.1.0). The association between the hypoxia score and the levels of immune cell infiltration was quantified using the Pearson correlation coefficient.

Drug sensitivity analysis

IC50 values for drugs were determined from the Genomics of Drug Sensitivity in Cancer (GDSC) database using the R package "oncoPredict" (v0.2). Differences in IC50 values between high and low hypoxia groups were statistically analyzed using the Wilcoxon test to assess significance.

Statistical analysis

All statistical analysis were performed using R software (version 4.2.2). Differences among groups were evaluated using the Wilcoxon test, while correlations were assessed using either the Spearman or Pearson correlation coefficients. Survival differences were analyzed using the Log-rank test through Kaplan-Meier curves. All statistical tests were two-tailed, and a P-value of less than 0.05 was considered statistically significant.

Results

Association between tumor hypoxic microenvironment and survival outcomes in pancreatic cancer

The 200 hypoxia hallmark genes were retrieved from the Molecular Signatures Database [18]. We performed Kaplan–Meier survival analysis to explore the correlation between the tumor hypoxic microenvironment and survival outcomes in TCGA-PAAD. Our findings demonstrated that a highly tumor hypoxic microenvironment in pancreatic cancer correlated with adverse prognostic outcomes, including reduced overall survival (OS) and progression free survival (PFS) (Figs 1, 2A and 2B).

Annotation of cell types and hypoxic microenvironment scores

Based on scRNA-seq data from GSE155698 [12], we obtained gene expression profiles of 37,018 cells from 16 primary PDAC samples and 7,316 cells from 3 non-malignant pancreas samples for further analysis after rigorous quality control and batch correction. Cells were classified into 13 distinct types based on the expression of typical cell markers. These types included epithelial cells, neutrophils, macrophages, T cells, acinar cells, mast cells, plasma cells, NK cells, fibroblasts, pericytes, B cells, dendritic cells, and endothelial cells (Fig 2C). Key marker genes utilized for annotation are depicted in Fig 2D.

We subsequently calculated hypoxic microenvironment scores for all identified cell types using the AddmoduleScore and AUCell functions. The analysis revealed that macrophages exhibited the highest hypoxic microenvironment scores, significantly higher than those of other immune cell types (Fig 2E and 2F). This significant elevation among macrophages suggested that tumor hypoxic microenvironment may play a crucial role in modulating macrophage functions, thereby potentially impacting the progression of PDAC. To further dissect the sensitivity of macrophages to the tumor hypoxic microenvironment, we performed subclustering of macrophages and recalculated their hypoxic microenvironment scores. We identified a subgroup of macrophages displaying significantly elevated hypoxic microenvironment scores, which we termed “macrophage cluster1”, representing hypoxia-responsive macrophages. In contrast, other subgroups exhibited relatively lower sensitivity to tumor hypoxic microenvironment, and we designated these as “macrophage cluster2” (Fig 2G). Gene Set Enrichment Analysis (GSEA) conducted on these subtypes indicated a significantly more pronounced effect of hypoxia on macrophage cluster1, compared to cluster2 (Fig 2H), consistent with the hypoxic microenvironment scores calculated by AddmoduleScore (Fig 2I) and AUCell functions (Fig 2J).

Hub genes identification and hypoxia signature construction

To develop a risk signature, we initially identified differentially expressed genes (DEGs) between PDAC and non-malignant pancreas tissues within macrophage cluster1 in scRNA-seq data. A total of 882 DEGs were discerned, comprising 571 upregulated and 311 downregulated genes. These DEGs represented the specific alterations in hypoxia-responsive macrophages within pancreatic cancer. Subsequently, we evaluated the prognostic value of these 882 DEGs using univariate Cox regression analysis, identifying 23 genes associated with poor prognosis in the TCGA-PAAD cohort. Further refinement was conducted through Lasso Cox regression analysis to pinpoint hub genes. As illustrated in Fig 3A and 3B, the optimal Lambda is 0.0432, and 13 hypoxia-related genes were finally included in the construction of the

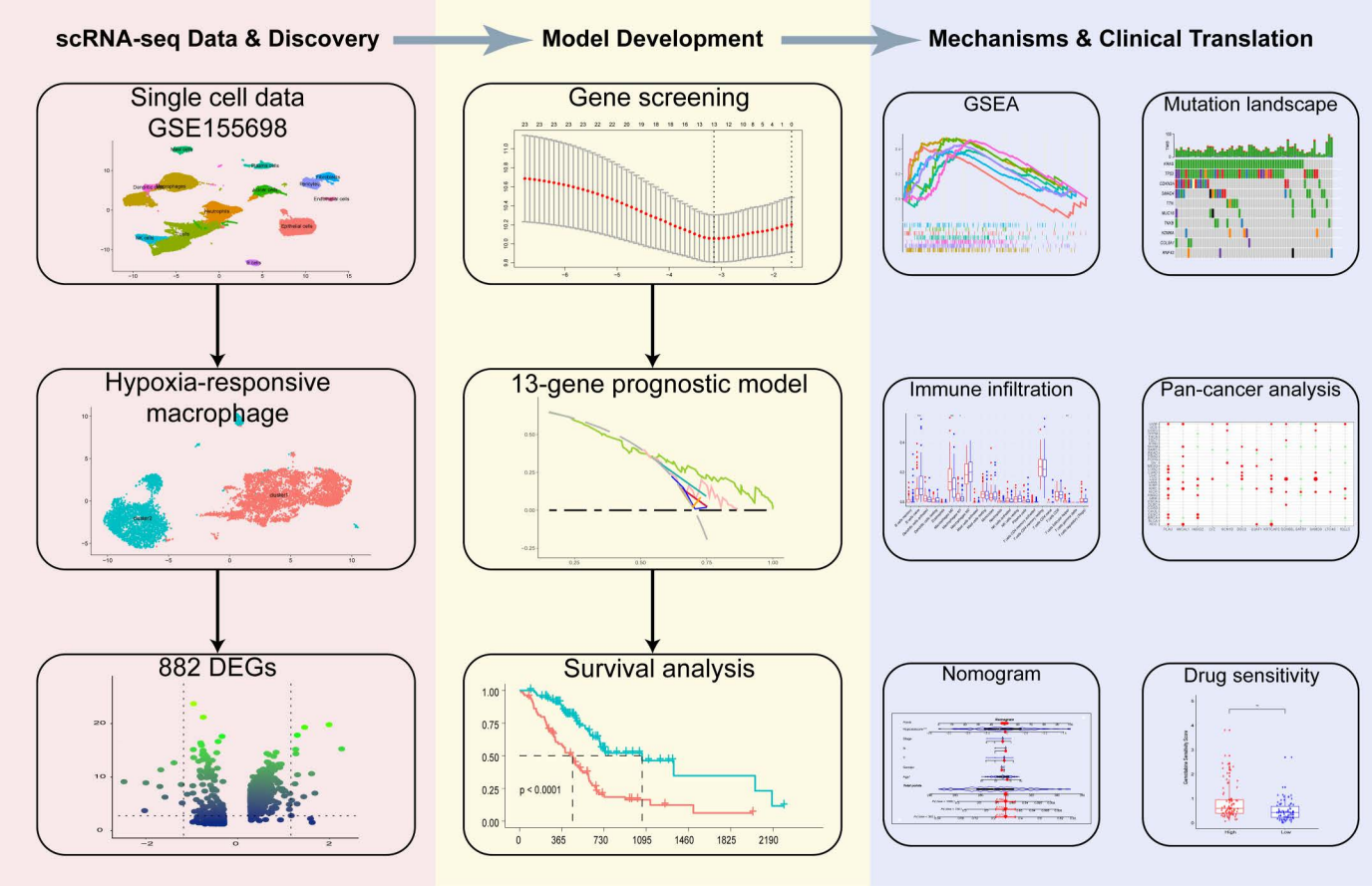

Fig 1. Flow chart of this study.

https://doi.org/10.1371/journal.pone.0322618.g001

hypoxia-related prognostic model, including *LYZ*, *SCN1B*, *PLAU*, *INSIG2*, *DSC2*, *MICAL1*, *U2AF1*, *KRTCAP2*, *DDX60L*, *SATB1*, *SAMD9*, *LTC4S*, *IGLL5*. The regression coefficients for each gene were detailed in S1 Table.

Hypoxia model predicts survival in TCGA cohort

To assess the prognostic significance of hypoxia in PDAC, we calculated a hypoxia score for each patient in the TCGA-PAAD cohort according to the hypoxia-related prognostic model. The hypoxia score represented the degree of hypoxic activity based on the hypoxia-related prognostic model. Patients were stratified into high and low hypoxia groups using the median hypoxia score as a cutoff. Fig 3C displayed the distribution of hypoxia scores and correlated them with patient survival status, indicating a higher mortality rate in the high hypoxia group. Kaplan−Meier survival curve demonstrated that patients with high hypoxia scores exhibited significantly worse outcomes, compared to those with low scores (Fig 3D). Additionally, to assess the predictive efficacy of our hypoxia model, time-dependent ROC curves for OS were generated. The area under the curve (AUC) values were 0.774 at 1 year, 0.727 at 2 years, and 0.711 at 3 years (Fig 3E). These values were significantly superior to those derived from clinicopathologic characteristics alone (Fig 3F), underscoring the model's excellent predictive capability. Decision curve analysis confirmed these findings, demonstrating that clinical interventions guided by the hypoxia model yielded greater benefits compared to those based solely on clinicopathologic characteristics (Fig 3G). Similar results were also observed in external validation cohorts, including PACA-CA (S1 Fig) and PACA-AU (S2 Fig).

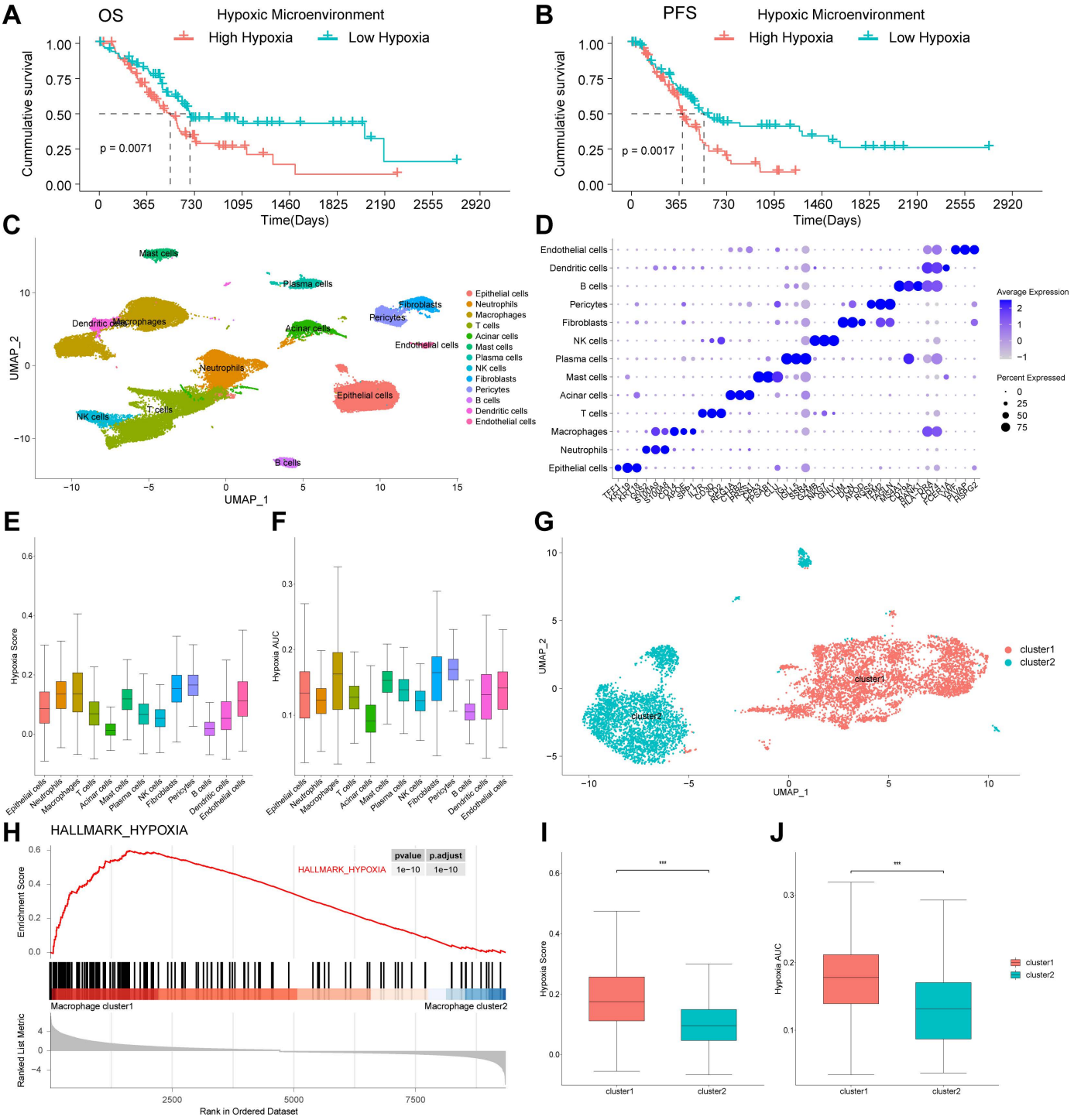

Fig 2. Single-cell RNA-sequencing analysis identified hypoxia-responsive macrophage subcluster. (A, B) Prognostic differences of the tumor hypoxic microenvironment in TCGA-PAAD cohort. **(C)** UMAP plot of 13 cell types in PDAC samples. **(D)** Key cell markers used to identify cell types. **(E, F)** Hypoxic microenvironment scores calculated by the AddModuleScore and AUCell function. **(G)** UMAP plot of macrophage subclusters. **(H)** GSEA showed that Hypoxia pathways was activated in macrophage cluster 1. **(I, J)** Hypoxic microenvironment scores calculated by the AddModuleScore and AUCell function in macrophage subclusters. Data sources: Panels A and B use the TCGA-PAAD cohort. Panels C-J use the dataset GSE155698.

https://doi.org/10.1371/journal.pone.0322618.g002

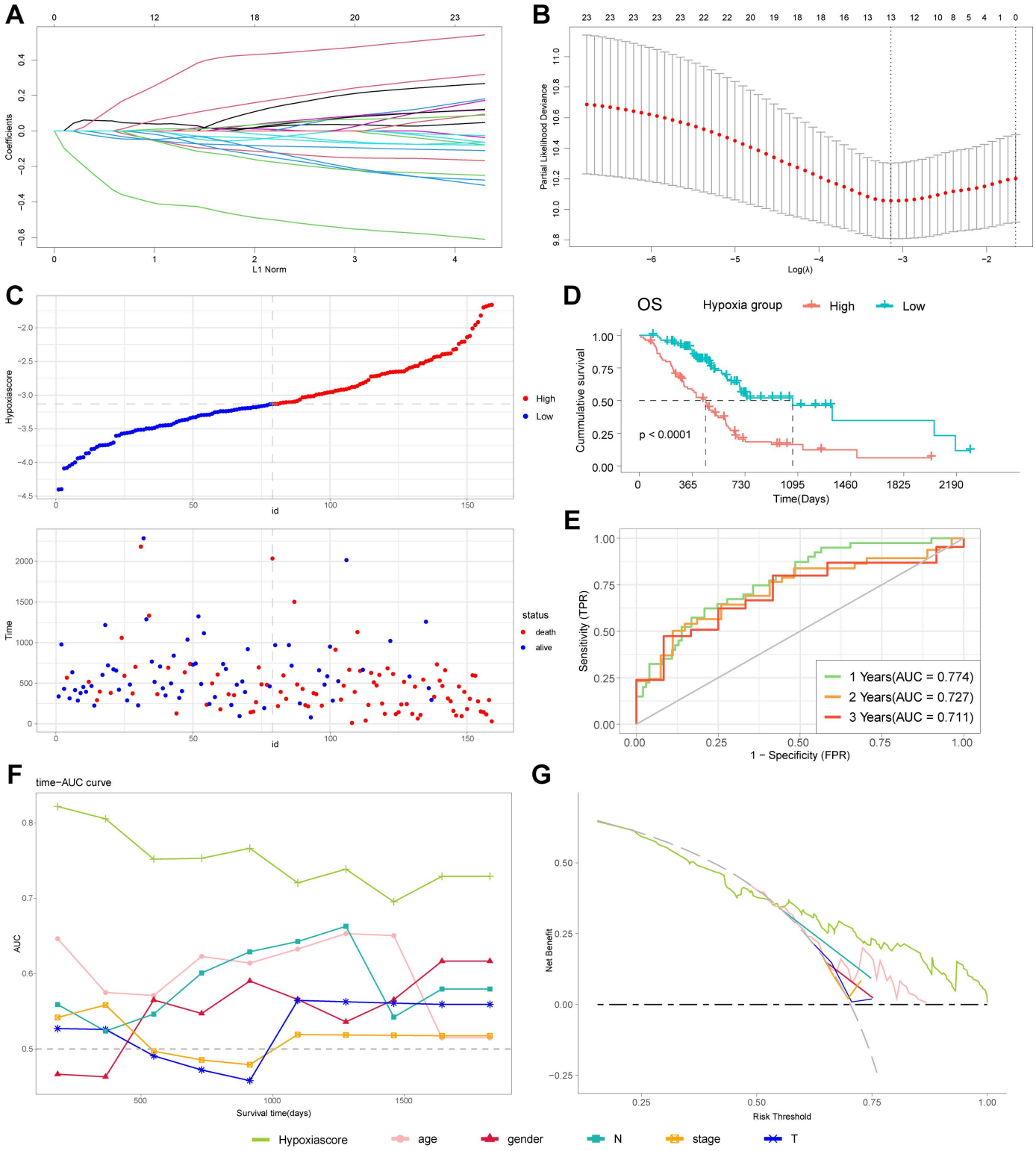

Fig 3. Construction of hypoxia-related prognostic model in TCGA cohort. (A, B) LASSO regression of hypoxia-related genes. **(C)** Relationship between survival status and hypoxia scores in TCGA cohort. **(D)** Kaplan−Meier curves of patients in high and low hypoxia groups. **(E)** ROC curves of hypoxia model for predicting the risk of death at 1, 2, and 3 years. **(F)** Time-AUC curves evaluating the predictive capacity of the hypoxia model and

clinicopathologic features. **(G)** Decision curve analysis evaluating the benefit rate of patients receiving clinical treatment based on the hypoxia model and clinicopathologic features. Data sources: Panels A-G use the TCGA-PAAD cohort.

https://doi.org/10.1371/journal.pone.0322618.g003

Analysis of clinicopathologic characteristics and nomogram construction

To ascertain whether the hypoxia score could independently influence prognosis, we collected clinicopathologic data from the TCGA-PAAD cohort. Multivariate Cox regression analysis confirmed that the hypoxia score was an independent prognostic factor for patients with pancreatic cancer (Fig 4A and S3 Fig). We compared clinicopathologic characteristics between the high and low hypoxia groups; a heatmap analysis indicated comparable features across both groups (Fig 4B). This suggested that the prognostic capability of the hypoxia model operated independently of clinicopathologic characteristics. Further analysis assessed the differences in hypoxia scores across various clinicopathologic factors; notably, higher N stages were associated with increased hypoxia scores (Fig 4C and S3 Fig). Additionally, stratified analysis based on clinicopathologic characteristics such as age<=65, age > 65, Male, T3-T4, N0, N1, Stage I-II, Stage III-IV demonstrated significant differences in survival outcomes between the high and low hypoxia groups (S4 Fig).

Finally, we developed a novel nomogram that integrated clinicopathologic features and hypoxia scores (Fig 4D). This nomogram exhibited strong predictive power for survival outcomes, as evidenced by the calibration curve (Fig 4E). Additionally, we conducted decision curve analysis to evaluate the nomogram's predictive efficacy, and the results demonstrated that the benefit rate for patients treated based on the nomogram exceeded those who were treated based on other clinicopathologic features alone (Fig 4F).

Gene set enrichment analysis

To elucidate the molecular mechanisms underlying the differences in cancer progression between high and low hypoxia groups, we performed Gene Set Enrichment Analysis (GSEA). The analysis revealed that the high hypoxia group exhibited significant enrichment in several critical pathways, including biosynthesis of amino acids, cell cycle, DNA replication, nucleocytoplasmic transport, nucleotide metabolism, protein processing in the endoplasmic reticulum, ribosome, E2F targets, G2M checkpoint, hypoxia and mitotic spindle (Fig 5A and 5B).

Mutation landscape analysis

Gene mutations play a critical role in tumor development and resistance to therapy. Analysis of single-nucleotide variants (SNV) indicated that missense mutations were the most common type in both the high and low hypoxia groups. Despite the top five most frequently mutated genes - *KRAS*, *TP53*, *CDKN2A*, *SMAD4* and *TTN* - being identical between the groups, the overall mutation rate was significantly higher in the high hypoxia group (90.28%), compared to the low hypoxia group (78.87%), as shown in Fig 5C and 5D. Further investigations revealed a higher incidence of copy number variation (CNV) gains in the high hypoxia group (Fig 5E). Additionally, the high hypoxia group harbored a remarkably greater tumor mutational burden (TMB) than the low hypoxia group (Fig 5F), and the hypoxia score was positively correlated with TMB, with a correlation coefficient of 0.28 and a statistically significant level ($P < 0.001$), as illustrated in Fig 5G. It is widely accepted that TMB is associated with poor prognosis, which may be the reason for the poor prognosis in the high hypoxia group.

Unveiling immune cell infiltration in hypoxia model

The immune microenvironment plays a crucial role in the prognosis of pancreatic cancer patients. We explored the association between hypoxia scores and immune cell infiltration using the ESTIMATE algorithm, which assesses the level of immune cell presence within tumors. Our analysis indicated that the high hypoxia group exhibited higher tumor purity and lower immune, stromal, and ESTIMATE scores (Fig 6A), suggesting a negative correlation between hypoxia scores and

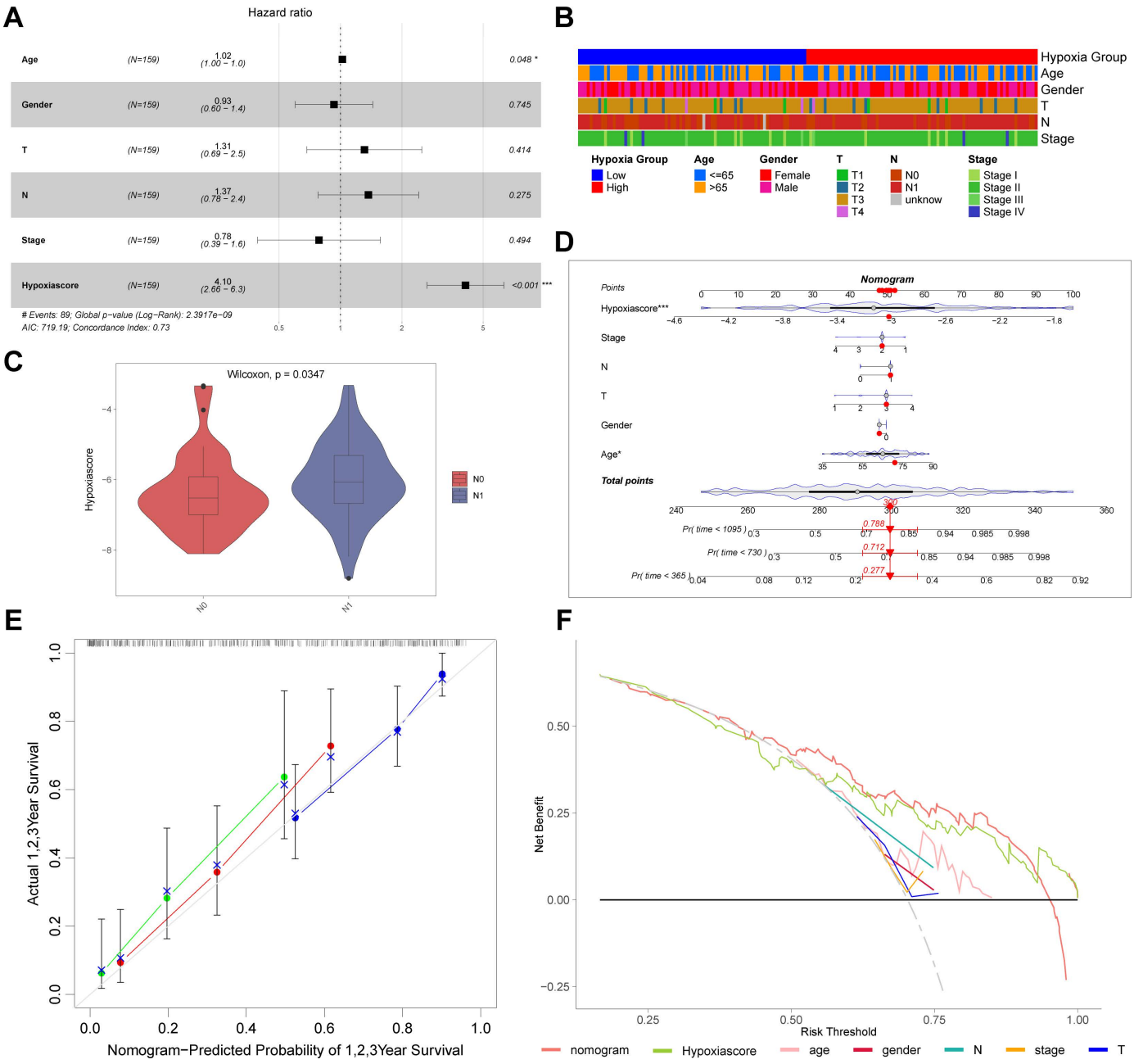

Fig 4. Clinicopathologic features of hypoxia model and construction of prognostic nomogram. (A) Multivariate Cox regression analysis of hypoxia model and clinicopathologic characteristics. **(B)** Heatmap of hypoxia groups and clinicopathologic characteristics. **(C)** Relationship between hypoxia score and N stage of patients. **(D)** Construction of the nomogram integrating the hypoxia model and clinicopathologic characteristics. **(E)** Calibration curves for 1, 2, and 3 years of nomogram. **(F)** Decision curve analysis evaluating the benefit rate of patients receiving clinical treatment based on the nomogram. Data sources: Panels A-F use the TCGA-PAAD cohort.

https://doi.org/10.1371/journal.pone.0322618.g004

immune cell infiltration. Moreover, the CIBERSORT algorithm was employed to further analyze the composition of immune cells in the TCGA-PAAD cohort. The findings demonstrated a reduced presence of anti-tumor immune cells such as naive B cells in the high hypoxia group, while macrophages M0 were predominantly enriched (Fig 6B), indicating that the

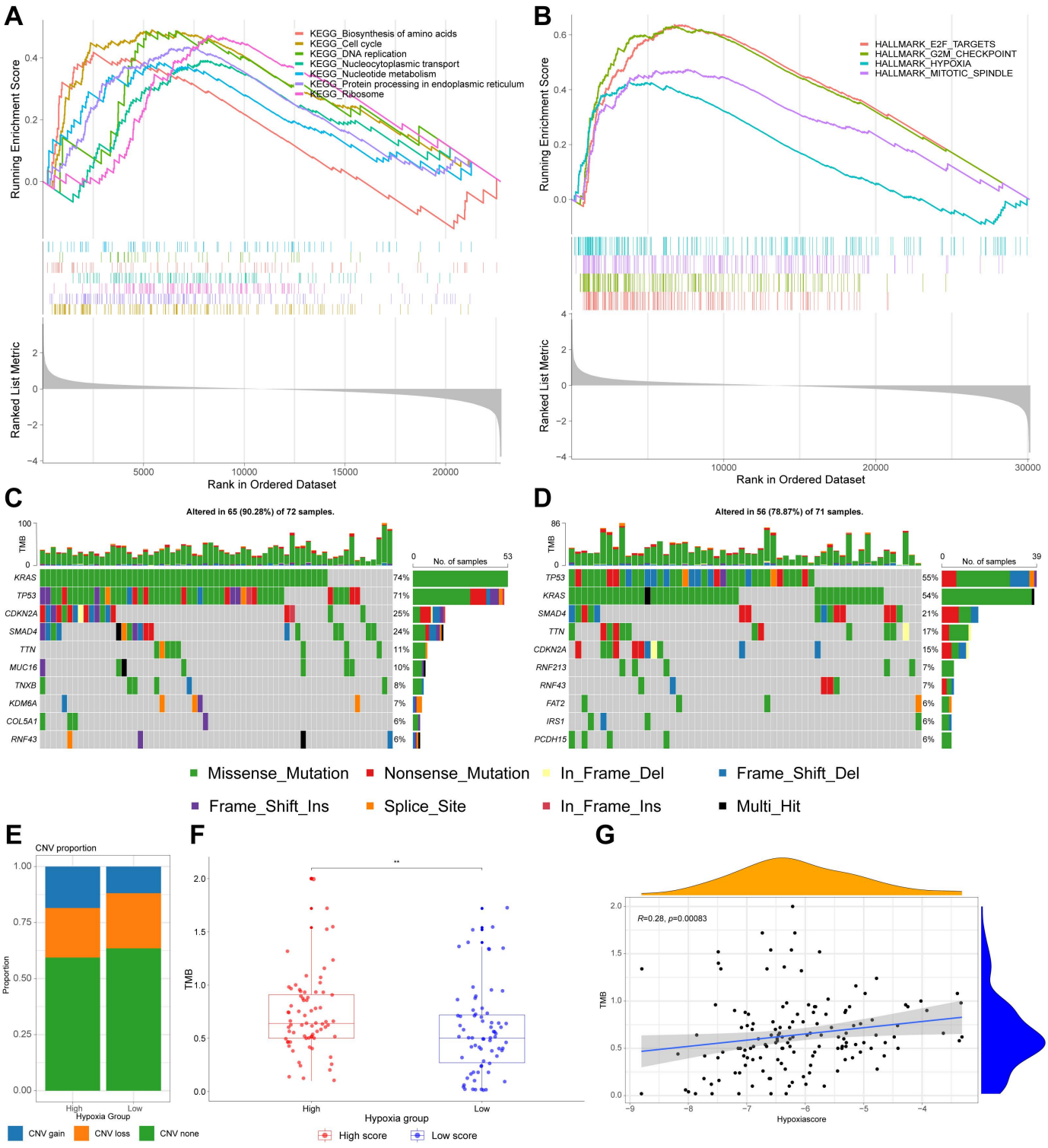

Fig 5. Molecular features of high and low hypoxia groups. (A, B) GSEA analysis revealed enriched signaling pathways in high hypoxia group. **(C, D)** Waterfall diagrams displaying SNV mutations of high and low hypoxia groups. **(E)** Differential proportion of CNV mutations in high and low hypoxia groups. **(F)** The comparison of TMB between high and low hypoxia groups. **(G)** The Pearson correlation analysis of TMB and hypoxia scores. Data sources: Panels A-G use the TCGA-PAAD cohort.

https://doi.org/10.1371/journal.pone.0322618.g005

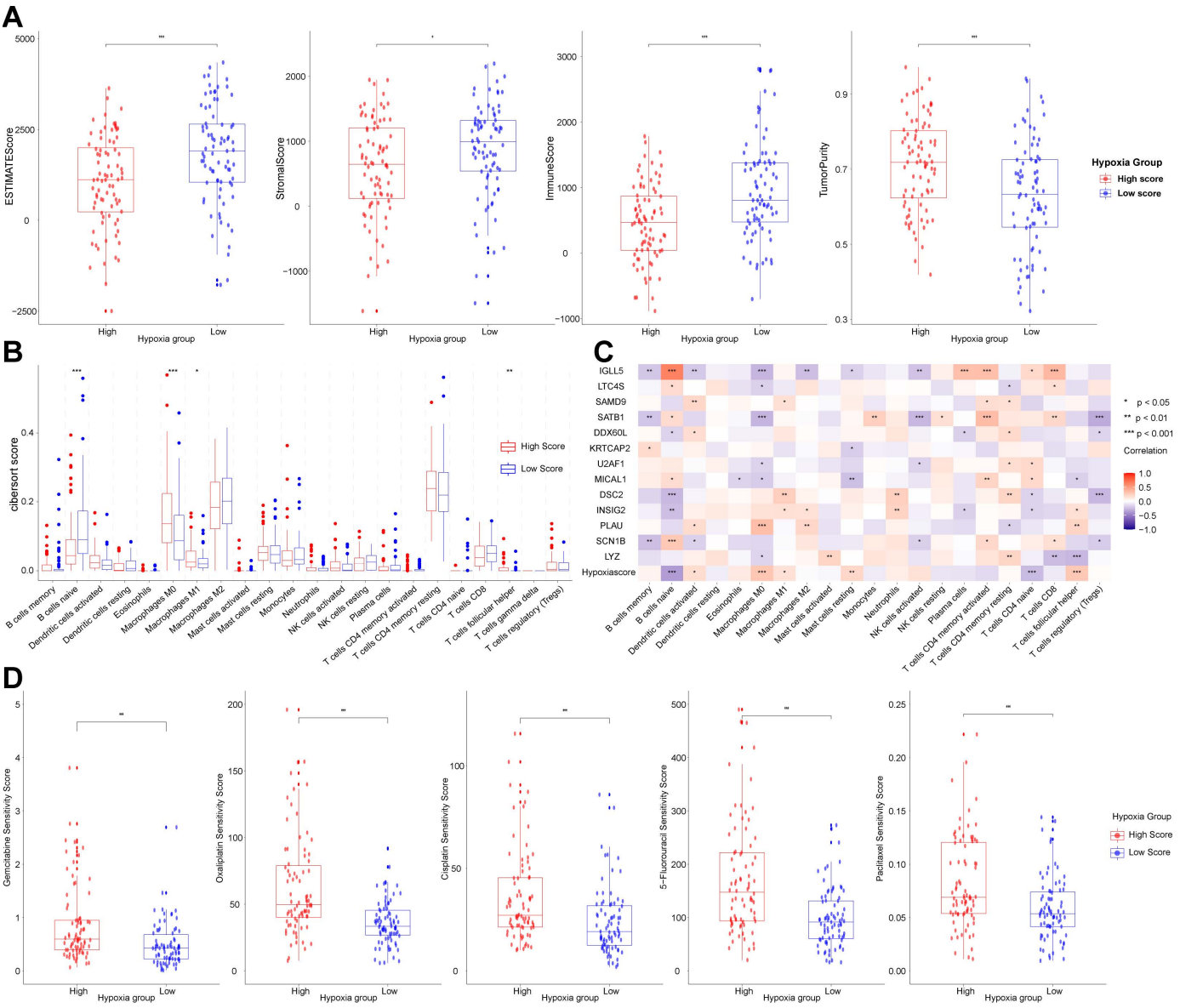

Fig 6. Relationship of hypoxia score with immune cell infiltration and chemotherapy sensitivity. (A) Differences among immune score, stromal score, ESTIMATE score, and tumor purity between different hypoxia groups. **(B)** Differential proportion of 22 immune cell subtypes between high and low hypoxia groups by the CIBERSORT algorithm. **(C)** Correlations between hypoxia-related genes and 22 immune cell subtypes. **(D)** Correlation analysis between hypoxia score and chemotherapy sensitivity. Data sources: Panels A-C use the TCGA-PAAD cohort. Panel D uses TCGA-PAAD cohort and GDSC dataset.

https://doi.org/10.1371/journal.pone.0322618.g006

hypoxia score was closely associated with an immunosuppressive tumor microenvironment. Correlation analysis between hypoxia-related genes and immune cell infiltration levels further substantiated these observations (Fig 6C).

Drug sensitivity analysis

We further explored the correlation between hypoxia score and chemotherapy sensitivity in pancreatic cancer. Utilizing the GDSC database, we assessed the response to chemotherapeutic agents. Notably, most drugs exhibited significant

differential responses between the high and low hypoxia groups, including key agents such as gemcitabine, oxaliplatin, cisplatin, 5-Fluorouracil and paclitaxel (Fig 6D). The elevated half-maximal inhibitory concentrations (IC50) of these drugs in the high hypoxia group suggested a diminished chemotherapy efficacy, which may contribute to the observed poorer prognosis in this group.

Pan-cancer analysis

We extended our analysis of hypoxia-related genes to a pan-cancer analysis. These genes were found to be significantly associated with various survival metrics across multiple cancer types, including overall survival (OS) (Fig 7A), disease specific survival (DSS) (Fig 7B), disease free survival (DFS) (Fig 7C), and progression free survival (PFS) (Fig 7D). Among these, *KRTCAP2* consistently emerged as a prognostic marker, with its expression linked to unfavorable outcomes across nearly all investigated cancer types, evidenced by a hazard ratio (HR) greater than 1. We further examined *KRTCAP2* expression levels in cancer tissues, compared to adjacent normal tissues, and across different stages of cancer progression. The results indicated a significant overexpression of *KRTCAP2* in cancer tissues across most cancer types (Fig 7E), with its abundant expression escalating in higher cancer stages (Fig 7F). Additionally, we also assessed the correlation between *KRTCAP2* expression and immune cell infiltration across various tumors and found *KRTCAP2* expression exhibited a negative correlation with several immune cell populations, including T cells gamma delta, $CD8^+$ T cells, $CD4^+$ memory activated T cells, neutrophils, monocytes, resting mast cells, M1 macrophages, and activated dendritic cells, across multiple tumor types. Conversely, a positive correlation was observed between *KRTCAP2* expression and the immunosuppressive Treg cells (Fig 7G), suggesting that *KRTCAP2* played a significant role in modulating the tumor immune microenvironment across diverse cancer types.

Discussion

The incidence of pancreatic cancer has been steadily increasing, with projections indicating that this upward trend will continue for the foreseeable future in both the United States and Europe [19]. Rahib *et al.* have projected that PDAC will become the second-leading cause of cancer-related mortality in the United States by 2030, despite rapid advancements in cancer therapeutics [20]. Therefore, identifying the factors contributing to pancreatic cancer drug resistance is imperative. Current literature highlights that pancreatic cancer is notably hypoxic compared to other cancer types, with the hypoxic microenvironment significantly enhancing drug resistance [21–23]. Despite these insights, current research has not yet revealed the specific cell types within the pancreatic cancer microenvironment that are most likely influenced by the local hypoxic microenvironment.

In this study, we observed that macrophages are disproportionately affected by hypoxic conditions compared to other immune cell types. Macrophages are indispensable components of the tumor microenvironment and play pivotal roles in the progression, metastasis, and therapeutic resistance of PDAC, thus posing significant challenges to developing effective treatment strategies targeting this deadly cancer [24,25]. Furthermore, the hypoxic microenvironment is known to promote macrophage polarization and their enrichment within tumors [26]. Our subgroup analysis further identified macrophage cluster1 as being particularly sensitive to tumor hypoxic microenvironment.

We conducted differential gene expression analysis on macrophage cluster 1 and combined COX and Lasso regression analysis to construct a novel hypoxia-related prognostic model, which contained 13 genes: *LYZ*, *SCN1B*, *PLAU*, *INSIG2*, *DSC2*, *MICAL1*, *U2AF1*, *KRTCAP2*, *DDX60L*, *SATB1*, *SAMD9*, *LTC4S* and *IGLL5*. Our hypoxia model shares common genes, such as *PLAU*, with other hypoxia-related prognostic models [27–29]. However, unlike other models constructed from bulk RNA data, our model was developed using single-cell data. This approach allowed us to identify the cell subpopulations with the highest hypoxia responsiveness in pancreatic cancer, providing a more precise localization of hypoxia-driven effects.

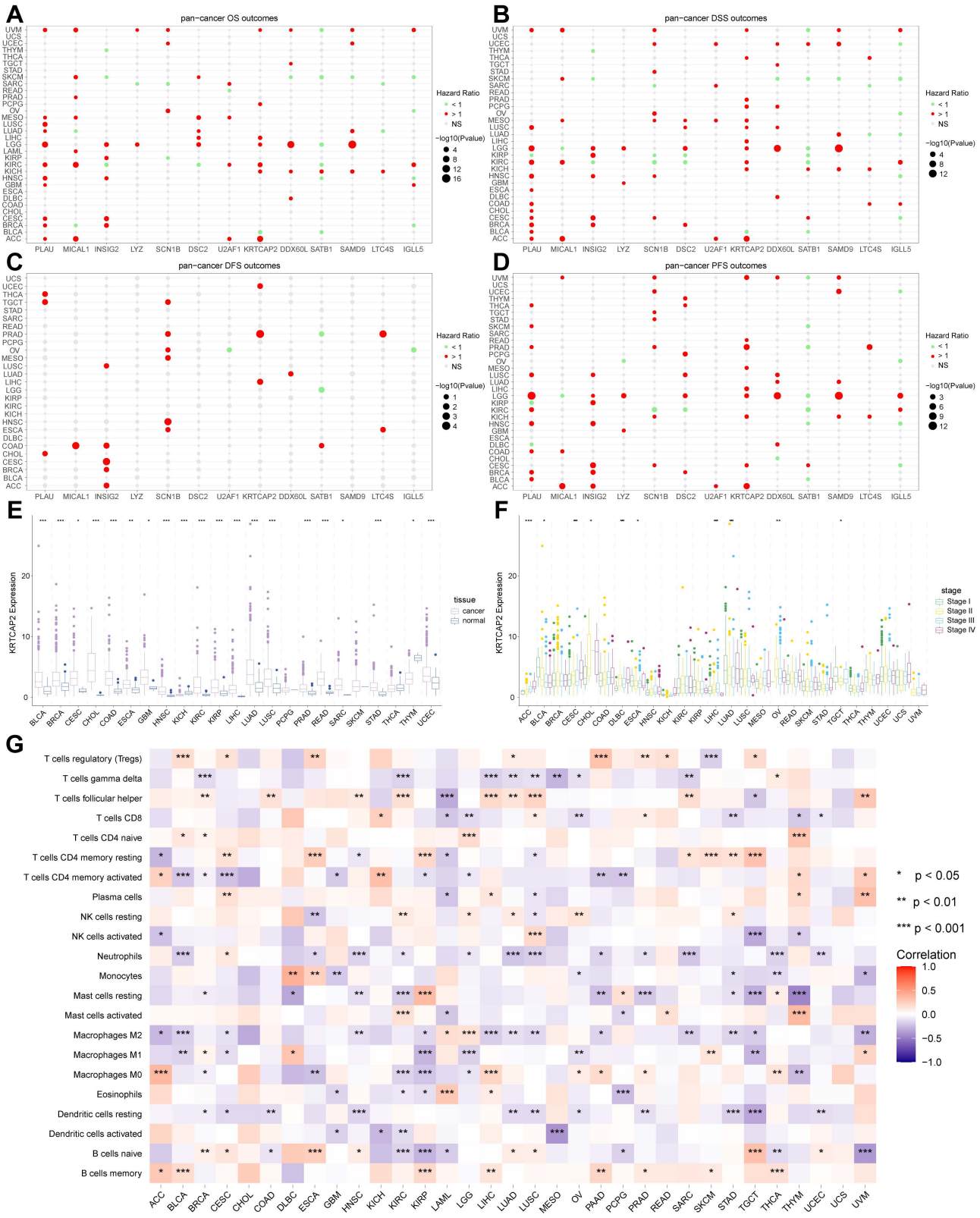
A
pan−cancer OS outcomes
B
pan−cancer DSS outcomes
C
pan−cancer DFS outcomes
D
pan−cancer PFS outcomes
PLAU MICAL1 INSIG2 LYZ SCN1B DSC2 U2AF1 KRTCAP2 DDX60L SATB1 SAMD9 LTC4S IGLL5
Hazard Ratio
< 1
> 1
NS
−log10(Pvalue)
E
KRTCAP2 Expression
tissue
cancer
normal
F
stage
Stage I
Stage II
Stage III
Stage IV
G
T cells regulatory (Tregs)
T cells gamma delta
T cells follicular helper
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
Plasma cells
NK cells resting
NK cells activated
Neutrophils
Monocytes
Mast cells resting
Mast cells activated
Macrophages M2
Macrophages M1
Macrophages M0
Eosinophils
Dendritic cells resting
Dendritic cells activated
B cells naive
B cells memory
* p < 0.05
** p < 0.01
*** p < 0.001
Correlation

Fig 7. Pan-cancer analysis of hypoxia-related genes. (A-D) Relationship between hypoxia-related genes with OS **(A)**, DSS **(B)**, DFS (C) and PFS (D) across multiple cancer types. **(E)** Differences in *KRTCAP2* expression levels between cancer and adjacent normal tissues across various tumor types. **(F)** Differences in *KRTCAP2* expression levels between different tumor stages across various tumor types. **(G)** The relationship between *KRTCAP2* and immune cell infiltration across various tumor types. Data sources: Panels A-G use the TCGA dataset.

https://doi.org/10.1371/journal.pone.0322618.g007

Based on our model, we calculated hypoxia score for each patient in the TCGA-PAAD cohort. Patients were then stratified into high and low hypoxia groups based on the median score. Our analysis revealed that patients in the high hypoxia group exhibited significantly worse prognostic outcomes compared to those in the low hypoxia group. Moreover, our hypoxia model demonstrated superior predictive performance for patient survival at 1, 2, and 3 years when compared to clinicopathologic characteristics. By integrating hypoxia scores with clinicopathologic features, our findings confirmed that the hypoxia score served as an independent prognostic factor. Interestingly, there were no significant differences in clinicopathologic features between the high and low hypoxia score groups. Moreover, within most subgroups defined by clinicopathologic characteristics, patients with high hypoxia scores generally exhibited worse prognoses compared to those with low scores. These findings underscored the predictive efficacy of the hypoxia model and was entirely independent of clinicopathologic features.

To further elucidate the relationship between the hypoxia microenvironment and tumor-infiltrating immune cells in pancreatic cancer, we employed several algorithms to assess immune infiltration states [30–32]. Our findings indicated a significant reduction in naïve B cells, which are well known for their anti-tumor immune responses, in patients with high hypoxia scores. This reduction potentially facilitated tumor immune evasion and progression, contributing to the substantially decreased survival rates observed in the high hypoxia group. Similar findings have been reported in studies of kidney renal clear cell carcinoma (KIRC), where cuproptosis signatures are closely associated with tumor immune cell infiltration and responses to immunotherapy, serving as a potential biomarker for predicting patient prognosis [33].

Pan-cancer analysis has emerged as a powerful approach to uncover biomarkers and therapeutic targets that play critical roles across multiple cancer types, providing insights into shared molecular mechanisms and prognostic factors [34–36]. In our study, pan-cancer analysis had further demonstrated that these 13 hypoxia-related genes, particularly the *KRTCAP2* gene, were not only significantly associated with the prognosis of pancreatic cancer but also played critical roles across various malignancies. Keratinocyte-associated protein 2 (*KRTCAP2*) is involved in N-glycosylation processes. Previous research has shown that elevated levels of KRTCAP2 protein are associated with reduced infiltration of $CD8^+$ T cells and $CD68^+$ macrophages in hepatocellular carcinoma tissues [37]. Additionally, *KRTCAP2* has been shown to regulate tumor cell proliferation, differentiation, and carcinogenesis in gastric cancer [38]. Elevated expression of *KRTCAP2* has also been associated with unfavorable prognosis in uveal melanoma, further underscoring its role as a prognostic marker [39]. Similar observations have been made with other pan-cancer biomarkers [40,41], where individual gene can demonstrate significant prognostic value and associations with immune cell infiltration across multiple cancer types. Moreover, when comparing the gene expression levels of *KRTCAP2* between tumor tissues and adjacent normal tissues, as well as across different cancer stages, it was found that *KRTCAP2* was consistently overexpressed in tumor tissues compared to adjacent normal tissues across a broad spectrum of cancers. Its expression also escalated with advancing tumor stages. Based on these findings, we propose that *KRTCAP2* serves as a potential tumor biomarker and represents a promising target for therapeutic intervention.

Despite the promising findings obtained, our study also has some limitations that need to be addressed. Firstly, additional investigations into the protein expression levels of hypoxia-related genes are necessary to complement our findings. Secondly, our study lacks supporting cellular and animal experiments to confirm the regulatory mechanisms in pancreatic

cancer. Therefore, more comprehensive research is essential to thoroughly assess the potential and applicability of our hypoxia model in future studies.

In conclusion, we developed a thirteen-gene hypoxia-related prognostic model for pancreatic cancer by integrating single-cell RNA sequencing and bulk RNA sequencing data. This prognostic signature not only served as an independent factor for predicting patient outcomes, but also offered important insights into immune cell infiltration and chemotherapy response. Further, it holds promise for identifying potential new biomarkers and therapeutic targets in pancreatic cancer.

Supporting information

S1 Fig. Validation of the hypoxia-related prognostic model in PACA-CA cohort. (A) Relationship between survival status and hypoxia score in PACA-CA cohort. (B) Kaplan−Meier curves of patients in high and low hypoxia groups. (C) ROC curves of hypoxia model for predicting the risk of death at 1, 2, and 3 years. (D) Time-AUC curves evaluating the predictive capacity of the hypoxia model and clinicopathologic features. (E) Decision curve analysis evaluating the benefit rate of patients receiving clinical treatment based on the hypoxia model and clinicopathologic features. Data sources: Panels A-E use the PACA-CA cohort.
(TIF)

S2 Fig. Validation of the hypoxia-related prognostic model in PACA-AU cohort. (A) Relationship between survival status and hypoxia score in PACA-AU cohort. (B) Kaplan−Meier curves of patients in high and low hypoxia groups. (C) ROC curves of hypoxia model for predicting the risk of death at 1, 2, and 3 years. (D) Time-AUC curves evaluating the predictive capacity of the hypoxia model and clinicopathologic features. (E) Decision curve analysis evaluating the benefit rate of patients receiving clinical treatment based on the hypoxia model and clinicopathologic features. Data sources: Panels A-E use the PACA-AU cohort.
(TIF)

S3 Fig. Clinicopathologic features of hypoxia model. (A, B) Multivariate Cox regression analysis of hypoxia model and clinicopathologic characteristics in PACA-CA (A) and PACA-AU (B) cohort. (C-F) Relationship between hypoxia score with age (C), gender (D), TNM stage (E) and T stage (F) of patients in TCGA-PAAD cohort. Data sources: Panel A use the PACA-CA cohort. Panel B use the PACA-AU cohort. Panels C-F use the TCGA-PAAD cohort.
(TIF)

S4 Fig. The prognostic value of hypoxia model in age (A, B), gender (C, D), T stage (E, F), N stage (G, H) and TNM stage (I, J), respectively. Data sources: Panels A-J use the TCGA-PAAD cohort.
(TIF)

S1 Table. The coefficient of each gene in the hypoxia model.
(XLSX)

Author contributions

Conceptualization: Heming Ge, Cenap Güngör.

Formal analysis: Heming Ge, Gerrit Wolters-Eisfeld, Thilo Hackert, Yuqiang Li.

Methodology: Heming Ge, Gerrit Wolters-Eisfeld, Yuqiang Li.

Supervision: Cenap Güngör.

Writing – original draft: Heming Ge.

Writing – review & editing: Gerrit Wolters-Eisfeld, Thilo Hackert, Cenap Güngör.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49. https://doi.org/10.3322/caac.21820 PMID: 38230766
2. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. Lancet. 2011;378(9791):607–20. https://doi.org/10.1016/S0140-6736(10)62307-0 PMID: 21620466
3. Chen L, Zhang X, Zhang Q, Zhang T, Xie J, Wei W, et al. A necroptosis related prognostic model of pancreatic cancer based on single cell sequencing analysis and transcriptome analysis. Front Immunol. 2022;13:1022420. https://doi.org/10.3389/fimmu.2022.1022420 PMID: 36275722
4. Sonkin D, Thomas A, Teicher BA. Cancer treatments: past, present, and future. Cancer Genet. 2024;286–287:18–24. https://doi.org/10.1016/j.cancergen.2024.06.002 PMID: 38909530
5. Yoon A-R, Hong J, Jung B-K, Ahn HM, Zhang S, Yun C-O. Oncolytic adenovirus as pancreatic cancer-targeted therapy: where do we go from here? Cancer Lett. 2023;579:216456. https://doi.org/10.1016/j.canlet.2023.216456 PMID: 37940067
6. Choudhry H, Harris AL. Advances in hypoxia-inducible factor biology. Cell Metab. 2018;27(2):281–98. https://doi.org/10.1016/j.cmet.2017.10.005 PMID: 29129785
7. Qiao Y, Jiang X, Li Y, Wang K, Chen R, Liu J, et al. Identification of a hypoxia-related gene prognostic signature in colorectal cancer based on bulk and single-cell RNA-seq. Sci Rep. 2023;13(1):2503. https://doi.org/10.1038/s41598-023-29718-2 PMID: 36781976
8. Qin Y, Zhu W, Xu W, Zhang B, Shi S, Ji S, et al. LSD1 sustains pancreatic cancer growth via maintaining HIF1α-dependent glycolytic process. Cancer Lett. 2014;347(2):225–32. https://doi.org/10.1016/j.canlet.2014.02.013 PMID: 24561118
9. Caronni N, La Terza F, Vittoria FM, Barbiera G, Mezzanzanica L, Cuzzola V, et al. IL-1β+ macrophages fuel pathogenic inflammation in pancreatic cancer. Nature. 2023;623(7986):415–22. https://doi.org/10.1038/s41586-023-06685-2 PMID: 37914939
10. Liu M, Ren Y, Zhou Z, Yang J, Shi X, Cai Y, et al. The crosstalk between macrophages and cancer cells potentiates pancreatic cancer cachexia. Cancer Cell. 2024;42(5):885–903 e4. https://doi.org/10.1016/j.ccell.2024.03.009 PMID:38608702
11. Bai R, Li Y, Jian L, Yang Y, Zhao L, Wei M. The hypoxia-driven crosstalk between tumor and tumor-associated macrophages: mechanisms and clinical treatment strategies. Mol Cancer. 2022;21(1):177. https://doi.org/10.1186/s12943-022-01645-2 PMID:36071472
12. Steele NG, Carpenter ES, Kemp SB, Sirihorachai VR, The S, Delrosario L, et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. Nat Cancer. 2020;1(11):1097–112. https://doi.org/10.1038/s43018-020-00121-4 PMID: 34296197
13. Chi H, Yang J, Peng G, Zhang J, Song G, Xie X, et al. Circadian rhythm-related genes index: a predictor for HNSCC prognosis, immunotherapy efficacy, and chemosensitivity. Front Immunol. 2023;14:1091218. https://doi.org/10.3389/fimmu.2023.1091218 PMID: 36969232
14. Liu H, Dong A, Rasteh AM, Wang P, Weng J. Identification of the novel exhausted T cell CD8 + markers in breast cancer. Sci Rep. 2024;14(1):19142. https://doi.org/10.1038/s41598-024-70184-1 PMID: 39160211
15. Liu H, Tang T. MAPK signaling pathway-based glioma subtypes, machine-learning risk model, and key hub proteins identification. Sci Rep. 2023;13(1):19055. https://doi.org/10.1038/s41598-023-45774-0 PMID: 37925483
16. Liu H, Tang T. A bioinformatic study of IGFBPs in glioma regarding their diagnostic, prognostic, and therapeutic prediction value. Am J Transl Res. 2023;15(3):2140–55. PMID: 37056850
17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453–7. https://doi.org/10.1038/nmeth.3337 PMID: 25822800
18. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–25. https://doi.org/10.1016/j.cels.2015.12.004 PMID: 26771021
19. Del Chiaro M, Sugawara T, Karam SD, Messersmith WA. Advances in the management of pancreatic cancer. BMJ. 2023;383:e073995. https://doi.org/10.1136/bmj-2022-073995 PMID: 38164628
20. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014;74(11):2913–21. https://doi.org/10.1158/0008-5472.CAN-14-0155 PMID: 24840647
21. Tan Z, Xu J, Zhang B, Shi S, Yu X, Liang C. Hypoxia: a barricade to conquer the pancreatic cancer. Cell Mol Life Sci. 2020;77(16):3077–83. https://doi.org/10.1007/s00018-019-03444-3 PMID: 31907561
22. McCarroll JA, Naim S, Sharbeen G, Russia N, Lee J, Kavallaris M, et al. Role of pancreatic stellate cells in chemoresistance in pancreatic cancer. Front Physiol. 2014;5:141. https://doi.org/10.3389/fphys.2014.00141 PMID: 24782785
23. Zhang H, Cao K, Xiang J, Zhang M, Zhu M, Xi Q. Hypoxia induces immunosuppression, metastasis and drug resistance in pancreatic cancers. Cancer Lett. 2023;571:216345. https://doi.org/10.1016/j.canlet.2023.216345 PMID: 37558084
24. Li M, Yang Y, Xiong L, Jiang P, Wang J, Li C. Metabolism, metabolites, and macrophages in cancer. J Hematol Oncol. 2023;16(1):80. https://doi.org/10.1186/s13045-023-01478-6 PMID:37491279
25. Ryu S, Lee EK. The pivotal role of macrophages in the pathogenesis of pancreatic diseases. Int J Mol Sci. 2024;25(11):5765. https://doi.org/10.3390/ijms25115765 PMID: 38891952

26. Park JE, Dutta B, Tse SW, Gupta N, Tan CF, Low JK, et al. Hypoxia-induced tumor exosomes promote M2-like macrophage polarization of infiltrating myeloid cells and microRNA-mediated metabolic shift. Oncogene. 2019;38(26):5158–73. https://doi.org/10.1038/s41388-019-0782-x PMID:30872795

27. Wu Y, Kou Q, Sun L, Hu X. Effects of anoxic prognostic model on immune microenvironment in pancreatic cancer. Sci Rep. 2023;13(1):9104. https://doi.org/10.1038/s41598-023-36413-9 PMID: 37277450

28. Wu Y, Zhou J, Kou Q, Sun L, Ma Y, Yang T, et al. Establishment of a prognostic model for pancreatic cancer based on hypoxia-related genes. Technol Cancer Res Treat. 2024;23:15330338241288687. https://doi.org/10.1177/15330338241288687 PMID: 39431298

29. Ren M, Feng L, Zong R, Sun H. Novel prognostic gene signature for pancreatic ductal adenocarcinoma based on hypoxia. World J Surg Oncol. 2023;21(1):257. https://doi.org/10.1186/s12957-023-03142-2 PMID: 37605192

30. Liu H, Weng J. A comprehensive bioinformatic analysis of cyclin-dependent kinase 2 (CDK2) in glioma. Gene. 2022;822:146325. https://doi.org/10.1016/j.gene.2022.146325 PMID: 35183683

31. Liu H, Li Y. Potential roles of Cornichon Family AMPA receptor auxiliary protein 4 (CNIH4) in head and neck squamous cell carcinoma. Cancer Biomark. 2022;35(4):439–50. https://doi.org/10.3233/CBM-220143 PMID:36404537

32. Li Y, Liu H. Clinical powers of Aminoacyl tRNA Synthetase Complex Interacting Multifunctional Protein 1 (AIMP1) for head-neck squamous cell carcinoma. Cancer Biomark. 2022;34(3):359–74. https://doi.org/10.3233/CBM-210340 PMID: 35068446

33. Liu H. Expression and potential immune involvement of cuproptosis in kidney renal clear cell carcinoma. Cancer Genet. 2023;274–275:21–5. https://doi.org/10.1016/j.cancergen.2023.03.002 PMID: 36963335

34. Liu H, Weng J, Huang CL, Jackson AP. Voltage-gated sodium channels in cancers. Biomark Res. 2024;12(1):70. https://doi.org/10.1186/s40364-024-00620-x PMID:39060933

35. Liu H, Tang T. Pan-cancer genetic analysis of disulfidptosis-related gene set. Cancer Genet. 2023;278–279:91–103. https://doi.org/10.1016/j.cancergen.2023.10.001 PMID: 37879141

36. Liu H, Tang T. Pan-cancer genetic analysis of cuproptosis and copper metabolism-related gene set. Front Oncol. 2022;12:952290. https://doi.org/10.3389/fonc.2022.952290 PMID: 36276096

37. Sun P, Zhang H, Shi J, Xu M, Cheng T, Lu B, et al. KRTCAP2 as an immunological and prognostic biomarker of hepatocellular carcinoma. Colloids Surf B Biointerfaces. 2023;222:113124. https://doi.org/10.1016/j.colsurfb.2023.113124 PMID: 36634487

38. Lee S, Yang H-K, Lee H-J, Park DJ, Kong S-H, Park SK. Systematic review of gastric cancer-associated genetic variants, gene-based meta-analysis, and gene-level functional analysis to identify candidate genes for drug development. Front Genet. 2022;13:928783. https://doi.org/10.3389/fgene.2022.928783 PMID: 36081994

39. Liu J, Lu J, Li W. A comprehensive prognostic and immunological analysis of a six-gene signature associated with glycolysis and immune response in uveal melanoma. Front Immunol. 2021;12:738068. https://doi.org/10.3389/fimmu.2021.738068 PMID: 34630418

40. Liu H, Weng J. A pan-cancer bioinformatic analysis of RAD51 regarding the values for diagnosis, prognosis, and therapeutic prediction. Front Oncol. 2022;12:858756. https://doi.org/10.3389/fonc.2022.858756 PMID: 35359409

41. Liu H, Dilger JP, Lin J. A pan-cancer-bioinformatic-based literature review of TRPM7 in cancers. Pharmacol Ther. 2022;240:108302. https://doi.org/10.1016/j.pharmthera.2022.108302 PMID:36332746

3. Summary

Spatial transcriptomics and mendelian randomization decipher the microenvironment in colorectal cancer:

i) Spatial transcriptomics analysis revealed that within the tumor thrombus associated microenvironment of colorectal cancer, despite the increased infiltration of $CD8^+$ T cells and macrophages, these immune populations exhibit significant dysfunction. $CD4^+$ T cells inhibit $CD8^+$ T cell activity, concomitant with enrichment of immunosuppressive subsets and impaired antigen-presenting capacity of cDCs.

ii) Large-scale genome-wide association study revealed robust causal relationships of five distinct microbial taxa and nine distinct immune cell phenotypes with CRC.

Single-cell RNA sequencing deciphers the microenvironment in PDAC:

iii) Integrating single-cell and bulk RNA sequencing data, a hypoxia-responsive macrophage-derived prognostic signature was developed to predict survival outcomes in PDAC. This 13-gene model independently predicts inferior overall survival and chemosensitivity. Comprehensive pan-cancer analysis further substantiated *KRTCAP2* as a pan-tumor biomarker correlating with diminished immune infiltration and adverse prognosis.

iv) Using single-cell RNA sequencing and functional validation, we identify a distinct malignant epithelial subcluster (c05) enriched in liver metastases and progressive disease PDAC, characterized by an immune-cold niche with disrupted cell communication. We further define *RARRES1* as a central resistance driver consistently upregulated across resistant groups and validated in experimental models.

3. Zusammenfassung

Räumliche Transkriptomik (spatial transcriptomics) und Mendelsche Randomisierung entschlüsseln das Mikromilieu im kolorektalen Karzinom:

i) Räumlich-transkriptomische Analysen offenbarten, dass im tumorthrombus-assoziierten Mikromilieu des kolorektalen Karzinoms trotz verstärkter Infiltration von $CD8^+$ T-Zellen und Makrophagen diese Immunpopulationen eine ausgeprägte Dysfunktion aufweisen. $CD4^+$ T-Zellen supprimieren die Aktivität von $CD8^+$ T-Zellen, begleitet von einer Anreicherung immunsuppressiver Subpopulationen sowie einer eingeschränkten Antigenpräsentationsfähigkeit konventioneller dendritischer Zellen (cDCs).

ii) Großangelegte genomweite Assoziationsstudien belegten robuste kausale Beziehungen von fünf spezifischen mikrobiellen Taxa und neun distinkten Immunzellphänotypen mit kolorektalen Karzinomen.

Einzelzell-RNA-Sequenzierungs-Analysen des Mikromilieu im duktalen Pankreaskarzinom (PDAC):

iii) Durch Integration von Einzelzell- und Bulk-RNA-Sequenzierungsdaten wurde eine hypoxieantwortende makrophagenabhängige prognostische Signatur zur Vorhersage von Überlebensendpunkten bei PDAC entwickelt. Dieses 13-Gen-Modell sagt unabhängig ein schlechteres Gesamtüberleben und Chemosensitivität voraus. Eine umfassende Pan-Krebs-Analyse validierte ferner *KRTCAP2* als Pan-Tumor-Biomarker, das mit reduzierter Immuninfiltration und ungünstiger Prognose korreliert.

iv) Mittels Einzelzell-RNA-Sequenzierung und funktioneller Validierung identifizierten wir einen distinkten malignen epithelialen Subcluster (c05), der in Lebermetastasen und progredientem PDAC angereichert ist und durch eine immunologisch “kalte” Nische mit gestörter Zellkommunikation charakterisiert ist. Wir definieren RARRES1 weiterhin als zentralen Resistenzfaktor, der konsistent über Resistenzgruppen hochreguliert ist und experimentell validiert wurde.

4. References

Al-Hujaily EM, Al-Sowayan BS, Alyousef Z, Uddin S, Alammari F (2022) Recruiting Immunity for the Fight against Colorectal Cancer: Current Status and Challenges. Int J Mol Sci. 23(22).

Argiles G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, Laurent-Puig P, Quirke P, Yoshino T, Taieb J, Martinelli E, Arnold D (2020) Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 31(10):1291-1305.

Bai R, Li Y, Jian L, Yang Y, Zhao L, Wei M (2022) The hypoxia-driven crosstalk between tumor and tumor-associated macrophages: mechanisms and clinical treatment strategies. Mol Cancer. 21(1):177.

Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 74(3):229-263.

Chen J, Zhao T, Li H, Xu W, Maas K, Singh V, Chen MH, Dorsey SG, Starkweather AR, Cong XS (2024) Multi-Omics Analysis of Gut Microbiota and Host Transcriptomics Reveal Dysregulated Immune Response and Metabolism in Young Adults with Irritable Bowel Syndrome. Int J Mol Sci. 25(6).

Chen L, Fu B (2022) T cell exhaustion assessment algorism in tumor microenvironment predicted clinical outcomes and immunotherapy effects in glioma. Front Genet. 13:1087434.

Chen L, Zhang X, Zhang Q, Zhang T, Xie J, Wei W, Wang Y, Yu H, Zhou H (2022) A necroptosis related prognostic model of pancreatic cancer based on single cell sequencing analysis and transcriptome analysis. Front Immunol. 13:1022420.

Chen Y, Wang D, Li Y, Qi L, Si W, Bo Y, Chen X, Ye Z, Fan H, Liu B, Liu C, Zhang L, Zhang X, Li Z, Zhu L, Wu A, Zhang Z (2024) Spatiotemporal single-cell analysis decodes cellular dynamics underlying different responses to immunotherapy in colorectal cancer. Cancer Cell. 42(7):1268-1285 e1267.

Choudhry H, Harris AL (2018) Advances in Hypoxia-Inducible Factor Biology. Cell Metab. 27(2):281-298.

Cornish AJ, Tomlinson IPM, Houlston RS (2019) Mendelian randomisation: A powerful and inexpensive method for identifying and excluding non-genetic risk factors for colorectal cancer. Mol Aspects Med. 69:41-47.

Davies NM, Holmes MV, Davey Smith G (2018) Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 362:k601.

de Vries NL, Mahfouz A, Koning F, de Miranda N (2020) Unraveling the Complexity of the Cancer Microenvironment With Multidimensional Genomic and Cytometric Technologies. Front Oncol. 10:1254.

Dilly J, Hoffman MT, Abbassi L, Li Z, Paradiso F, Parent BD, Hennessey CJ, Jordan AC, Morgado M, Dasgupta S, Uribe GA, Yang A, Kapner KS, Hambitzer FP, Qiang L, Feng H, Geisberg J, Wang J, Evans KE, Lyu H, Schalck A, Feng N, Lopez AM, Bristow CA, Kim MP, Rajapakshe KI, Bahrambeigi V, Roth JA, Garg K, Guerrero PA, Stanger BZ, Cristea S, Lowe SW, Baslan T, Van Allen EM, Mancias JD, Chan E, Anderson A, Katlinskaya YV, Shalek AK, Hong DS, Pant S, Hallin J, Anderes K, Olson P, Heffernan TP, Chugh S, Christensen JG, Maitra A, Wolpin BM, Raghavan S, Nowak JA, Winter PS, Dougan SK, Aguirre AJ (2024) Mechanisms of Resistance to Oncogenic KRAS Inhibition in Pancreatic Cancer. Cancer Discov. 14(11):2135-2161.

Ding J, Xie Y, Liu Z, Zhang Z, Ni B, Yan J, Zhou T, Hao J (2025) Hypoxic and Acidic Tumor Microenvironment-Driven AVL9 Promotes Chemoresistance of Pancreatic Ductal Adenocarcinoma via the AVL9-IkappaBalpha-SKP1 Complex. Gastroenterology. 168(3):539-555 e535.

Du P, Liao Y, Zhao H, Zhang J, Muyiti, Keremu, Mu K (2020) ANXA2P2/miR-9/LDHA axis regulates Warburg effect and affects glioblastoma proliferation and apoptosis. Cell Signal. 74:109718.

El Tekle G, Andreeva N, Garrett WS (2024) The Role of the Microbiome in the Etiopathogenesis of Colon Cancer. Annu Rev Physiol. 86:453-478.

Güngör C, Hofmann BT, Wolters-Eisfeld G, Bockhorn M (2014) Pancreatic cancer. Br J Pharmacol. 171(4):849-858.

Hielscher A, Gerecht S (2015) Hypoxia and free radicals: role in tumor progression and the use of engineering-based platforms to address these relationships. Free Radic Biol Med. 79:281-291.

Ho WJ, Jaffee EM, Zheng L (2020) The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. Nat Rev Clin Oncol. 17(9):527-540.

Hu Y, Zhou P, Deng K, Zhou Y, Hu K (2024) Targeting the gut microbiota: a new strategy for colorectal cancer treatment. J Transl Med. 22(1):915.

Kwon JY, Vera RE, Fernandez-Zapico ME (2025) The multi-faceted roles of cancer-associated fibroblasts in pancreatic cancer. Cell Signal. 127:111584.

Li C, Xu X, Wei S, Jiang P, Xue L, Wang J, Senior C (2021) Tumor-associated macrophages: potential therapeutic strategies and future prospects in cancer. J Immunother Cancer. 9(1).

Liu S, Guo W, Shi J, Li N, Yu X, Xue J, Fu X, Chu K, Lu C, Zhao J, Xie D, Wu M, Cheng S, Liu S (2012) MicroRNA-135a contributes to the development of portal vein tumor thrombus by promoting metastasis in hepatocellular carcinoma. J Hepatol. 56(2):389-396.

Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, Smyla JK, Anderson JM, Kim HJ, Bentrem DJ, Talamonti MS, Iacobuzio-Donahue CA, Hollingsworth MA, Yeh JJ (2015) Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 47(10):1168-1178.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 12(5):453-457.

Nitschke C, Markmann B, Tolle M, Kropidlowski J, Belloum Y, Goetz MR, Schluter H, Kwiatkowski M, Sinn M, Izbicki J, Pantel K, Güngör C, Uzunoglu FG, Wikman H (2022) Characterization of RARRES1 Expression on Circulating Tumor Cells as Unfavorable Prognostic Marker in Resected Pancreatic Ductal Adenocarcinoma Patients. Cancers (Basel). 14(18)

Park SS, Lee YK, Choi YW, Lim SB, Park SH, Kim HK, Shin JS, Kim YH, Lee DH, Kim JH, Park TJ (2024) Cellular senescence is associated with the spatial evolution toward a higher metastatic phenotype in colorectal cancer. Cell Rep. 43(3):113912.

Qiao Y, Jiang X, Li Y, Wang K, Chen R, Liu J, Du Y, Sun L, Li J (2023) Identification of a hypoxia-related gene prognostic signature in colorectal cancer based on bulk and single-cell RNA-seq. Sci Rep. 13(1):2503.

Qin Y, Zhu W, Xu W, Zhang B, Shi S, Ji S, Liu J, Long J, Liu C, Liu L, Xu J, Yu X (2014) LSD1 sustains pancreatic cancer growth via maintaining HIF1alpha-dependent glycolytic process. Cancer Lett. 347(2):225-232.

Shi Y, Zhang Q, Bi H, Lu M, Tan Y, Zou D, Ge L, Chen Z, Liu C, Ci W, Ma L (2022) Decoding the multicellular ecosystem of vena caval tumor thrombus in clear cell renal cell carcinoma by single-cell RNA sequencing. Genome Biol. 23(1):87.

Siegel RL, Giaquinto AN, Jemal A (2024) Cancer statistics, 2024. CA Cancer J Clin. 74(1):12-49.

Singhal A, Styers HC, Rub J, Li Z, Torborg SR, Kim JY, Grbovic-Huezo O, Feng H, Tarcan ZC, Sahin Ozkan H, Hallin J, Basturk O, Yaeger R, Christensen JG, Betel D, Yan Y, Chio IIC, de Stanchina E, Tammela T (2024) A Classical Epithelial State Drives Acute Resistance to KRAS Inhibition in Pancreatic Cancer. Cancer Discov. 14(11):2122-2134.

Smith GD, Ebrahim S (2003) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 32(1):1-22.

Song X, Xie D, Tan F, Zhou Y, Li Y, Zhou Z, Pei Q, Pei H (2021) Intravascular emboli relates to immunosuppressive tumor microenvironment and predicts prognosis in stage III colorectal cancer. Aging (Albany NY). 13(16):20609-20628.

Sonkin D, Thomas A, Teicher BA (2024) Cancer treatments: Past, present, and future. Cancer Genet. 286-287:18-24.

Tao J, Yang G, Zhou W, Qiu J, Chen G, Luo W, Zhao F, You L, Zheng L, Zhang T, Zhao Y (2021) Targeting hypoxic tumor microenvironment in pancreatic cancer. J Hematol Oncol. 14(1):14.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 32(4):381-386.

Vincent A, Herman J, Schulick R, Hruban RH, Goggins M (2011) Pancreatic cancer. Lancet. 378(9791):607-620.

Wang S, Zheng R, Li J, Zeng H, Li L, Chen R, Sun K, Han B, Bray F, Wei W, He J (2024) Global, regional, and national lifetime risks of developing and dying from gastrointestinal cancers in 185 countries: a population-based systematic analysis of GLOBOCAN. Lancet Gastroenterol Hepatol. 9(3):229-237.

Werba G, Weissinger D, Kawaler EA, Zhao E, Kalfakakou D, Dhara S, Wang L, Lim HB, Oh G, Jing X, Beri N, Khanna L, Gonda T, Oberstein P, Hajdu C, Loomis C, Heguy A, Sherman MH, Lund AW, Welling TH, Dolgalev I, Tsirigos A, Simeone DM (2023) Single-cell RNA sequencing reveals the effects of chemotherapy on human pancreatic adenocarcinoma and its tumor microenvironment. Nat Commun. 14(1):797.

Won EJ, Park H, Chang SH, Kim JH, Kwon H, Cho YS, Yoon TJ (2021) One-shot dual gene editing

for drug-resistant pancreatic cancer therapy. Biomaterials. 279:121252.

Xiang X, Wang J, Lu D, Xu X (2021) Targeting tumor-associated macrophages to synergize tumor immunotherapy. Signal Transduct Target Ther. 6(1):75.

Yoon AR, Hong J, Jung BK, Ahn HM, Zhang S, Yun CO (2023) Oncolytic adenovirus as pancreatic cancer-targeted therapy: Where do we go from here? Cancer Lett. 579:216456.

5. List of abbreviations

IARC: International Agency for Research on Cancer

TME: tumor microenvironment

CRC: colorectal cancer

TT: tumor thrombus

NTT: non-tumor thrombus

CMS: consensus molecular subtypes

Tregs: regulatory T cells

M-MDSCs: monocytic myeloid-derived suppressor cells

PMN-MDSCs: polymorphonuclear myeloid-derived suppressor cells

tDCs: tolerogenic dendritic cells

cDCs: conventional dendritic cells

GM: gut microbial

MR: Mendelian Randomization

IVs: instrumental variables

GWAS: genome-wide association studies

PDAC: Pancreatic ductal adenocarcinoma

TAMs: tumor-associated macrophages

scRNA-seq: single-cell RNA sequencing

AUC: area under the curve

TMB: tumor mutational burden

CNV: copy number variations

PR: Partial Response

SD: Stable Disease

PD: Progressive Disease

NMF: Non-negative matrix factorization

SNVs: single nucleotide variants

GEM-R: gemcitabine-resistant L3.6pl cells

6. Figures

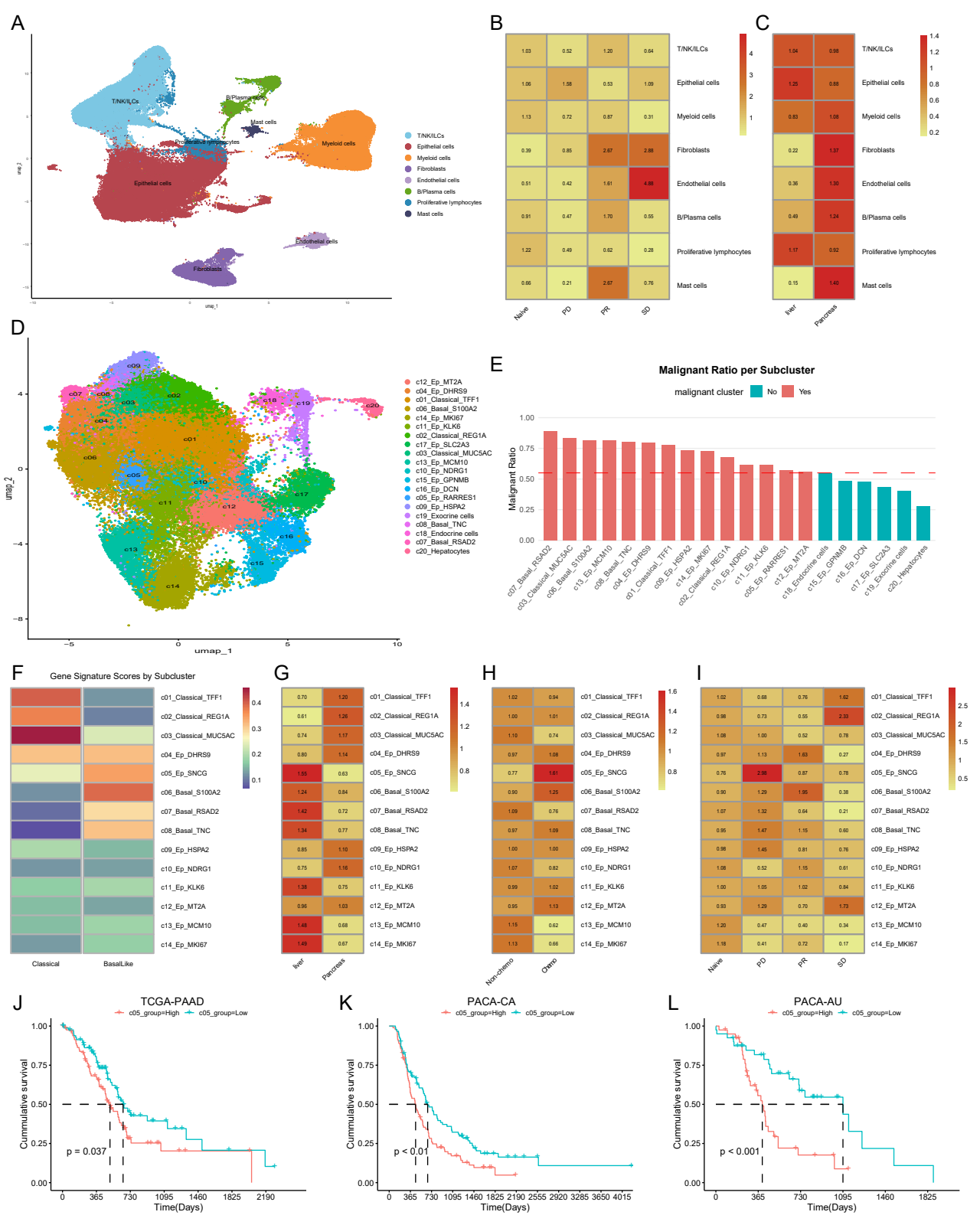
A
T/NK/ILCs
B/Plasma cells
Mast cells
Proliferative lymphocytes
Myeloid cells
Epithelial cells
Endothelial cells
Fibroblasts
umap_1
umap_2
B
T/NK/ILCs
Epithelial cells
Myeloid cells
Fibroblasts
Endothelial cells
B/Plasma cells
Proliferative lymphocytes
Mast cells
Naive
PD
PR
SD
C
liver
Pancreas
D
c12_Ep_MT2A
c04_Ep_DHRS9
c01_Classical_TFF1
c06_Basal_S100A2
c14_Ep_MKI67
c11_Ep_KLK6
c02_Classical_REG1A
c17_Ep_SLC2A3
c03_Classical_MUC5AC
c13_Ep_MCM10
c10_Ep_NDRG1
c15_Ep_GPNMB
c16_Ep_DCN
c05_Ep_RARRES1
c09_Ep_HSPA2
c19_Exocrine cells
c08_Basal_TNC
c18_Endocrine cells
c07_Basal_RSAD2
c20_Hepatocytes
umap_1
umap_2
E
Malignant Ratio per Subcluster
malignant cluster
No
Yes
Malignant Ratio
F
Gene Signature Scores by Subcluster
Classical
BasalLike
G
liver
Pancreas
H
Non-chemo
Chemo
I
Naive
PD
PR
SD
J
TCGA-PAAD
c05_group=High
c05_group=Low
Cummulative survival
p = 0.037
Time(Days)
K
PACA-CA
p < 0.01
L
PACA-AU
p < 0.001

Figure 1. Identification of a chemoresistance-associated malignant epithelial subcluster (c05) in PDAC.

(A) UMAP plot of 8 cell types from single-cell RNA sequencing (scRNA-seq) data in 27 PDAC samples.

(B) Proportion of different cell types across treatment response groups: Treatment-naive (Naive), Partial Response (PR), Stable Disease (SD), and Progressive Disease (PD).

(C) Proportion of different cell types in primary pancreatic cancers versus liver metastases.

(D) UMAP plot showing 20 distinct epithelial subclusters (c01-c20).

(E) CNV scores across epithelial subclusters. c01-c14 exhibit high CNV scores, confirming their malignant origin. Dashed line indicates the CNV threshold for malignancy.

(F) Heatmap of established PDAC subtype marker genes across malignant subclusters (c01-c14). c01-c03: Classical subtype; c06-c08: Basal-like subtype; c04-c05: Mixed subtype expression; c09-c14: No defined subtype markers.

(G) Proportion of malignant subclusters (c01-c14) in primary pancreatic cancers versus liver metastases.

(H) Proportion of malignant subclusters (c01-c14) in chemotherapy-exposed samples versus treatment-naive samples.

(I) Proportion of malignant subclusters (c01-c14) across treatment response groups: Naive, PR, SD, PD.

(J-L) Kaplan-Meier overall survival curves for PDAC patients stratified by high and low expression of the c05 subcluster gene signature in independent cohorts: (J) TCGA-PAAD, (K) PACA-CA, (L) PACA-AU.

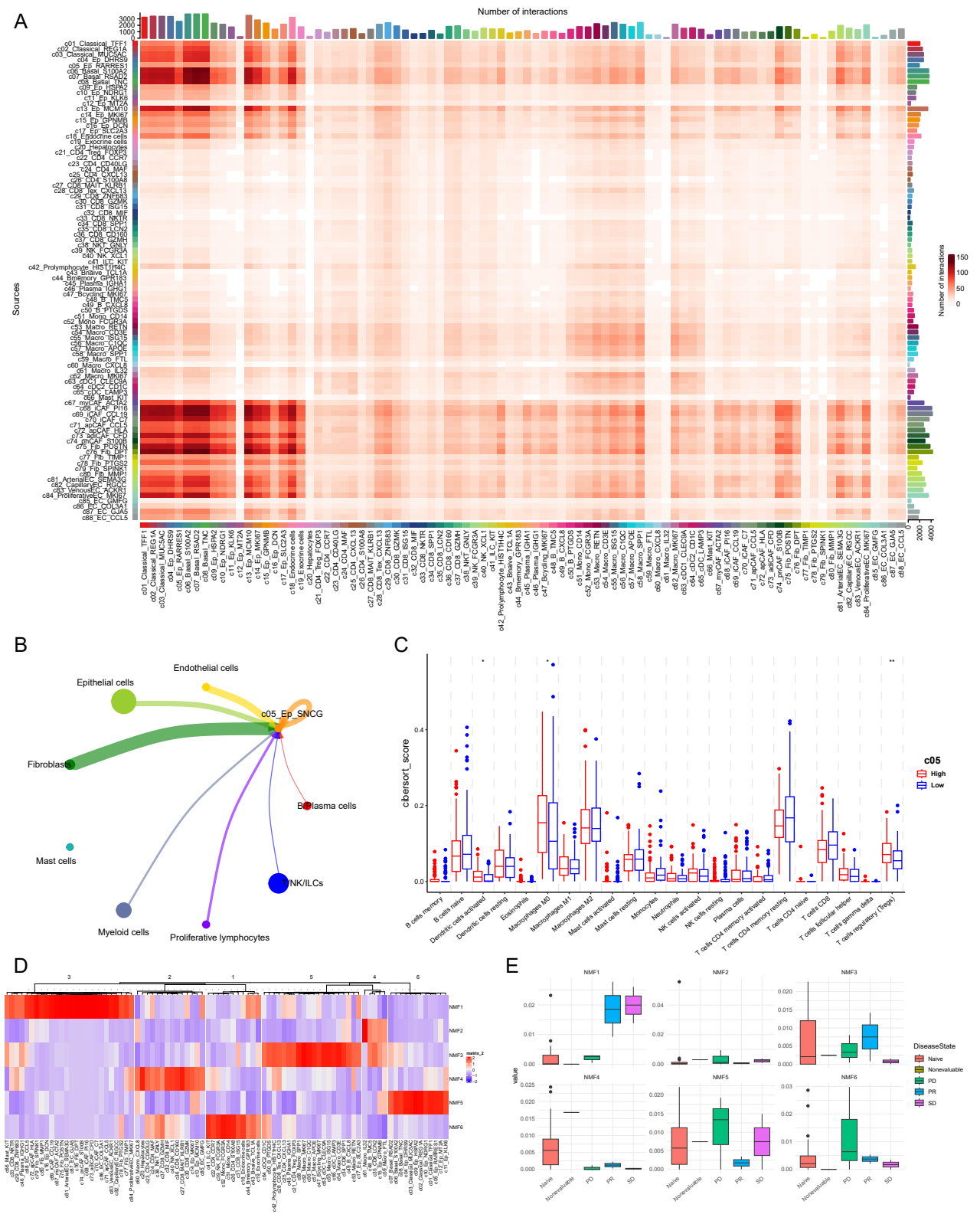
A
Number of interactions
Sources
Number of interactions
B
Endothelial cells
Epithelial cells
c05_Ep_SNCG
Fibroblasts
B Plasma cells
Mast cells
TNK/ILCs
Myeloid cells
Proliferative lymphocytes
C
cibersort_score
c05
High
Low
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
D
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
matrix_z
E
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
value
DiseaseState
Naive
Nonevaluable
PD
PR
SD

Figure 2. The c05 subcluster exhibited an immune-cold phenotype and disrupted cell-cell communication.

(A) Cell-cell communication networks revealed significantly weaker overall signaling strength in the c05 malignant epithelial subcluster compared to other malignant epithelial subclusters.

(B) Cell-cell communication networks showed minimal incoming signaling to the c05 subcluster from immune cell populations (Mast, Myeloid, T/NK/ILC, B/Plasma).

(C) Deconvolution of bulk RNA-seq data (TCGA-PAAD) using CIBERSORT demonstrated that samples with high c05 signature gene expression exhibited a significantly increased proportion of immunosuppressive regulatory T cells (Tregs) and a lack of anti-tumor effector immune cell infiltration.

(D) NMF analysis of bulk RNA-seq data identified an NMF5 program significantly enriched for the c05 cells. This NMF5 program strongly correlated with an immune-cold tumor microenvironment phenotype. Conversely, the NMF1 program was enriched for immune cell signatures.

(E) NMF program enrichment analysis confirmed that samples exhibiting progressive disease (PD) were significantly enriched for the c05-associated, immune-cold NMF5 program. In contrast, samples showing treatment response (Partial Response, PR; Stable Disease, SD) were primarily enriched for the immune-rich NMF1 program.

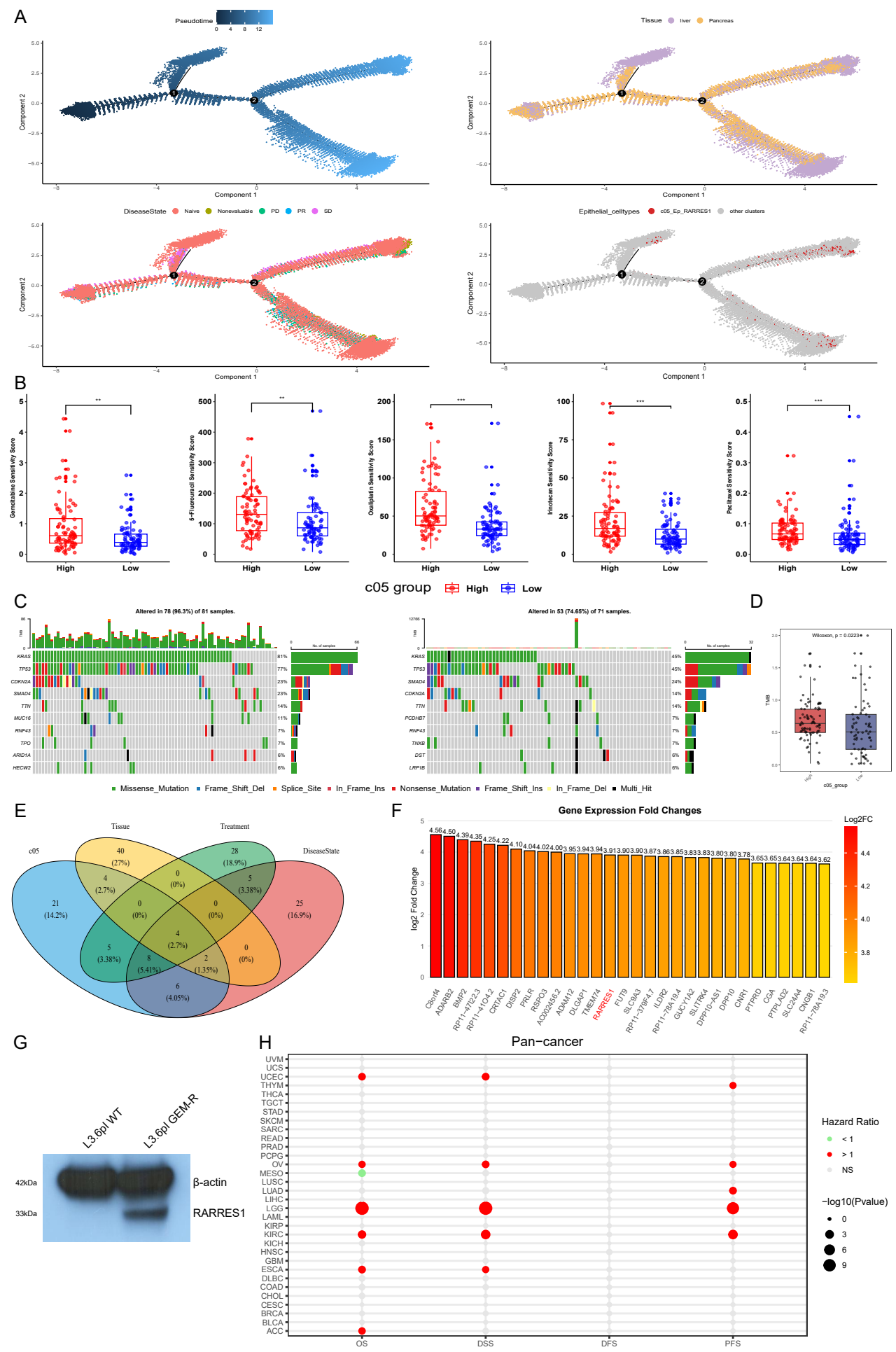
A
Pseudotime
Tissue
liver
Pancreas
DiseaseState
Naive
Nonevaluable
PD
PR
SD
Epithelial_celltypes
c05_Ep_RARRES1
other clusters
Component 1
Component 2
B
Gemcitabine Sensitivity Score
5-Fluorouracil Sensitivity Score
Oxaliplatin Sensitivity Score
Irinotecan Sensitivity Score
Paclitaxel Sensitivity Score
High
Low
c05 group
C
Altered in 78 (96.3%) of 81 samples.
Altered in 53 (74.65%) of 71 samples.
KRAS
TP53
CDKN2A
SMAD4
TTN
MUC16
RNF43
TPO
ARID1A
HECW2
PCDHB7
TNKB
DST
LRP1B
Missense_Mutation
Frame_Shift_Del
Splice_Site
In_Frame_Ins
Nonsense_Mutation
Frame_Shift_Ins
In_Frame_Del
Multi_Hit
D
Wilcoxon, p = 0.0223
TMB
c05_group
E
c05
Tissue
Treatment
DiseaseState
F
Gene Expression Fold Changes
log2 Fold Change
Log2FC
C8orf4
ADARB2
BMP2
RP11-477J2.3
RP11-410A4.2
CRTAC1
DISP2
PRLR
RSPO3
AC002456.2
ADAM12
DLGAP1
TMEM74
RARRES1
FUT9
SLC9A3
RP11-379F4.7
ILDR2
RP11-78A19.4
GUCY1A2
SLITRK4
DPP10-AS1
DPP10
CNR1
PTPRD
CGA
PTPLAD2
SLC24A4
CNGB1
RP11-78A19.3
G
L3.6pl WT
L3.6pl GEM-R
42kDa
33kDa
β-actin
RARRES1
H
Pan−cancer
UVM
UCS
UCEC
THYM
THCA
TGCT
STAD
SKCM
SARC
READ
PRAD
PCPG
OV
MESO
LUSC
LUAD
LIHC
LGG
LAML
KIRP
KIRC
KICH
HNSC
GBM
ESCA
DLBC
COAD
CHOL
CESC
BRCA
BLCA
ACC
OS
DSS
DFS
PFS
Hazard Ratio
< 1
> 1
NS
−log10(Pvalue)

Figure 3. The c05 subcluster exhibited multi-dimensional chemoresistance features and identified RARRES1 as a central resistance driver.

(A) Pseudotime trajectory analysis localized the c05 subcluster and PD-derived cells at terminal endpoints, indicating a differentiated, treatment-resistant state in PDAC progression.

(B) Drug sensitivity profile demonstrated significantly reduced sensitivity to first-line chemotherapies (gemcitabine, 5-fluorouracil, oxaliplatin, irinotecan, paclitaxel) in samples with high c05 signature expression.

(C, D) SNV analysis revealed elevated *KRAS* mutation frequency (81% vs. 45%) and increased TMB in high c05-expression samples compared to low-expression counterparts.

(E) Intersection of differentially expressed genes (DEGs) from four comparisons: (i) c05 vs. other malignant subclusters, (ii) PD vs. Naive/PR/SD, (iii) metastases vs. primary tumors, and (iv) chemo-treated vs. naive samples, identifies consistent upregulation of *SNCG*, *CST6*, *CAPS*, and *RARRES1*.

(F) Transcriptome profile confirmed significantly elevated *RARRES1* expression in GEM-R compared to wild type.

(G) Western blot showed increased RARRES1 protein abundance in GEM-R cells compared to wild type line.

(H) Pan-cancer survival analysis established RARRES1 overexpression as a prognostic biomarker for reduced overall survival across multiple malignancies.

7. Declaration of the contribution to the publications

Publication: **Ge H**, Pei Z, Zhou Z, Pei Q, Güngör C, Zheng L, Liu W, Li F, Zhou J, Xiang Y, Pei H, Li Y, Liu W (2024). *Spatial transcriptomics deciphers the immunosuppressive microenvironment in colorectal cancer with tumour thrombus*. Clin Transl Med. 14(12):e70112.

Contribution of Heming Ge: Study Design; Statistical Analysis; Data Interpretation; Manuscript Preparation.

Contribution of co-authors: Study Design; Data Collection; Statistical Analysis; Data Interpretation; Manuscript Preparation; Funds Collection.

The first author, Heming Ge, contributed mainly to this article.

Publication: Zheng L, Li Y, Güngör C, **Ge H** (2025). *Gut microbiota influences colorectal cancer through immune cell interactions: a Mendelian randomization study*. Discov Oncol. 16(1):747.

Contribution of Heming Ge: Study Design; Statistical Analysis; Manuscript Preparation.

Contribution of co-authors: Study Design; Statistical Analysis; Data Interpretation; Manuscript Preparation; Funds Collection.

Publication: **Ge H**, Wolters-Eisfeld G, Hackert T, Li Y, Güngör C (2025). *Development of a hypoxia-responsive macrophage prognostic model using single-cell and bulk RNA sequencing in pancreatic cancer*. PLoS One. 20(5):e0322618.

Contribution of Heming Ge: Study Design; Data Collection; Statistical Analysis; Data Interpretation; Manuscript Preparation.

Contribution of co-authors: Study Design; Data Collection; Statistical Analysis; Data Interpretation; Manuscript Preparation; Funds Collection.

The first author, Heming Ge, contributed mainly to this article.

8. Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.

Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots“) verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG“ aus September 2023 wurde dabei beachtet.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Datum Unterschrift

9. Acknowledgement

The completion of this dissertation marks the end of a long and memorable journey of doctoral research at the University of Hamburg. I extend my deepest appreciation to all those who have guided, encouraged and supported me along the way.

The sincerest gratitude is to my supervisor professor Dr. Cenap Güngör for his invaluable guidance throughout my research. I am profoundly grateful for his patience, empathy, and academic professions. He consistently created opportunities for me to develop as a researcher - providing expert mentorship in academic writing, dedicating significant time to reviewing manuscripts, and offering constructive feedback that elevated my work. Being part of his research team stands as a distinct honor in my academic journey.

I sincerely thank the members of my dissertation committee for their time and advice during my defense. Their incisive feedback significantly strengthened my dissertation.

I would like to acknowledge the support and encouragement provided by my friends and colleagues in the laboratory since the beginning of my study.

I wish to thank my parents for their patience, love and care throughout my education and life. Their constant encouragement and support sustained me through every challenge I faced.

Finally, I would like to appreciate the China Scholarship Council (CSC) for its generous financial support, which made my overseas studies possible.