

# Integrative bioinformatics analysis identifies C1orf198 as a novel prognostic biomarker in colorectal cancer (#119038)

1

First submission

## Guidance from your Editor

Please submit by **22 Jul 2025** for the benefit of the authors (and your token reward) .



### Structure and Criteria

Please read the 'Structure and Criteria' page for guidance.



### Custom checks

Make sure you include the custom checks shown below, in your review.



### Author notes

Have you read the author notes on the [guidance page](#)?



### Raw data check

Review the raw data.



### Image check

Check that figures and images have not been inappropriately manipulated.

All review materials are strictly confidential. Uploading the manuscript to third-party tools such as Large Language Models is not allowed.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

## Files

Download and review all files from the [materials page](#).

5 Figure file(s)

2 Raw data file(s)



## Custom checks

## Human participant/human tissue checks



Have you checked the authors [ethical approval statement](#)?



Does the study meet our [article requirements](#)?



Has identifiable info been removed from all files?



Were the experiments necessary and ethical?



# Structure and Criteria

## Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).

## Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).

### BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [Peerj standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [Peerj policy](#)).

### EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

### VALIDITY OF THE FINDINGS

-  **Impact and novelty is not assessed.** Meaningful replication encouraged where rationale & benefit to literature is clearly stated.
-  All underlying data have been provided; they are robust, statistically sound, & controlled.
-  Conclusions are well stated, linked to original research question & limited to supporting results.



The best reviewers use these techniques

## Tip

## Example

**Support criticisms with evidence from the text or from other sources**

*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*

**Give specific suggestions on how to improve the manuscript**

*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*

**Comment on language and grammar issues**

*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*

**Organize by importance of the issues, and number your points**

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

**Please provide constructive criticism, and avoid personal opinions**

*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*

**Comment on strengths (as well as weaknesses) of the manuscript**

*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

# Integrative bioinformatics analysis identifies C1orf198 as a novel prognostic biomarker in colorectal cancer

Changjiang Yang<sup>1</sup>, Xuhua Geng<sup>2</sup>, Zihan Zhao<sup>Corresp. 3, 4</sup>

<sup>1</sup> Department of Gastroenterological surgery, Peking University People's Hospital, Beijing, China

<sup>2</sup> Boston University, Boston, United States

<sup>3</sup> Interventional Center of Valvular Heart Disease, Beijing Anzhen Hospital, Capital Medical University, Beijing, China, Beijing, China

<sup>4</sup> Aerospace center hospital, Peking University Aerospace School of Clinical Medicine, Beijing, China

Corresponding Author: Zihan Zhao

Email address: 1911210739@pku.edu.cn

Colorectal cancer (CRC) is a leading cause of cancer-related mortality, necessitating the identification of novel prognostic biomarkers to improve patient management. In this study, we integrated bioinformatics analyses and experimental validation to explore the role of C1orf198 in CRC. Data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) revealed significantly upregulated C1orf198 mRNA in CRC tissues compared to normal counterparts, confirmed by immunohistochemistry (IHC) in clinical samples. High C1orf198 expression correlated with advanced tumor stages (T, N, M) and poor survival outcomes, including shorter overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI). Functional enrichment analyses highlighted involvement in extracellular matrix organization, cell adhesion, and oncogenic signaling pathways such as PI3K-AKT and focal adhesion. Immune infiltration analysis showed positive correlations with stromal/immune scores and M2 macrophage infiltration, linking C1orf198 to tumor microenvironment (TME) remodeling. Notably, C1orf198 was strongly associated with cytokines CXCL12 and receptors CXCR5, which mirrored its immune cell correlations. Collectively, our findings identify C1orf198 as a potential prognostic biomarker in CRC, implicating its role in TME modulation and oncogenic progression.

# Integrative Bioinformatics Analysis Identifies C1orf198 as a Novel Prognostic Biomarker in Colorectal Cancer

Changjiang Yang<sup>2</sup>, Xuhua Geng<sup>3</sup> and Zihan Zhao<sup>1,4,\*</sup>

<sup>1</sup> Aerospace center hospital, Peking University Aerospace School of Clinical Medicine, Beijing, China.

<sup>2</sup> Department of Gastroenterological surgery, Peking University People's Hospital, Beijing, China.

<sup>3</sup> School of public health, Boston University, Beijing, China.

<sup>4</sup> Interventional Center of Valvular Heart Disease, Beijing Anzhen Hospital, Capital Medical University, Beijing, China.

\*Corresponding to: Zihan Zhao, Email: zzh960407@pku.edu.cn.

## Abstract:

Colorectal cancer (CRC) is a leading cause of cancer-related mortality, necessitating the identification of novel prognostic biomarkers to improve patient management. In this study, we integrated bioinformatics analyses and experimental validation to explore the role of C1orf198 in CRC. Data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) revealed significantly upregulated C1orf198 mRNA in CRC tissues compared to normal counterparts, confirmed by immunohistochemistry (IHC) in clinical samples. High C1orf198 expression correlated with advanced tumor stages (T, N, M) and poor survival outcomes, including shorter overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI). Functional enrichment analyses highlighted involvement in extracellular matrix organization, cell adhesion, and oncogenic signaling pathways such as PI3K-AKT and focal adhesion. Immune infiltration analysis showed positive correlations with stromal/immune scores and M2 macrophage infiltration, linking C1orf198 to tumor microenvironment (TME) remodeling. Notably, C1orf198 was strongly associated with cytokines CXCL12 and receptors CXCR5, which mirrored its immune cell correlations. Collectively, our findings identify C1orf198 as a potential prognostic biomarker in CRC, implicating its role in TME modulation and oncogenic progression.

**Keywords:** C1orf198, Colorectal Cancer, Immune Infiltration, Prognosis, metastasis

## 1. Introduction

Colorectal cancer (CRC), ranked among the most common malignant tumors globally, poses a significant threat to human health<sup>1</sup>. Although considerable advancements have been achieved in diagnostic and therapeutic approaches, treatment outcomes for a substantial portion of CRC patients still remain suboptimal<sup>2</sup>. The discovery of novel biomarkers and clarification of the underlying molecular mechanisms driving CRC progression are pivotal for enhancing clinical management and improving patient prognosis<sup>3</sup>.

The study of the association between uncharacterized proteins and tumors has gradually become a hotspot in the field of tumor biology. These proteins account for over 10% of the genome, and

their functions remain unclear<sup>4</sup>. However, increasing evidence suggests that they play critical roles in tumor initiation, progression, and therapeutic resistance<sup>5</sup>. C1orf198, an understudied gene located on chromosome, has been associated with cellular mechanisms including chromatin organization and immune regulation. However, its involvement in cancer development remains poorly understood. Previous studies have suggested that C1orf198 plays a role in the initiation and progression of gastric<sup>6</sup> and breast cancers<sup>7</sup>, but its relationship with CRC remains unclear. On the other hand, emerging evidence suggests that genes involved in tumor microenvironment (TME) remodeling, including extracellular matrix (ECM) components and immune cell crosstalk, are critical for CRC progression. Dysregulation of cell adhesion molecules, ECM-receptor interactions, and immune microenvironment has been linked to metastasis and therapeutic resistance in CRC.

Here, we hypothesized that C1orf198 contributes to CRC pathogenesis by influencing TME components and oncogenic signaling. Using multi-omics datasets and experimental validation, we aimed to investigate C1orf198 expression patterns, clinical correlations, functional pathways, and immune infiltration associations in CRC. Identifying C1orf198 as a prognostic biomarker could provide insights into novel therapeutic targets for CRC.

## 2. Materials and Methods

### Acquisition and Processing of Data

RNA-seq data along with matched clinical profiles from CRC tumors and paired adjacent normal tissues were obtained from The Cancer Genome Atlas (TCGA, <https://cancergenome.nih.gov/>), a prominent public cancer genomics resource. Additionally, the datasets GSE113513, GSE89076, GSE22598, and GSE110224 utilized in this study were retrieved from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>), a widely accessed functional genomics repository that supports MIAME-compliant data submissions.

### Patient Samples and Clinical Specimens

Supplied by Shanghai Outdo Biotech Company (Shanghai, China), the colorectal cancer (CRC) tissue microarray (HCoLA160CS01) comprised 80 paired tumor and adjacent normal tissue samples. Ethical approval for the study protocols was granted by the company's Ethics Committee under the approval ID: YB M-05-02. All participants provided WRITTEN informed consent prior to their involvement in the study.

# **Immunohistochemistry (IHC) Staining Protocol**

Processing of CRC and normal tissue sections followed this protocol: after embedding in paraffin, specimens underwent deparaffinization with dimethylbenzene and sequential rehydration through ethanol gradients. Antigen retrieval was performed using a sodium citrate buffer via microwave treatment at 95°C. Endogenous peroxidase activity was blocked by incubating specimens in 3% hydrogen peroxide for 10 minutes. To reduce non-specific binding, tissues were then incubated for one hour in a blocking solution containing 10% fetal bovine serum. The primary antibody against C1orf198 (1:100 dilution, CST #33418) was applied and incubated overnight at 4°C. After washing, sections were treated with an HRP-conjugated anti-rabbit secondary antibody in the dark. Immunostaining was visualized using DAB, followed by counterstaining with hematoxylin. Slides were dehydrated, mounted, and prepared for microscopic evaluation.

C1orf198 expression was assessed using two criteria: the percentage of positively stained cells and staining intensity. Positively stained cells were categorized into four groups based on their proportion: category 0 (0–10%), category 1 (10–40%), category 2 (40–70%), and category 3 (>70%). Staining intensity received numerical scores: 1 for weak, 2 for moderate, and 3 for strong visibility. The final immunohistochemistry (IHC) score for each sample was calculated by summing the cell positivity percentage score and the intensity score, resulting in a combined score ranging from 0 to 6. Samples were classified as low expression (0–3 points) or high expression (4–6 points). All procedures were conducted in strict accordance with standardized guidelines to ensure methodological consistency and regulatory compliance.

# **Survival Analysis**

Survival outcomes, including overall survival (OS), disease-specific survival (DSS), progression-free interval (PFI), and disease-free survival (DFS), were evaluated using the Kaplan-Meier method with log-rank testing. Patients were stratified into low- and high-expression groups based on the median C1orf198 expression level as the cutoff threshold. Cox proportional hazards regression models were applied to assess associations between clinical/pathological characteristics and prognostic endpoints, incorporating these survival metrics into the analysis.

# **Functional Enrichment Analysis**

Differential gene expression analysis between C1orf198 low and C1orf198 high tissues was performed with significance thresholds set as Fold change > 1.5 and FDR<0.05 to identify DEGs

(differentially expressed genes). The ClusterProfiler package (R v3.6.3) was utilized to conduct Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, alongside gene set enrichment analysis (GSEA). GO annotations were categorized into molecular functions (MF), cellular components (CC), and biological processes (BP). For GSEA, pathway enrichment was evaluated using normalized enrichment scores (NES) and adjusted p-values, with reference gene sets derived from c2.cp.kegg.v2022.1.Hs.symbols.gmt (KEGG) and c5.go.all.v2022.1.Hs.symbols.gmt (GO) databases. Pathways were considered significantly enriched when meeting criteria of adjusted  $p < 0.05$  and  $FDR < 0.25$ .

### Immune Infiltration Analysis

The ESTIMATE algorithm was utilized to compute immune and stromal scores for CRC samples. We leveraged the GSVA package in R to perform single-sample gene set enrichment analysis (ssGSEA), aiming to explore literature-supported associations between SPOCD1 and the hallmark gene signatures of 24 distinct immune cell types<sup>8</sup>. The CIBERSORT was also applied to examine the link between C1orf198 and infiltrating immune cells. Spearman's correlation analysis was conducted to examine the association between C1orf198 expression levels and the degrees of immune cell infiltration. Differences in immune cell composition across the low- and high-expression groups of C1orf198 were evaluated using the Wilcoxon rank-sum test.

### Statistical Methods

Data from TCGA were analyzed using R (v4.2.1). To compare C1orf198 expression levels between tumor and normal tissues, Wilcoxon rank-sum tests (for independent samples) and signed-rank tests (for paired samples) were applied. Associations between C1orf198 expression and clinicopathological characteristics were evaluated using Welch's one-way ANOVA, followed by Bonferroni post-hoc tests for multi-group comparisons (or t-tests for two-group analyses). Pearson's chi-square test was used to assess correlations between C1orf198 expression and clinical factors, with Fisher's exact test employed for small sample sizes to ensure analytical validity. The prognostic value of C1orf198 was evaluated using Kaplan-Meier curves. All statistical tests were two-tailed, with significance defined as  $P \leq 0.05$  to maintain consistency in inferential analysis.

## 3. Results

### CRC is characterized by upregulated C1orf198 mRNA and protein level



The mRNA and protein expression levels of C1orf198 were analyzed in pan-cancer and CRC tissues. Using TCGA data, comparative analysis of C1orf198 mRNA expression in pan-cancer and normal tissues revealed expression differences across multiple cancer types (Figure 1A, B). TCGA-based box-plot and paired-sample analysis showed that C1orf198 mRNA expression was significantly up-regulated in CRC tissues compared with normal tissues (Figure 1C, D). To comprehensively evaluate the diagnostic utility of C1orf198 in CRC, a ROC analysis was meticulously conducted (Figure 1E). This outcome implies that C1orf198 has potential as a diagnostic biomarker for CRC, capable of partially differentiating tumor-bearing states from normal conditions. Analyses of GEO datasets (GSE113513, GSE89076, GSE22598, GSE110224) confirmed the up-regulation of C1orf198 mRNA in CRC versus normal samples (Figure 1F-I). Subsequently, we performed IHC staining on clinical samples of CRC. IHC score analysis indicated that C1orf198 protein expression was significantly higher in CRC than in normal tissues ( $p < 0.001$ ) (Figure 1J). Representative immunohistochemistry images visually corroborated the expression discrepancy between normal and CRC tissues (Figure 1K). Collectively, these results demonstrate that C1orf198 is prominently overexpressed at both mRNA and protein levels in CRC tissues, suggesting its potential role in CRC tumorigenesis.

Figure 1 The levels of C1orf198 protein and mRNA expression in pan-cancer and CRC relative to normal samples. (A, B) Comparative analysis of C1orf198 mRNA expression in pan-cancer and normal tissues using TCGA. (C, D) TCGA-based box-plot and paired-sample analysis showing that C1orf198 mRNA expression was up-regulated in CRC tissues versus normal tissues. (E) ROC analysis evaluating the diagnostic value of C1orf198 in CRC. (F-I) Analyses of GEO datasets (GSE113513, GSE89076, GSE22598, GSE110224) revealed up-regulation of C1orf198 mRNA expression in CRC compared with normal samples. (J) IHC score analysis showing that C1orf198 protein expression was up-regulated in CRC compared with normal samples. (K) Representative immunohistochemistry images of C1orf198 in normal and CRC tissues (ns denotes no significance, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

## Correlation between C1orf198 Expression and Clinicopathologic Parameters and Prognosis in CRC

Based on the TCGA database, the correlations between C1orf198 expression and clinicopathologic parameters as well as prognosis in CRC were analyzed. Initially, the association between C1orf198 expression and clinicopathologic parameters was explored. Results demonstrated that C1orf198

expression showed no significant association with age or gender, but was significantly correlated with T stage, N stage, M stage, pathologic stage, OS event, DSS event, and PFI event (Figure 2A-I). Further Kaplan-Meier analysis based on TCGA data revealed that up-regulated C1orf198 expression was associated with shortened OS, DSS, and PFI times (Log-rank  $P=0.004$ ,  $0.002$ ,  $<0.001$ ) (Figure 2J-L). Additionally, Kaplan-Meier curves from GEO datasets (GSE17536 and GSE14333) indicated that up-regulated C1orf198 expression was related to reduced DSS/DFS times (Log-rank  $P=0.001$ ,  $<0.001$ ,  $<0.001$ ) (Figure 2M-O). These findings suggest that high C1orf198 expression may be associated with poor prognosis in CRC.

Figure 2. TCGA database-based correlations between C1orf198 expression and clinicopathologic parameters and prognosis in CRC. (A-I) Analysis of the association between C1orf198 expression and clinicopathologic parameters in CRC using TCGA data, demonstrating associations with age (A), gender (B), T stage (C), N stage (D), M stage (E), pathologic stage (F), OS event (G), DSS event (H), and PFI event (I). (J-L) Kaplan-Meier curves based on TCGA data showing that up-regulated C1orf198 expression was associated with reduced OS, DSS, and PFI times. (M-O) Kaplan-Meier curves from GEO datasets (GSE17536 and GSE14333) indicating that up-regulated C1orf198 expression was related to reduced DSS/DFS times. (ns represents no significance, \*  $p<0.05$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ ).

## Biological Correlation and Pathway Analysis of C1orf198 in CRC

To further investigate the biological functions of C1orf198 in CRC, we analyzed its co-expressed genes and enriched functional pathways. We first selected the top 20 genes most significantly correlated with C1orf198 expression (Figure 3A), visualized in a heatmap, to explore the molecular mechanisms of C1orf198-mediated biological processes in CRC. Figure 3C presents a volcano plot highlighting DEGs between low- and high-C1orf198-expression subgroups, where red and blue dots denote genes upregulated or downregulated in the high-expression group, respectively. This analysis identified 1151 upregulated and 690 downregulated genes, forming the basis for subsequent functional enrichment. Co-expressed gene analysis (Figure 3B) further linked C1orf198 to processes like "myeloid cell differentiation", "regulation of epithelial cell proliferation" and "regulation of small GTPase mediated signal transduction," while GO and KEGG pathway analyses of these upregulated genes revealed significant enrichment in biological processes such as "extracellular matrix organization", "extracellular structure organization" and "cell junction assembly", "cell-cell adhesion via plasma-membrane adhesion molecules", cellular

components such as " cell-cell junction ", " collagen trimer " and " collagen-containing extracellular matrix ", as well as molecular functions like "integrin binding" and "extracellular matrix structural constituent". KEGG pathways including "ECM-receptor interaction," "PI3K-AKT signaling," "ECM-receptor interaction",and "Focal adhesion" were prominently enriched, suggesting their involvement in tumor microenvironment remodeling and oncogenic signaling (Figure 3D). GSEA (Figure 3E–G) further illuminated distinct functional landscapes: Hallmark gene sets (Figure 3E) showed enrichment in "Angiogenesis" and "Epithelial-Mesenchymal Transition (EMT)," aligning with pro-tumor microenvironment remodeling. GO-related gene sets (Figure 3F) emphasized "Wnt signaling pathway" and "integrin/collagen-mediated signaling," while KEGG GSEA (Figure 3G) underscored activation of "Calcium signaling" and "Focal adhesion" pathways. These integrative analyses clarify that C1orf198 influences CRC progression through coordinated regulation of ECM-adhesion networks, oncogenic signaling (e.g., PI3K-AKT, Wnt signaling pathway), and tumor vasculature formation, with potential roles in EMT and microenvironmental crosstalk.

Figure 3. (A) Heatmap depicting the top 20 genes in CRC with significant correlation (positive) to C1orf198, where red and blue represent high and low expression-related correlation levels, respectively. (B) GO analysis of co-expressed genes with C1orf198, illustrating biological processes and molecular functions associated with these co-expressed genes. (C) Volcano plot showing DEGs between low and high C1orf198 expression subgroups in CRC. Red denotes up-regulated genes in high C1orf198 expression, while blue represents down-regulated genes. (D) GO and KEGG Pathway enrichment analyses for up-regulated and down-regulated DEGs, highlighting significantly enriched biological pathways and functions. (E-G) GSEA plots. (E) Hallmark gene sets, (F) GO-related gene sets, and (G) KEGG Pathway-related gene sets, showing pathways significantly correlated with C1orf198 expression in CRC.

### Correlation between C1orf198 Expression and Immune Cells in CRC

To explore the correlation between C1orf198 and immune cells in CRC, the ESTIMATE algorithm was employed (Figure 4A-C). Results showed positive correlations between C1orf198 expression and the ESTIMATE score ( $R=0.220$ ,  $P<0.001$ ), stromal score ( $R=0.290$ ,  $P<0.001$ ), and immune score ( $R=0.198$ ,  $P=0.006$ ), indicating its association with tumor microenvironment components. Using the ssGSEA, C1orf198 was found to significantly correlate with multiple immune cells. Notably, C1orf198 expression exhibited a strong correlation with macrophage infiltration ( $R=0.345$ ,  $P<0.001$ ), particularly with M2 macrophages ( $R=0.156$ ,  $P<0.001$ ) with CIBERSORT,

which serve as key components of the TME (Figure 4D–E). M2 macrophages are known to modulate multiple pro-tumor processes, including angiogenesis, extracellular matrix (ECM) remodeling, cancer cell proliferation, metastasis, immunosuppressive signaling, chemotherapeutic resistance, and reduced responsiveness to immune checkpoint blockade therapy. These associations suggest that C1orf198 may interact with M2 macrophage-driven pathways to influence CRC progression and tumor-immune crosstalk. Further analysis (Figure 4F) demonstrated a positive correlation between C1orf198 and macrophages ( $R=0.345$ ,  $P<0.001$ ), while Figure 4G showed its correlation with M2 macrophages ( $R=0.158$ ,  $P<0.001$ ). Associations between C1orf198 expression and macrophage markers CD163 ( $R=0.360$ ,  $P<0.001$ ) and MRC1 ( $R=0.323$ ,  $P<0.001$ ) were also observed (Figure 4H–I). Additionally, C1orf198 expression correlated with other immune checkpoints CD274 ( $R=0.220$ ,  $P<0.001$ ) and PDCD1 ( $R=0.208$ ,  $P<0.001$ ) (Figure 4J, K), further underscoring its role in immune-cell-related processes. Collectively, these results suggest that C1orf198 may influence immune-cell-related processes in CRC, potentially impacting tumor-immune interactions.

Figure 4 The correlation between C1orf198 and immune cells in CRC. (A–C) The ESTIMATE algorithm was used to analyze the association of C1orf198 expression with the ESTIMATE score, stromal score, and immune score. (D, E) The relationship between C1orf198 gene expression and immune cell infiltration in CRC was explored via the ssGSEA and CIBERSORT tool. (F) Correlation analysis between C1orf198 and macrophages. (G) Correlation analysis between C1orf198 and M2 macrophages. (H, I) Associations between C1orf198 expression levels and macrophage markers (CD163, MRC1). (J, K) Associations between C1orf198 expression levels and immune checkpoints (CD274, PDCD1) (ns represents no significance, \*  $p<0.05$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ ).

### Correlation Analysis of C1orf198, Cytokines and Immune-related Factors in CRC

Given the potential of cancer cells to modulate immune cell polarization via chemokines and their receptors, this study investigated the relationship between C1orf198 expression and chemokine/receptor profiles sourced from the TISIDB database. (Figure 5A, B, C). In general, the highest correlation was observed for CXCL12 ( $R=0.390$ ) among cytokines and for CXCR5 ( $R=0.367$ ) among cytokine receptors. We thus further investigated the relationship between CXCL12, CXCR5, and immune cells, and the results were similar to those of the C1orf198 gene itself. CXCL12 significantly correlated with macrophages with ssGSEA ( $R=0.541$ ,  $P<0.001$ ) and

macrophages M2 cells with CIBERSORT( $R=0.267$ ,  $P<0.001$ ) (Figure 5D, E). Similarly, CXCR4 expression strongly correlated with macrophages ( $R=0.651$ ,  $P<0.001$ ) and macrophages M2 cells ( $R=0.255$ ,  $P<0.001$ ) (Figure 5F, G).

Associations between CXCL12, CXCR4 and macrophage markers (CD163, MRC1) were observed (Figure 5H-I). CXCL12 positively correlated with CD163 ( $R=0.598$ ) and MRC1 ( $R=0.573$ ); CXCR4 also showed significant positive correlations with CD163 ( $R=0.701$ ) and MRC1 ( $R=0.626$ ). Finally, correlations between CXCL12, CXCR4 and immune checkpoints (CD274, PDCD1) were found (Figure 5J-K). CXCL12 correlated with CD274 ( $R=0.369$ ) and PDCD1 ( $R=0.288$ ); CXCR4 correlated with CD274 ( $R=0.579$ ) and PDCD1 ( $R=0.520$ ), suggesting their roles in regulating immune checkpoints and tumor immune escape.

Figure 5 (A-C) Correlation analysis of C1orf198 gene expression with cytokines (CCL, CXCL, CCR, CXCR families) in CRC. (D, E) Correlation analysis between CXCL12 expression and immune cell infiltration in CRC. (F, G) Correlation analysis between CXCR4 expression and immune cell infiltration in CRC. (H, I) Associations between the expression levels of cytokines (CXCL12, CXCR4) and macrophage markers (CD163, MRC1). (J, K) Associations between the expression levels of cytokines (CXCL12, CXCR4) and immune checkpoints (CD274, PDCD1).

#### 4. Discussion

CRC represents a biologically diverse malignancy marked by perturbed cytokine signaling networks, which disrupt multiple cellular pathways and drive tumor initiation, progression, and the acquisition of aggressive phenotypic traits<sup>9</sup>. Deciphering the complex molecular mechanisms underpinning CRC pathogenesis is pivotal for advancing early detection methods, optimizing treatment regimens, and enhancing the capacity to modulate disease progression<sup>10</sup>. A critical next step involves identifying novel biological indicators associated with immune cell infiltration landscapes and unraveling the underlying molecular pathways that dictate responses to immunotherapeutic interventions. This approach aims to deepen our understanding of CRC heterogeneity and inform the development of personalized strategies to improve clinical outcomes<sup>11,12</sup>. Currently, researchers have successfully detected and functionally annotated nearly 90% of predicted human proteins. However, approximately 10% remain uncharacterized or poorly annotated, particularly regarding their structural features, biological functions, and roles in disease

contexts. Among this understudied group are chromosome-specific open-reading frame proteins, commonly referred to as CxORFx or ORF proteins. These proteins are encoded by ORF genes, whose naming convention integrates a chromosome number (designated as "Cx") with an open-reading frame identifier ("ORFx"), directly reflecting their genomic location<sup>13</sup>. Presently, increasing evidence suggests that uncharacterized ORF genes may be involved in the occurrence and development of cancer<sup>5,14,15</sup>. As mentioned above, previous studies have shown that C1orf198 plays a role in the development and progression of gastric and breast cancers, but its relationship with CRC remains unclear.

Our study demonstrates that C1orf198 is significantly upregulated in CRC at both mRNA and protein levels, consistent across TCGA, GEO datasets, and clinical IHC samples. High C1orf198 expression correlated with aggressive clinicopathologic features and poor survival outcomes, establishing it as a potential prognostic indicator. These findings align with prior studies linking ECM and adhesion-related genes to CRC metastasis, suggesting C1orf198 may promote tumor progression by enhancing cell motility and invasive capacity. Functional enrichment analyses revealed C1orf198-associated genes were enriched in biological processes such as ECM organization, cell adhesion molecule activity, and PI3K-AKT signaling—pathways critical for tumor cell proliferation, migration, and resistance to apoptosis<sup>16</sup>. The involvement of focal adhesion and ECM-receptor interaction pathways highlights C1orf198's role in mediating cell-matrix interactions, which are pivotal for metastatic dissemination. GSEA further implicated angiogenesis and EMT, key processes in tumor invasion and TME remodeling. TME is a complex ecosystem of immune cells, stromal cells, and signaling molecules that profoundly influences cancer progression and therapeutic response. Our study revealed a significant association between C1orf198 expression and TME composition, particularly immune cell infiltration and stromal remodeling. Using the ESTIMATE algorithm, we observed positive correlations between C1orf198 levels and both stromal scores ( $R=0.290$ ,  $P<0.001$ ) and immune scores ( $R=0.198$ ,  $P=0.006$ ), indicating its role in shaping the structural and immune landscapes of the TME. A key finding was the strong correlation between C1orf198 expression and macrophage infiltration, especially M2-type macrophages—critical mediators of tumor-promoting inflammation. Analysis via the ssGSEA and CIBERSORT tool demonstrated a robust association with macrophage abundance ( $R=0.268$ ,  $P<0.001$ ) especially M2-Macrophage ( $R=0.156$ ,  $P<0.001$ ), further validated by direct correlation with macrophage markers CD163 ( $R=0.360$ ,  $P<0.001$ ) and MRC1 ( $R=0.323$ ,

R values are below 50%; not sure how robust this associations.

P<0.001). M2 macrophages are known to secrete growth factors (e.g., VEGF) and matrix-degrading enzymes, fostering angiogenesis, ECM remodeling, and cancer cell invasion. Additionally, they suppress anti-tumor immune responses by producing interleukin-10 (IL-10) and transforming growth factor- $\beta$  (TGF- $\beta$ ), inhibiting T-cell activation and promoting regulatory T-cell (Treg) recruitment<sup>17</sup>. The enrichment of C1orf198 in M2-related pathways suggests it may drive macrophage polarization toward an immunosuppressive phenotype, creating a niche conducive to tumor progression and resistance to immunotherapy. C1orf198 expression also correlated with immune checkpoint molecules PD-L1 (CD274, R=0.220, P<0.001) and PD-1 (PDCD1, R=0.208, P<0.001), key regulators of T-cell exhaustion. This association implies a potential mechanism by which C1orf198 facilitates tumor immune escape: upregulated PD-L1/PD-1 signaling dampens cytotoxic T-cell responses, allowing tumors to evade immune surveillance<sup>18</sup>. Clinically, high C1orf198 expression may predict poor response to immune checkpoint inhibitors, as observed in patients with immunosuppressive TMEs. The correlation analysis extended to cytokine-receptor networks revealed that C1orf198 was strongly linked to the CXCL12/CXCR5 axis, which is central to immune cell trafficking and TME organization<sup>19</sup>. CXCL12, a chemokine highly correlated with C1orf198 (R=0.390), promotes macrophage recruitment and M2 polarization<sup>20</sup>, while its receptor CXCR5 (R=0.367) modulates lymphocyte homing to lymphoid tissues<sup>21</sup>. Functional follow-up showed that CXCL12 and CXCR5 mirrored C1orf198's associations with M2 macrophages and immune checkpoints, suggesting a coordinated role in orchestrating TME crosstalk. These interactions likely drive a feedforward loop where C1orf198 upregulates chemokine signaling, attracting immunosuppressive cells while repelling cytotoxic T cells, thereby fostering a pro-tumor microenvironment. In summary, C1orf198 emerges as a pivotal node in TME biology, integrating ECM remodeling, immune cell recruitment, and checkpoint signaling to promote a pro-tumor microenvironment. These findings deepen our understanding of CRC-immune interactions and highlight C1orf198 as a candidate for stratified therapy based on TME characteristics.

While our study provides robust bioinformatics and clinical validation, mechanistic investigations (e.g., in vitro/in vivo functional assays) are needed to confirm C1orf198's direct role in CRC cell behavior and immune modulation. Additionally, exploring C1orf198 as a therapeutic target—alone or in combination with immune checkpoint inhibitors—warrants further study, particularly in subsets of patients with high TME infiltration.

## 5. Conclusions

In conclusion, this study identifies C1orf198 as a novel prognostic biomarker in CRC, strongly associated with tumor progression, ECM remodeling, and immunosuppressive TME formation. Our findings highlight its potential as a therapeutic target for disrupting oncogenic signaling and immune evasion, paving the way for personalized CRC treatment strategies. Further functional studies are essential to unravel the precise mechanisms by which C1orf198 influences CRC pathogenesis and immune crosstalk.

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                                         |
|------|-----------------------------------------|
| CRC  | Colorectal cancer                       |
| TCGA | The Cancer Genome Atlas                 |
| GEO  | Gene Expression Omnibus                 |
| IHC  | Immunohistochemistry                    |
| OS   | Overall survival                        |
| DSS  | Disease-specific survival               |
| PFI  | Progress free interval                  |
| TME  | tumor microenvironment                  |
| ECM  | Extracellular matrix                    |
| DFS  | Disease-free survival                   |
| DEGs | Differentially expressed genes          |
| GO   | Gene Ontology                           |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| GSEA | Gene set enrichment analysis            |
| MF   | Molecular functions                     |
| CC   | Cellular components                     |
| BP   | Biological processes                    |
| NES  | Normalized enrichment scores            |
| ROC  | Receiver operating characteristic       |
| AUC  | Area under the curve                    |

## Declarations

**Ethics approval:** Supplied by Shanghai Outdo Biotech Company (Shanghai, China), the colorectal cancer (CRC) tissue microarray (HColA160CS01) comprised 80 paired tumor and adjacent normal tissue samples. Ethical approval for the study protocols was granted by the company's Ethics Committee under the approval ID: YB M-05-02.

**Consent for publication:** Not applicable

**Availability of data and materials:** The complete dataset exists in accessible, public repositories. The information utilized and/or examined in this research can be accessed through the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and The Cancer Genome Atlas (TCGA) network (<https://cancergenome.nih.gov/>).



**Competing interests:** The authors have no conflicts of interest to declare 318

**Funding:** Not applicable 319

**Acknowledgements** Not applicable 320

# References 322

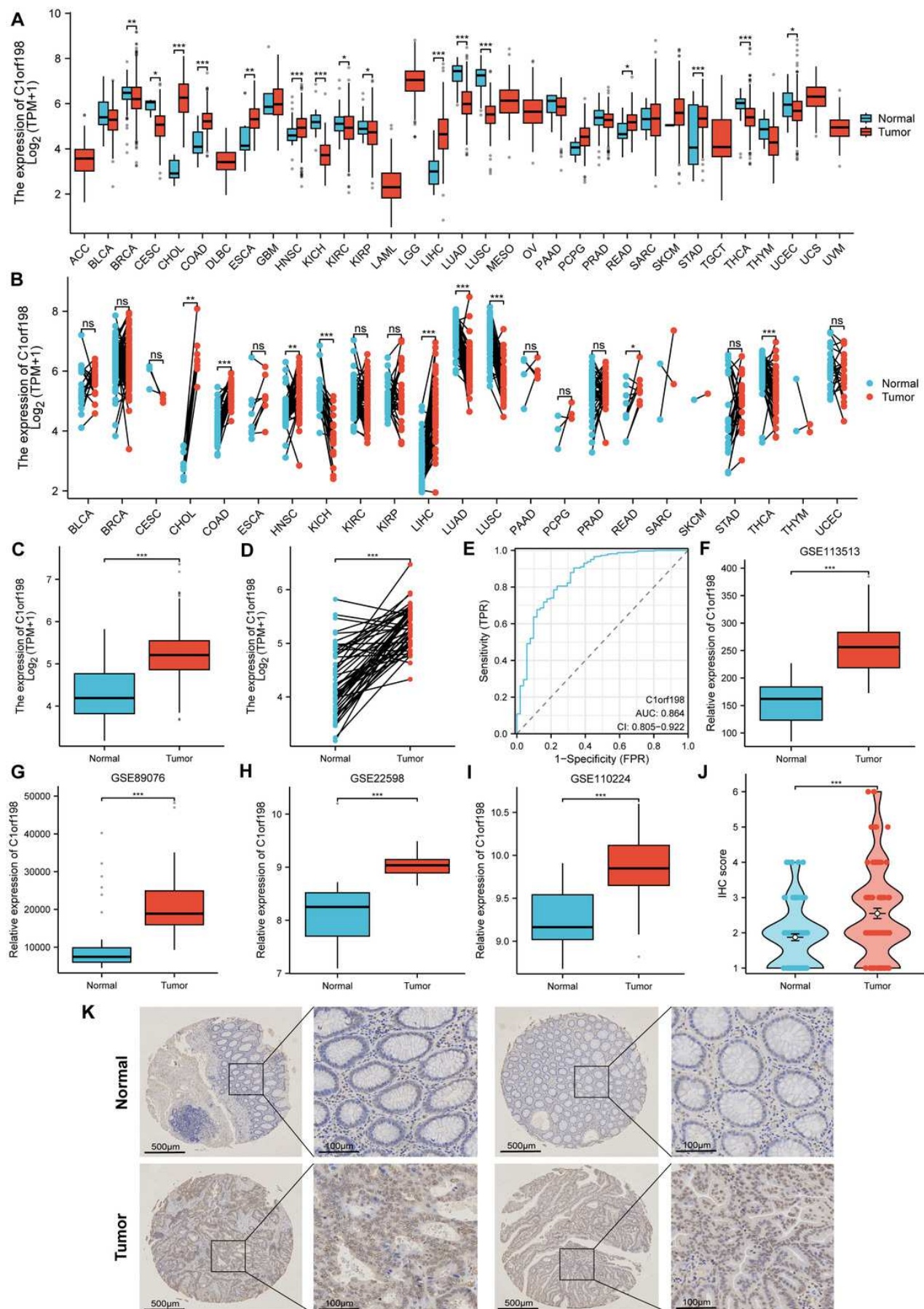
1. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet (London, England)*. 2019;394(10207):1467-1480. 323 324 325
2. Baidoun F, Elshiwly K, Elkeraie Y, et al. Colorectal Cancer Epidemiology: Recent Trends and Impact on Outcomes. *Current drug targets*. 2021;22(9):998-1009. 326 327
3. Li J, Ma X, Chakravarti D, Shalapour S, DePinho RA. Genetic and biological hallmarks of colorectal cancer. *Genes & development*. 2021;35(11-12):787-820. 328 329
4. Duek P, Mary C, Zahn-Zabal M, Bairoch A, Lane L. Functionathon: a manual data mining workflow to generate functional hypotheses for uncharacterized human proteins and its application by undergraduate students. *Database : the journal of biological databases and curation*. 2021;2021. 330 331 332 333
5. Ma J, Liu C, Yang D, et al. C1orf106, an innate immunity activator, is amplified in breast cancer and is required for basal-like/luminal progenitor fate decision. *Science China Life sciences*. 2019;62(9):1229-1242. 334 335 336
6. Wang Z, Chen G, Wang Q, Lu W, Xu M. Identification and validation of a prognostic 9-genes expression signature for gastric cancer. *Oncotarget*. 2017;8(43):73826-73836. 337 338 339
7. Lee H, Lee S, Jeong D, Kim SJ. Ginsenoside Rh2 epigenetically regulates cell-mediated immune pathway to inhibit proliferation of MCF-7 breast cancer cells. *Journal of ginseng research*. 2018;42(4):455-462. 340 341 342
8. Bindea G, Mlecnik B, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39(4):782-795. 343 344 345
9. Ahmad R, Singh JK, Wunnavu A, Al-Obeed O, Abdulla M, Srivastava SK. Emerging trends in colorectal cancer: Dysregulated signaling pathways (Review). *International journal of molecular medicine*. 2021;47(3). 346 347 348
10. Abedizadeh R, Majidi F, Khorasani HR, Abedi H, Sabour D. Colorectal cancer: a comprehensive review of carcinogenesis, diagnosis, and novel strategies for classified treatments. *Cancer metastasis reviews*. 2024;43(2):729-753. 349 350 351
11. Zhou Y, Tao L, Qiu J, et al. Tumor biomarkers for diagnosis, prognosis and targeted therapy. *Signal transduction and targeted therapy*. 2024;9(1):132. 352 353

12. Underwood PW, Ruff SM, Pawlik TM. Update on Targeted Therapy and Immunotherapy for Metastatic Colorectal Cancer. *Cells*. 2024;13(3). 354 355
  13. Ershov P, Yablokov E, Mezentsev Y, Ivanov A. Uncharacterized Proteins CxORFx: Subinteractome Analysis and Prognostic Significance in Cancers. *International journal of molecular sciences*. 2023;24(12). 356 357 358
  14. Gao J, Wang Y, Zhang W, et al. C20orf27 Promotes Cell Growth and Proliferation of Colorectal Cancer via the TGF $\beta$ R-TAK1-NF $\kappa$ B Pathway. *Cancers*. 2020;12(2). 359 360
  15. Tang Y, Liao S, Liu G, et al. Advanced single-cell pooled CRISPR screening identifies C19orf53 required for cell proliferation based on mTORC1 regulators. *Cell biology and toxicology*. 2022;38(1):43-68. 361 362 363
  16. Yu L, Wei J, Liu P. Attacking the PI3K/Akt/mTOR signaling pathway for targeted therapeutic treatment in human cancer. *Seminars in cancer biology*. 2022;85:69-94. 364 365
  17. Li M, Yang Y, Xiong L, Jiang P, Wang J, Li C. Metabolism, metabolites, and macrophages in cancer. *Journal of hematology & oncology*. 2023;16(1):80. 366 367
  18. Jiang X, Wang J, Deng X, et al. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. *Molecular cancer*. 2019;18(1):10. 368 369
  19. Hussain M, Adah D, Tariq M, Lu Y, Zhang J, Liu J. CXCL13/CXCR5 signaling axis in cancer. *Life sciences*. 2019;227:175-186. 370 371
  20. Babazadeh S, Nassiri SM, Siavashi V, Sahlabadi M, Hajinasrollah M, Zamani-Ahmadm Mahmudi M. Macrophage polarization by MSC-derived CXCL12 determines tumor growth. *Cellular & molecular biology letters*. 2021;26(1):30. 372 373 374
  21. Elzein SM, Zimmerer JM, Han JL, Ringwald BA, Bumgardner GL. CXCR5(+)CD8(+) T cells: A Review of their Antibody Regulatory Functions and Clinical Correlations. *Journal of immunology (Baltimore, Md : 1950)*. 2021;206(12):2775-2783. 375 376 377 378
- 379

# Figure 1

Figure 1 The levels of C1orf198 protein and mRNA expression in pan-cancer and CRC relative to normal samples.

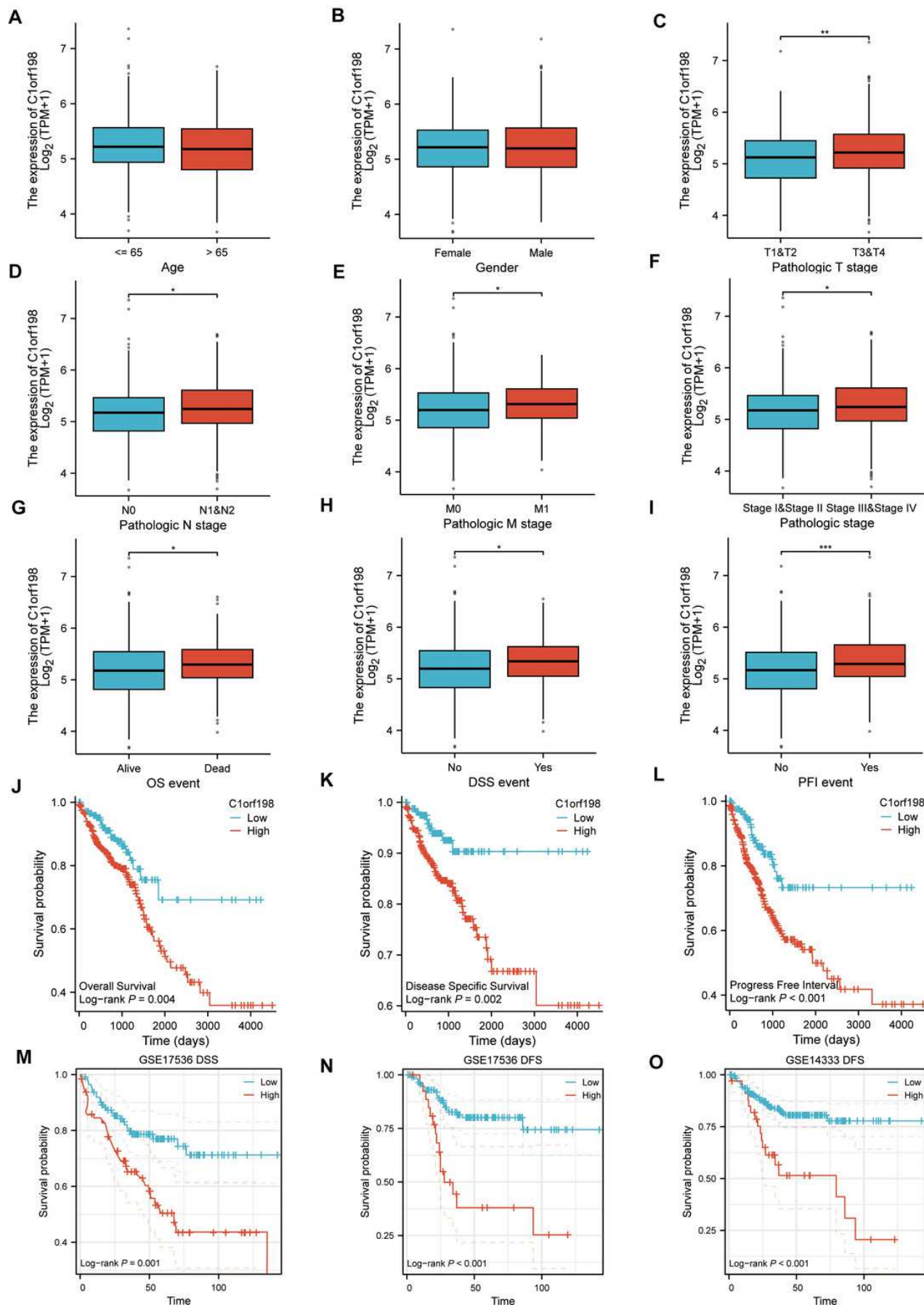
(A, B) Comparative analysis of C1orf198 mRNA expression in pan-cancer and normal tissues using TCGA. (C, D) TCGA-based box-plot and paired-sample analysis showing that C1orf198 mRNA expression was up-regulated in CRC tissues versus normal tissues. (E) ROC analysis evaluating the diagnostic value of C1orf198 in CRC. (F-I) Analyses of GEO datasets (GSE113513, GSE89076, GSE22598, GSE110224) revealed up-regulation of C1orf198 mRNA expression in CRC compared with normal samples. (J) IHC score analysis showing that C1orf198 protein expression was up-regulated in CRC compared with normal samples. (K) Representative immunohistochemistry images of C1orf198 in normal and CRC tissues (ns denotes no significance, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).



# Figure 2

Figure 2. TCGA database-based correlations between C1orf198 expression and clinicopathologic parameters and prognosis in CRC.

(A-I) Analysis of the association between C1orf198 expression and clinicopathologic parameters in CRC using TCGA data, demonstrating associations with age (A), gender (B), T stage (C), N stage (D), M stage (E), pathologic stage (F), OS event (G), DSS event (H), and PFI event (I). (J-L) Kaplan-Meier curves based on TCGA data showing that up-regulated C1orf198 expression was associated with reduced OS, DSS, and PFI times. (M-O) Kaplan-Meier curves from GEO datasets (GSE17536 and GSE14333) indicating that up-regulated C1orf198 expression was related to reduced DSS/DFS times. (ns represents no significance, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).

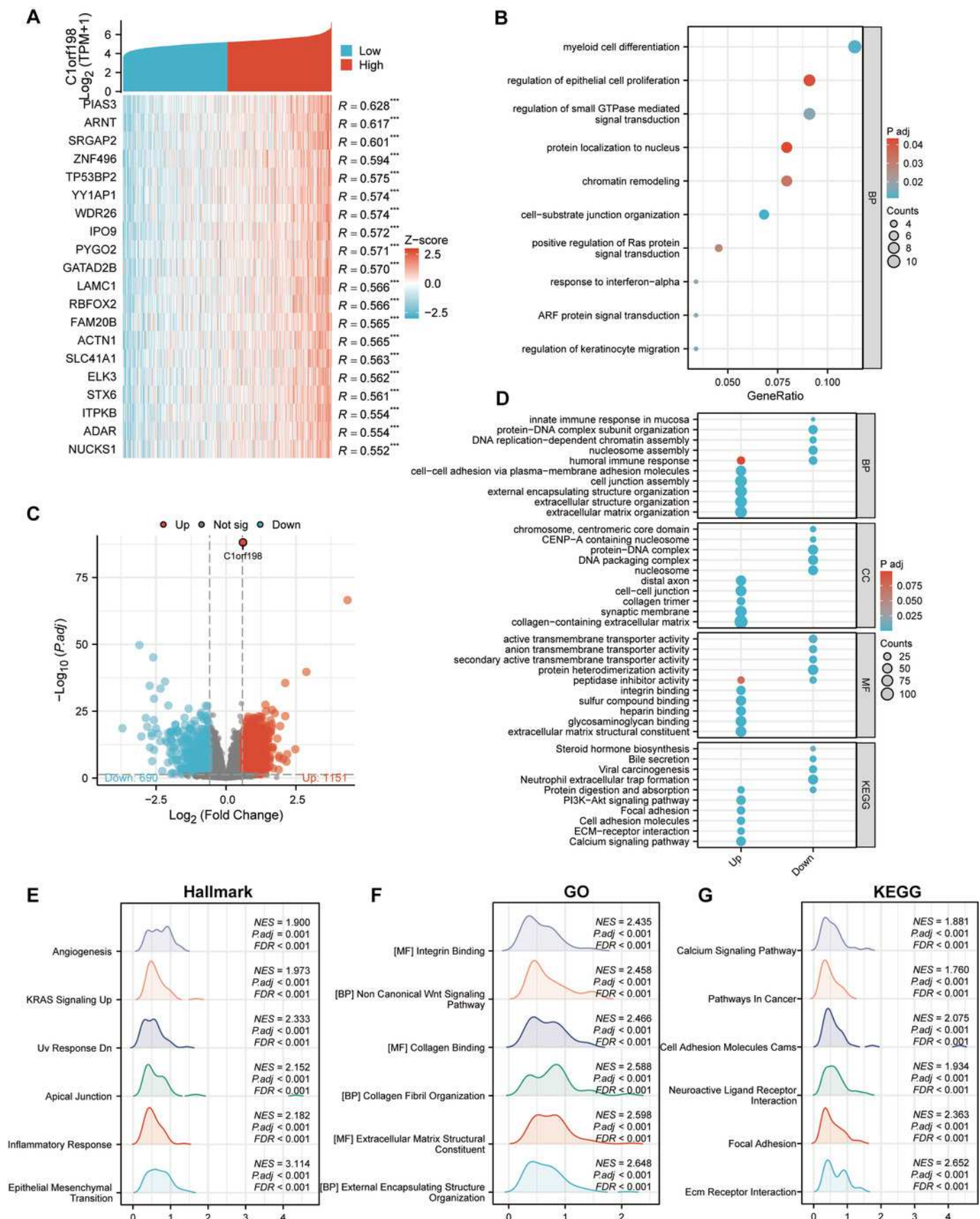


# Figure 3

Figure 3 Comprehensive analysis of C1orf198-correlated gene expression, functional enrichment, and pathway associations in CRC

Figure 3. (A) Heatmap depicting the top 20 genes in CRC with significant correlation (positive) to C1orf198, where red and blue represent high and low expression-related correlation levels, respectively. (B) GO analysis of co-expressed genes with C1orf198, illustrating biological processes and molecular functions associated with these co-expressed genes. (C) Volcano plot showing DEGs between low and high C1orf198 expression subgroups in CRC. Red denotes up-regulated genes in high C1orf198 expression, while blue represents down-regulated genes. (D) GO and KEGG Pathway enrichment analyses for up-regulated and down-regulated DEGs, highlighting significantly enriched biological pathways and functions. (E-G) GSEA plots. (E) Hallmark gene sets, (F) GO-related gene sets, and (G) KEGG Pathway-related gene sets, showing pathways significantly correlated with C1orf198 expression in CRC.



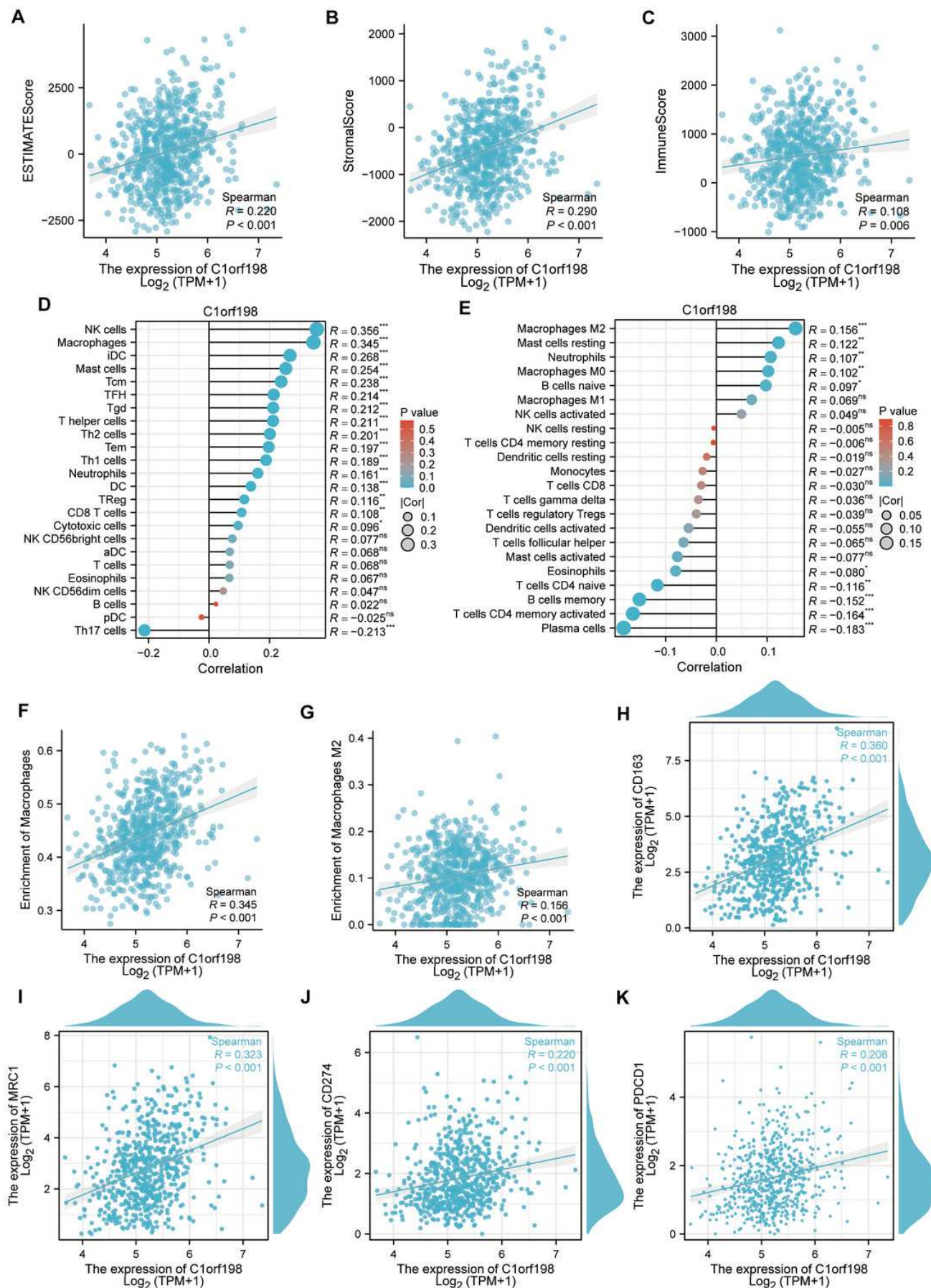




# Figure 4

Figure 4 The correlation between C1orf198 and immune cells in CRC.

(A-C) The ESTIMATE algorithm was used to analyze the association of C1orf198 expression with the ESTIMATE score, stromal score, and immune score. (D, E) The relationship between C1orf198 gene expression and immune cell infiltration in CRC was explored via the ssGSEA and CIBERSORT tool. (F) Correlation analysis between C1orf198 and macrophages. (G) Correlation analysis between C1orf198 and M2 macrophages. (H, I) Associations between C1orf198 expression levels and macrophage markers (CD163, MRC1). (J, K) Associations between C1orf198 expression levels and immune checkpoints (CD274, PDCD1) (ns represents no significance, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ).



# Figure 5

Figure 5. Correlation analysis of C1orf198 expression with cytokine families (CCL/CXCL/CCR/CXCR), associations with immune cell infiltration, macrophage markers (CD163, MRC1), and immune checkpoints (CD274, PDCD1) in CRC.

Figure 5 (A-C) Correlation analysis of C1orf198 gene expression with cytokines (CCL, CXCL, CCR, CXCR families) in CRC. (D, E) Correlation analysis between CXCL12 expression and immune cell infiltration in CRC. (F, G) Correlation analysis between CXCR4 expression and immune cell infiltration in CRC. (H, I) Associations between the expression levels of cytokines (CXCL12, CXCR4) and macrophage markers (CD163, MRC1). (J, K) Associations between the expression levels of cytokines (CXCL12, CXCR4) and immune checkpoints (CD274, PDCD1).



