

Identification of gut microbes-related molecular subtypes and their biomarkers in colorectal cancer

Xuliang Liu^{1,*}, Guolin Zhang^{2,*}, Shiyao Li^{3,*}, Yuechuan Liu¹, Kexin Ma¹, Liming Wang^{1,4,5}

¹Department of General Surgery, Division of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, China

²Department of Cardiovascular Medicine, The Second Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, China

³Department of Respiratory Medicine, The Second Affiliated Hospital of Dalian Medical University, Dalian, Liaoning, China

⁴Engineering Research Center for New Materials and Precision Treatment Technology of Malignant Tumors Therapy, The Second Affiliated Hospital, Dalian Medical University, Dalian, Liaoning, China

⁵Engineering Technology Research Center for Translational Medicine, The Second Affiliated Hospital, Dalian Medical University, Dalian, Liaoning, China

*Equal contribution

Correspondence to: Liming Wang, Kexin Ma; email: wangbcc259@163.com, <https://orcid.org/0000-0003-0914-6113>; forevermkx@163.com, <https://orcid.org/0009-0005-7864-0416>

Keywords: colorectal cancer, gut microbes, immune infiltration, IL7, BCL10

Received: July 27, 2023

Accepted: December 6, 2023

Published: January 29, 2024

Copyright: © 2024 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

The role of gut microbes (GM) and their metabolites in colorectal cancer (CRC) development has attracted increasing attention. Several studies have identified specific microorganisms that are closely associated with CRC occurrence and progression, as well as key genes associated with gut microorganisms. However, the extent to which gut microbes-related genes can serve as biomarkers for CRC progression or prognosis is still poorly understood. This study used a bioinformatics-based approach to synthetically analyze the large amount of available data stored in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. Through this analysis, this study identified two distinct CRC molecular subtypes associated with GM, as well as CRC markers related to GM. In addition, these new subtypes exhibit significantly different survival outcomes and are characterized by distinct immune landscapes and biological functions. Gut microbes-related biomarkers (GMRBs), IL7 and BCL10, were identified and found to have independent prognostic value and predictability for immunotherapeutic response in CRC patients. In addition, a systematic collection and review of prior research literature on GM and CRC provided additional evidence to support these findings. In conclusion, this paper provides new insights into the underlying pathological mechanisms by which GM promotes the development of CRC and suggests potentially viable solutions for individualized prevention, screening, and treatment of CRC.

INTRODUCTION

Colorectal cancer (CRC) ranks as the third most prevalent cancer worldwide and represents a significant global health challenge, being the second leading cause of cancer-related mortality. The incidence and mortality

rates of CRC are anticipated to rise substantially, with projections estimating 3.2 million new cases and 1.6 million deaths by 2040, particularly in countries with a high Human Development Index [1–3]. As with many diseases, CRC is the result of the accumulation of multifactorial perturbations that include genetic,

epigenetic, and environmental aspects [4]. In particular, environmental factors such as consumption of carcinogenic foods, lack of physical activity, and cigarette smoking are known to assume a pivotal role in the intricate web of CRC progression [5].

Among the environmental factors, there has been increasing recognition of the role of the collective microbial community in the tumor environment. Our colon harbors about 30 trillion bacteria, which constantly interact with our intestinal epithelium. Alterations in the quantity of these microbes can disrupt gut physiology and lead to diseases [6]. Enrichment of several gut microbes (GM), such as *Fusobacterium nucleatum*, *Peptostreptococcus anaerobius*, and *enterotoxigenic Bacteroides fragilis*, has been found to enhance the development of CRC by promoting inflammation [7], DNA damage [8], tumor growth [9], and immune evasion [9]. Conversely, certain bacteria, primarily probiotics including *Bifidobacterium adolescentis* (ATCC 15703) [10] and *Faecalibacterium prausnitzii* (DSM 17677) [11, 12], are observed to be diminished in individuals with CRC and are believed to confer a safeguarding influence against CRC. In addition, there is growing evidence that dysbiosis of gut microbial ecology can influence the development of CRC by affecting the functional state of host cells. Interacting with Toll-like receptors, *Bifidobacterium* leads to the release of interleukin 10 (IL10) and expression of forkhead box P3 (FOXP3) regulatory T cells (Tregs) [13]. Meanwhile, *Lactobacillus rhamnosus GG* suppresses Th17 cell expression and cytokine secretion of interleukin 23(IL23) and interleukin 17(IL17). It also promotes a shift in macrophage phenotype from pro-inflammatory M1 to immunosuppressive M2 by inhibiting nuclear factor kappa B (NF- κ B) subunit 1 and signal transducer and activator of transcription 3(STAT3) signaling [14]. Additionally, *Lactobacillus reuteri* activates aryl hydrocarbon receptor (AHR) and influences the levels of interleukin 6(IL6), interleukin 12(IL12), tumor necrosis factor (TNF) and interleukin 7(IL7), as well as macrophage and T-lymphocyte activation status [15]. A recent investigation has revealed that *Bifidobacteria*, known for their protective role against CRC, possess the capability to modulate dendritic cell (DC) maturation as well as tumor-specific T-cell responses. This intriguing finding suggests that *Bifidobacteria* may exert an impactful influence on the therapeutic efficacy of anti-CRC immunization and anti-programmed death ligand 1(PDL1) therapy [16].

Due to the intricate interplay between CRC and gut microbiome, existing studies have predominantly focused on exploring the use of particular microbiota and their metabolites as predictive markers of clinical

response or cancer progression through microbiome analysis [17, 18]. However, the contribution of gut microbes-related genes (GMRGs) to the pathogenesis and prognosis of CRC remains unknown, and their potential as immunotherapy targets is yet to be fully understood. In our analysis, we conducted a comprehensive literature review of correlations between gut microbiota and host genes using the PubMed database, resulting in the identification of 164 GMRGs. Based on these genes, we developed a new subtype of CRC, revealing significant differences between various subtypes in terms of survival prognosis, functional enrichment, immune infiltration, and immunotherapy efficacy. Furthermore, we employed univariate and multivariate Cox regression analysis for identifying hub genes highly associated with CRC patient prognosis among GMRGs. By analyzing the expressed core genes in healthy humans, colorectal inflammation patients, and CRC patients, we evaluated their effectiveness as predictive indicators for the occurrence, development, and outcome of CRC.

The core objective of this study is to identify GM-related molecular subtypes and biomarkers to understand the complexity of CRC in terms of molecular biology, immunology, drug sensitivity, survival prognosis, and disease dynamics, aiming to promote more precise treatment planning to improve the survival rate and quality of life of patients.

RESULTS

Identification of GM molecular subtypes in CRC

Figure 1 shows how the study works. A total of 164 GMRGs were searched and screened in the PubMed database, and their expression levels were compared between 332 cancer and 41 normal samples from the TCGA-COAD dataset (Supplementary Figure 1). A noticeable discrepancy in transcript levels was observed between the two subgroups, as shown in the heatmap. The principal component analysis (PCA) plot further confirmed the existence of differential expression of GMRGs. Using a consensus clustering algorithm, the 332 CRC patients were classified into two subtypes with optimal cluster stability ($K = 2$) (Supplementary Figure 2). Among them, 203 patients were assigned to the C1 cluster and 129 to the C2 cluster. The Kaplan-Meier (K-M) analysis revealed that overall survival (OS) was significantly longer in subtype C1 compared to subtype C2 ($P = 0.043$, Figure 2A). The transcriptomes of the two subtypes were significantly different, as shown by PCA (Figure 2B) and heatmap (Figure 2C). To validate the results, the GSE87211 dataset was used as a validation set, which showed similar results (Figure 2D–2F). Similarly, the

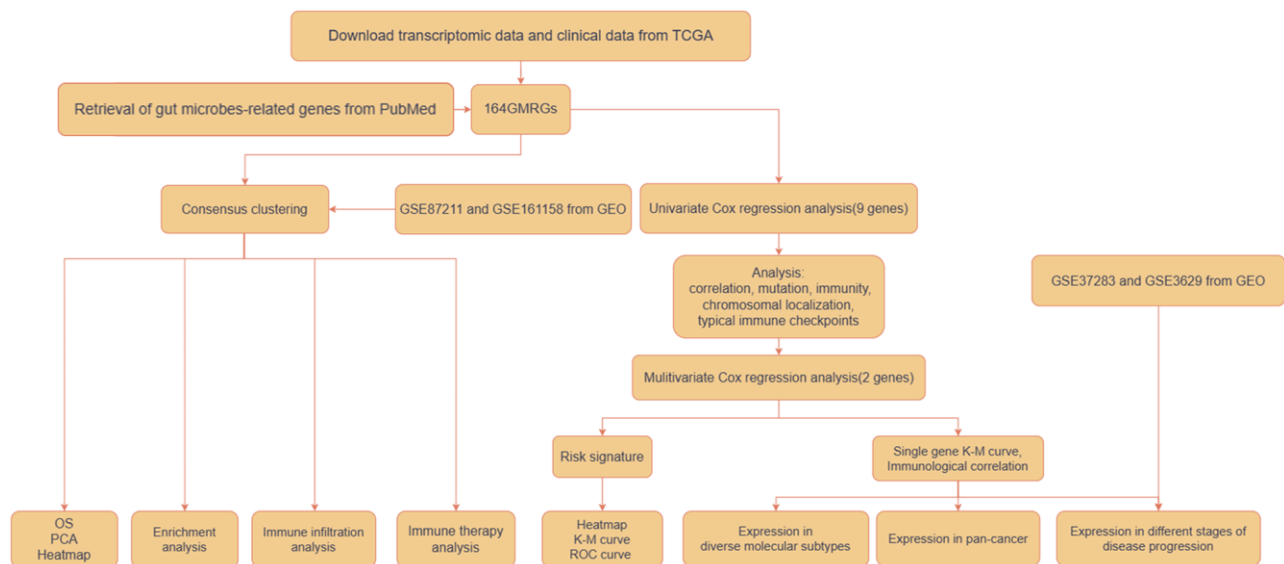


Figure 1. Workflow of the analytic process.

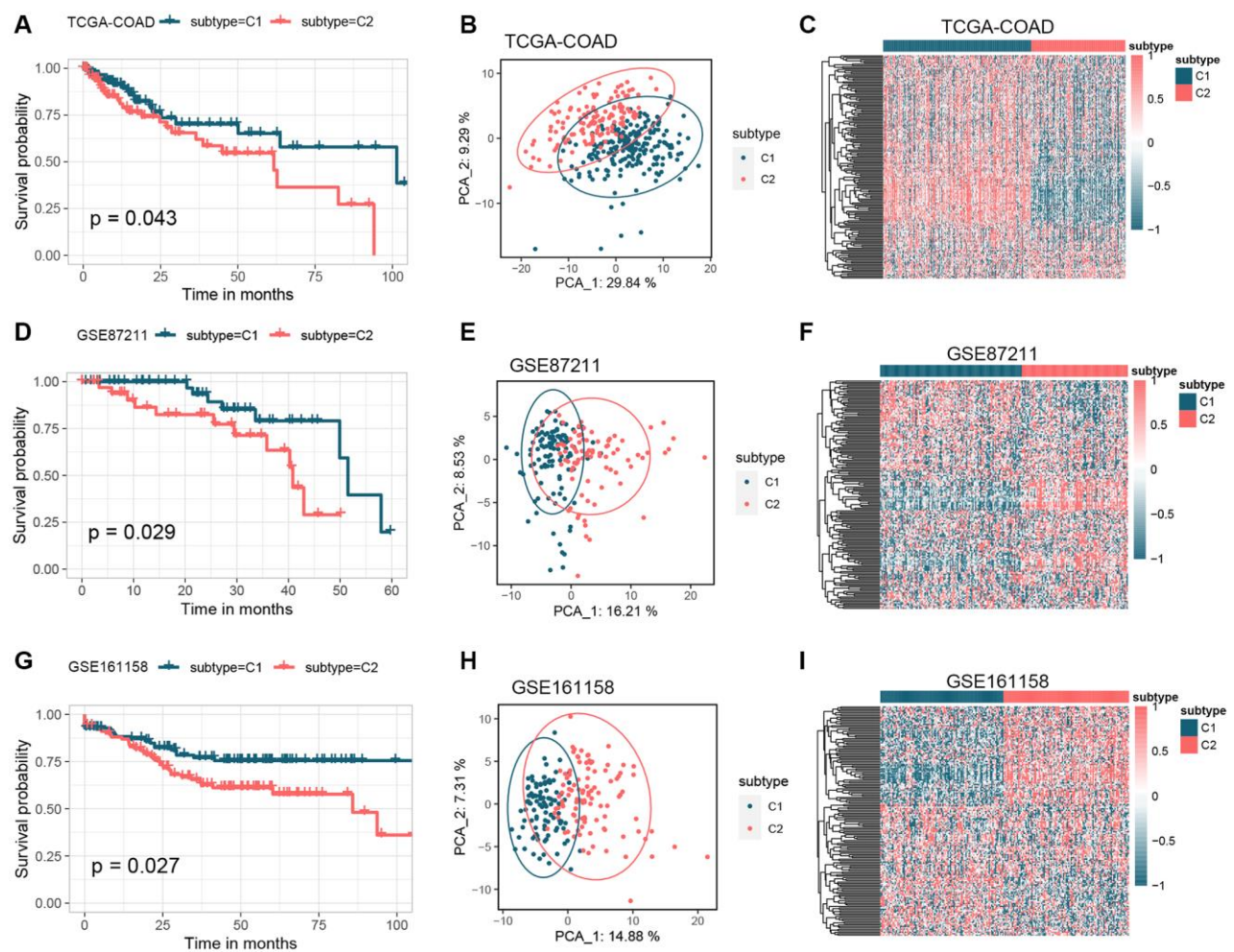


Figure 2. Identification of GM molecular subtypes in CRC. (A–C) K-M analysis, PCA analysis, and heatmap between subtype C1 and subtype C2 in the TCGA-COAD cohort. (D–F) K-M analysis, PCA analysis, and heatmap between subtype C1 and subtype C2 in the GSE87211 cohort. (G–I) K-M analysis, PCA analysis, and heatmap between subtype C1 and subtype C2 in the GSE161158 cohort. Abbreviations: GM: gut microbes; CRC: colorectal cancer; K-M: Kaplan-Meier; PCA: principal component analysis.

GSE161158 dataset was used as a test set, and the consensus clustering results confirmed that the C1 subtype had a better prognosis than the C2 subtype ($P = 0.027$, Figure 2G). The PCA (Figure 2H) and heatmap (Figure 2I) analyses also demonstrated significant differences between the two subtypes.

Functional annotation of GM molecular subtypes

To understand the functional patterns of two colon cancer subtypes based on GMRGs expression, we analyzed their distribution across the consensus molecular subtypes (CMS) of CRC (Figure 3A). We observed that C1 subtypes were more prevalent in CMS1 and CMS4 than C2 subtypes, while C2 subtypes were more prevalent in CMS2. To gain insight into the potential associations between these two classifications and aid in clinical translation, we conducted functional enrichment. We performed a differential analysis of the two subtypes to obtain differentially expressed genes (DEGs) in the different types (Supplementary Figure 3). Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome enrichment analysis revealed that GMRGs were primarily enriched in the developmental maturation of the immune system and intra- and extracellular signaling, among others (Figure 3B, 3C). Using the Gene Ontology (GO) database for gene set enrichment analysis (GSEA) analysis (Figure 3D), we found that enriched pathways in the C1 subtype were mainly associated with the regulation of inflammatory immune response, cellular responses to various stimuli, and angiogenesis regulation. Meanwhile, enriched pathways in the C2 subtype were mainly associated with Wnt signaling, epithelial-mesenchymal transition, apoptosis, and metabolic regulation. We further performed gene set variation analysis (GSVA) analysis using the GO database (Figure 3E), which showed that enriched pathways in the C1 subtype were mainly associated with immune regulation, including macrophage fusion, T cell and plasma cell differentiation, and binding of vascular endothelial growth factor and its regulation. The enriched pathways in the C2 subtype were mainly associated with ribosome assembly, biological enzyme activity, and cell cycle regulation.

Evaluation of immune features and immunotherapy for GM molecular subtypes

To further characterize tumor microenvironment (TME) in GM subtypes, we conducted CIBERSORT analysis to evaluate the distribution of 22 tumor-infiltrating immune cells (TICs) in the COAD cohort (Figure 4A). Significant disparities in the abundance of distinct immune cell populations were observed between the C1 and C2 subtypes. The C1 subtype showed higher levels

of M1 macrophages, T cells follicular helper, and B cells native compared to the C2 subtype. In contrast, regulatory T cells, plasma cells, and monocytes were more abundant in the C2 subtype compared to the C1 subtype (Figure 4B). To further evaluate TME, the ESTIMATE algorithm was employed to calculate the ESTIMATE score, immune score, and stromal score. The findings revealed that the C1 subtype exhibited a significantly elevated ESTIMATE score ($P < 0.001$), immune score ($P < 0.001$), and stromal score ($P < 0.001$) in comparison to the C2 subtype (Figure 4C). Subsequently, the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm was utilized to predict the response to immunotherapy in CRC patients with varying levels of immune infiltration and GM subtypes. By stratifying responders and non-responders based on the median TIDE score, a higher proportion of responders was observed in the C2 subtype characterized by low immune infiltration and low TIDE scores (Figure 4D). This suggests that the C2 subtype, characterized by lower immune infiltration, may exhibit a more favorable response to immunotherapy. Furthermore, we examined the expression of immune checkpoint markers as potential indicators of immunosuppression. The analysis revealed a higher proportion of samples in the C1 subtype with high expression of programmed cell death 1 (PD1) and cytotoxic T-lymphocyte associated protein 4 (CTLA4), while the C2 subtype had a higher proportion of samples with low expression of immune checkpoint markers (Figure 4E). Lastly, we performed drug sensitivity predictions to identify potential therapeutic options for each subtype. The analysis showed that the C2 subtype was more sensitive to Vorinostat and Sorafenib, while the C1 subtype exhibited higher sensitivity to Bortezomib and Temsirolimus (Figure 4F).

Identification and analysis of prognosis-related GMRGs

We conducted univariate Cox regression analysis on 164 GMRGs and identified 9 prognostic genes in the TCGA-COAD cohort ($P < 0.05$) (Supplementary Table 1). Subsequently, we examined the correlation between these 9 genes (Figure 5A) to assess their association. We also analyzed the somatic mutation frequencies of these 9 GMRGs and found relatively high mutation rates in the TCGA-COAD samples (Figure 5B). Among the samples, 61 (13.41%) had mutations in the GMRGs, with MTOR showing the highest mutation frequency (7%), followed by MMP9, NPC1L1, PKN2, PTGS2, SULT2B1, BCL10, STAT3, and IL7. Furthermore, we explored the relationship between these 9 prognostic genes and immune cells to investigate the interplay between gene expression and immune functions (Figure 5C). We also determined the chromosomal locations of

these genes (Figure 5D). To examine whether changes in gene expression directly impact checkpoint activity or vice versa, we assessed the correlation between the 9

prognostic genes and immune checkpoints (Figure 5E). Interestingly, *SULT2B1* exhibited a significant negative correlation with most of the checkpoints, while *MMP9*

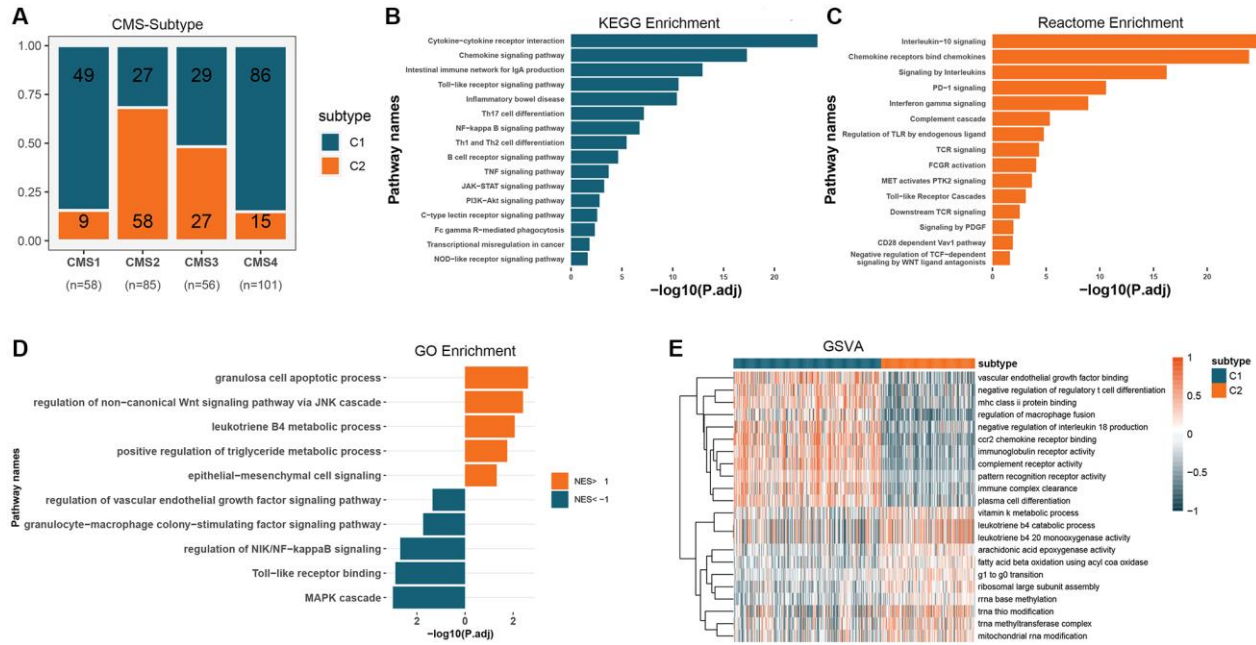


Figure 3. Functional expression patterns of GM molecular subtypes in CRC. (A) Distribution of GM molecular subtypes in CRC consensus subtypes. (B) KEGG enrichment analysis of DEGs in GM molecular subtypes. (C) Reactome enrichment analysis of DEGs in GM molecular subtypes. (D) GSEA analysis of GM molecular subtypes using the GO database. (E) GSEA analysis of GM molecular subtypes using the GO database. Abbreviations: GM: gut microbes; CRC: colorectal cancer; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: differentially expressed genes; GSEA: gene set enrichment analysis; GO: Gene Ontology; GSEA: gene set variation analysis.

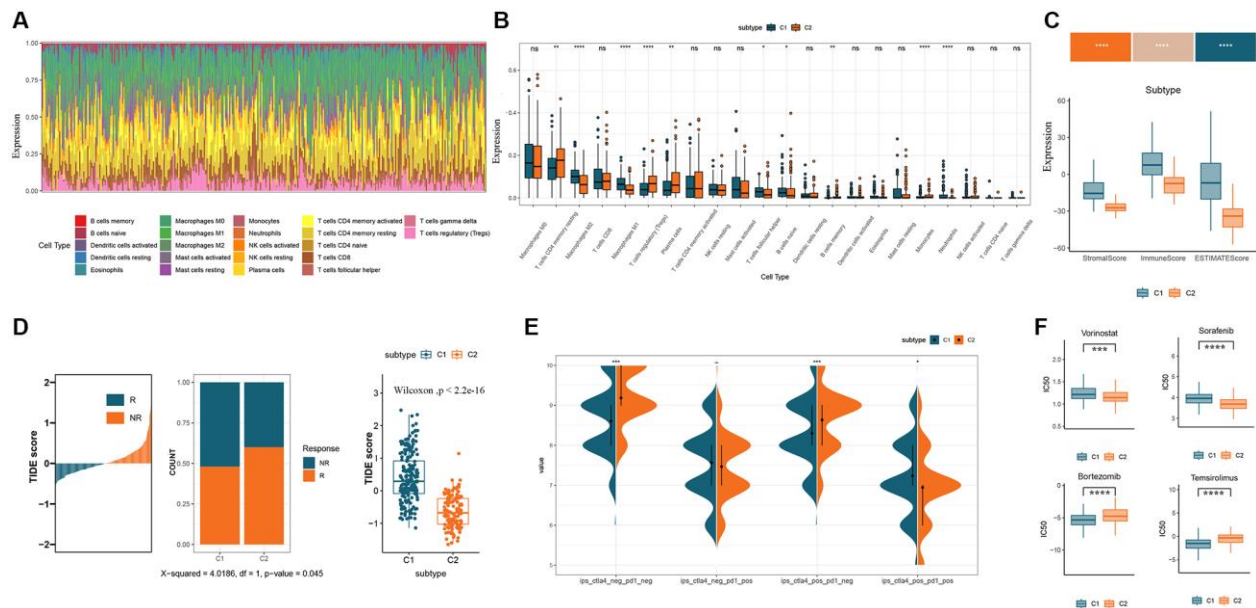


Figure 4. Immunological characterization and evaluation of immunotherapy for GM molecular subtypes. (A) Relative content distribution of 22 TICs in COAD samples. (B) Box plots showing the difference in abundance of 22 TICs between the two subtypes. (C) Comparison of stromal, immune, and ESTIMATE scores in the two subtypes. (D) Estimation of immunotherapy effect in the two subtypes using the TIDE algorithm. (E) Immune checkpoint expression in two GM subtypes. (F) Box plots of the estimated IC50 of chemotherapy drugs between the two GM subtypes. Abbreviations: GM: gut microbes; TICs: tumor-infiltrating immune cells; TIDE: Tumor Immune Dysfunction and Exclusion; IC50: semi-inhibitory concentration.

showed a strong positive correlation with most of the immune checkpoints.

Construction and evaluation of GM risk signature

The prognostic signature based on 2 GMRGs was established using stepwise multivariate Cox regression analysis (Figure 6A). The prognostic model formula was as follows: risk score = $(-0.044870355898046 \times \text{expression level of IL7}) + (-0.00841862944423366 \times \text{expression level of BCL10})$. Using the median risk score, CRC patients were divided into high-risk ($n = 166$) and low-risk ($n = 166$) groups. The high-risk group showed a lower long-term survival rate compared to the low-risk group ($p = 0.0034$, Figure 6B). The sensitivity and specificity of the prognostic characteristics were assessed using receiver operating characteristic (ROC) curves for 5-, 10-, and 30-month overall survival rates,

with area under the curves (AUCs) of 0.745, 0.726, and 0.699, respectively (Figure 6C). Forest plots were used to present the results of the clinical characteristics analysis (Supplementary Figure 4). The relationship between risk models, GM molecular subtypes, CRC consensus molecular subtypes, and the end state was analyzed to gain a deeper understanding of disease mechanisms (Figure 6D).

GMRBs survival analysis and immune correlation analysis

To investigate the potential prognostic significance of the two GMRBs identified by multivariate Cox regression analysis, we conducted a correlation analysis between risk scores and the expression levels of GMRBs. Notably, a significant negative correlation was observed between IL7 and BCL10 and risk scores

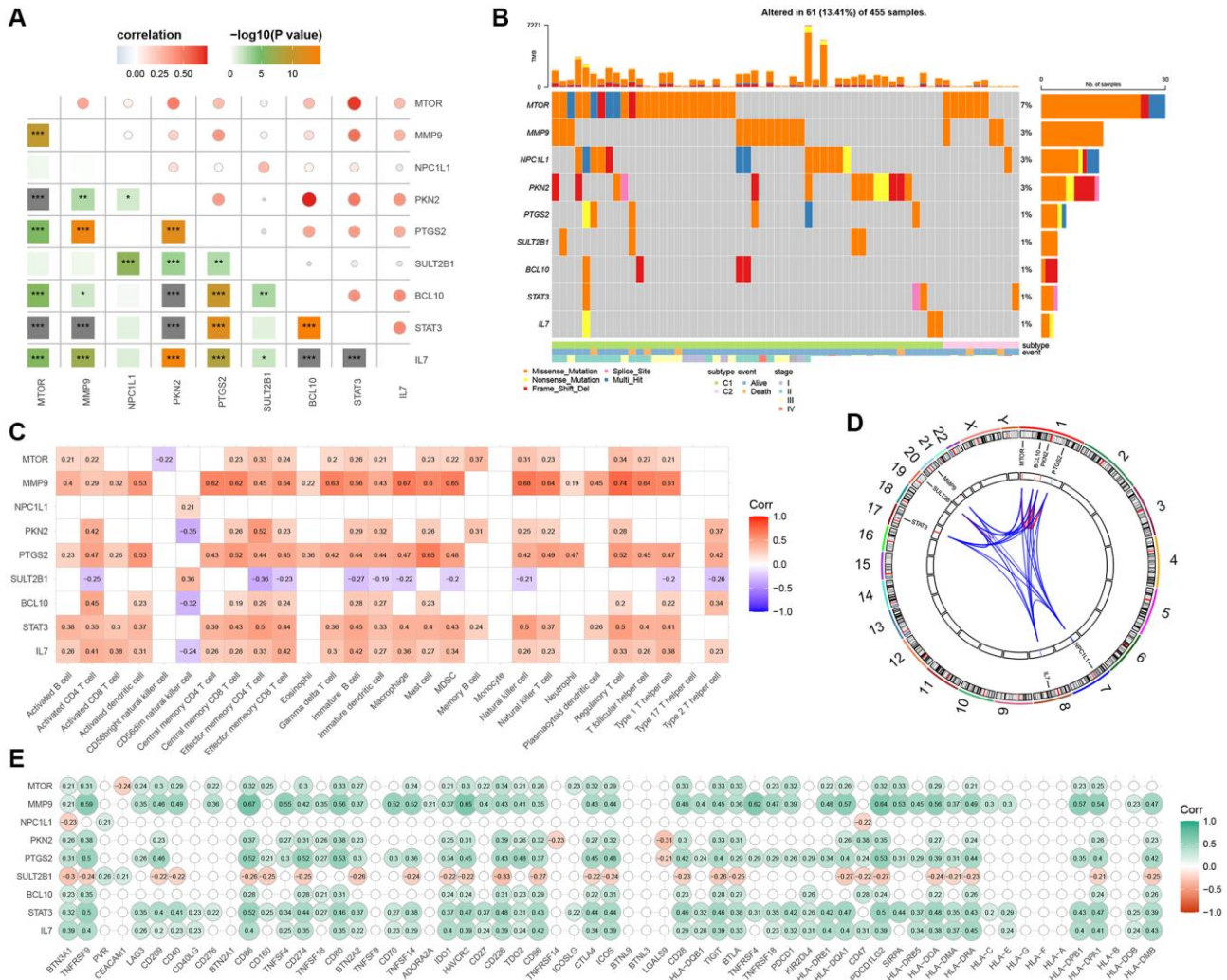


Figure 5. Discovery and investigation of GMRGs linked to prognostic outcomes. (A) Correlation analysis of 9 prognostic genes obtained by univariate Cox regression analysis. (B) Mutation landscape of 9 prognostic genes. (C) Correlation analysis of 9 prognostic genes with immune cells. (D) Display of 9 prognostic genes locations on the chromosome. (E) Correlation analysis of 9 prognostic genes with immune checkpoints. Abbreviation: GMRGs: gut microbes-related genes.

(Figure 7). Based on the median gene expression as a threshold, patients were categorized into high and low-expression groups. Subsequent survival analysis was performed for each group, and K-M survival curves were generated to visually represent these findings (Figure 8A, 8B). The prognostic characteristics of both IL7 and BCL10 consistently demonstrated a better prognosis in the high-expression group. To further validate the prognostic characteristics, we performed an analysis in an independent GEO cohort, confirming the favorable prognosis associated with high expression of both IL7 and BCL10 (Supplementary Figure 5). Moreover, to elucidate the underlying biological

mechanisms, we investigated their correlation with immune cell populations. Remarkably, both IL7 and BCL10 exhibited a strong association with CD4 T cells (Figure 8C). Moreover, both GMRBs displayed a positive correlation with M1 cell expression (Figure 8D), indicating the potential involvement of GMRBs in its activation or recruitment.

Expression analysis of GMRBs across diverse molecular subtypes, cancer types, and disease stages

Differential expression analysis of the two GMRBs was performed across various molecular subtype systems,

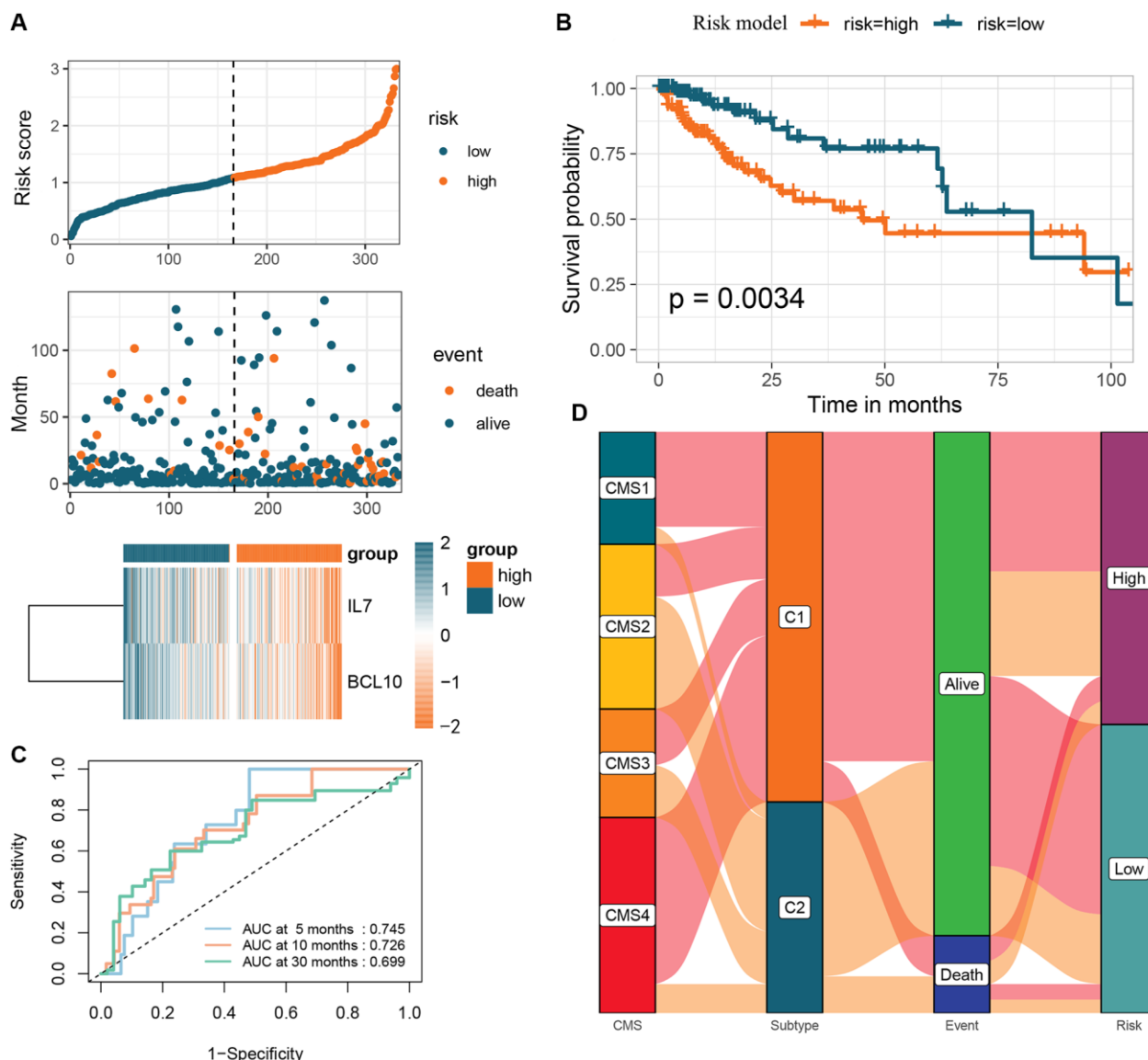


Figure 6. Creation of a risk profile through stepwise multiple Cox regression analysis based on two GMRGs. (A) Analysis of GMRG risk scores: distribution of risk scores among CRC patients; survival status and duration of CRC patients; heatmap illustrating the expression of 2 genes. (B) K-M survival analysis of CRC patients in the high- and low-risk groups in the TCGA-COAD cohort. (C) ROC calibration curves predicting overall survival at 5, 10, and 30 months in the high- and low-risk groups. (D) Sankey diagram illustrating the differential distribution of CRC patients in CRC consensus molecular subtypes, GM subtypes, survival outcomes, and high- and low-risk groups. Abbreviations: GMRGs: gut microbes-related genes; CRC: colorectal cancer; K-M: Kaplan-Meier; ROC: receiver operating characteristic; GM: gut microbes.

including consensus clustering (Figure 9A), high- and low-risk clustering (Figure 9B), and CRC consensus subtypes (Figure 9C). The analysis revealed higher

expression levels of both IL7 and BCL10 in the C1 subtype compared to the C2 subtype. Furthermore, in the high-risk group, their expression was lower than in

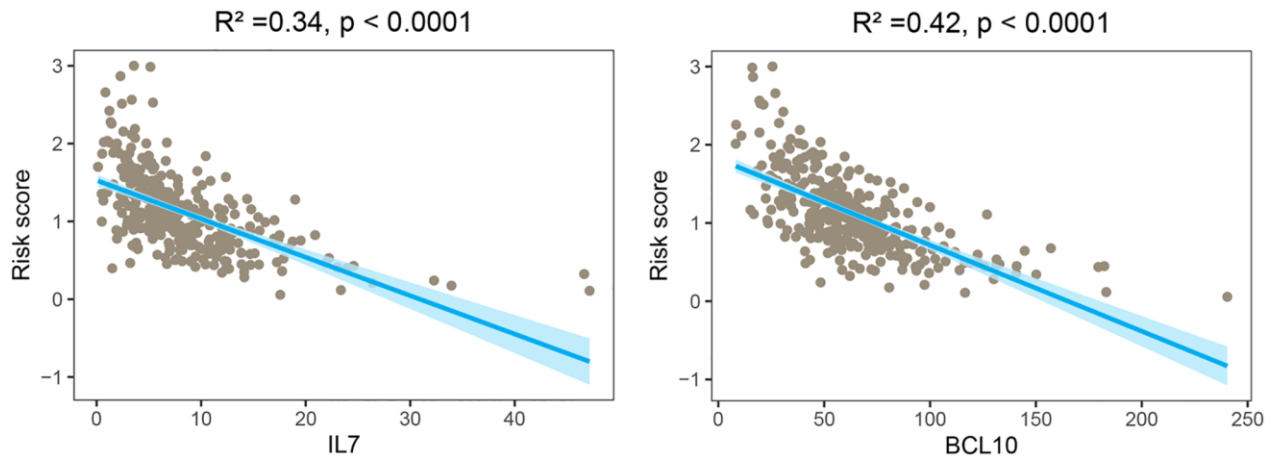


Figure 7. Correlation analysis between risk scores and expression levels of the two GMRBs. The scatter plot from the correlation analysis revealed a significant negative correlation between the risk score and the expression levels of the two GMRBs (IL7 and BCL10). Abbreviation: GMRBs gut microbes-related biomarkers.

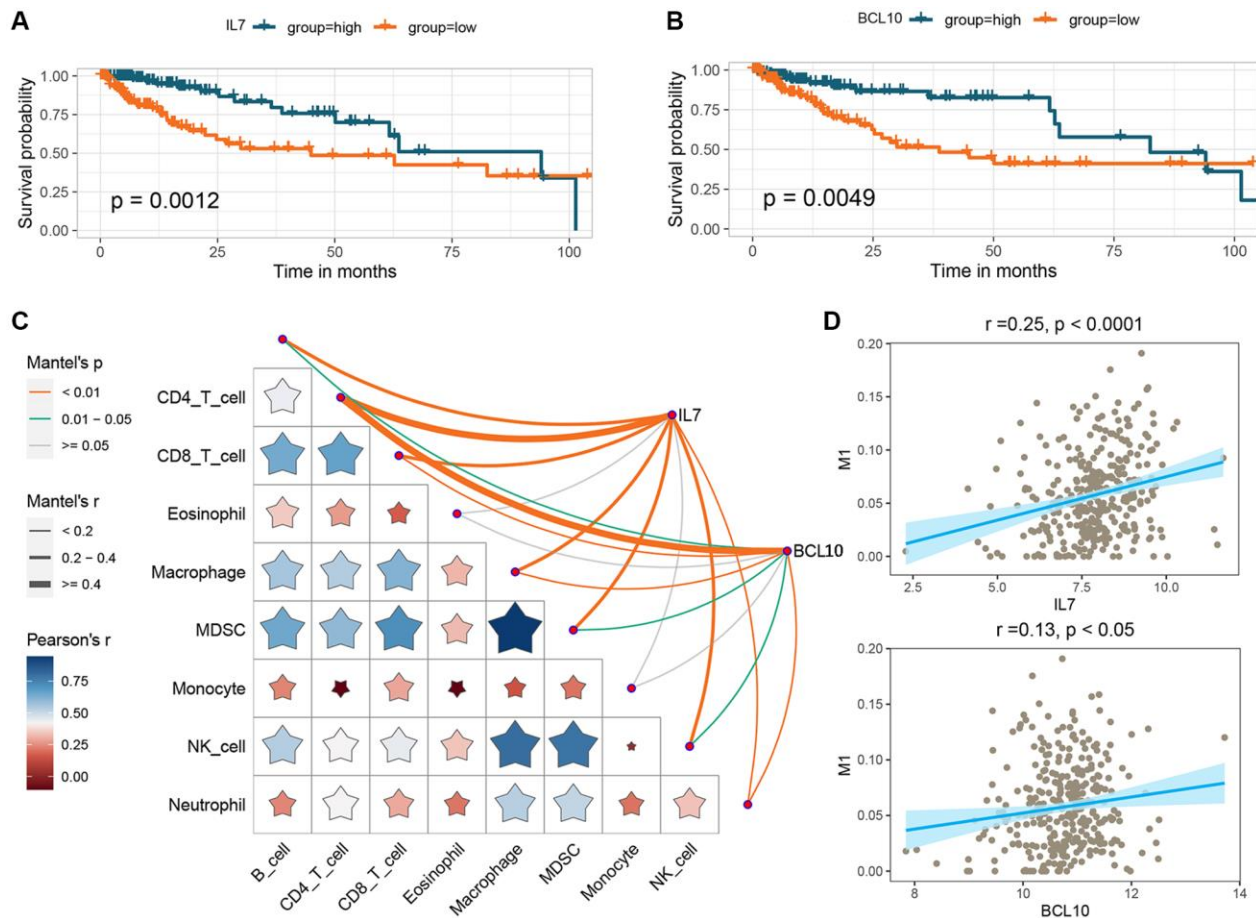


Figure 8. Analysis of the prognostic impact of GMRBs and their correlation with immune response. (A) K-M analysis of OS in CRC patients based on IL7 expression. (B) K-M analysis of OS in CRC patients based on BCL10 expression. (C) Correlation analysis of GMRBs with immune cell infiltration. (D) Correlation analysis of GMRBs with M1 macrophages. Abbreviations: GMRBs: gut microbes-related biomarkers; K-M: Kaplan-Meier; OS: overall survival; CRC: colorectal cancer.

the low-risk group. Notably, the CMS1 subtype, characterized by a strong immune profile in CRC consensus subtypes, exhibited the highest expression of GMRBs. Pan-cancer analysis revealed that GMRBs showed differential expression in different cancers (Figure 9D, 9E). Interestingly, in the COAD, DLBC, PAAD, READ, and STAD datasets, both IL7 and BCL10 were expressed higher in the cancer group than in the normal group. Additionally, differential expression data of the GMRBs were obtained for healthy individuals, patients with inflammatory bowel disease (IBD), and CRC patients using the GSE37283, TCGA-COAD, and GSE3629 datasets. Both IL7 and BCL10 exhibited elevated expression levels in both IBD and CRC patients compared to healthy individuals (Figure 9F, 9G). When comparing the IBD and CRC groups, both genes showed higher expression levels in the CRC group (Figure 9H). Finally, in the TCGA-COAD dataset, the expression levels of the GMRBs

were analyzed in patients at different stages of CRC. The analysis showed that the expression of GMRBs showed a decreasing trend at different stages of the cancer (Figure 9I).

DISCUSSION

This study provides a comprehensive analysis of the role of GMRGs in CRC, including their subtyping, prognostic significance, immune-related characteristics, and potential as predictive markers. Initially, we examined the differential expression of 164 GMRGs in the TCGA-COAD dataset, comparing cancer and normal groups. Using a consensus clustering algorithm, we identified two distinct GM clusters, namely subtype C1 and subtype C2, in the TCGA-COAD cohort. These clusters were further validated using the GEO cohort. Through comprehensive functional enrichment and immune analysis, we embarked on a profound

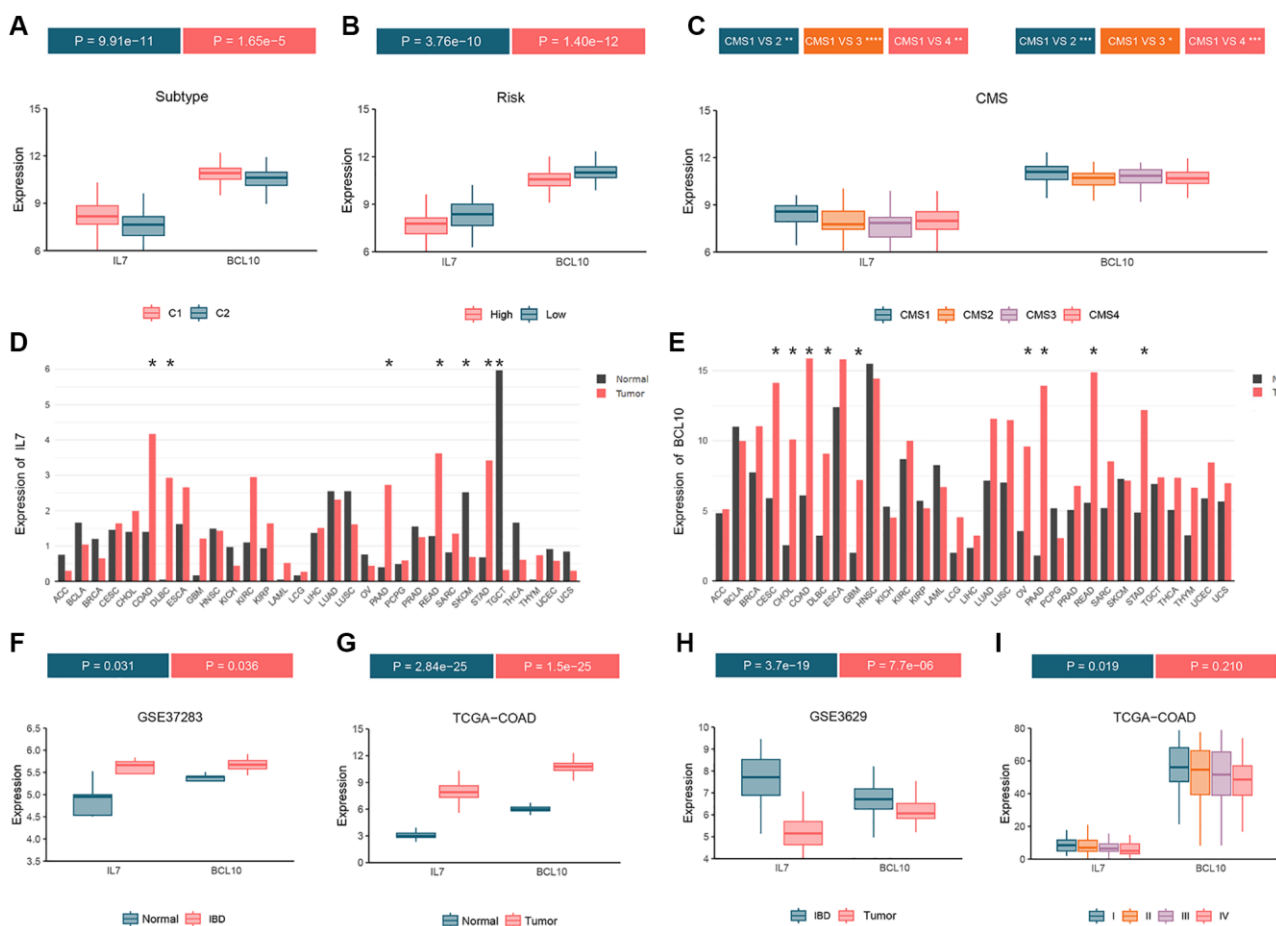


Figure 9. Analyze the expression status of GMRBs in different molecular subtypes, cancer types and disease stages. (A) GMRBs expression in molecular subtypes obtained by consensus clustering. (B) GMRBs expression in high-risk and low-risk groups. (C) GMRBs expression in CRC consensus molecular subtypes. (D, E) Expression of IL7 and BCL10 in various cancer types. (F) Differential expression of GMRBs between healthy individuals and patients with IBD. (G) Differential expression of GMRBs between healthy individuals and patients with CRC. (H) Differential expression of GMRBs between patients with IBD and patients with CRC. (I) Differential expression of GMRBs in CRC patients at different disease stages. Abbreviations: GMRBs: gut microbes-related biomarkers; CRC: colorectal cancer; IBD: inflammatory bowel disease.

exploration of the intricate biological functionalities and immunological characteristics exhibited by each cancer subtype. By conducting univariate Cox regression analysis, we found that nine genes (MTOR, MMP9, NPC1L1, PKN2, PTGS2, SULT2B1, BCL10, STAT3, IL7) exhibited significant associations with OS within CRC patients. We also examined characteristics such as mutations and immunity in these genes. Further analysis using stepwise multifactorial Cox regression identified two selected OS-related GMRBs, namely IL7 and BCL10. Prognostic models were developed based on these GMRBs, allowing for clinical survival predictions. Additionally, we investigated the association of IL7 and BCL10 with prognosis and immunity. We explored their expression levels across different staging systems, cancer types, disease stages, and CRC stages. In summary, our study revealed novel insights into the GM molecular subtypes of CRC and identified IL7 and BCL10 as predictive GM markers for CRC.

Molecular subtyping of tumors has garnered considerable attention in cancer research due to its relevance in prognostic prediction and understanding the composition of the TME. However, there remains a limited number of comprehensive studies investigating the role of GMRGs in molecular subtypes of CRC. Previous research by Zheng et al. focused on stem cell-related molecular subtypes associated with CRC prognosis [19], while Dai et al. established a senescence-related subtype using 91 senescence-related genes [20]. Additionally, Yuan et al. employed a non-negative matrix factorization clustering algorithm to identify energy metabolism-related CRC subtypes [21]. In this study, we aimed to identify distinct GM molecular subtypes (C1 and C2) based on their expression profiles in CRC. Our findings revealed a strong association between the C1 subtype and CMS1 and CMS4 subtypes, representing immune activation and stromal invasion, respectively, as defined by Justin et al. [22]. Furthermore, enrichment analysis demonstrated that C1 subtypes have significant regulatory effects on immune responses and immune system processes, particularly in Toll-like receptor (TLR) signaling, mitogen-activated protein kinase kinase kinase 14 (MAP3K14)/NF- κ B signaling, and vascular endothelial growth factor signaling pathways. TLR signaling is critical for innate and adaptive immunity, as it plays a crucial role in the development and maturation of the human immune system [23]. This receptor detects pathogen-associated molecular patterns, activating the immune system and providing opportunities for the use of tailored adjuvants with specific immune effects in targeted cancer therapy [24].

Regarding TME patterns, the C1 subtype exhibited a substantial enrichment of M1 macrophages known for

their pro-inflammatory and anti-tumor properties [25]. These M1 macrophages possess enhanced phagocytic activity and secrete pro-inflammatory cytokines and chemokines that facilitate the recruitment and activation of cytotoxic T cells and natural killer cells, effectively targeting and eliminating malignant cells [26]. Additionally, M1 macrophages promote tumor antigen presentation and support the generation of a robust anti-tumor immune response [27]. Stimulators of the classical pathway, such as bacterial components, interferon-gamma, lipopolysaccharide, and TLRs, promote the polarization of M1 macrophages. These polarized M1 macrophages induce oncogenic effects by releasing pro-inflammatory cytokines (e.g., interleukin 1 and IL12) and cytotoxic agents (e.g., reactive oxygen species and TNF) [28, 29]. In addition, the C2 subtype exhibited a significant enrichment of Tregs, which play a complex role in the progression and immune evasion of CRC [30]. Tregs, characterized by FOXP3 expression, are recruited to CRC sites via chemokine signaling and exert immunosuppressive functions by suppressing effector T cell responses and hindering anti-tumor immune surveillance [31]. Mechanistically, Tregs exert their suppressive effects through the secretion of immunosuppressive cytokines such as IL10 and transforming growth factor- β (TGF- β), as well as direct cell-to-cell contact involving immune checkpoint molecules such as CTLA4 [32, 33]. The abundance of Tregs within CRC tumors is associated with a poor prognosis and reduced patient survival [34]. Notably, our study found that the outcome of immune infiltration in both GM molecular subtypes was consistent with prognosis, with patients characterized by TME activation in the C1 subtype exhibiting a longer survival time than those with TME suppression in the C2 subtype.

Based on GM molecular subtypes, we identified 2 core GMRGs (IL7 and BCL10). IL7 is an important cytokine that can be regulated by gut microorganisms and is involved in human immune response and lymphocyte development [35–37]. Intestinal symbionts with preventive or inhibitory effects on colorectal cancer, *Bifidobacterium* [38] and *Lactobacillus reuteri* [39], have been disclosed to have significant regulatory effects on host IL7 expression [35, 40]. Among them, a specific subspecies of *Lactobacillus reuteri*, *clade IV strains*, was able to specifically activate human myeloid cells and significantly increase their IL7 expression [40]. Current research substantiates the vital role of IL7 throughout all phases of T-cell development [41, 42] and highlights its importance in the preservation of quiescent T cells and the establishment and sustenance of memory T cell populations [43, 44]. The restoration of T cell homeostasis through the activation of the suppressor of cytokine signaling (SOCS)/signal

transducer and activator of transcription 5 (STAT5) pathway [45] has been demonstrated for IL7, alongside its role in guiding T cell migration, homing and activation through the induction of chemokine and integrin expression [46, 47]. Activated T cells then mediate the tumor antigen recognition process of CRC immunotherapy, a key process that greatly determines the efficacy and availability of antigen-targeted immunotherapy for cancer patients [48, 49]. In particular, it should be emphasized that owing to extensive and potent biological effects of IL7 on T-cell survival, development, proliferation, and memory maintenance, several studies have employed it as a molecular adjunct to augment the antigenicity elicited by cancer vaccines and sustain long-term memory responses against cancer. Zhao et al. reported persistent and active IL7 production in cancer vaccines modified with the IL7 gene, with undetectable levels in tumor cells [50]. The supernatant from transfected cancer cells was found to enhance T-cell proliferation, with a volume of 10 μ l corresponding to 1 ng of recombinant human IL7 protein [46]. Whole tumor cell-based vaccines expressing IL7 demonstrated vigorous preventative and therapeutical outcomes in murine models, inhibiting tumorigenesis and extending survival time. These vaccines also increased T-lymphocytes of the CD4 and CD8 subsets infiltration within tumor fields [50, 51]. In a Phase-II clinical trial, subcutaneous injection of IL7 in oncology patients under treatment with the FDA-approved tumoral vaccine sipuleucel-T resulted in the augmentation of CD4 and CD8 T lymphocytes [52, 53]. Enhanced levels of intracellular cytokines, such as interleukin 2(IL2), TNF- α , interferon-gamma (IFN- γ), and IL6, were observed within the memory subpopulation, indicating that IL7 may stimulate a memory immune response targeting cancer. Furthermore, Jeong et al. investigated the delivery of a fusion protein containing human macrophage migration inhibitory factor (MIF) and IL7 through *Mycobacterium smegmatis*, utilized as a bacterial-based cancer vaccine [54]. This vaccine elicited anti-tumor immune responses through the recruitment of potent T cells and reduced the infiltration of myeloid-derived suppressor cells (MDSCs) within the TME. Notably, combining this vaccine with PDL1 immunotherapy led to heightened antitumor effects. Furthermore, in our study, we consistently observed elevated IL7 expression in the immune inflammation-activated phenotype C1. This higher expression was significantly associated with CD4 T cell infiltration in TME and correlated with a favorable prognosis in the high-expression group. When comparing the normal, IBD, and cancer groups, we found that IL7 levels were higher in the cancer group compared to the control group but lower than in the IBD group. In conclusion, the above findings not only reveal that IL7 regulated by

GM has the potential to serve as a cancer vaccine adjuvant and a novel immune checkpoint in CRC, but also emphasize the great feasibility of in-depth study of certain specific gut-microbial subspecies including *Lactobacillus reuteri clade IV strains* for the realization of microecological therapies for CRC.

BCL10, a critical member of the caspase activation and recruitment domain (CARD) family, plays a pivotal role in mediating inflammatory responses [55]. Through interactions with CARD11 and mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1), it forms the CARD11-BCL10-MALT1 (CBM) complex [56]. Activation of the CBM complex downstream of B-cell receptor (BCR) or T-cell receptor (TCR) signaling in B and T cells initiates the NF- κ B and mitogen-activated protein kinase (MAPK) pathways, leading to a robust phenotypic response [57, 58]. In particular, hyperactivation of the NF- κ B pathway permits tumor-selective reprogramming of the chemokine microenvironment, and this reprogramming greatly enhances the recruitment of effector CD8(+) T cells, which in turn leads to the generation of potent anti-CRC effects [59, 60]. Notably, butyrate produced by intestinal *Roseburia intestinalis* can bind directly to T cell surface receptors, which would likely induce CD8 T cell cell activation through CBM complex activation of NF- κ B signaling [61]. In addition, Aberrant CBM function resulting from BCL10 mutations is a key factor in a range of diseases, including B-cell lymphoma, autoimmune diseases, and IBD [62–64]. Xia et al. reported frequent mutations in BCL10 signaling mediators, which chronically activate the CBM signaling amplification complex, thereby contributing to an unfavorable outcome in activated B-cell-like diffuse large B-cell lymphoma (ABC-DLBCL) [62]. Similarly, Zhou et al. demonstrated a significant positive correlation between higher expression levels of BCL10 and favorable survival outcomes in colon cancer patients [65]. Furthermore, Borthakur et al. found that carrageenan, a high molecular weight sulfated polygalactan, significantly increases BCL10 secretion in patients with IBD [66]. Collectively, these studies underscore BCL10 as a valuable prognostic indicator in CRC patients. The enriched NF- κ B and MAPK pathways, along with its association with T and B cells, suggest that BCL10 may exert its effects by inducing an inflammatory response enhancing killing and inhibiting tumor cell proliferation, further emphasizing its potential significance.

Although our study has yielded encouraging results, there are several specific limitations that should be addressed. Firstly, while we utilized open data sources including TCGA and GEO to construct, validate, and test GM subtypes, we acknowledge the necessity for

external independent large datasets with comprehensive coverage to facilitate further in-depth investigations. Secondly, the lack of complete and reliable experimental evidence poses challenges in validating the actual impact of hub genes in clinical immunotherapy and comprehending the influence of microecological modulators on hub gene expression levels. In addition, the analysis of the relationship between GM and host genes in the paper relied primarily on existing literature, and further experiments may be needed to determine the specific mechanisms linking GM, GMRBs, and CRC. Furthermore, it is imperative to acknowledge that the exact mechanism underlying the role of GM in CRC remains unclear, and the fluctuation in the amount and ratio of pathogenic and non-pathogenic GMs in different genera is often implicated in the development and progression of CRC through its involvement in the role of GM. Therefore, further research in this area is warranted. Future studies should endeavor to elucidate the intricate mechanisms underlying the involvement of these two GMRBs in CRC progression and to meticulously evaluate their efficacy in the context of clinical immunotherapy and microecological adjuvant therapy for CRC.

In conclusion, this paper presents novel CRC subtypes and biomarkers associated with GM, thereby providing new potentially feasible options for patient risk stratification and immunotherapy.

MATERIALS AND METHODS

Data acquisition

We downloaded RNA sequencing (RNA-seq) information and clinical information of CRC patients from The Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>) database, which was processed to obtain 332 cancer samples as well as 41 normal samples. Then, the GSE87211, GSE161158, GSE37283, and GSE3629 datasets were collected from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) project. The GSE87211 and GSE161158 datasets were used as external validation sets, while the GSE37283 and GSE3629 datasets contained the expression profiles of patients with colitis. In addition, somatic mutation data were acquired from TCGA, containing 455 CRC samples.

Consensus clustering of GMRGs

We employed a range of keyword search strategies in the PubMed database, including “Medical Subject Headings (MeSH)”, “probiotics”, “microorganisms”, “microbiome”, “microbiome”, “bacteria”, and “fungi” in conjunction with “host”, “human”, “intestinal”,

“colitis”, “Crohn’s disease”, “ulcerative colitis”, “colorectal”, and “colon cancer” term combinations to retrieve related articles. The reference lists of relevant articles were also reviewed, resulting in 164 GMRGs, as detailed in Supplementary Table 2. Based on the GMRGs expression levels, we applied consensus unsupervised clustering to categorize patients into separate GM molecular subtypes using the “ConsensusClusterPlus” R package [67]. Then, we compared the overall survival (OS) of GM molecular subtypes with the assistance of the R packages “survminer” and “survival”. Principal component analysis (PCA) and heatmap visualization were conducted to show the classification of GM molecular subtypes. Subsequently, it was further validated and tested on external datasets GSE87211 and GSE16115.

Association analysis of GM molecular subtypes with consensus molecular subtypes of CRC

To classify the CRC patient samples into consensus subtypes, we employed the “Biobase” and “CMScaller” packages in R. We then visualized the proportion of GM molecular subtypes in each consensus subtype using the “tidyverse” and “ggplot2” packages. This approach allowed for a comprehensive exploration of the characteristics of the GM molecular subtypes and their relationship with consensus subtypes.

Differential expression analysis and functional enrichment of GM molecular subtypes

The limma, DESeq2, and edgeR packages in R were utilized to determine differentially expressed genes (DEGs) within dissimilar GM molecular subtypes. The significant criterion for identifying DEGs was set to $|\log_2(\text{FoldChange})| > (\text{mean}(\text{abs}(\log_2\text{FoldChange})) + 2 \times \text{sd}(\text{abs}(\log_2\text{FoldChange})))$ and adjusted P -value < 0.05 . The “ReactomePA” and “org.Hs.eg.db” R packages were used for the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome enrichment analyses of DEGs.

Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

The gseGO function of “clusterProfiler” package was deployed to examine the underlying mechanism of diverse GM molecular subtypes, subsequently identifying enriched Gene Ontology (GO) pathways. To discern dissimilar biological functions among GM molecular subtypes, GSVA was executed on “c5.go.bp.v7.5.symbols” using the “GSVA” R package. Visualization of results was carried out via implementation of the “pheatmap” R package.

Immune cell infiltration analysis

We used CIBERSORT deconvolution algorithm for quantifying abundance of tumor-infiltrating immune cells (TICs) in each CRC sample [68]. Specifically, we downloaded the TICs gene expression signature matrix on the CIBERSORT website platform (<https://cibersortx.stanford.edu/>) and employed it in our analysis. This allowed us to accurately determine the composition and quantity of TICs present in each sample and provided valuable insights into the immune microenvironment of CRC. Then, we used the R package “ESTIMATE” to assess the degree of infiltration of tumor fibrous tissue and immune cells to obtain a comprehensive assessment of the overall TME.

Evaluation of immunotherapy response and immune checkpoints

Therapy using immune checkpoint blockade is emerging as a promising strategy for cancer treatment [69]. In this study, we employed the Tumor Immune Dysfunction and Exclusion (TIDE) model to investigate the key mechanisms underlying tumor immune evasion and to forecast the therapeutic response of GM subtypes using expression profile data [70]. Moreover, we estimated the expression of programmed cell death 1 (PD1) and cytotoxic T-lymphocyte associated protein 4 (CTLA4) immune checkpoint molecules in different subtypes, which can regulate self-immune response by modulating T cell activity. This analysis allowed us to determine the sensitivity of different subtypes to immune checkpoint inhibitor therapy [71].

Drug susceptibility analysis

For the evaluation of the therapeutic efficacy of the drugs in both subgroups, we utilized the “pRRophetic” R package to compute semi-inhibitory concentration (IC50) values corresponding to the drugs.

Identification and analysis of prognosis-related GMRGs

We performed rigorous analyses using the TCGA-COAD dataset to determine GMRGs associated with prognosis, utilizing a univariate Cox regression analysis. To further assess the degree of relationship among these genes, we calculated Pearson correlation coefficients using the highly regarded “Hmisc” R package. The post-screening genes were then subjected to rigorous visualization and analysis using the powerful “maftool” package. In addition, we employed advanced R packages, such as “psych”, “GSVA”, and “ggcorrplot”, to comprehensively investigate the association between prognostic GMRGs and

tumor-infiltrating immune cells. To elucidate the chromosomal localization of prognosis-associated GMRGs, we utilized the “RCircos” package. Finally, we conducted a thorough examination of the association between prognosis-related GMRGs and common immune checkpoints, providing a critical understanding of the potential impact of these genes on the immune response to cancer.

Construction and evaluation of GM risk signature

On the basis of prognosis-associated genes by univariate Cox regression analysis, we performed multivariate Cox regression analysis on the TCGA-COAD dataset to construct a prediction model associated with GM and to obtain GMRBs. All CRC patients were stratified into high- and low-risk groups according to the median risk score calculated from TCGA-COAD data. The log-rank test was applied to investigate the difference in OS between the high- and low-risk groups, and the sensitivity and specificity of the features were evaluated by time-dependent receiver operating characteristic (ROC) curve analysis. The association between molecular consensus subtypes, consensus clustering subtypes, endpoint survival status, and risk models was also visualized by Sankey diagram.

GMRBs survival analysis and immune correlation analysis

To investigate the potential prognostic implications of two GMRBs identified through multivariate Cox regression analysis, we correlated risk scores with the expression levels of these genes. Patients were divided into high- and low-expression cohorts according to median gene-expressed levels. Survival analysis was subsequently conducted for each group, and Kaplan-Meier (K-M) survival curves were constructed to visualize these results. To gain further insights into the underlying biological mechanisms driving the prognostic role of these genes, we examined their correlation with representative immune cells using the “ggcor” package.

Expression analysis of GMRBs across diverse molecular subtypes, cancer types, and disease stages

We performed differential expression profiling of two GMRBs in various molecular subtype systems, encompassing consensus clustering, high-low risk clustering, and CRC consensus subtypes. This analysis was carried out employing the “ggpubr” and “ggsignif” software packages, with the goal of elucidating the association between GMRBs and the prognostic value within different subtypes. Furthermore, a comprehensive pan-cancer investigation of the GMRBs

was conducted using the web-based tool “GEPIA” (<http://gepia.cancer-pku.cn/>). Additionally, we acquired differential expression data for the GMRBs in healthy individuals, patients diagnosed with V, and individuals with CRC, utilizing the GSE37283, TCGA-COAD, and GSE3629 datasets. Finally, within the TCGA-COAD dataset, we performed expression profiling of the GMRBs across distinct stages of CRC.

Statistical analysis

Data analyses were carried out with R software v4.2.0. Pearson’s coefficient or Spearman’s coefficient was used to calculate the correlation between variables. Continuous variables fitted to a normal distribution between two groups were compared using *t*-tests. Analysis of variance was employed to compare more than two groups. Count data between groups were tested by χ^2 . All statistical calculations were evaluated by the following standards: **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

Abbreviations

GM: gut microbes; CRC: colorectal cancer; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; GMRBs: gut microbes-related biomarkers; IL10: interleukin 10; FOXP3: forkhead box P3; Tregs: regulatory T cells; IL23: interleukin 23; IL17: interleukin 17; NF- κ B: nuclear factor kappa B; STAT3: signal transducer and activator of transcription 3; AHR: aryl hydrocarbon receptor; IL6: interleukin 6; IL12: interleukin 12; TNF: tumor necrosis factor; IL7: interleukin 7; DC: dendritic cell; PDL1: programmed death ligand 1; GMRGs: gut microbes-related genes; PCA: principal component analysis; K-M: Kaplan-Meier; OS: overall survival; CMS: consensus molecular subtypes; DEGs: differentially expressed genes; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; GSEA: gene set enrichment analysis; GSVA: gene set variation analysis; TME: tumor micro-environment; TICs: tumor-infiltrating immune cells; TGF- β : transforming growth factor- β ; TIDE: Tumor Immune Dysfunction and Exclusion; PD1: programmed cell death 1; CTLA4: cytotoxic T-lymphocyte associated protein 4; ROC: receiver operating characteristic; AUCs: area under the curves; IBD: inflammatory bowel disease; TLR: Toll-like receptor; MAP3K14: mitogen-activated protein kinase kinase 14; TGF- β : transforming growth factor- β ; SOCS: suppressor of cytokine signaling; STAT5: signal transducer and activator of transcription 5; IL2: interleukin 2; IFN- γ : interferon-gamma; MIF: migration inhibitory factor; MDSCs: myeloid-derived suppressor cells; IL13: interleukin 13; CARD: caspase activation and recruitment domain; MALT1: mucosa-associated lymphoid tissue lymphoma translocation

protein 1; CBM: CARD11-BCL10-MALT1; BCR: B-cell receptor; TCR: T-cell receptor; MAPK: mitogen-activated protein kinase; ABC-DLBCL: activated B-cell-like diffuse large B-cell lymphoma; RNA-seq: RNA sequencing; MeSH: Medical Subject Headings.

AUTHOR CONTRIBUTIONS

L-MW and K-XM contributed to the conception and design of the study. X-LL was the major contributor to writing the manuscript. G-LZ was involved in the writing of the manuscript, contributing more to the research and methodological studies. SY-L is mainly responsible for data management and validation. YC-L contributed to the supervision and resources. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to all individuals and institutions who contributed to this research.

CONFLICTS OF INTEREST

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

FUNDING

This work was supported by the National Natural Science Foundation of China (81972749), the Science and Technology Project of Liaoning (2021JH2/10300020), and the Innovative Teams Project in Key Areas of Dalian (2021RT01).

REFERENCES

1. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J, Murphy N, Bray F. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023; 72:338–44. <https://doi.org/10.1136/gutjnl-2022-327736> PMID:36604116
2. Araghi M, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, Arnold M. Global trends in colorectal cancer mortality: projections to the year 2035. *Int J Cancer*. 2019; 144:2992–3000. <https://doi.org/10.1002/ijc.32055> PMID:30536395
3. Liu Y, Zhang C, Wang Q, Wu K, Sun Z, Tang Z, Zhang B. Temporal Trends in the Disease Burden of Colorectal Cancer with Its Risk Factors at the Global and

- National Level from 1990 to 2019, and Projections Until 2044. *Clin Epidemiol.* 2023; 15:55–71.
<https://doi.org/10.2147/CLEP.S388323>
 PMID:36659904
4. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343:78–85.
<https://doi.org/10.1056/NEJM200007133430201>
 PMID:10891514
 5. Peto R. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer.* 2011 (Suppl 2); 105:S1.
<https://doi.org/10.1038/bjc.2011.473>
 PMID:22158311
 6. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, et al, and MetaHIT Consortium. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature.* 2010; 464:59–65.
<https://doi.org/10.1038/nature08821>
 PMID:20203603
 7. Chung L, Thiele Orberg E, Geis AL, Chan JL, Fu K, DeStefano Shields CE, Dejea CM, Fathi P, Chen J, Finard BB, Tam AJ, McAllister F, Fan H, et al. *Bacteroides fragilis* Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe.* 2018; 23:203–14.e5.
<https://doi.org/10.1016/j.chom.2018.01.007>
 PMID:29398651
 8. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ β -catenin signaling via its FadA adhesin. *Cell Host Microbe.* 2013; 14:195–206.
<https://doi.org/10.1016/j.chom.2013.07.012>
 PMID:23954158
 9. Yang Y, Weng W, Peng J, Hong L, Yang L, Toiyama Y, Gao R, Liu M, Yin M, Pan C, Li H, Guo B, Zhu Q, et al. *Fusobacterium nucleatum* Increases Proliferation of Colorectal Cancer Cells and Tumor Development in Mice by Activating Toll-Like Receptor 4 Signaling to Nuclear Factor- κ B, and Up-regulating Expression of MicroRNA-21. *Gastroenterology.* 2017; 152:851–66.e24.
<https://doi.org/10.1053/j.gastro.2016.11.018>
 PMID:27876571
 10. Lin Y, Fan L, Qi Y, Xu C, Jia D, Jiang Y, Chen S, Wang L. *Bifidobacterium adolescentis* induces Decorin⁺ macrophages via TLR2 to suppress colorectal carcinogenesis. *J Exp Clin Cancer Res.* 2023; 42:172.
<https://doi.org/10.1186/s13046-023-02746-6>
 PMID:37464382
 11. Dikeocha IJ, Al-Kabsi AM, Chiu HT, Alshawsh MA. *Faecalibacterium prausnitzii* Ameliorates Colorectal Tumorigenesis and Suppresses Proliferation of HCT116 Colorectal Cancer Cells. *Biomedicines.* 2022; 10:1128.
<https://doi.org/10.3390/biomedicines10051128>
 PMID:35625865
 12. Obuya S, Elkholy A, Avuthu N, Behring M, Bajpai P, Agarwal S, Kim HG, El-Nikhely N, Akinyi P, Orwa J, Afaq F, Abdalla M, Michael A, et al. A signature of *Prevotella copri* and *Faecalibacterium prausnitzii* depletion, and a link with bacterial glutamate degradation in the Kenyan colorectal cancer patients. *J Gastrointest Oncol.* 2022; 13:2282–92.
<https://doi.org/10.21037/jgo-22-116>
 PMID:36388691
 13. Konieczna P, Groeger D, Ziegler M, Frei R, Ferstl R, Shanahan F, Quigley EM, Kiely B, Akdis CA, O'Mahony L. *Bifidobacterium infantis* 35624 administration induces Foxp3 T regulatory cells in human peripheral blood: potential role for myeloid and plasmacytoid dendritic cells. *Gut.* 2012; 61:354–66.
<https://doi.org/10.1136/gutjnl-2011-300936>
 PMID:22052061
 14. Sichetti M, De Marco S, Pagiotti R, Traina G, Pietrella D. Anti-inflammatory effect of multistrain probiotic formulation (*L. rhamnosus*, *B. lactis*, and *B. longum*). *Nutrition.* 2018; 53:95–102.
<https://doi.org/10.1016/j.nut.2018.02.005>
 PMID:29674267
 15. Pernomian L, Duarte-Silva M, de Barros Cardoso CR. The Aryl Hydrocarbon Receptor (AHR) as a Potential Target for the Control of Intestinal Inflammation: Insights from an Immune and Bacteria Sensor Receptor. *Clin Rev Allergy Immunol.* 2020; 59:382–90.
<https://doi.org/10.1007/s12016-020-08789-3>
 PMID:32279195
 16. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, Benyamin FW, Lei YM, Jabri B, Alegre ML, Chang EB, Gajewski TF. Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science.* 2015; 350:1084–9.
<https://doi.org/10.1126/science.aac4255>
 PMID:26541606
 17. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, Chan FKL, Kristiansen K, Sung JY, Wong SH, Yu J. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across

- populations and universal bacterial markers. *Microbiome*. 2018; 6:70.
<https://doi.org/10.1186/s40168-018-0451-2>
PMID:29642940
18. Sadrekarimi H, Gardanova ZR, Bakhshesh M, Ebrahimzadeh F, Yaseri AF, Thangavelu L, Hasanpoor Z, Zadeh FA, Kahrizi MS. Emerging role of human microbiome in cancer development and response to therapy: special focus on intestinal microflora. *J Transl Med*. 2022; 20:301.
<https://doi.org/10.1186/s12967-022-03492-7>
PMID:35794566
19. Zheng H, Liu H, Li H, Dou W, Wang J, Zhang J, Liu T, Wu Y, Liu Y, Wang X. Characterization of stem cell landscape and identification of stemness-relevant prognostic gene signature to aid immunotherapy in colorectal cancer. *Stem Cell Res Ther*. 2022; 13:244.
<https://doi.org/10.1186/s13287-022-02913-0>
PMID:35681225
20. Dai JJ, Fu YY, Zhong XQ, Cen W, Ye MF, Chen XH, Pan YF, Ye LC. Identification of Senescence-Related Subtypes, the Development of a Prognosis Model, and Characterization of Immune Infiltration and Gut Microbiota in Colorectal Cancer. *Front Med (Lausanne)*. 2022; 9:916565.
<https://doi.org/10.3389/fmed.2022.916565>
PMID:35721059
21. Yuan Y, Chen J, Wang J, Xu M, Zhang Y, Sun P, Liang L. Development and Clinical Validation of a Novel 4-Gene Prognostic Signature Predicting Survival in Colorectal Cancer. *Front Oncol*. 2020; 10:595.
<https://doi.org/10.3389/fonc.2020.00595>
PMID:32509568
22. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015; 21:1350–6.
<https://doi.org/10.1038/nm.3967>
PMID:26457759
23. Lim KH, Staudt LM. Toll-like receptor signaling. *Cold Spring Harb Perspect Biol*. 2013; 5:a011247.
<https://doi.org/10.1101/cshperspect.a011247>
PMID:23284045
24. Kaur A, Baldwin J, Brar D, Salunke DB, Petrovsky N. Toll-like receptor (TLR) agonists as a driving force behind next-generation vaccine adjuvants and cancer therapeutics. *Curr Opin Chem Biol*. 2022; 70:102172.
<https://doi.org/10.1016/j.cbpa.2022.102172>
PMID:35785601
25. Mosser DM, Edwards JP. Exploring the full spectrum of macrophage activation. *Nat Rev Immunol*. 2008; 8:958–69.
<https://doi.org/10.1038/nri2448>
PMID:19029990
26. Li Y, Chen Z, Han J, Ma X, Zheng X, Chen J. Functional and Therapeutic Significance of Tumor-Associated Macrophages in Colorectal Cancer. *Front Oncol*. 2022; 12:781233.
<https://doi.org/10.3389/fonc.2022.781233>
PMID:35186730
27. Shapouri-Moghaddam A, Mohammadian S, Vazini H, Taghadosi M, Esmaeili SA, Mardani F, Seifi B, Mohammadi A, Afshari JT, Sahebkar A. Macrophage plasticity, polarization, and function in health and disease. *J Cell Physiol*. 2018; 233:6425–40.
<https://doi.org/10.1002/jcp.26429>
PMID:29319160
28. Tariq M, Zhang J, Liang G, Ding L, He Q, Yang B. Macrophage Polarization: Anti-Cancer Strategies to Target Tumor-Associated Macrophage in Breast Cancer. *J Cell Biochem*. 2017; 118:2484–501.
<https://doi.org/10.1002/jcb.25895>
PMID:28106295
29. Wunderlich CM, Ackermann PJ, Ostermann AL, Adams-Quack P, Vogt MC, Tran ML, Nikolajev A, Waisman A, Garbers C, Theurich S, Mauer J, Hövelmeyer N, Wunderlich FT. Obesity exacerbates colitis-associated cancer via IL-6-regulated macrophage polarisation and CCL-20/CCR-6-mediated lymphocyte recruitment. *Nat Commun*. 2018; 9:1646.
<https://doi.org/10.1038/s41467-018-03773-0>
PMID:29695802
30. Fanelli GN, Naccarato AG, Scatena C. Recent Advances in Cancer Plasticity: Cellular Mechanisms, Surveillance Strategies, and Therapeutic Optimization. *Front Oncol*. 2020; 10:569.
<https://doi.org/10.3389/fonc.2020.00569>
PMID:32391266
31. Sui H, Zhang L, Gu K, Chai N, Ji Q, Zhou L, Wang Y, Ren J, Yang L, Zhang B, Hu J, Li Q. YYFZBJS ameliorates colorectal cancer progression in Apc^{Min/+} mice by remodeling gut microbiota and inhibiting regulatory T-cell generation. *Cell Commun Signal*. 2020; 18:113.
<https://doi.org/10.1186/s12964-020-00596-9>
PMID:32677955
32. Wang S, Xia P, Chen Y, Qu Y, Xiong Z, Ye B, Du Y, Tian Y, Yin Z, Xu Z, Fan Z. Regulatory Innate Lymphoid Cells Control Innate Intestinal Inflammation. *Cell*. 2017; 171:201–16.e18.
<https://doi.org/10.1016/j.cell.2017.07.027>
PMID:28844693

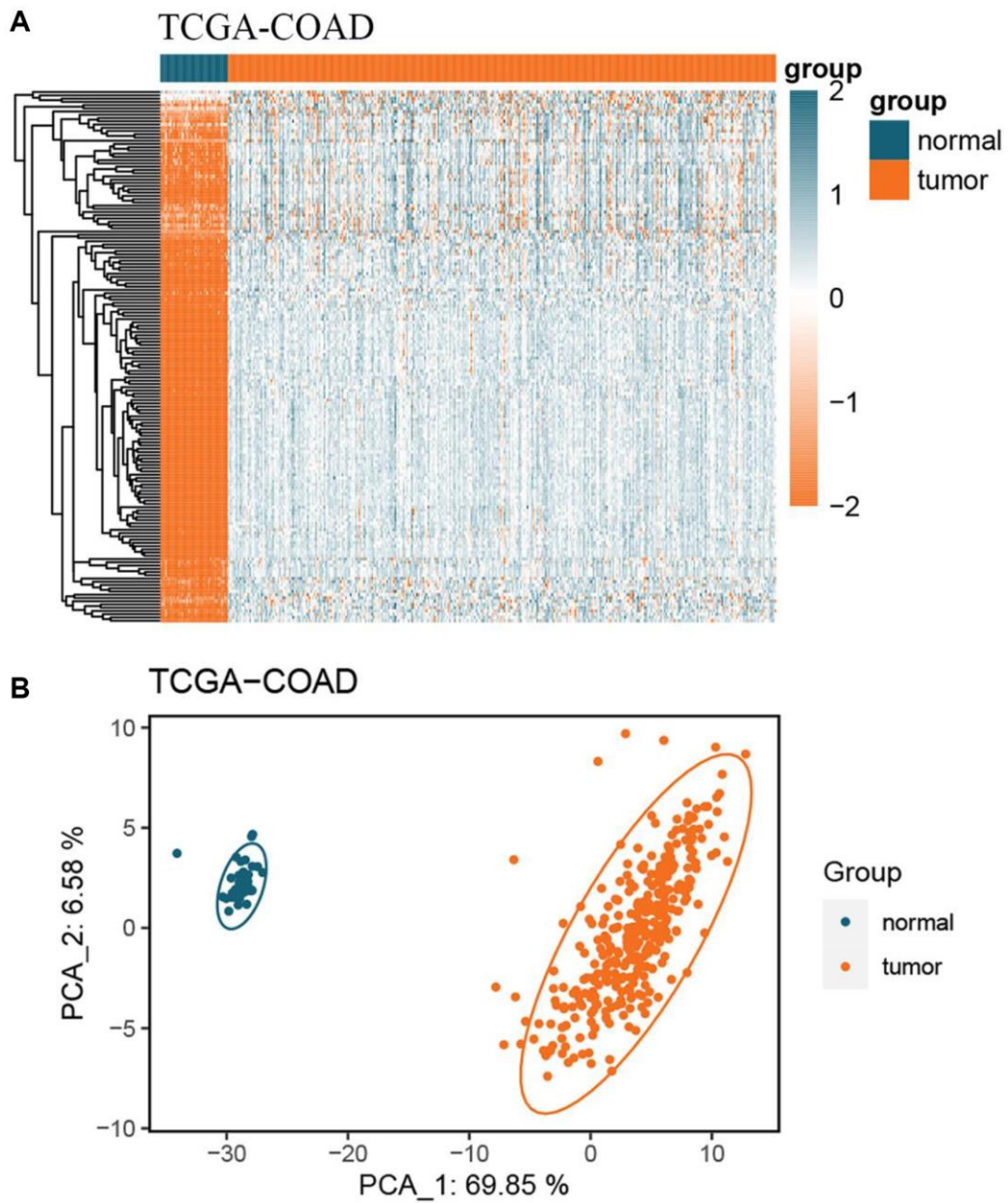
33. Sakaguchi S, Mikami N, Wing JB, Tanaka A, Ichiyama K, Ohkura N. Regulatory T Cells and Human Disease. *Annu Rev Immunol*. 2020; 38:541–66.
<https://doi.org/10.1146/annurev-immunol-042718-041717>
PMID:[32017635](https://pubmed.ncbi.nlm.nih.gov/32017635/)
34. Ye L, Zhang T, Kang Z, Guo G, Sun Y, Lin K, Huang Q, Shi X, Ni Z, Ding N, Zhao KN, Chang W, Wang J, et al. Tumor-Infiltrating Immune Cells Act as a Marker for Prognosis in Colorectal Cancer. *Front Immunol*. 2019; 10:2368.
<https://doi.org/10.3389/fimmu.2019.02368>
PMID:[31681276](https://pubmed.ncbi.nlm.nih.gov/31681276/)
35. Zhang J, Wan S, Gui Q. Comparison of safety, effectiveness and serum inflammatory factor indexes of *Saccharomyces boulardii* versus *Bifidobacterium* triple viable in treating children with chronic diarrhea: a randomized trial. *Transl Pediatr*. 2021; 10:1677–85.
<https://doi.org/10.21037/tp-21-195>
PMID:[34295782](https://pubmed.ncbi.nlm.nih.gov/34295782/)
36. Lundström W, Fewkes NM, Mackall CL. IL-7 in human health and disease. *Semin Immunol*. 2012; 24:218–24.
<https://doi.org/10.1016/j.smim.2012.02.005>
PMID:[22410365](https://pubmed.ncbi.nlm.nih.gov/22410365/)
37. Barata JT, Durum SK, Seddon B. Flip the coin: IL-7 and IL-7R in health and disease. *Nat Immunol*. 2019; 20:1584–93.
<https://doi.org/10.1038/s41590-019-0479-x>
PMID:[31745336](https://pubmed.ncbi.nlm.nih.gov/31745336/)
38. Asadollahi P, Ghanavati R, Rohani M, Razavi S, Esghaei M, Talebi M. Anti-cancer effects of *Bifidobacterium* species in colon cancer cells and a mouse model of carcinogenesis. *PLoS One*. 2020; 15:e0232930.
<https://doi.org/10.1371/journal.pone.0232930>
PMID:[32401801](https://pubmed.ncbi.nlm.nih.gov/32401801/)
39. Gao C, Ganesh BP, Shi Z, Shah RR, Fultz R, Major A, Venable S, Lugo M, Hoch K, Chen X, Haag A, Wang TC, Versalovic J. Gut Microbe-Mediated Suppression of Inflammation-Associated Colon Carcinogenesis by Luminal Histamine Production. *Am J Pathol*. 2017; 187:2323–36.
<https://doi.org/10.1016/j.ajpath.2017.06.011>
PMID:[28917668](https://pubmed.ncbi.nlm.nih.gov/28917668/)
40. Spinler JK, Sontakke A, Hollister EB, Venable SF, Oh PL, Balderas MA, Saulnier DM, Mistretta TA, Devaraj S, Walter J, Versalovic J, Highlander SK. From prediction to function using evolutionary genomics: human-specific ecotypes of *Lactobacillus reuteri* have diverse probiotic functions. *Genome Biol Evol*. 2014; 6:1772–89.
<https://doi.org/10.1093/gbe/evu137>
PMID:[24951561](https://pubmed.ncbi.nlm.nih.gov/24951561/)
41. Akashi K, Kondo M, Weissman IL. Role of interleukin-7 in T-cell development from hematopoietic stem cells. *Immunol Rev*. 1998; 165:13–28.
<https://doi.org/10.1111/j.1600-065x.1998.tb01226.x>
PMID:[9850848](https://pubmed.ncbi.nlm.nih.gov/9850848/)
42. Hong C, Luckey MA, Park JH. Intrathymic IL-7: the where, when, and why of IL-7 signaling during T cell development. *Semin Immunol*. 2012; 24:151–8.
<https://doi.org/10.1016/j.smim.2012.02.002>
PMID:[22421571](https://pubmed.ncbi.nlm.nih.gov/22421571/)
43. Fry TJ, Mackall CL. Interleukin-7: from bench to clinic. *Blood*. 2002; 99:3892–904.
<https://doi.org/10.1182/blood.v99.11.3892>
PMID:[12010786](https://pubmed.ncbi.nlm.nih.gov/12010786/)
44. Ribeiro D, Melão A, van Boxtel R, Santos CI, Silva A, Silva MC, Cardoso BA, Coffey PJ, Barata JT. STAT5 is essential for IL-7-mediated viability, growth, and proliferation of T-cell acute lymphoblastic leukemia cells. *Blood Adv*. 2018; 2:2199–213.
<https://doi.org/10.1182/bloodadvances.2018021063>
PMID:[30185437](https://pubmed.ncbi.nlm.nih.gov/30185437/)
45. Li HB, Tong J, Zhu S, Batista PJ, Duffy EE, Zhao J, Bailis W, Cao G, Kroehling L, Chen Y, Wang G, Broughton JP, Chen YG, et al. m⁶A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature*. 2017; 548:338–42.
<https://doi.org/10.1038/nature23450>
PMID:[28792938](https://pubmed.ncbi.nlm.nih.gov/28792938/)
46. Cimbri R, Vassena L, Arthos J, Cicala C, Kehrl JH, Park C, Sereti I, Lederman MM, Fauci AS, Lusso P. IL-7 induces expression and activation of integrin $\alpha 4\beta 7$ promoting naive T-cell homing to the intestinal mucosa. *Blood*. 2012; 120:2610–9.
<https://doi.org/10.1182/blood-2012-06-434779>
PMID:[22896005](https://pubmed.ncbi.nlm.nih.gov/22896005/)
47. Ponte R, Rancez M, Figueiredo-Morgado S, Dutrieux J, Fabre-Mersseman V, Charmeteau-de-Muylder B, Guilbert T, Routy JP, Cheyrier R, Couëdel-Courteille A. Acute Simian Immunodeficiency Virus Infection Triggers Early and Transient Interleukin-7 Production in the Gut, Leading to Enhanced Local Chemokine Expression and Intestinal Immune Cell Homing. *Front Immunol*. 2017; 8:588.
<https://doi.org/10.3389/fimmu.2017.00588>
PMID:[28579989](https://pubmed.ncbi.nlm.nih.gov/28579989/)
48. Leko V, Rosenberg SA. Identifying and Targeting Human Tumor Antigens for T Cell-Based Immunotherapy of Solid Tumors. *Cancer Cell*. 2020; 38:454–72.
<https://doi.org/10.1016/j.ccell.2020.07.013>
PMID:[32822573](https://pubmed.ncbi.nlm.nih.gov/32822573/)
49. Feng M, Zhao Z, Yang M, Ji J, Zhu D. T-cell-based immunotherapy in colorectal cancer. *Cancer Lett*. 2021; 498:201–9.

- <https://doi.org/10.1016/j.canlet.2020.10.040>
PMID:33129958
50. Zhao L, Mei Y, Sun Q, Guo L, Wu Y, Yu X, Hu B, Liu X, Liu H. Autologous tumor vaccine modified with recombinant new castle disease virus expressing IL-7 promotes antitumor immune response. *J Immunol*. 2014; 193:735–45.
<https://doi.org/10.4049/jimmunol.1400004>
PMID:24943214
51. Gu YZ, Fan CW, Lu R, Shao B, Sang YX, Huang QR, Li X, Meng WT, Mo XM, Wei YQ. Forced co-expression of IL-21 and IL-7 in whole-cell cancer vaccines promotes antitumor immunity. *Sci Rep*. 2016; 6:32351.
<https://doi.org/10.1038/srep32351>
PMID:27571893
52. Pachynski RK, Morishima C, Szmulewitz R, Harshman L, Appleman L, Monk P, Bitting RL, Kucuk O, Millard F, Seigne JD, Fling SP, Maecker HT, Duault C, et al. IL-7 expands lymphocyte populations and enhances immune responses to sipuleucel-T in patients with metastatic castration-resistant prostate cancer (mCRPC). *J Immunother Cancer*. 2021; 9:e002903.
<https://doi.org/10.1136/jitc-2021-002903>
PMID:34452927
53. Higano CS, Small EJ, Schellhammer P, Yasothan U, Gubernick S, Kirkpatrick P, Kantoff PW. Sipuleucel-T. *Nat Rev Drug Discov*. 2010; 9:513–4.
<https://doi.org/10.1038/nrd3220>
PMID:20592741
54. Jeong H, Lee SY, Seo H, Kim BJ. Recombinant *Mycobacterium smegmatis* delivering a fusion protein of human macrophage migration inhibitory factor (MIF) and IL-7 exerts an anticancer effect by inducing an immune response against MIF in a tumor-bearing mouse model. *J Immunother Cancer*. 2021; 9:e003180.
<https://doi.org/10.1136/jitc-2021-003180>
PMID:34389619
55. Zhang T, Sun J, Wang L, Yao H, Guo Z, Wu W, Li Y, Wang L, Song L. BCL10 regulates the production of proinflammatory cytokines by activating MAPK-NF- κ B/Rel signaling pathway in oysters. *Fish Shellfish Immunol*. 2022; 120:369–76.
<https://doi.org/10.1016/j.fsi.2021.12.009>
PMID:34906687
56. Yin H, Karayel O, Chao YY, Seeholzer T, Hamp I, Plettenburg O, Gehring T, Zielinski C, Mann M, Krappmann D. A20 and ABIN-1 cooperate in balancing CBM complex-triggered NF- κ B signaling in activated T cells. *Cell Mol Life Sci*. 2022; 79:112.
<https://doi.org/10.1007/s00018-022-04154-z>
PMID:35099607
57. Thome M, Charton JE, Pelzer C, Hailfinger S. Antigen receptor signaling to NF-kappaB via CARMA1, BCL10, and MALT1. *Cold Spring Harb Perspect Biol*. 2010; 2:a003004.
<https://doi.org/10.1101/cshperspect.a003004>
PMID:20685844
58. Meininger I, Krappmann D. Lymphocyte signaling and activation by the CARMA1-BCL10-MALT1 signalosome. *Biol Chem*. 2016; 397:1315–33.
<https://doi.org/10.1515/hsz-2016-0216>
PMID:27420898
59. Muthuswamy R, Berk E, Junecko BF, Zeh HJ, Zureikat AH, Normolle D, Luong TM, Reinhart TA, Bartlett DL, Kalinski P. NF- κ B hyperactivation in tumor tissues allows tumor-selective reprogramming of the chemokine microenvironment to enhance the recruitment of cytolytic T effector cells. *Cancer Res*. 2012; 72:3735–43.
<https://doi.org/10.1158/0008-5472.CAN-11-4136>
PMID:22593190
60. Naito Y, Saito K, Shiiba K, Ohuchi A, Saigenji K, Nagura H, Ohtani H. CD8+ T cells infiltrated within cancer cell nests as a prognostic factor in human colorectal cancer. *Cancer Res*. 1998; 58:3491–4.
PMID:9721846
61. Kang X, Liu C, Ding Y, Ni Y, Ji F, Lau HCH, Jiang L, Sung JJ, Wong SH, Yu J. *Roseburia intestinalis* generated butyrate boosts anti-PD-1 efficacy in colorectal cancer by activating cytotoxic CD8+ T cells. *Gut*. 2023; 72:2112–22.
<https://doi.org/10.1136/gutjnl-2023-330291>
PMID:37491158
62. Juilland M, Thome M. Role of the CARMA1/BCL10/MALT1 complex in lymphoid malignancies. *Curr Opin Hematol*. 2016; 23:402–9.
<https://doi.org/10.1097/MOH.0000000000000257>
PMID:27135977
63. Turvey SE, Durandy A, Fischer A, Fung SY, Geha RS, Gewies A, Giese T, Greil J, Keller B, McKinnon ML, Neven B, Rozmus J, Ruland J, et al. The CARD11-BCL10-MALT1 (CBM) signalosome complex: Stepping into the limelight of human primary immunodeficiency. *J Allergy Clin Immunol*. 2014; 134:276–84.
<https://doi.org/10.1016/j.jaci.2014.06.015>
PMID:25087226
64. Borthakur A, Bhattacharyya S, Alrefai WA, Tobacman JK, Ramaswamy K, Dudeja PK. Platelet-activating factor-induced NF-kappaB activation and IL-8 production in intestinal epithelial cells are Bcl10-dependent. *Inflamm Bowel Dis*. 2010; 16:593–603.
<https://doi.org/10.1002/ibd.21092>
PMID:19714753

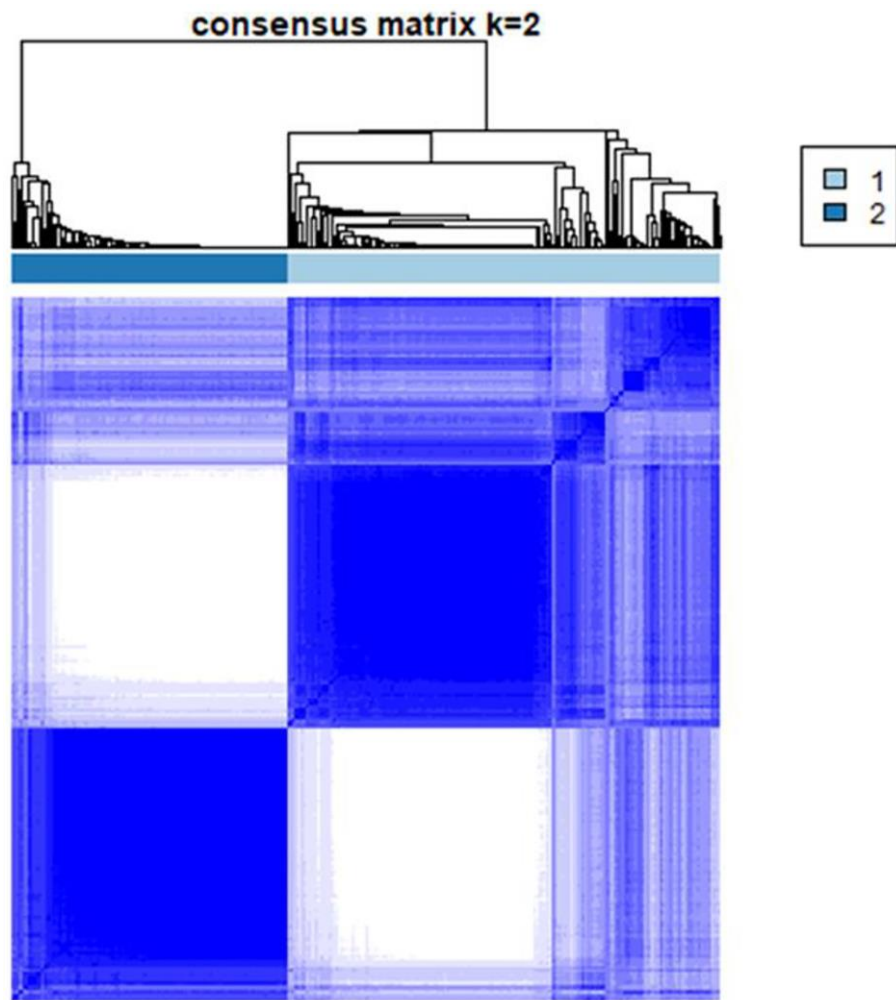
65. Zhou Y, Guo Y, Wang Y. Identification and validation of a seven-gene prognostic marker in colon cancer based on single-cell transcriptome analysis. *IET Syst Biol.* 2022; 16:72–83.
<https://doi.org/10.1049/syb2.12041>
PMID:[35352485](https://pubmed.ncbi.nlm.nih.gov/35352485/)
66. Borthakur A, Bhattacharyya S, Dudeja PK, Tobacman JK. Carrageenan induces interleukin-8 production through distinct Bcl10 pathway in normal human colonic epithelial cells. *Am J Physiol Gastrointest Liver Physiol.* 2007; 292:G829–38.
<https://doi.org/10.1152/ajpgi.00380.2006>
PMID:[17095757](https://pubmed.ncbi.nlm.nih.gov/17095757/)
67. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics.* 2010; 26:1572–3.
<https://doi.org/10.1093/bioinformatics/btq170>
PMID:[20427518](https://pubmed.ncbi.nlm.nih.gov/20427518/)
68. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M, West RB, Plevritis SK, Alizadeh AA. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med.* 2015; 21:938–45.
<https://doi.org/10.1038/nm.3909>
PMID:[26193342](https://pubmed.ncbi.nlm.nih.gov/26193342/)
69. Yap TA, Parkes EE, Peng W, Moyers JT, Curran MA, Tawbi HA. Development of Immunotherapy Combination Strategies in Cancer. *Cancer Discov.* 2021; 11:1368–97.
<https://doi.org/10.1158/2159-8290.CD-20-1209>
PMID:[33811048](https://pubmed.ncbi.nlm.nih.gov/33811048/)
70. Qiu C, Shi W, Wu H, Zou S, Li J, Wang D, Liu G, Song Z, Xu X, Hu J, Geng H. Identification of Molecular Subtypes and a Prognostic Signature Based on Inflammation-Related Genes in Colon Adenocarcinoma. *Front Immunol.* 2021; 12:769685.
<https://doi.org/10.3389/fimmu.2021.769685>
PMID:[35003085](https://pubmed.ncbi.nlm.nih.gov/35003085/)
71. Kornepati AVR, Vadlamudi RK, Curiel TJ. Programmed death ligand 1 signals in cancer cells. *Nat Rev Cancer.* 2022; 22:174–89.
<https://doi.org/10.1038/s41568-021-00431-4>
PMID:[35031777](https://pubmed.ncbi.nlm.nih.gov/35031777/)

SUPPLEMENTARY MATERIALS

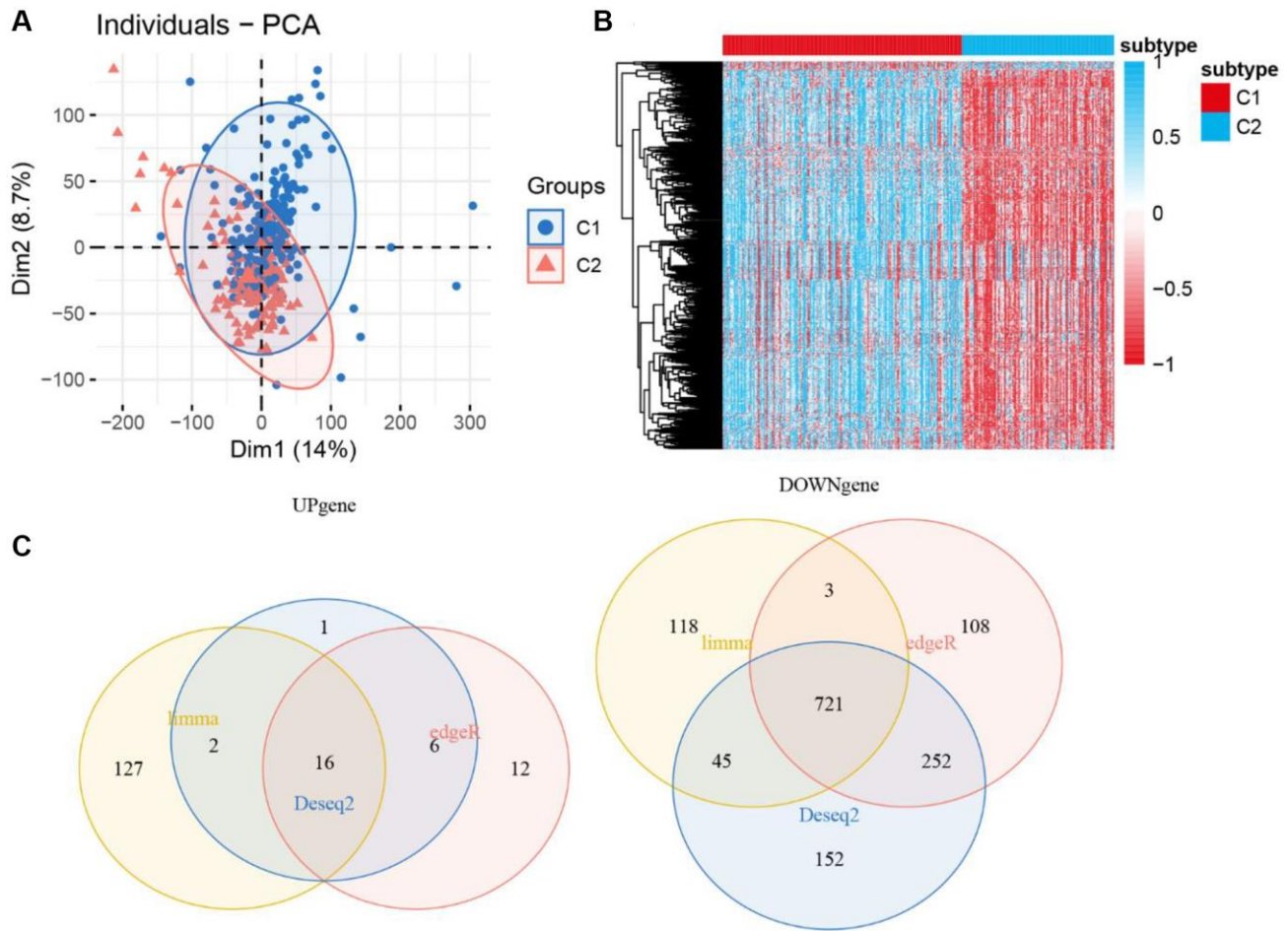
Supplementary Figures



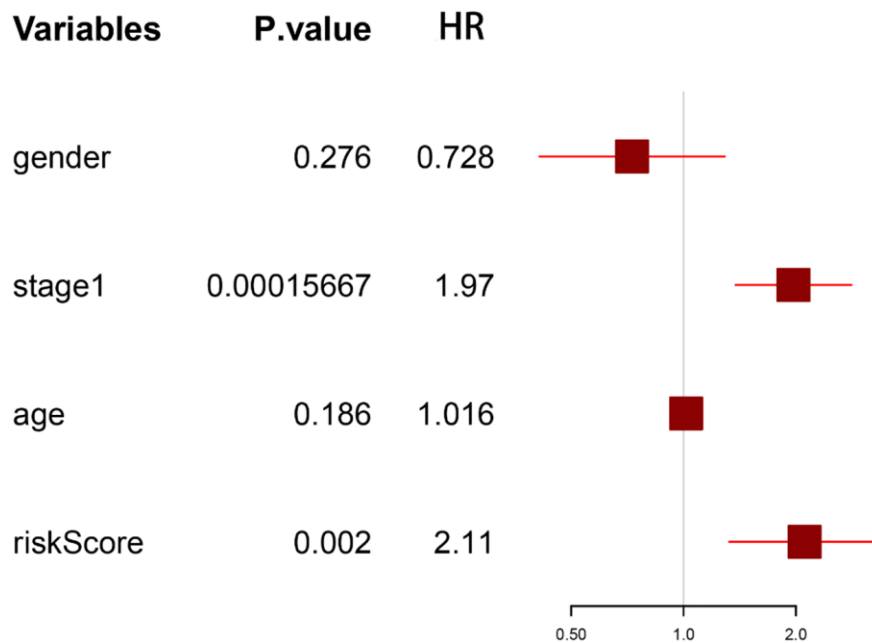
Supplementary Figure 1. Expression analysis of GMRGs in the TCGA-COAD cohort. (A) PCA plot of GMRGs. (B) Heatmap of GMRGs. Abbreviations: GMRGs: gut microbes-related genes; PCA: principal component analysis.



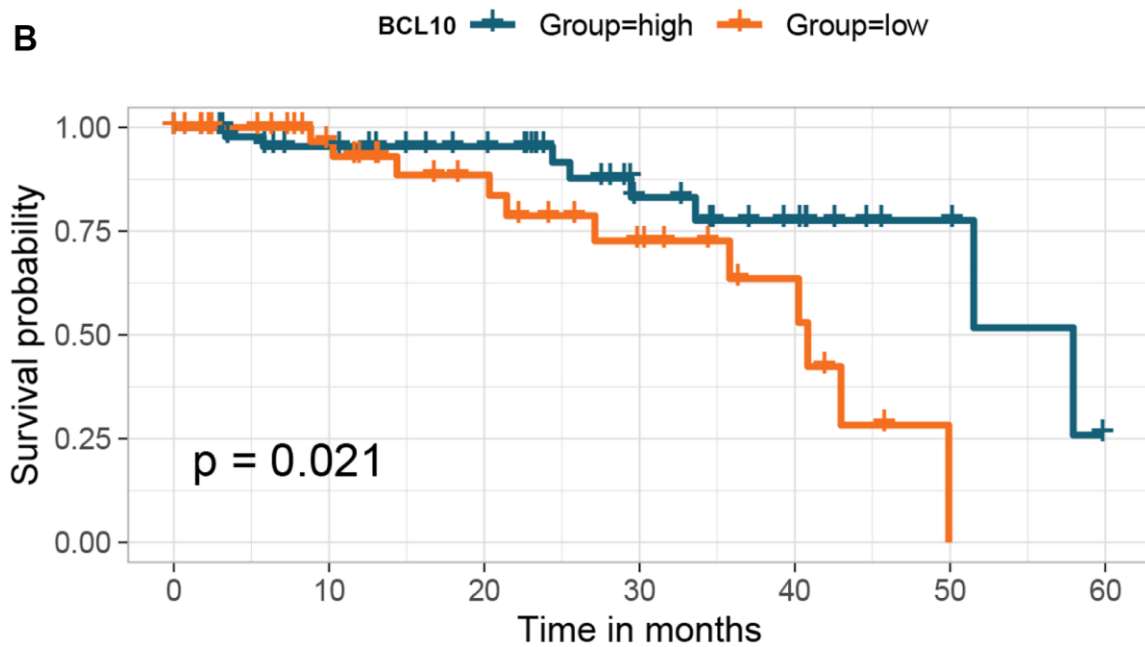
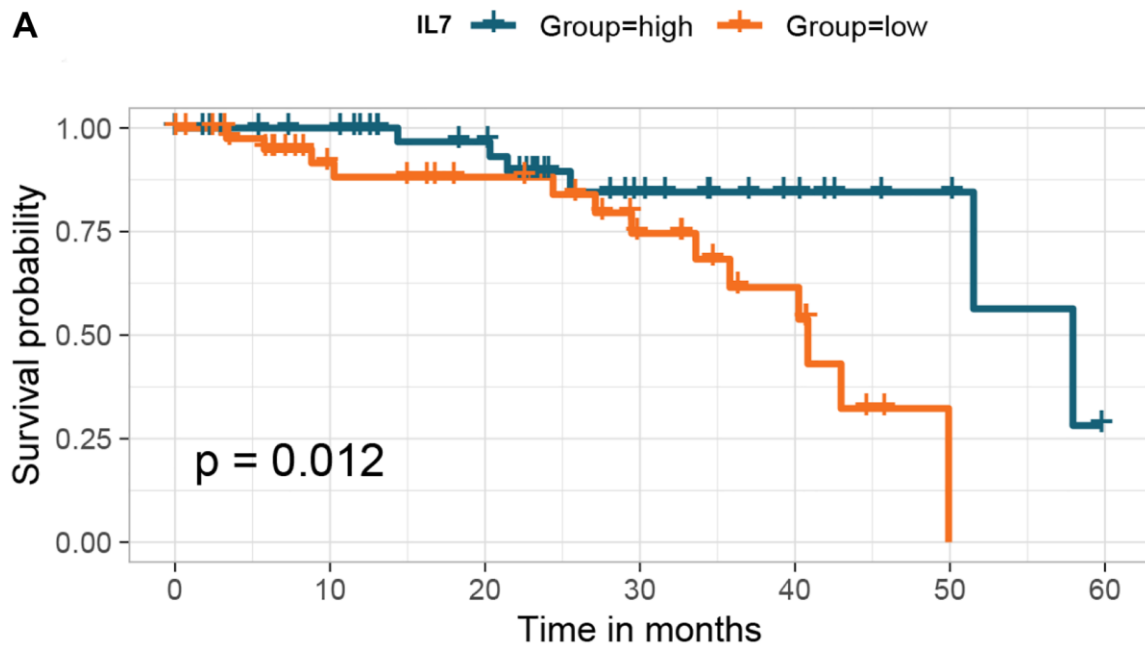
Supplementary Figure 2. Construction of GM molecular subtypes using TCGA-COAD samples. The two GM subtypes identified by the consensus clustering. Abbreviation: GM: gut microbes.



Supplementary Figure 3. DEGs between GM molecular subtypes. (A) PCA plot of DEGs. (B) Heatmap of DEGs. (C) Venn diagram of DEGs. Abbreviations: DEGs: differentially expressed genes; GM: gut microbes; PCA: principal component analysis.



Supplementary Figure 4. Clinical value of risk signature. Forest plot of clinical characteristics and risk scores.



Supplementary Figure 5. K-M survival analysis based on GMRBs expression in the GSE87211 cohort. (A) K-M survival analysis between high and low expression groups of IL7. **(B)** K-M survival analysis between high and low expression groups of BCL10. Abbreviations: K-M: Kaplan-Meier; GMRBs: gut microbes-related biomarkers.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 2.

Supplementary Table 1. Univariate Cox regression analysis identified 9 prognostic genes in the TCGA-COAD cohort ($P < 0.05$).

GENE	HR	P
MTOR	2.256	0.007
MMP9	2.012	0.025
NPC1L1	0.433	0.007
PKN2	1.853	0.038
PTGS2	2.064	0.027
SULT2B1	0.552	0.045
BCL10	2.248	0.007
STAT3	2.104	0.019
IL7	2.090	0.019

Supplementary Table 2. 164 GMRGs obtained using a keyword search strategy in the PubMed database.