Original Article

Development and validation of a prognosis prediction model for overall survival in correlation between butyrate metabolism and gastric cancer prognosis: Mendelian randomization and transcriptomics analysis

Renjun Gu[1,2#], Kun Mei[3#], Zilu Chen[4#], Yan Huang[5], Fangyu Wang[1]

[1]Department of Gastroenterology and Hepatology, Jinling Hospital, Medical School of Nanjing University, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China; [2]School of Chinese Medicine & School of Integrated Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, China; [3]Department of Cardiothoracic Surgery, The Third Affiliated Hospital of Soochow University, Changzhou, China; [4]Department of Ultrasound, Nanjing University of Chinese Medicine, Nanjing, China; [5]Department of Ultrasound, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, China

Contributions: (I) Conception and design: R Gu, K Mei, F Wang; (II) Administrative support: Y Huang, F Wang; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: R Gu, K Mei, Z Chen; (V) Data analysis and interpretation: K Mei, Z Chen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work as co-first authors.

Correspondence to: Fangyu Wang, MD, PhD. Department of Gastroenterology and Hepatology, Jinling Hospital, Medical School of Nanjing University, Affiliated Hospital of Nanjing University of Chinese Medicine, No. 305 Zhongshan East Road, Nanjing 210002, China. Email: wangf65@nju.edu.cn.

Background: Gastric cancer (GC) remains a leading cause of cancer-related mortality due to its late diagnosis and poor prognosis. Butyrate metabolism (BM) has demonstrated significant roles in tumor biology, but its prognostic implications in GC remain unexplored. We aimed to investigate the effect of butyrate metabolic biomarkers on the prognosis of GC.

Methods: We acquired datasets from The Cancer Genome Atlas and Gene Expression Omnibus. Differential BM-related genes (BMGs) were identified using weighted gene co-expression network analysis (WGCNA). Patients were stratified into subtypes, and a prognostic model was constructed using least absolute shrinkage and selection operator (LASSO) regression. Mendelian randomization (MR) analysis was conducted using genetic variants as instrumental variables to establish causal links between BM and GC prognosis.

Results: Our model demonstrated robust prognostic accuracy with an area under the receiver operating characteristic (ROC) curve of 0.716. Transcriptomic analysis identified two key BMGs, SMC2 and HSPB1, with significant implications for GC survival. However, MR analysis provided no evidence of a causal association between BM and GC.

Conclusions: We identified two butyrate metabolic prognostic genes, namely, structural maintenance of chromosome 2 and heat shock protein beta-1, as the prognostic markers for GC. Furthermore, MR indicated no causal association between the butyrate metabolic pathway and GC.

Keywords: Butyrate metabolism (BM); gastric cancer (GC); prognosis; Mendelian randomization (MR); bioinformatics

Submitted Apr 24, 2024. Accepted for publication Dec 17, 2024. Published online Feb 26, 2025.
doi: 10.21037/tcr-24-677
View this article at: https://dx.doi.org/10.21037/tcr-24-677

Introduction

Each year, gastric cancer (GC) affects more than 1 million people globally. It causes approximately 800,000 deaths, thereby being recognized as the fifth most prevalent cancer worldwide (1,2). The stomach is a sac-like organ; therefore, obstruction caused by the luminal growth of early tumors does not cause symptoms. Furthermore, the incidence of GC is high, the prognosis is poor, the metastasis is extensive, the effect of radiotherapy and chemotherapy is poor, and the metabolic and immune microenvironment disorders are easy to occur (3,4). The rapid development of medical technology has led to the advancement of comprehensive GC treatment; it includes surgical resection, chemotherapy, radiotherapy, targeted immunotherapy, and multi-disciplinary standardized and individualized treatment (5). However, the 5-year survival rate of GC is only approximately 20% (6). This low survival rate can be attributed to the low diagnosis rate of early GC and the chemical drug resistance of advanced GC. The prognosis of patients with GC is principally based on the tissue type and the differentiation degree of GC (7,8). Nonetheless, the prognosis of patients diagnosed with GC can vary substantially, resulting in inconsistent outcomes even among patients of similar age. Variations in the genetic traits and risk factors result in such discrepancies, leading to diverse disease progression and treatment responses. Moreover, the routine utilization of tumor markers serves as an adjunct method for GC diagnosis (9). Tumor markers are produced by cancer cells or by other cells in the body as a response to cancer. Typically, they are found in the blood or urine. They serve as indicators to assess the likelihood of primary or secondary tumors. However, not all patients with cancer present with elevated levels of tumor markers. Additionally, certain non-cancerous conditions can increase specific tumor marker levels that exceed the reference ranges (10,11). Therefore, determining accurate and novel prognostic markers of GC will facilitate describing the molecular characteristics of patients with GC and identifying new therapeutic targets (12).

In recent years, numerous prognostic biomarkers have been identified for GC, aiming to enhance the accuracy of prognosis and guide personalized treatment strategies. Notable biomarkers include human epidermal growth factor receptor 2 (HER2), programmed death-ligand 1 (PD-L1), p53, and vascular endothelial growth factor (VEGF), which are involved in cellular proliferation, immune response modulation, and angiogenesis in GC (13). Circulating tumor markers such as carbohydrate antigen 72-4 (CA72-4), carbohydrate antigen 19-9 (CA19-9), and carcinoembryonic antigen (CEA) have also been explored as potential prognostic factors, though their sensitivity and specificity are often limited (14). Prognosis prediction models, such as the American Joint Committee on Cancer (AJCC), tumor node metastasis (TNM) classification staging system and gene expression-based models, have been widely employed to predict patient outcomes. However, these models frequently fail to account for the complex molecular interactions within the tumor microenvironment and the genetic heterogeneity of GC, leading to variable prognostic outcomes. Despite the identification of these biomarkers and models, their clinical utility remains limited due to several factors. First, many of the proposed biomarkers lack validation in independent cohorts, which undermines their generalizability. Second, most models do not incorporate the tumor microenvironment's impact on prognosis, particularly the role of immune cells and metabolic factors, which are critical for predicting patient responses to therapy. Lastly, many studies have been unable to establish causality between molecular markers and disease progression, relying instead on observational correlations that may be influenced by confounding factors.

To overcome these limitations, this study employs Mendelian randomization (MR) analysis, a method that

Highlight box

Key findings

- This study identifies two butyrate metabolism genes, SMC2 and HSPB1, as prognostic markers for gastric cancer (GC). The prognostic model developed demonstrated an area under the curve of 0.716, indicating its robust predictive capability. Mendelian randomization analysis revealed no direct causal association between BM and GC prognosis.

What is known and what is new?

- BM has been implicated in various cancer types, but its role in GC prognosis remains unclear.
- The study not only identifies potential prognostic markers linked to BM but also establishes a comprehensive predictive model for GC prognosis while clarifying the lack of causality between BM and GC progression.

What is the implication, and what should change now?

- These findings suggest that while BM-related genes can serve as prognostic markers, they may not be direct therapeutic targets in GC. Future research should focus on exploring alternative pathways involving BM and the gut microbiome's role in GC progression to develop more effective therapeutic strategies.

leverages genetic variants as instrumental variables to explore the causal relationships between exposures—such as butyrate metabolism (BM)—and disease outcomes like GC prognosis (15-17). Unlike traditional observational studies that are prone to confounding, MR provides a robust framework for establishing causality, thus addressing the controversy surrounding whether butyrate metabolism-related genes (BMGs) are merely correlated with GC prognosis or play a direct causal role in its progression.

While the prognostic significance analysis of BMG focuses on identifying correlations between BMG expression and patient survival, it does not establish whether these genes directly influence disease progression. In contrast, MR analysis is used to determine whether there is a causal relationship between BM and GC prognosis, providing a more definitive answer to whether BMG can be considered a therapeutic target. The combination of these two approaches allows for a more comprehensive understanding of BMG's role in GC, addressing both the association and the causality questions.

Butyrate is a short-chain fatty acid; it is a major metabolic product of the intestinal flora fermentation. It maintains the normal functioning of the intestinal epithelial cells and plays a mediating role in regulating host energy homeostasis by the intestinal flora. Butyric acid plays an anti-inflammatory role by regulating intestinal microbiota directly (18). This is because butyric acid increases short-chain fatty acid-producing bacteria and reduces endotoxin-secreting bacteria. Butyrate treatment can reduce high-fat diet-induced pro-inflammatory markers in mice, such as tumor necrosis factor alpha and interleukin-1β (19). Changes in the intestinal microbiome may be related to the close interaction among chronic gastrointestinal diseases, intestinal microbiome, and GC. Researchers have detected numerous *Klebsiella pneumoniae* and *Shigella enterobacteria* in the gastric mucosa of patients with GC (20,21). BM is central to several tumors. Additionally, the combination of butyrate and G protein-coupled receptor GPR109A can inhibit colon tumors (22). Exploring butyric acid metabolism in GC will enhance our comprehension of the underlying molecular mechanisms and pave the way for innovative and improved treatment approaches.

In this study, we screened characteristic genes through weighted gene co-expression network analysis (WGCNA) and categorized patients with GC and characteristic genes to observe the differences in prognosis and immune efficacy among GC subtypes. Subsequently, we analyzed the differential genes among subtypes by univariate Cox regression and least absolute shrinkage and selection operator (LASSO) regression to construct a prognostic model based on the genes related to butyric acid metabolism. This step helped us predict and analyze the prognosis of patients with GC comprehensively. In addition, we adopted the MR approach to investigate the causal relationship between BMG and GC progression. We present this article in accordance with the TRIPOD and STROBE-MR reporting checklists (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-677/rc).

Methods

Information processing and data sets collection

Initially, we extracted 349 genes associated with BM from the Molecular Signature Database (table available online at https://cdn.amegroups.cn/static/public/tcr-24-677-1.xlsx). Subsequently, the expression matrix and corresponding clinical data sets of GC samples were acquired from The Cancer Genome Atlas (TCGA) database (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). Additionally, to supplement our analysis, the GC expression matrix and clinical data were acquired from the Gene Expression Omnibus (GEO) database (GSE84437, Platforms: GPL6947) (https://www.ncbi.nlm.nih.gov/). The TCGA included 443 patients with GC, and the clinical data included survival information, age, gender, tumor stage, etc. GSE84437 included 483 patients with GC, and the clinical data included survival information, age, sex, and tumor stage. BMG data (GWAS ID: met-d-bOHbutyrate) were downloaded from the Genome-Wide Association Studies (GWAS) database (https://gwas.mrcieu.ac.uk/), which consists of 12,321,875 single nucleotide polymorphisms (SNPs) from a sample of 113,595 Europeans. Furthermore, GC data sets (GWAS ID: bbj-a-119) were downloaded, which consisted of 8,885,324 SNPs from a sample of 202,308 East Asians.

Identification of differential genes associated with BMG

First, the TCGA-stomach adenocarcinoma (STAD) data set was combined with GSE84437 through the "sva" package (v.1.50.0). Consequently, the "limma" package (v.3.58.1) was used to identify the combined data set for differential genes in GC tissues and the control group [|log fold change (FC)| ≥0.585, P<0.05]. Differential genes correlated with BMG were identified through the "Veen" package (v.1.12.0).

Additionally, we generated the leading 50 differential genes through a heatmap and Volcano plot.

WGCNA

The "WGCNA" package (v.1.72-5) was performed on BMGs; soft-threshold power β was set to 5 (scale-free R^2=0.9). Subsequently, the data were converted into a topological overlap matrix, and each module was set to consist of at least 10 genes. Finally, the modules were stratified, and similar modules were merged.

Prognostic identification and immune characteristics of BMG-related subtypes

We classified the patients with GC into different subtypes using the "ConsensusClusterPlus" package (v.1.66.0). The difference between the two subtypes was identified by a principal component analysis (PCA). Subsequently, the differences in clinical information and prognosis production between the subtypes were analyzed. We assessed the differences between the immune cells of different subtypes.

Differential analysis of the subtypes and prognostic types

The differential genes of different BMG subtypes were identified. A univariate Cox regression was used to determine the prognosis of the differential genes. We genotyped the prognosis-related differential genes and assessed their clinical differences and prognosis.

Constructing and validating the prognostic model

To explore the role of BMG in GC, we screened the differential prognostic genes by the LASSO ("glmnet" package v.4.1-8) regression algorithm. Additionally, the risk-scoring formula of BMG was calculated. To evaluate the accuracy of the model, the combined data sets of TCGA-GEO were assigned randomly, and the prognosis of each data sets was evaluated. We have randomly assigned the TCGA dataset at a ratio of 7:3 in order to verify the accuracy of the model.

Constructing a nomogram prognostic model

The "RMS" package (v.6.7-1) was used to construct a nomogram prognostic model after combining the age, sex, stage, and risk scores of the patients with GC. Additionally, we established calibration curves to evaluate the accuracy of the nomogram prognosis model for 1-, 3-, and 5-year predictions.

Immune infiltration analysis

First, we used the "CIBERSOER" (v.0.1.0) algorithm to calculate the 22 types of immune cell infiltration data of patients with GC. Furthermore, the correlation between the prognostic model genes and immune cells was determined. Subsequently, we evaluated the difference in tumor microenvironment among the subtypes of GC by calculating the immune microenvironment score of the patients. Finally, the microsatellite instability among different subtypes was evaluated.

Screening for SNP exposure and outcome

All SNPs were screened with the thresholds R^2=0.001 and kb =10,000 to avoid linkage imbalances. Additionally, we set $P<5\times10^{-8}$ to screen the eligible SNPs that satisfied MR. The F statistic was calculated to estimate the sample overlap effects and weak tool bias, considering relatively loose thresholds where F <10 suggests bias.

Tissue sample acquisition

A total of 3 STAD tissues and matched normal tissues were collected from the Gastrointestinal Surgery Department of The Third Affiliated Hospital of Soochow University in April 2018 to October 2018. Patients received no neoadjuvant chemotherapy or radiotherapy. Consents was obtained from the study participants prior to study commencement. Clinicopathological characteristics of these patients were also collected. The collected tissues were frozen in liquid nitrogen until RNA extraction.The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This research involving human participants was reviewed and approved by the Ethics Committee of The Third Affiliated Hospital of Soochow University. Written informed consent was obtained from all the patients before their participation in the study.

Quantitative real-time polymerase chain reaction (qRT-PCR)

Total RNA was extracted from the STAD tissues and

matched normal tissues using RNA extraction kit (Vazyme, Nanjing, China). Subsequently, the total RNA from each tissue was reverse-transcribed into cDNA using the PrimeScript RT reagent Kit (Takara, Dalian, China, DRR037A). Then the synthetic cDNA was used as a template for quantitative experiments, and Bio-Rad CFX96 Real-time PCR assay system (Bio-Rad, Hercules, CA, USA) was used. The results were analyzed using the comparative Ct method, normalizing the Ct values of each gene to the corresponding actin beta (ACTB) Ct readings. All data were presented as the mean ± standard deviation (SD) of three independent experiments.

Immunofluorescence

Immunofluorescence was fulfilled with essential antibodies against SMC2, HSPB1. To begin with, paraffin-embedded GC tissues were deparaffinized and rehydrated. After antigen recovered tissues were recolored with essential antibodies overnight at 4 ℃, they were in this way recolored with fluorescein-labeled auxiliary antibodies at 37 ℃ for 1 h. After recoloring with 4',6-diamidino-2-phenylindole (DAPI), the target proteins were visualized with filtering microscope.

MR analysis

This paper employed the "harmonize_data" function within the TwoSampleMR package to normalize effect estimates. The key MR techniques included MR-Egger13, Weighted median 14, and inverse variance weighted (IVW) utilizing both multiplicative random effects and fixed effects 15. Emphasis in the primary analysis was placed on the IVW method, which requires SNPs to fully comply with the three principles of MR research in order to obtain correct causal estimates. The MR-Egger regression intercept term was used to assess the possible presence of horizontal pleiotropy, where deviation from zero ($P<0.05$) indicates directional pleiotropy.

The primary analysis focused on the IVW method, which relies on SNPs that fully satisfy the principles of MR to generate accurate causal estimates. The MR-Egger regression intercept term was used to detect potential horizontal pleiotropy, with deviations from zero ($P<0.05$). The weighted median method assesses causality by aggregating the majority of SNPs. If the IVW method yields consistent results without additional heterogeneity, it can be considered positive, even if other methods do not yield significant results, showing that the directionality of β-values is consistent across methods and pleiotropy or heterogeneity is absent.

This paper visualized findings using scatter plots, forest plots, and funnel plots. In the scatter plot, the IVW method remained the primary focus, with a small intercept indicating minimal confounding and reliable results. A positive slope indicated a risk factor, while a negative slope indicated a protective factor. The forest plot assessed the diagnostic efficacy of each SNP locus in predicting exposure factors, focusing on the position of the IVW. Funnel plots evaluated randomness, with symmetric distribution along the IVW line indicating conformity to Mendel's second law of random grouping.

Sensitivity analysis

The MR analysis was primarily performed by the IVW approach, which assumes the absence of average pleiotropic effect. Cochran's Q statistic was first computed to evaluate the heterogeneity induced by different genetic variants in the fixed-effect IVW method, with a P value <0.05 indicating the presence of heterogeneity. Subsequently, a test for horizontal pleiotropy was conducted to determine whether the IVW estimates were influenced by confounding factors. A P value greater than 0.05 suggested the absence of horizontal pleiotropic effects, indicating no confounding factors in the study. To evaluate the stability of results, a leave one-out (LOO) analysis was conducted to assess if the overall effect was driven by a single SNP.

Statistical analysis

The R 4.2.0 programming dialect was utilized for statistical analyses. We performed the Spearman's rank test to determine between-group factually critical contrasts. The test displays information through cruel and SD. $P<0.05$ indicated statistical significance. Additionally, we performed two-sample MR studies. In this study, the IVW, MR-Egger regression, weighted median, weighted model, and simple mode were used for a causal analysis.

Results

Screening of differential genes associated with BMG

The combined TCGA-STAD and GSE84437 intersected with the genes correlated with BMG. Finally, we acquired 209 genes associated with BMG through differential

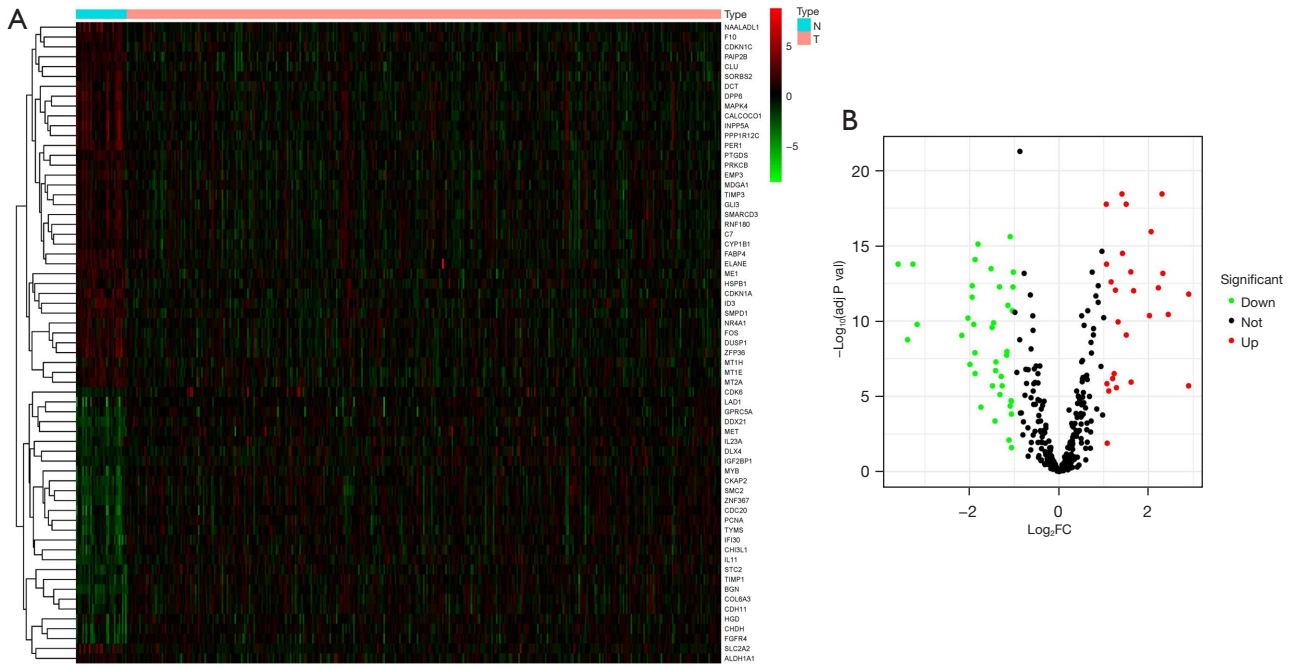

Figure 1 BMG differential analysis. (A) Heat map of the leading 50 genes in BMGs. (B) Volcano plot of the leading 50 genes in BMGs. FC, fold change; BMG, butyrate metabolism-related gene.

analysis of the butyrate genes. We visualized the heatmaps and volcanic plots of the leading 50 BMG genes (*Figure 1*).

WGCNA analysis

We performed a cluster analysis of the 209 BMG genes. We set the soft-threshold power at 5 (R^2=0.9) to construct a scale-free network. After removing the outliers, we constructed a sample clustering tree (*Figure 2A*,*2B*). Subsequently, the adjacency matrix was converted into a topological overlap matrix. The blue model demonstrated the highest correlation with GC. Therefore, we selected 56 genes in the blue module for further analysis (*Figure 2C*).

Gene mutations associated with BMG in GC

We identified somatic mutations in 56 genes related to BMG in patients with GC from the TCGA data sets (*Figure 3*). Of 433 samples, 130 genes were mutated, with a mutation rate of 30.02% and a mutation frequency ranging from 0% to 5%. The *SORBS2* gene had the maximum mutations (5%).

Prognostic regulatory network of BMGs

Prognostic correlation networks were constructed for 56 genes to detect correlations among the genes (*Figure 4*). Dual-specificity phosphatase 1 (DUSP1), heat shock protein beta-1 (HSPB1), mitogen-activated protein kinase 4 (MAPK4), and structural maintenance of chromosomes 2 (SMC2) accounted for a large proportion. SMC2 was identified as the favorable factor, whereas DUSP1, HSPB1, and MAPK4 were identified as the risk factors.

Identification of subtypes related to BMG in GC

We classified the patients with GC into subtypes A and B based on 56 genes (*Figure 5A*). PCA indicated that we classified the patients into two clusters (*Figure 5B*). Furthermore, differences between the immune cells of subtypes A and B were analyzed. Immune cells of subtype B displayed higher functional activity than cells of subtype A (*Figure 5C*). In addition, BMGs were predominantly expressed in subtype A (*Figure 5D*). The Kaplan-Meier survival curve indicated that the prognosis of subtype A was better than that of subtype B (*Figure 5E*), suggesting that the

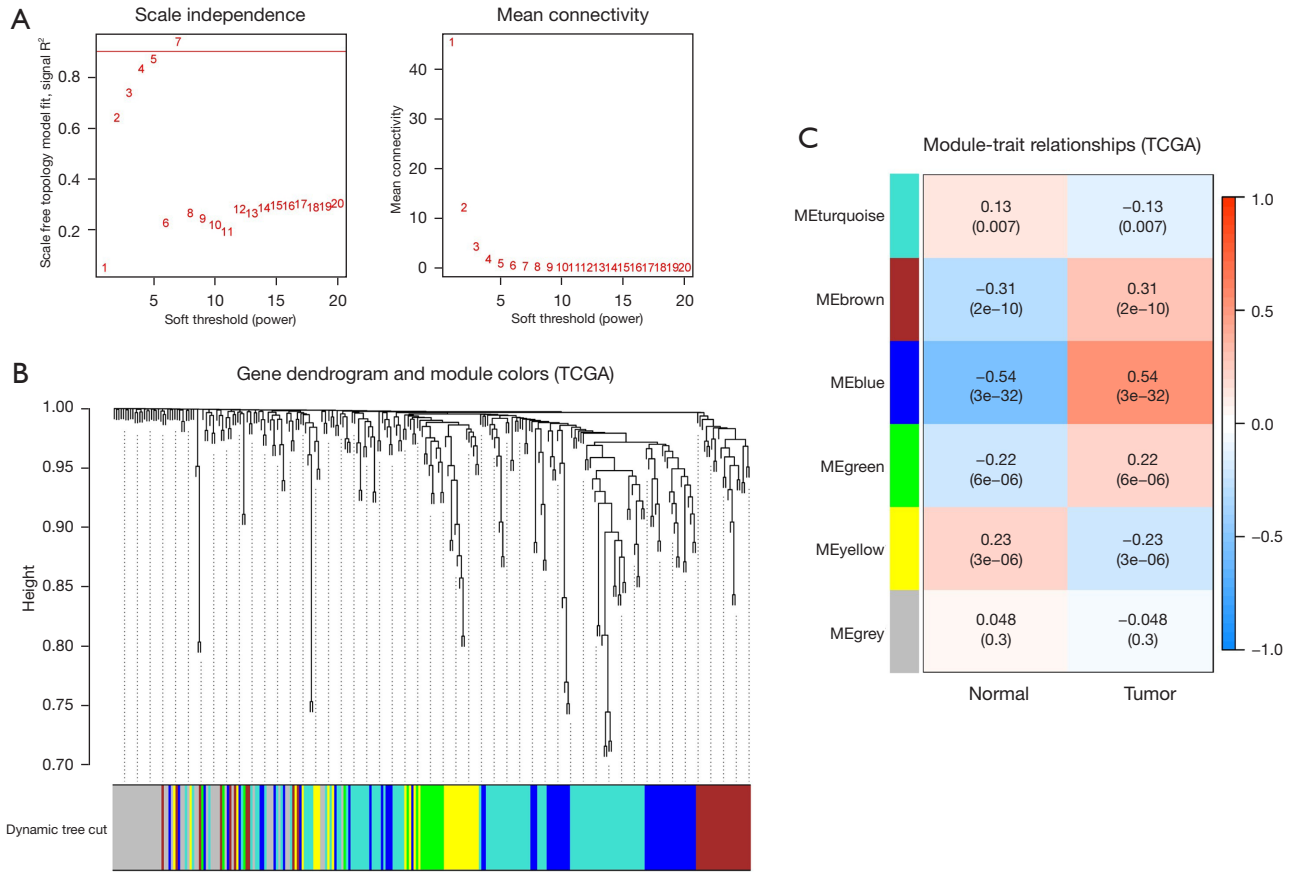

Figure 2 WGCNA analysis. (A) Scale-free exponential analysis of soft threshold powers. (B) Gene clustering tree. (C) Module gene correlation heatmap. TCGA, The Cancer Genome Atlas; ME, module eigengene; WGCNA, weighted gene co-expression network analysis.

genes associated with BMG may serve as protective factors for GC. We analyzed the pathway differences between the subtypes. Subtype A was primarily enriched in the basal transcription factor, oocyte meiosis, and proteasome. By contrast, subtype B was primarily enriched in dilated cardiomyopathy and hypertrophic cardiomyopathy (*Figure 6*).

Screening of prognostic genes in BMGs

We acquired 17 differential genes through a comparative difference analysis of subtypes A and B. Ten BMG-related prognostic genes were acquired through the univariate prognostic analysis of the 17 subtypes of differential genes (*Figure 7A*). Furthermore, we classified patients with GC and BMG prognostic genes and the subtypes A and B (*Figure 7B,7C*). The BMG prognostic genes were highly expressed in subtype A (*Figure 7D*). Additionally, subtype A demonstrated a better prognosis in the survival curve (*Figure 7E*), thereby indicating that BMGs may be a potential protective factor for GC.

Constructing a BMG prognostic model

Our findings indicated a robust correlation among the BMG subtypes, BMG prognostic subtypes, and survival prognosis (*Figure 8A*). Utilizing the LASSO algorithm, we developed a prognostic model for BMGs (*Figure 8B*). Table S1 describes the model formulation. We divided the TCGA-GEO data sets into training and validation sets to assess the model's accuracy (*Figure 8C-8E*). This step facilitated the verification of the BMG prognostic model. We observed promising prognostic potential through the risk scores, survival status profiles, and prognostic gene heat maps associated with BMGs. Notably, both groups

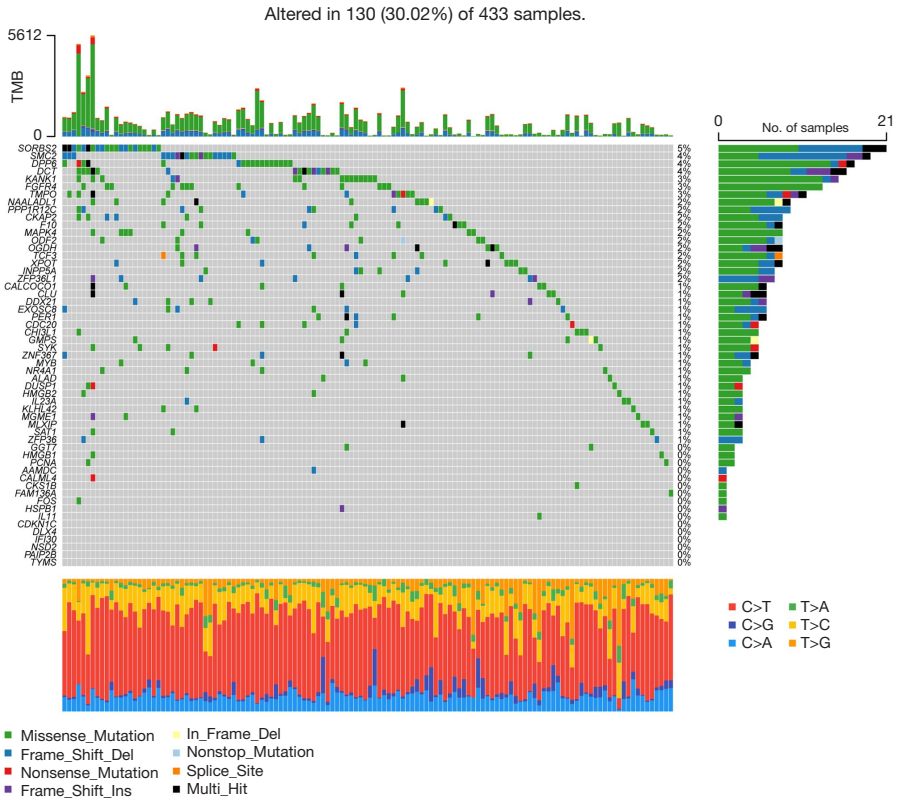

Figure 3 BMG mutation map of TCGA data sets. TMB, tumor mutational burden; BMG, butyrate metabolism-related gene; TCGA, The Cancer Genome Atlas.

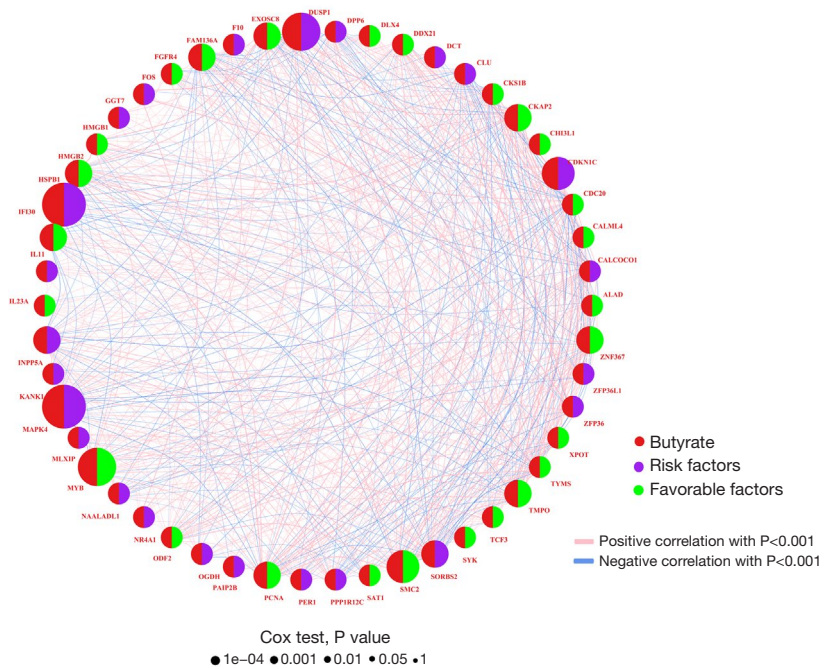

Figure 4 BMG prognostic correlation network. BMG, butyrate metabolism-related gene.

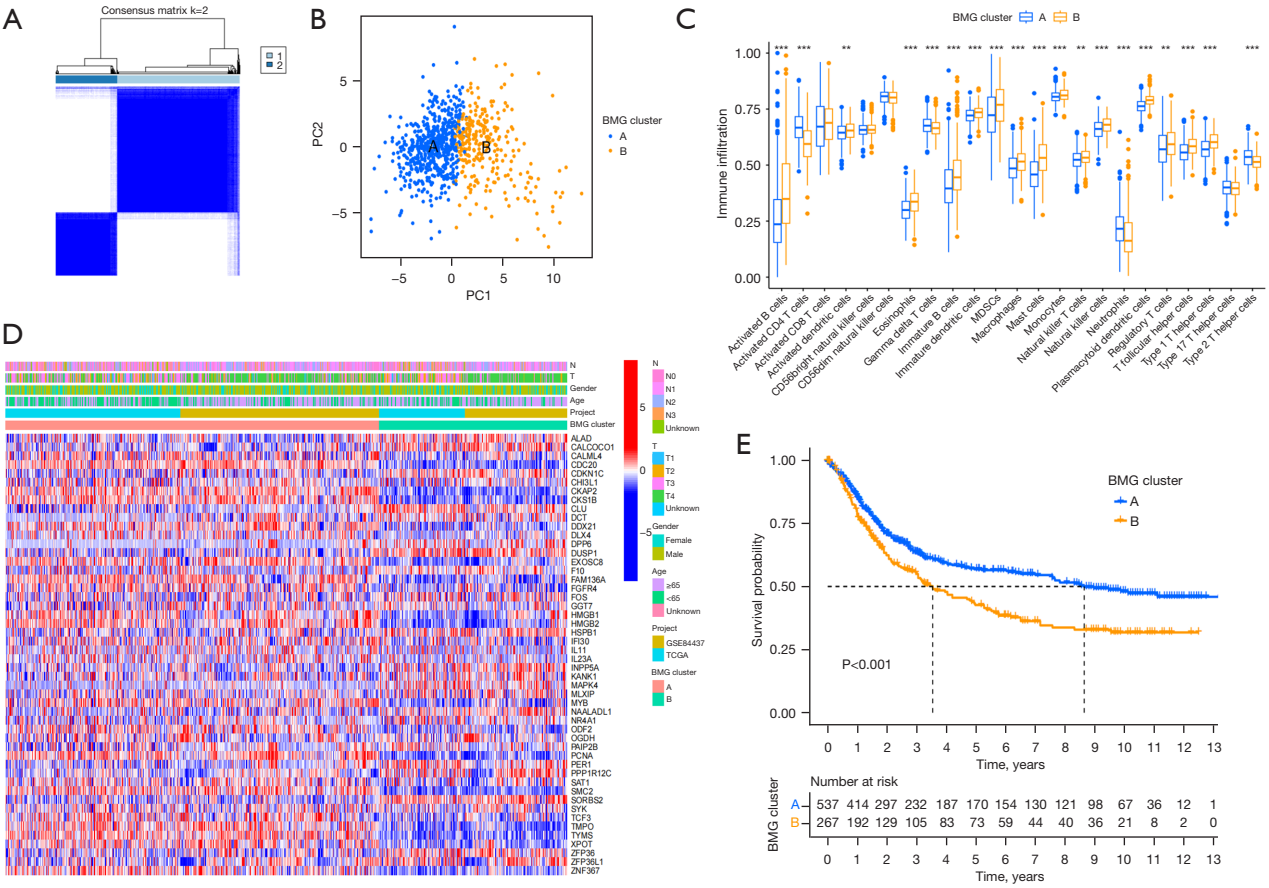

Figure 5 Consensus clustering of the genes associated with butyrate metabolism. (A) Consensus clustering typing matrix. (B) PCA typing of two subtypes. (C) Differential typing of immune cells between two subtypes. (D) Complex heat map of BMG clinical features. (E) KM survival curves of two subtypes. **, P<0.01; ***, P<0.001. BMG, butyrate metabolism-related gene; MDSC, myeloid-derived suppressor cells; TCGA, The Cancer Genome Atlas; PCA, principal component analysis; KM, Kaplan-Meier.

consistently demonstrated better prognosis in the low-risk group than that in the high-risk group, emphasizing the viability of BMGs as a strong prognostic factor for GC.

Nomogram construction and evaluation

We constructed a nomogram prognostic model using the clinical information (age, sex, and stage) of the patients to explore its clinical application (*Figure 9A*). The calibration curve suggests that the predicted and actual values for 1, 3, and 5 years were close to agreement (*Figure 9B*). Additionally, we verified the accuracy of the model through the receiver operating characteristic (ROC) curve, whose area under the ROC value was 0.716 (*Figure 9C*).

Immune infiltration analysis

We acquired 22 types of immune cell infiltration data through the CIBERSOER algorithm. Subsequently, we evaluated the immune cell correlation of BMG prognostic model genes (*Figure 10A*). HSPB1 was positively correlated with T cell regulation and natural killer cell activation. SMC2 was positively correlated with the cluster of differentiation 4 (CD4) T cell activation (*Figure 10B*). Results of the immune microenvironment indicated substantial immunosuppression in the high-risk group (*Figure 10C*). Additionally, microsatellite instability analysis showed a higher proportion of stability groups [microsatellite stability (MSS)] in high-risk groups

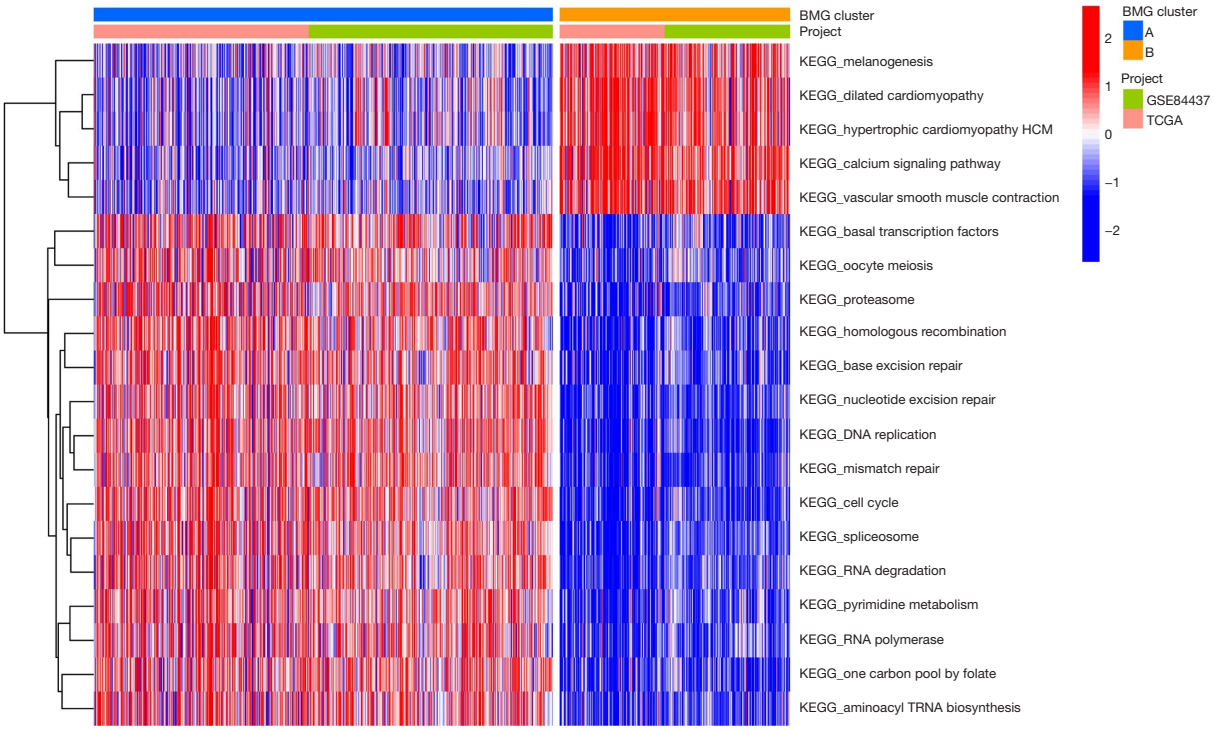

Figure 6 GSVA results. BMG, butyrate metabolism-related gene; TCGA, The Cancer Genome Atlas; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSVA, gene set variation analysis.

(*Figure 10D*). The high-frequency microsatellite instability group differed from the other two groups, and patients in the high-frequency microsatellite instability group also had the lowest scores (*Figure 10E*).

MR analysis

We conducted an MR study to explore the causal relationship between the groups. This step helped us explore the association between BMG and GC. First, SNPs strongly correlated with the exposure factors were selected as the instrumental variables for MR analysis. We set the filtering condition at $P<5\times10^{-8}$ (*Figure 11*). Subsequently, we conducted a linkage disequilibrium analysis and a F-number test. Fourteen SNPs were selected for further analysis (Table S2). Our results provided no evidence for an association between BMG and GC (*Table 1* and *Figure 12*).

Validation of prognostic genes

We collected tissue samples and adjacent tissues of 3 GC patients, and identified by PCR analysis that, the primer sequence is described in Table S3, SMC2 was highly expressed in tumor tissues, while HSPB1 was low expressed in tumor tissues (*Figure 13A*). We further detected the protein level expression of prognostic genes by immunofluorescence, and found that SMC2 was highly expressed in tumor tissues, while HSPB1 was low expressed in tumor tissues (*Figure 13B*), which is consistent with our previous results.

Discussion

Butyrate plays a multifaceted and intricate role in cancer. Researchers have demonstrated the substantial anti-tumor effects of butyrate across various cancer types. For instance, butyrate impedes bladder cancer migration by inducing autophagy (23) and suppresses the proliferation of breast cancer cells by generating reactive oxygen species (24). Additionally, the combination of butyrate and cisplatin enhances the therapeutic effects for GC (25). Therefore, butyrate serves as a potential biological target or therapeutic agent for GC. However, researchers have not elucidated the underlying regulatory mechanism and role of the butyrate

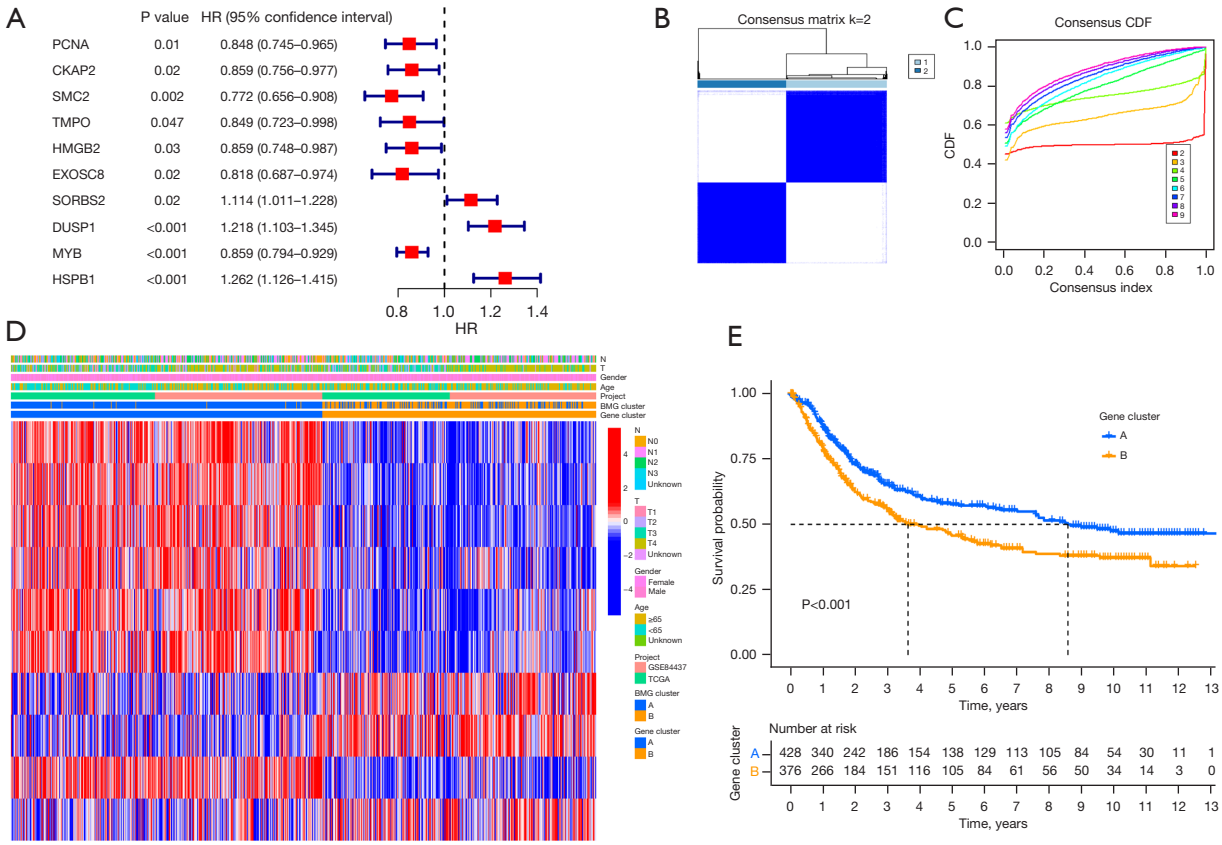

Figure 7 BMG prognostic genotyping. (A) Univariate Cox prognostic phenotype forest map. (B,C) Consensus clustering matrix. (D) Complex heat maps of clinical features of BMG prognostic classification. (E) KM survival curve of the prognostic subtypes. HR, hazard ratio; CDF, cumulative distribution function; BMG, butyrate metabolism gene; TCGA, The Cancer Genome Atlas; KM, Kaplan-Meier.

metabolic pathway in GC.

We identified 209 differentially expressed genes related to BM, followed by the screening of 56 genes with the strongest correlation through co-expression networks. We classified the patients into subtypes A and B based on gene expression. Compared with subtype A, subtype B demonstrated a worse prognosis and an effect on immunotherapy, thereby indicating that subtype B can be used to predict the immunotherapy situation in patients with GC. We acquired 17 differential genes associated with subtypes A and B and 10 differential genes associated with BM and prognosis by univariate Cox regression. This step helped us elucidate the role of BMGs in GC prognosis. We conducted retyping based on the expression of 10 differential genes. These genes are central to regulating GC treatment and prognosis.

The diagnostic model of BMG prognostic genes suggested the potential role of the butyrate metabolic pathway in the early diagnosis and prognosis of GC. The model was constructed using the LASSO algorithm and consisted of two genes, namely, SMC2 and HSPB1. SMC2 is the core subunit of the agglutination enzyme and is central to chromosome organization and separation. SMC2 is a member of the SMC family and is involved in several biological processes, such as apoptosis, chromosome separation defects, and DNA damage (26,27). Researchers have reported a notable upregulation of SMC2 in patients diagnosed with colorectal cancer, GC, lymphoma, and specific forms of neuroblastoma (28-30). SMC2 is a risk biomarker for pancreatic cancer (31). Je *et al.* (32) reported that 22% of the patients with GC and microsatellite instability-high (MSI-H) demonstrated changes in both

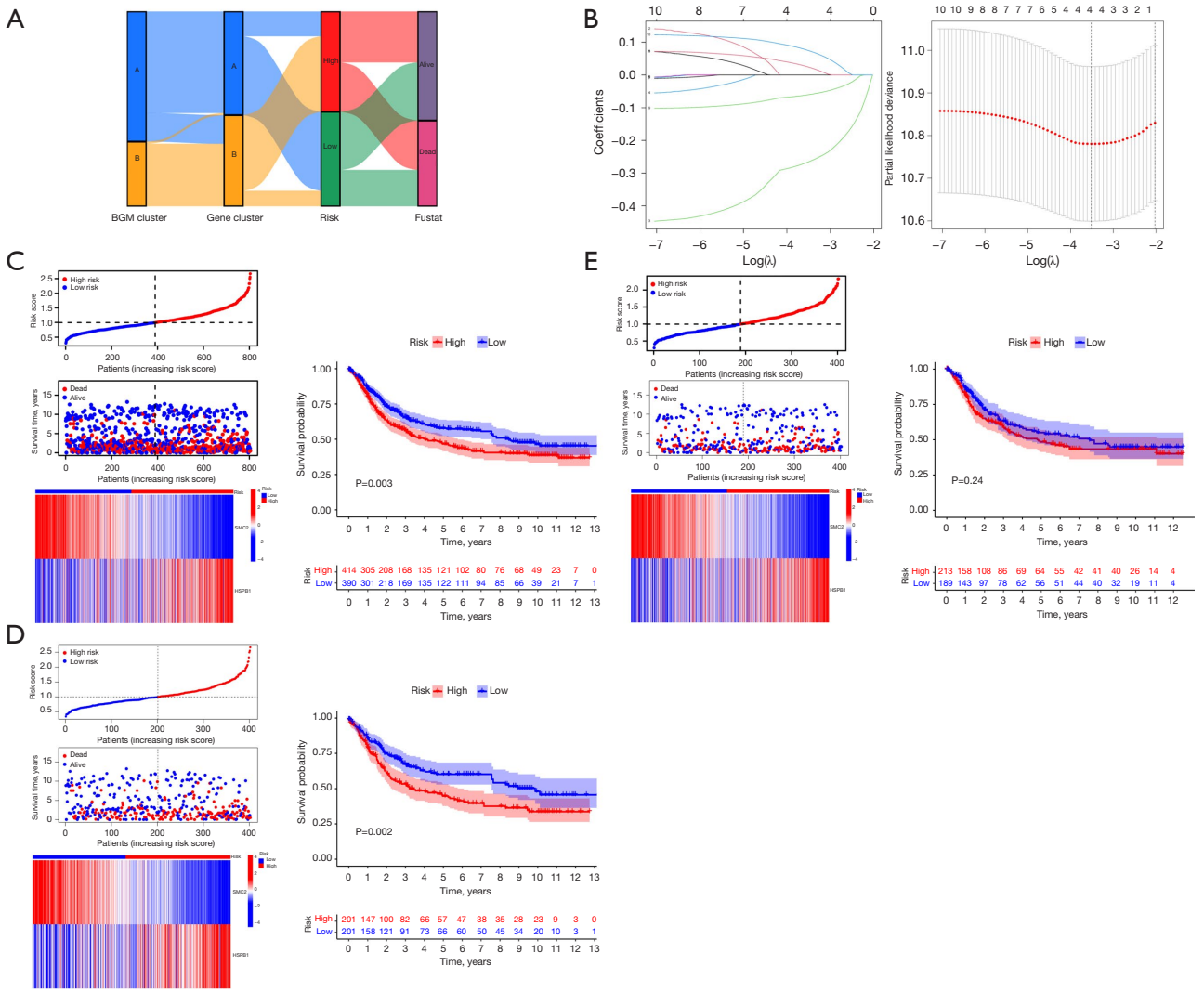

Figure 8 Construction of the prognostic model. (A) Associations among BMG subtypes, BMG prognostic subtypes, and prognosis survival. (B) LASSO prognostic model construction. (C-E) Evaluation of the prognostic models. BMG, butyrate metabolism-related gene; LASSO, least absolute shrinkage and selection operator.

displacement mutation and the loss of expression of the *SMC2* gene. Thereby, changes in the *SMC2* gene may be related to GC pathogenesis. Heat-shock 27-kDa (HSP27) protein 1 is a ubiquitin-binding protein involved in the proteasomal degradation of proteins under stress. HSP27 knockdown inhibits the differentiation of human red blood cells and leukemia cell lines (33). HSPB1 plays an important role in various cancers, including GC (34,35). In GC, HSPB1 can form a p38/pHSPB1 cascade with p38/MAPKAPK2 to promote tumor growth and metastasis (36). Ser15 of HSPB1 is phosphorylated by T-LAK cell-originated protein kinase (TOPK), promoting the proliferation and transfer of GC through the FYN/TOPK/HSPB1 axis (37). The research status of previous studies is consistent with our results, thereby confirming the accuracy and scientific validity of the model.

Similarly, we used MR to evaluate the causal relationship between BGM and GC. MR indicated no causal

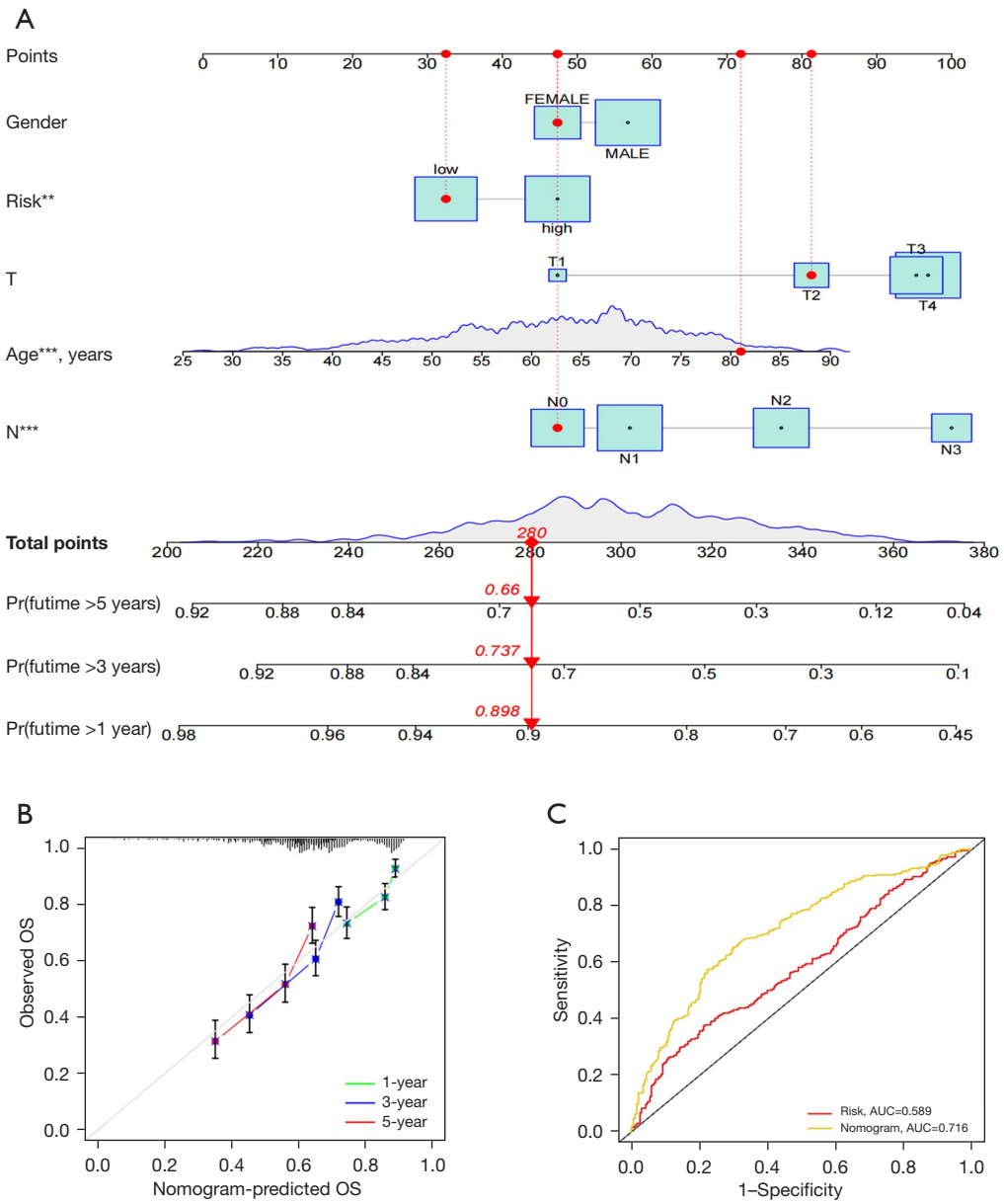

Figure 9 Nomogram construction and evaluation. (A) Nomogram construction. (B) Calibration curve of the nomogram model. (C) ROC curve of the nomogram model. **, P<0.01; ***, P<0.001. OS, overall survival; AUC, area under the plasma concentration-time curve; ROC, receiver operating characteristic.

relationship; however, butyrate is a metabolite of intestinal flora. Furthermore, the intestinal flora is negatively correlated with GC regulation (38). Therefore, butyrate may act as an intermediate regulator between the intestinal flora and GC. We intend to explore the regulatory relationship among the intestinal flora, BGM, and GC in the future.

Conclusions

We identified two BG prognostic genes, namely, SMC2 and HSPB1, as the prognostic markers of GC. Furthermore,

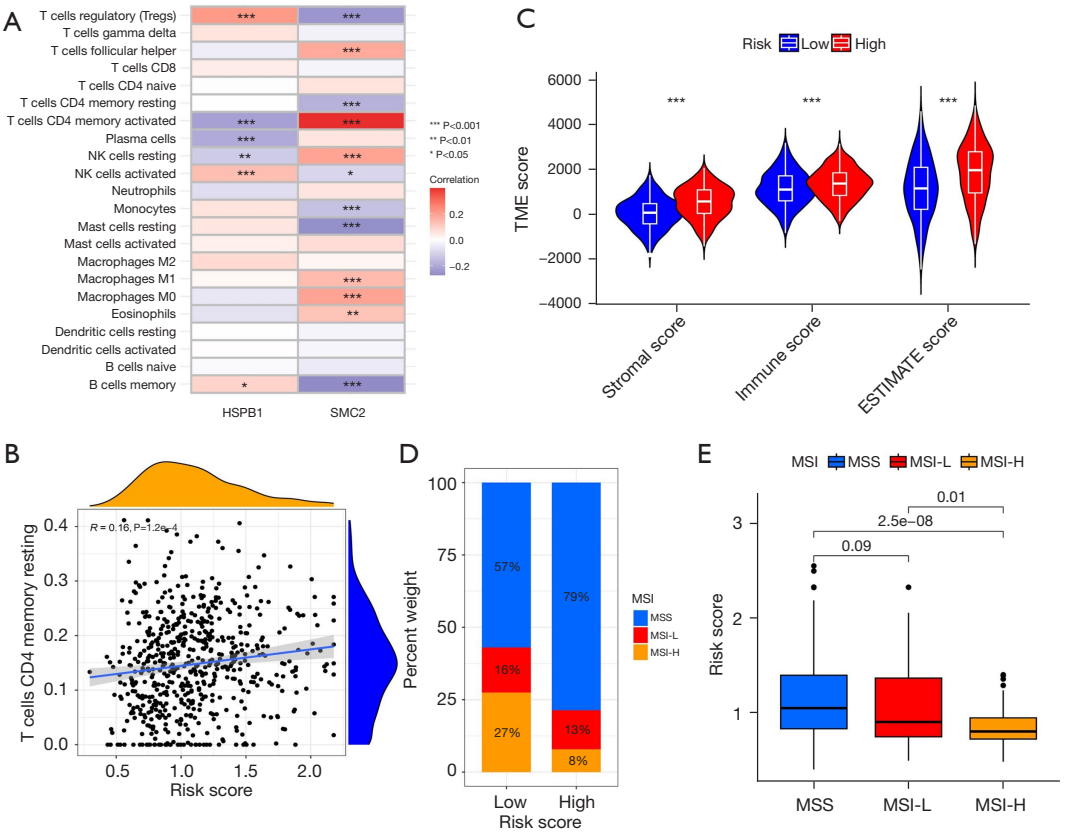

Figure 10 Immune infiltration analysis. (A) Immunocell correlation analysis of prognostic genes. (B) Dot plot of SMC2-mediated CD4 T cell activation. (C) TME differential analysis. (D,E) MSI analysis. *, P<0.05; **, P<0.01; ***, P<0.001. CD4, cluster of differentiation 4; SMC2, structural maintenance of chromosomes protein 2; TME, tumor microenvironment; MSI, microsatellite instability; MSS, microsatellite stability; MSI-L, microsatellite instability-low; MSI-H, microsatellite instability-high.

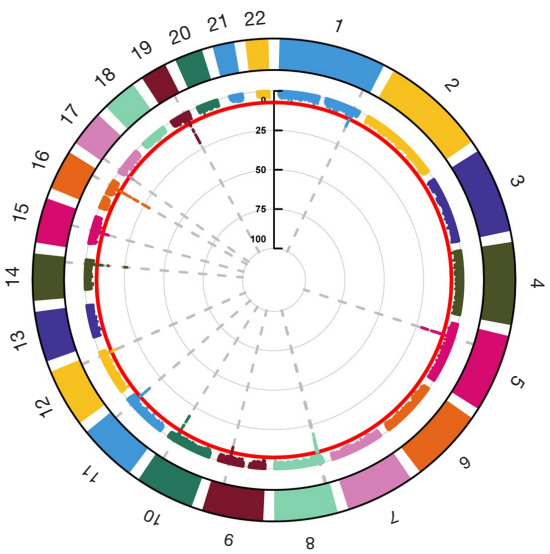

Figure 11 SNP correlation analysis. Each color represents one chromosome. The red line represents the threshold P<5e−8. Y-axis is the statistical significance P value associated with this site [$-Log_{10}(P)$]. SNP, single nucleotide polymorphism.

Table 1 MR analysis of the causal effect of butyrate metabolism on gastric cancer

Case	Outcome	Exposure	Method	P value	OR	95% CI	
						Lower	Upper
1	GC	Butyrate	Inverse variance weighted	0.15	0.78	0.55	1.10
2	GC	Butyrate	MR Egger	0.50	0.61	0.16	2.42
3	GC	Butyrate	Weighted median	0.37	0.81	0.50	1.29
4	GC	Butyrate	Simple mode	0.81	1.11	0.47	2.63
5	GC	Butyrate	Weighted mode	0.77	1.12	0.53	2.38

MR, Mendelian randomization; OR, odds ratio; CI, confidence interval; GC, gastric cancer.

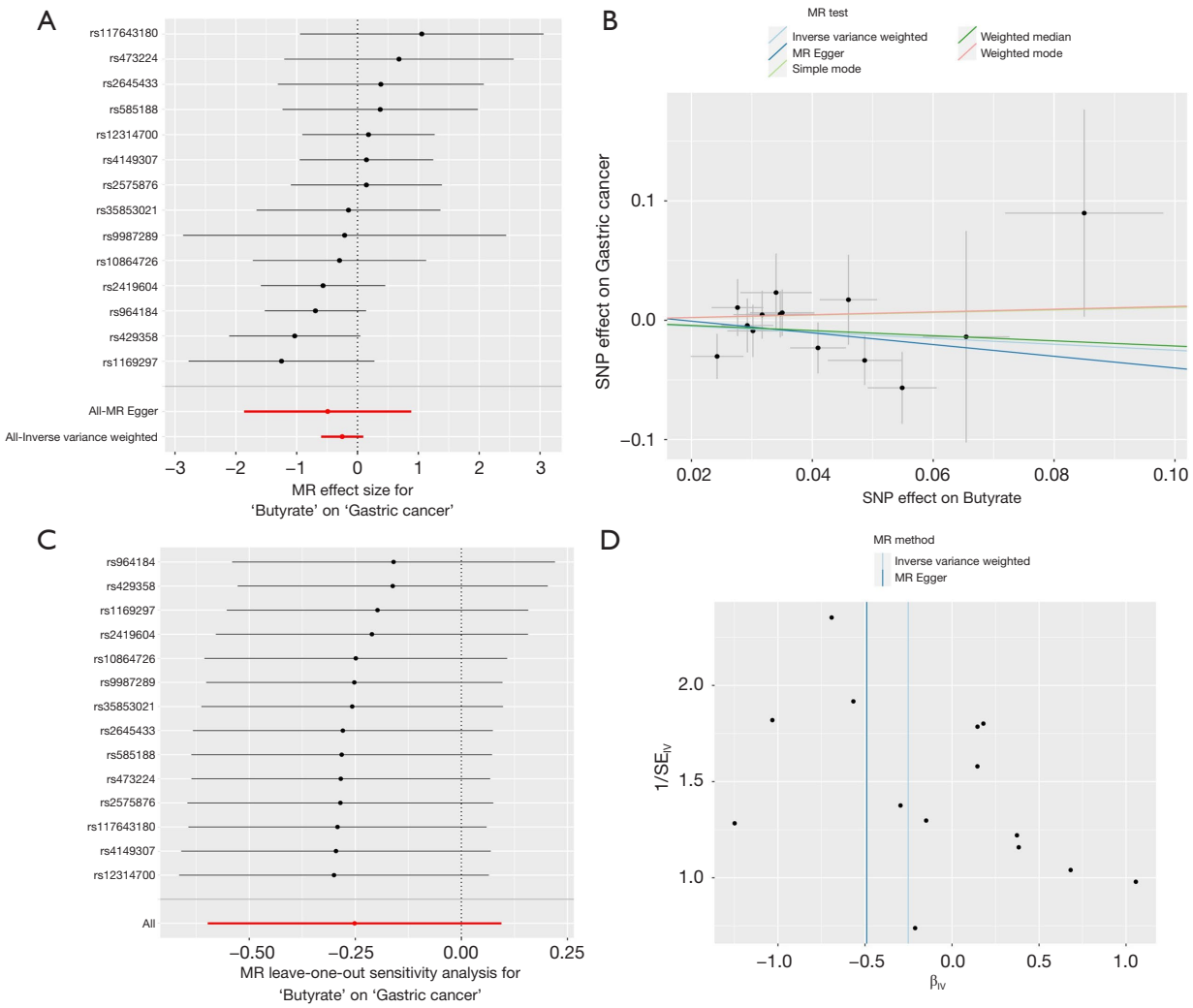

Figure 12 MR analysis. (A) MR analysis of forest maps. (B) MR analysis scatter plot. (C) Sensitivity analysis. (D) MR analysis funnel diagram. MR, Mendelian randomization; SNP, single nucleotide polymorphism; SE, standard error; IV, instrumental variable.

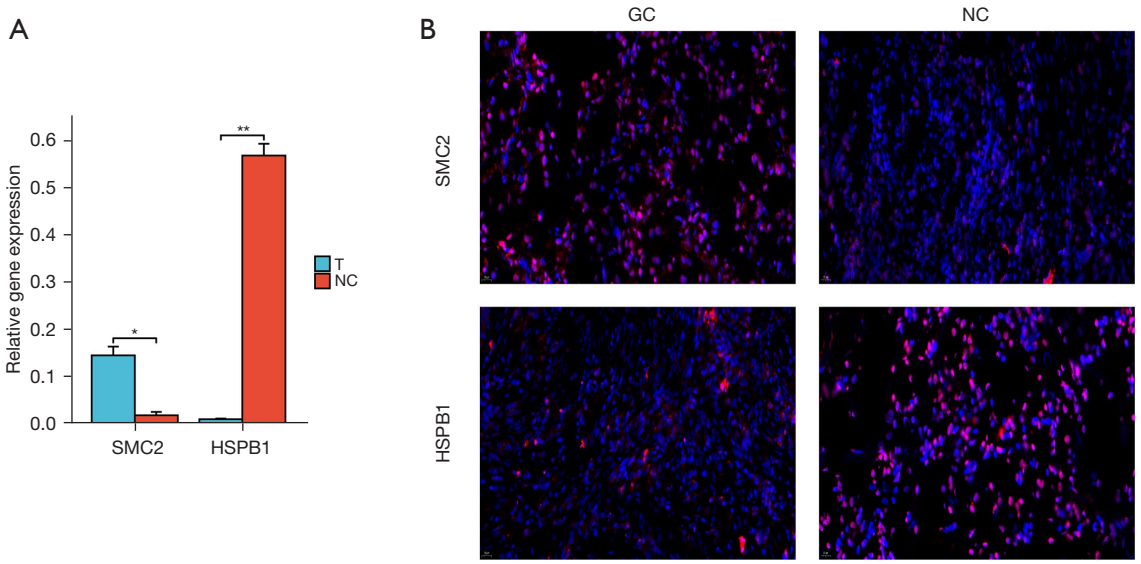

Figure 13 Experimental verification. (A) PCR validation. (B) Immunofluorescence validation. The magnification is ×40. *, P<0.05; **, P<0.01. T, tumor; NC, negative control; GC, gastric cancer; PCR, polymerase chain reaction.

we observed no causal association between the butyrate metabolic pathways and GC through MR.

Acknowledgments

We would like to thank Dr. Jin Feng of The Third Affiliated Hospital of Soochow University for providing clinical sample support for this study.

Footnote

Reporting Checklist: The authors have completed the TRIPOD and STROBE-MR reporting checklists. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-677/rc

Data Sharing Statement: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-677/dss

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-677/prf

Funding: The study was funded by the Nanjing Medical Key Foundation (No. ZKX21028), Key Project of TCM Technology Development in Jiangsu Province (No. ZD202227), The 75th Batch of China Postdoctoral Science Foundation Projects (No. 2024M754279), Natural Science Foundation of Jiangsu Province (No. BK20240738), Basic Science (Natural Science) Research Project in Universities of Jiangsu Province (No. 24KJB360004), Jiangsu Province Chinese Medicine Science and Technology Development Plan Youth Talent Project (No. QN202206), Nanjing University of Chinese Medicine Luo Linxiu Teacher Development Fund Project (No. LLX202310), Jiangsu Province Traditional Chinese Medicine Science and Technology Development Program (No. MS2023001).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-677/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This research involving human participants was reviewed and approved by the Ethics Committee of The Third Affiliated Hospital of Soochow University. Written informed consent was obtained from all the patients before their participation in the study.

References

1. Smyth EC, Nilsson M, Grabsch HI, et al. Gastric cancer. Lancet 2020;396:635-48.
2. Thrift AP, El-Serag HB. Burden of Gastric Cancer. Clin Gastroenterol Hepatol 2020;18:534-42.
3. Cui MY, Yi X, Zhu DX, et al. Aberrant lipid metabolism reprogramming and immune microenvironment for gastric cancer: a literature review. Transl Cancer Res 2021;10:3829-42.
4. Xu W, Ding H, Zhang M, et al. The prognostic role of fatty acid metabolism-related genes in patients with gastric cancer. Transl Cancer Res 2022;11:3593-609.
5. Ajani JA, D'Amico TA, Bentrem DJ, et al. Gastric Cancer, Version 2.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2022;20:167-92.
6. Machlowska J, Baj J, Sitarz M, et al. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. Int J Mol Sci 2020;21:4012.
7. Lordick F, Carneiro F, Cascinu S, et al. Gastric cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol 2022;33:1005-20.
8. Muro K, Van Cutsem E, Narita Y, et al. Pan-Asian adapted ESMO Clinical Practice Guidelines for the management of patients with metastatic gastric cancer: a JSMO-ESMO initiative endorsed by CSCO, KSMO, MOS, SSO and TOS. Ann Oncol 2019;30:19-33.
9. Shimada H, Noie T, Ohashi M, et al. Clinical significance of serum tumor markers for gastric cancer: a systematic review of literature by the Task Force of the Japanese Gastric Cancer Association. Gastric Cancer 2014;17:26-33.
10. Sun Z, Zhang N. Clinical evaluation of CEA, CA19-9, CA72-4 and CA125 in gastric cancer patients with neoadjuvant chemotherapy. World J Surg Oncol 2014;12:397.
11. Shibata C, Nakano T, Yasumoto A, et al. Comparison of CEA and CA19-9 as a predictive factor for recurrence after curative gastrectomy in gastric cancer. BMC Surg 2022;22:213.
12. Zhu H, Luo M, Wang P, et al. Comprehensive bioinformatics analysis for MEF2 family genes in gastric cancer. Transl Cancer Res 2022;11:4057-69.
13. Garganese G, Inzani F, Fragomeni SM, et al. The Vulvar Immunohistochemical Panel (VIP) Project: Molecular Profiles of Vulvar Squamous Cell Carcinoma. Cancers (Basel) 2021;13:6373.
14. Guo X, Peng Y, Song Q, et al. A Liquid Biopsy Signature for the Early Detection of Gastric Cancer in Patients. Gastroenterology 2023;165:402-413.e13.
15. Larsson SC, Butterworth AS, Burgess S. Mendelian randomization for cardiovascular diseases: principles and applications. Eur Heart J 2023;44:4913-24.
16. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ 2021;375:n2233.
17. Feng X, Zhao T, Liu X, et al. Serum iron status and the risk of lung cancer: a two-sample Mendelian-randomization study. J Thorac Dis 2023;15:6291-300.
18. Zhai S, Qin S, Li L, et al. Dietary butyrate suppresses inflammation through modulating gut microbiota in high-fat diet-fed mice. FEMS Microbiol Lett 2019;366:fnz153.
19. Mollica MP, Mattace Raso G, Cavaliere G, et al. Butyrate Regulates Liver Mitochondrial Function, Efficiency, and Dynamics in Insulin-Resistant Obese Mice. Diabetes 2017;66:1405-18.
20. Jo HJ, Kim J, Kim N, et al. Analysis of Gastric Microbiota by Pyrosequencing: Minor Role of Bacteria Other Than Helicobacter pylori in the Gastric Carcinogenesis. Helicobacter 2016;21:364-74.
21. Wu ZF, Zou K, Wu GN, et al. A Comparison of Tumor-Associated and Non-Tumor-Associated Gastric Microbiota in Gastric Cancer Patients. Dig Dis Sci 2021;66:1673-82.
22. Encarnação JC, Abrantes AM, Pires AS, et al. Revisit dietary fiber on colorectal cancer: butyrate and its role on prevention and treatment. Cancer Metastasis Rev 2015;34:465-78.
23. Wang F, Wu H, Fan M, et al. Sodium butyrate inhibits migration and induces AMPK-mTOR pathway-dependent autophagy and ROS-mediated apoptosis via the miR-139-5p/Bmi-1 axis in human bladder cancer cells. FASEB J 2020;34:4266-82.
24. Salimi V, Shahsavari Z, Safizadeh B, et al. Sodium butyrate promotes apoptosis in breast cancer cells through reactive oxygen species (ROS) formation and mitochondrial impairment. Lipids Health Dis 2017;16:208.
25. Li Y, He P, Liu Y, et al. Combining Sodium Butyrate

With Cisplatin Increases the Apoptosis of Gastric Cancer In Vivo and In Vitro via the Mitochondrial Apoptosis Pathway. Front Pharmacol 2021;12:708093.
26. Nishide K, Hirano T. Overlapping and non-overlapping functions of condensins I and II in neural stem cell divisions. PLoS Genet 2014;10:e1004847.
27. Li X, Song G, Zhao Y, et al. Functions of SMC2 in the Development of Zebrafish Liver. Biomedicines 2021;9:1240.
28. Montero S, Seras-Franzoso J, Andrade F, et al. Intracellular Delivery of Anti-SMC2 Antibodies against Cancer Stem Cells. Pharmaceutics 2020;12:185.
29. Strunnikov AV. One-hit wonders of genomic instability. Cell Div 2010;5:15.
30. Xu K, Qiao JY, Zhao BW, et al. Maternal SMC2 is essential for embryonic development via participating chromosome condensation in mice. J Cell Physiol 2023;238:2535-45.
31. Gong L, Zhang D, Lei Y, et al. Transcriptome-wide association study identifies multiple genes and pathways associated with pancreatic cancer. Cancer Med 2018;7:5727-32.
32. Je EM, Yoo NJ, Lee SH. Mutational and expressional analysis of SMC2 gene in gastric and colorectal cancers with microsatellite instability. APMIS 2014;122:499-504.
33. de Thonel A, Vandekerckhove J, Lanneau D, et al. HSP27 controls GATA-1 protein level during erythroid cell differentiation. Blood 2010;116:85-96.
34. Sun X, Ou Z, Xie M, et al. HSPB1 as a novel regulator of ferroptotic cancer cell death. Oncogene 2015;34:5617-25.
35. Wang L, Wu S, He H, et al. CircRNA-ST6GALNAC6 increases the sensitivity of bladder cancer cells to erastin-induced ferroptosis by regulating the HSPB1/P38 axis. Lab Invest 2022;102:1323-34.
36. Shi DB, Ma RR, Zhang H, et al. GAGE7B promotes tumor metastasis and growth via activating the p38δ/pMAPKAPK2/pHSP27 pathway in gastric cancer. J Exp Clin Cancer Res 2019;38:124.
37. Peng S, Yin Y, Zhang Y, et al. FYN/TOPK/HSPB1 axis facilitates the proliferation and metastasis of gastric cancer. J Exp Clin Cancer Res 2023;42:80.
38. Xie N, Wang Z, Shu Q, et al. Association between Gut Microbiota and Digestive System Cancers: A Bidirectional Two-Sample Mendelian Randomization Study. Nutrients 2023;15:2937.

Cite this article as: Gu R, Mei K, Chen Z, Huang Y, Wang F. Development and validation of a prognosis prediction model for overall survival in correlation between butyrate metabolism and gastric cancer prognosis: Mendelian randomization and transcriptomics analysis. Transl Cancer Res 2025;14(2):743-760. doi: 10.21037/tcr-24-677