

Development of a Microbiome-Based Prognostic Signature for Survival Prediction and Therapeutic Stratification in Colorectal Cancer

Shiyi Xu^{1†}, Xuejie Gao^{1†}, Hong Ye^{1*}

¹Department of Digestive & Nutrition, Fujian Children's Hospital (Fujian Branch of Shanghai Children's Medical Center), College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian, China

†: These authors contributed equally to this work.

*: Corresponding author.

E-mail:

Shiyi Xu (co-first author): xushiyl_nlxx@163.com

Xuejie Gao (co-first author): xjgao_ped@fjmu.edu.cn

Hong Ye (corresponding author): yhfjsfy99@163.com

CITATION

Author Name. (2025). Article title.
Journal of Infrastructure, Policy and
Development. 9(x): X.
<https://doi.org/10.24294/jipdxxxx>

ARTICLE INFO

Received: Date Month Year

Revised:

Accepted: Date Month Year

Available online: Date Month Year

COPYRIGHT



Copyright © 2025 by author(s).

Psycho-Oncologie is published by
PiscoMed Publishing Pte.Ltd.This
work is licensed under the Creative
Commons Attribution (CC BY) license.
[https://creativecommons.org/licenses/
by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: The intratumoral bacteria effect have been highlighted recently. In this study, we developed a microbiome-based signature using TCGA data to predict overall survival (OS) of colorectal cancer. We identified five microbial species through univariate Cox regression and refined the model with multivariate Cox regression and survival random forest algorithms. Based on the median score, we classified samples into high- and low-risk groups, revealing significantly poorer prognosis in the high-risk group ($p < 0.01$). Additionally, we found no significant correlation between the signature and clinical factors such as age, sex, or TNM stage, but it retained independent prognostic value in multivariate analysis ($p < 0.01$). We also analyzed genomic mutations, with CSMD3 showing significant differences between risk groups. Differentially expressed genes (DEGs) were enriched in ion channels, especially calcium channels, and immune-related pathways. Copy number variation analysis revealed more amplifications in the high-risk group. Furthermore, immune cell infiltration analysis using multiple algorithms highlighted significant differences between risk groups, particularly in naïve B cells, CD4+ T cells, and CD8+ T cells. Drug response analysis identified several drugs, including Gemcitabine and Sorafenib, with greater sensitivity in the low-risk group. Our study provides a novel tool and insight into the utilization of intratumoral microbiota for colorectal cancer.

Keywords: CRC, microbiota, prognosis, drug sensitivity

Funding: This study was supported by: 1. The Province-level special subsidy funds for the finance department in Fujian Province (No: Fujian Finance Index (2023) 834); 2. Research Fund of Fujian Maternity and Child Health Hospital (grant number YCXM 20-31).

Introduction

Colorectal cancer has caused huge burden to human health and economy. Before the 25 years ago, it is already the fourth leading lethal cancer¹ in younger aged population (50 years old or younger), and it have become the in first cause of cancer-related death among tumors in men and the 2nd in women² in the United States, 2024. In China, it was estimated that the fourth leading death cancer in 20203, according to a survey across 30 countries. The precision medicine and targeted therapy is a hotspot of current investigations, which stratify the colorectal cancers using different clinical and biological information, including metabolome⁴, genome⁵, micro-environmental⁵, microbiome⁵ etc. information to predict the necessity and response to specific drugs.

Among the predicting and prognostic biomarkers, a multiple features based signature have been emphasized, due to the higher reproducibility and robustness across cohorts. Mammaprint, a 70-gene based signature, was now used for guiding the radiotherapy in early staged breast cancer^{6,7}. In colorectal cancer, OncotypeDX, which integrates the transcriptomic information of genes, was also reported as an effective prognostic biomarker^{7,8}. For example, Sifei et al. reported a signature using senescence related genes could discriminate the clinical outcomes in colorectal cancer, the area under the ROC curve (AUC) of one-year survival reached 0.7319. Another study developed the model using oxidative stress genes and also validated its performance, the signature was also associated with multiple biological and clinical indicators¹⁰. Multi-feature-based signatures have demonstrated significant potential in predicting clinical outcomes and guiding treatment decisions in various cancers. Their ability to integrate diverse biological data enhances their predictive accuracy and clinical utility, suggesting that such signatures could play a pivotal role in personalized cancer treatment and prognosis in the future.

Genomic, transcriptomic proteomic and metabolomics revealed the process of tumor, which reflected the inherit feature of tumor. On the other hand, tumor is a complex biological process, involving both inherit and environmental interactions, affecting prognosis and drug response of cancers¹¹. The microbiome has shown a unique signature of on-set colorectal cancer, along with metabolome pattern, reflecting the interplay of microbiome and metabolome¹². Similarly, *Fusobacterium nucleatum*, a bacterium, was revealed as key for formulation of colorectal cancer-specific niche¹³. Another study revealed that it facilitates the resistance to Oxaliplatin chemotherapy by targeting TLR4 and MYD88 innate immune signaling pathway¹⁴. In 2023, Lejia et al. reported an intra-tumoural microbiome signature for predicting the survival of hepatocellular carcinoma post-surgery using 16S rDNA sequencing, associating with clinical indicators¹⁵. While investigations developing prognostic signature using microbiome is still lacking, except for reports regarding function of several species¹⁶. In summary, multi-omics data have revealed the intrinsic characteristics of tumors and their complex biological processes, particularly highlighting the critical role of the microbiome in tumor development, chemotherapy resistance, and prognosis. However, research on the microbiome as a prognostic biomarker remains limited.

In this work, we developed a signature using identified several prognosis-related bacteria in TCGA, quantified according to bulk RNA sequencing data. We also evaluated the performance of the signature, and assayed the clinical, genomic, micro-environmental and putative drug sensitivity. This study provides novel insight and tool for predicting the clinical outcome of colorectal cancer.

Result

Microbiome based signature development using TCGA dataset

In this study, we utilized TCGA samples and the estimated intratumoral microbiota abundance to identify candidate features associated with overall survival (OS) through univariate Cox regression analysis. We then applied multivariate Cox regression and survival random forest algorithms to derive a combination of five microbial species to construct a linear regression model. The model was defined as follows: $\text{score} = 0.145 * \text{Candidatus_Aquiluna} + 0.0677 * \text{Ferriphaselus} - 0.162 * \text{Aeromonas} - 0.0993 * \text{Tateyamaria} - 0.0041 * \text{Niabella}$. The species names indicate the relative abundance of corresponding bacteria. The coefficients for each microbial species in the linear model are presented in Fig. 1a, while the hazard ratios (HR) and p-values from the multivariate Cox regression for each species are shown in Fig. 1b. Based on the median score of samples in the TCGA dataset, we classified samples into high- and low-risk groups, with scores above the median categorized as high-risk and those below as low-risk. Evaluation of OS between these groups revealed a significantly poorer prognosis in the high-risk group ($p < 0.01$, Fig. 1c), with a similar trend observed for progression-free survival (PFS; $p < 0.001$, Fig. 1d). The detailed score values and survival outcomes for each sample, along with a heatmap showing the abundance of the corresponding microbial species across samples was also visualized (Fig. 1e). The high risk samples were depleted with most bacteria species. These findings indicate that the intratumoral microbiota composition is closely associated with patient prognosis.

Association of the signature and clinical indicators

We further evaluated the correlation between the identified microbial signature and key clinical indicators to understand its broader applicability in cancer prognosis. Our analysis indicated no significant correlation between this signature and patient age, sex, or TNM staging (Fig. 2a), suggesting that the signature functions independently of these conventional clinical factors. To further explore the clinical relevance of the signature, we conducted a multivariate Cox regression analysis, incorporating age, sex, and TNM staging as covariates. The results confirmed that the microbial signature retains significant prognostic power as an independent factor, with a strong association with overall survival (Fig. 2b, $p < 0.01$), emphasizing the robustness of the signature in predicting patient outcomes, irrespective of standard clinical characteristics. Additionally, we examined the relationship between the signature and immunotherapy-related markers, including PD-1 and PD-L1 expression, which are often associated with the tumor immune microenvironment and treatment response. We found no significant correlation between the signature and these markers (Fig. 2c), implying that the prognostic value of the signature is distinct from that of the immune microenvironment. To facilitate its application in clinical settings, we constructed a 5-year overall

survival nomogram that integrates this microbial signature. The nomogram highlights the signature's potential as a valuable biomarker, capable of contributing to personalized risk stratification for long-term survival (Fig. 2d). These results underscore the relevance of the signature as a prognostic tool that may aid in refining therapeutic decisions and enhancing personalized cancer management.

Genome feature of the signature

We analyzed the correlation between the identified microbial signature and genomic mutations, finding no significant differences in mutation profiles between the high- and low-risk groups, except for CSMD3, which exhibited a significant variation ($p < 0.05$, Fig. 3a). To further characterize the high- and low-risk groups, we identified differentially expressed genes (DEGs) using the TCGA dataset. Gene Ontology (GO) and KEGG pathway enrichment analyses were performed on these DEGs, revealing significant enrichment in metal ion channels, particularly calcium ion channels (Fig. 3b-e). Additionally, Gene Set Enrichment Analysis (GSEA) demonstrated that the signature is linked to various cancer-related signaling pathways, with a strong emphasis on immune-related pathways. Specifically, enrichment was observed in pathways related to ANTIGEN PROCESSING AND PRESENTATION and NATURAL KILLER CELL MEDIATED CYTOTOXICITY (Fig. 3f). We also conducted a comparative analysis of copy number variations (CNVs) between the high- and low-risk groups, finding that the high-risk group exhibited more frequent genomic amplifications (Fig. 3g). This suggests that genomic instability, as indicated by increased amplifications, may contribute to the poorer prognosis observed in high-risk patients, further supporting the clinical relevance of this signature in cancer prognostication.

Immune infiltration and the signature

We utilized multiple algorithms to estimate immune cell infiltration abundance within the TCGA dataset, including Quantiseq, XCELL, CIBERSORT, and TIMER. These algorithms perform relative quantification of various immune cells based on the expression profiles and specific gene signatures associated with immune cell types. This approach enabled us to obtain a comprehensive assessment of immune cell infiltration patterns in the dataset. We then identified differentially infiltrated immune cells between the high- and low-risk groups and visualized these differences using a heatmap (Fig. 4a). Our analysis revealed that various algorithms consistently identified significant differences in immune cell infiltration between risk groups. Specifically, naïve B cells, CD4⁺ T cells, and CD8⁺ T cells showed marked differences in infiltration levels between high- and low-risk groups (Fig. 4b). These results suggest distinct immune cell profiles in association with risk stratification, highlighting the potential role of immune infiltration in prognosis.

Signature and drug response

Using the OncoPredict algorithm, we evaluated the IC50 values of 198 drugs across a range of TCGA samples. This analysis allowed us to calculate the differences in drug sensitivity between high- and low-risk groups, identifying several drugs with significantly different IC50 values (Fig. 5a). The majority of these drugs showed greater sensitivity in the low-risk group, indicating that patients in this group may respond more effectively to certain therapies. Among the drugs with the most pronounced differences were Gemcitabine, Nilotinib, Dabrafenib, Sorafenib, and Topotecan, all of which demonstrated significantly lower IC50 values in the low-risk group compared to the high-risk group (Fig. 5b). These findings suggest that these drugs could be more effective in patients classified as low-risk based on our model, highlighting the potential for personalized treatment strategies.

Method and Material

Candidate bacteria identification and signature construction

The correlation between relative bacterial abundance and overall survival and progression-free survival was analyzed with Cox univariate regression. The bacteria significantly associated with both overall survival and progression-free survival was retained as candidate bacteria. Afterward, a Cox multivariate regression model was constructed using the identified bacteria abundance and overall survival. The formula was calculated as $\text{score} = \sum (b_i \times \text{abundance}_i)$, where b_i is the intratumoral bacteria abundance and a_i is coefficient evaluated with Cox multivariate regression model. Score value of each TCGA sample was calculated, and the sample in TCGA was separated into high and low-risk samples according to the median value of TCGA dataset.

Genome and transcriptome association analyses

The genome mutation, transcription and clinical information data of TCGA dataset was retrieved from cBioportal database¹⁷. During mutation analyses, only non-synonymous mutation causing protein sequence alteration was used. For each gene, the mutated sample number and total sample number in both high and low-risk groups were identified and the mutational difference groups were calculated with Fisher's exact test, and $p < 0.05$ were considered as statistically significant. The detailed mutational information of both groups of the top frequently mutated genes were visualized with R package 'GenVisR'¹⁸. For Gene Set Enrichment Analyses and Gene Ontology enrichment analyses, the differentially expressed genes between high and low-risk groups were identified with R package 'limma'¹⁹. The differentially expressed genes were identified as adjusted p value < 0.01 and $|\log_2 \text{fold change}| > 0.4$. The enriched pathways were identified and visualized with R package 'clusterprofiler'²⁰ and 'GSEABase'. For copy number variation analyses, the differential copy number of genes between groups

were evaluated with Wilcox rank sum test, by comparing the copy numbers (-2, -1, 0, 1, 2). The differential genes were and corresponding groups were visualized with R package ‘pheatmap’.

Immune infiltration and the signature

The infiltration abundance of immune cells was estimated according to the gene expression profile of TCGA data. We employed several algorithms in this study, including XCELL21, QUANTISEQ22Francesca</author><author>Mayer, Clemens</author><author>Plattner, Christina</author><author>Laschober, Gerhard</author><author>Rieder, Dietmar</author><author>Hackl, Hubert</author><author>Krogsdam, Anne</author><author>Loncova, Zuzana</author><author>Posch, Wilfried</author><author>Wilflingseder, Doris</author><author>Sopper, Sieghart</author><author>Ijsselsteijn, Marieke</author><author>Brouwer, Thomas P.</author><author>Johnson, Douglas</author><author>Xu, Yaomin</author><author>Wang, Yu</author><author>Sanders, Melinda E.</author><author>Estrada, Monica V.</author><author>Ericsson-Gonzalez, Paula</author><author>Charoentong, Pornpimol</author><author>Balko, Justin</author><author>de Miranda, Noel Filipe da Cunha Carvalho</author><author>Trajanoski, Zlatko</author></authors></contributors><titles><title>Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data</title><secondary-title>Genome Medicine</secondary-title></titles><periodical><full-title>Genome Medicine</full-title></periodical><pages>34</pages><volume>11</volume><number>1</number><dates><year>2019</year><pub-dates><date>2019/05/24</date></pub-dates></dates><isbn>1756-994X</isbn><urls><related-urls><url>https://doi.org/10.1186/s13073-019-0638-6</url></related-urls></urls><electronic-resource-num>10.1186/s13073-019-0638-6</electronic-resource-num></record></Cite></EndNote>, EPIC, COBERSORT23, TIMER24, ect. Default parameters were used during the analyses, and it is accordant with the TIMER 2.0 website provided TCGA dataset. The association was calculated using Student’s t-test to assay the infiltration difference of different cell types. The results were visualized with heatmap using R package ‘pheatmap’ and vioplot using R package ‘vioplot’, $p < 0.05$ was considered as statistically significant.

Drug IC50 prediction

In this step, ‘oncopredict’25 algorithm was used to predict the IC50 values of TCGA samples according to gene expression data of TCGA dataset. The transcriptome data of TCGA dataset was prepared according to the algorithm, using GDSC2 as training set. The low varying gene threshold was set as 0.2, min number of sample as 10 and used raw data. The IC50 values of 198 drugs were compared between low and high-risk samples, and adjusted p value < 0.01 was considered as statistically significant.

Discussion

Intratumoral microbiome were revealed to exist in many types of cancers and each cancer type has its own distinct microbiome composition²⁶. Most of these bacteria were intracellular, existing in both cancer and immune cells. The role of intratumoral microbiome have been emphasized recently, due to its roles in multiple aspects. For example, Leore et al. reported that specific intratumoral bacteria participate in drug resistance in pancreatic ductal carcinoma²⁷. *Fusobacterium nucleatum* was shown to promote progression phenotypes of breast cancer, including proliferation and metastasis²⁸. Other studies also revealed that *Fusobacterium nucleatum* modulates the signaling pathways of colorectal cancer, via miRNA or long non-coding RNAs^{29,30}. Similar to the heterogeneity of colorectal in genomic and transcriptomic levels, the intratumoral microbiome is also highly heterogeneous across individuals³¹. Thus, the microbiota based signature may provide valuable insight into the prognosis of colorectal cancer.

The signature is consisting of five bacteria species, named *Candidatus Aquiluna*, *Ferriphaselus*, *Aeromonas*, *Tateyamaria* and *Niabella*. In other words, the score value is determined by the relative abundance of these bacteria species. It was noticed that previous report revealed that *Aeromonas* was significantly have higher abundance in colorectal cancer fecal samples, compared to the control³². Besides, mouse model revealed that when treated with traditional Chinese Medicine Pien Tze Huang, the *Aeromonas* was depleted in fecal samples³³. Despite that little is known about the function of *Aeromonas* in intratumoral microbiome. The investigations regarding the effect of other microbiome species is still limited, especially intratumoral effect.

The limitation of this study include the followings. First of all, this study lacks validation dataset. The advantage of multiple feature based model is robustness and reproducibility. This study needs further validation. Secondly, a better way for the intratumoral microbiome abundance estimation is 16s DNA sequencing, which is more convincing than this study. Lastly, the function candidate bacteria species is not validated, not mentioning the detailed mechanism of the bacteria types.

Acknowledgments

The authors would like to thank Liping Zhu from the Medical Research Center, Fujian Maternity and Child Health Hospital for supporting in the data interpretation.

References

- 1 Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J Clin* 73, 233-254, doi:10.3322/caac.21772 (2023).
- 2 Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J Clin* 74, 12-49, doi:10.3322/caac.21820 (2024).
- 3 Qi, J. et al. National and subnational trends in cancer burden in China, 2005-20: an analysis of national mortality surveillance data. *Lancet Public Health* 8, e943-e955, doi:10.1016/s2468-2667(23)00211-6 (2023).
- 4 Hirose, Y. & Taniguchi, K. Intratumoral metabolic heterogeneity of colorectal cancer. *Am J Physiol Cell Physiol* 325, C1073-c1084, doi:10.1152/ajpcell.00139.2021 (2023).
- 5 Senchukova, M. A. Genetic heterogeneity of colorectal cancer and the microbiome. *World J Gastrointest Oncol*

15, 443-463, doi:10.4251/wjgo.v15.i3.443 (2023).

6 Cardoso, F. et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* 375, 717-729, doi:10.1056/NEJMoa1602253 (2016).

7 Andre, F. et al. Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. *J Clin Oncol* 40, 1816-1837, doi:10.1200/jco.22.00069 (2022).

8 Marisa, L. et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 10, e1001453, doi:10.1371/journal.pmed.1001453 (2013).

9 Yu, S., Chen, M., Xu, L., Mao, E. & Sun, S. A senescence-based prognostic gene signature for colorectal cancer and identification of the role of SPP1-positive macrophages in tumor senescence. *Front Immunol* 14, 1175490, doi:10.3389/fimmu.2023.1175490 (2023).

10 Wang, X. et al. A novel oxidative stress- and ferroptosis-related gene prognostic signature for distinguishing cold and hot tumors in colorectal cancer. *Front Immunol* 13, 1043738, doi:10.3389/fimmu.2022.1043738 (2022).

11 Xiao, Y. & Yu, D. Tumor microenvironment as a therapeutic target in cancer. *Pharmacol Ther* 221, 107753, doi:10.1016/j.pharmthera.2020.107753 (2021).

12 Kong, C. et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. *Gut* 72, 1129-1142, doi:10.1136/gutjnl-2022-327156 (2023).

13 Zepeda-Rivera, M. et al. A distinct *Fusobacterium nucleatum* clade dominates the colorectal cancer niche. *Nature* 628, 424-432, doi:10.1038/s41586-024-07182-w (2024).

14 Yu, T. et al. *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* 170, 548-563.e516, doi:10.1016/j.cell.2017.07.008 (2017).

15 Sun, L. et al. Intratumoural microbiome can predict the prognosis of hepatocellular carcinoma after surgery. *Clin Transl Med* 13, e1331, doi:10.1002/ctm2.1331 (2023).

16 Villéger, R. et al. Microbial markers in colorectal cancer detection and/or prognosis. *World J Gastroenterol* 24, 2327-2347, doi:10.3748/wjg.v24.i22.2327 (2018).

17 Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6, pl1, doi:10.1126/scisignal.2004088 (2013).

18 Skidmore, Z. L. et al. GenVisR: Genomic Visualizations in R. *Bioinformatics* 32, 3012-3014, doi:10.1093/bioinformatics/btw325 (2016).

19 Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47, doi:10.1093/nar/gkv007 (2015).

20 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16, 284-287, doi:10.1089/omi.2011.0118 (2012).

21 Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology* 18, 220, doi:10.1186/s13059-017-1349-1 (2017).

22 Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Medicine* 11, 34, doi:10.1186/s13073-019-0638-6 (2019).

23 Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* 12, 453-457, doi:10.1038/nmeth.3337 (2015).

24 Li, T. et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 48, W509-w514, doi:10.1093/nar/gkaa407 (2020).

25 Maeser, D., Gruener, R. F. & Huang, R. S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* 22, doi:10.1093/bib/bbab260 (2021).

26 Nejman, D. et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 368, 973-980, doi:10.1126/science.aay9189 (2020).

27 Geller, L. T. et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* 357, 1156-1160, doi:10.1126/science.aah5043 (2017).

28 Parhi, L. et al. Breast cancer colonization by *Fusobacterium nucleatum* accelerates tumor growth and metastatic progression. *Nat Commun* 11, 3259, doi:10.1038/s41467-020-16967-2 (2020).

29 Guo, S. et al. Exosomes derived from *Fusobacterium nucleatum*-infected colorectal cancer cells facilitate tumour metastasis by selectively carrying miR-1246/92b-3p/27a-3p and CXCL16. *Gut*, doi:10.1136/gutjnl-2020-321187 (2020).

30 Yang, Y. 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* 152, 851-866.e824, doi:10.1053/j.gastro.2016.11.018 (2017).

31 Liu, W. et al. Microbial Community Heterogeneity Within Colorectal Neoplasia and its Correlation With Colorectal Carcinogenesis. *Gastroenterology* 160, 2395-2408, doi:10.1053/j.gastro.2021.02.020 (2021).

32 Yang, T. W. et al. Enterotype-based Analysis of Gut Microbiota along the Conventional Adenoma-Carcinoma Colorectal Cancer Pathway. *Sci Rep* 9, 10923, doi:10.1038/s41598-019-45588-z (2019).

33 Gou, H. et al. Traditional Medicine Pien Tze Huang Suppresses Colorectal Tumorigenesis Through Restoring Gut Microbiota and Metabolites. *Gastroenterology* 165, 1404-1419, doi:10.1053/j.gastro.2023.08.052 (2023).

Figure Legend

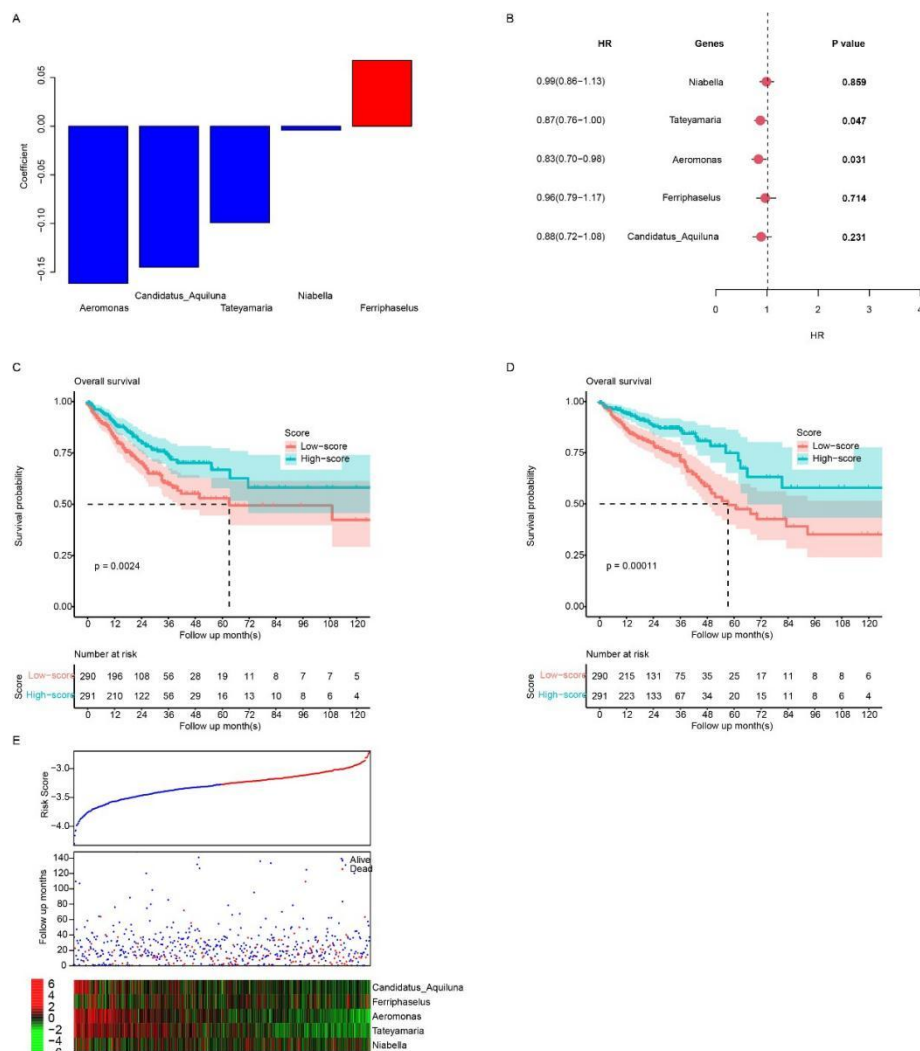


Figure 1. Prognostic microbial species and survival analysis. (A) Coefficients for each microbial species in the multivariate Cox regression model. Positive coefficients indicate a higher risk, while negative coefficients indicate a protective effect. (B) Hazard ratios (HR) and p-values for each microbial species, highlighting significant associations with overall survival. (C) Kaplan-Meier curves showing overall survival stratified by the microbiome signature, where high-risk patients exhibit significantly poorer prognosis compared to low-risk patients ($p < 0.01$). (D) Progression-free survival (PFS) curves showing similar trends. (E) Heatmap displaying the abundance of the five identified microbial species across TCGA samples and survival information, highlighting species depletion in high-risk samples(F1).

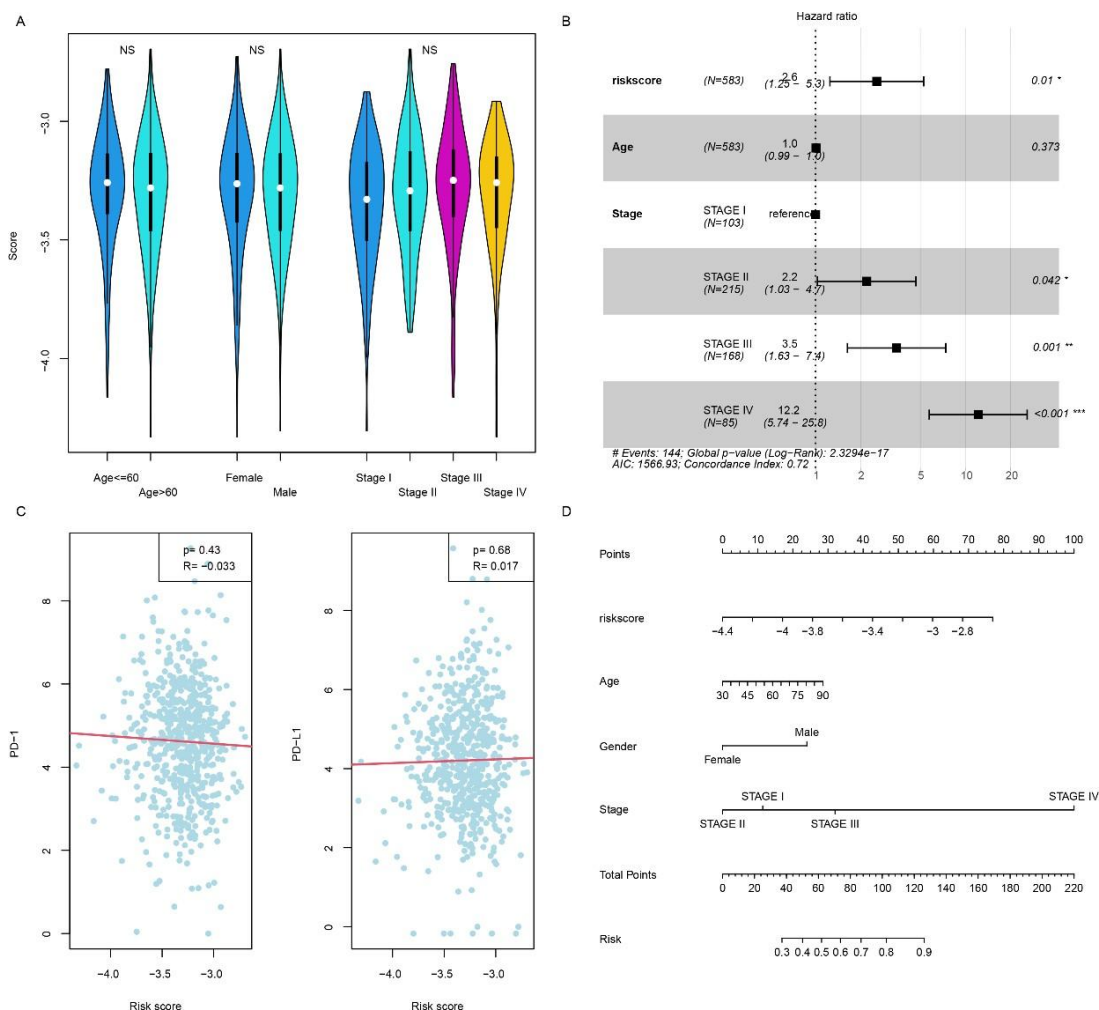


Figure 2. Correlation of microbiome signature with clinical factors and immune-related markers. (A) Risk score stratified by clinical factors including age, sex, and TNM stage, showing no significant correlations. (B) Multivariate Cox regression incorporating clinical factors (age, sex, TNM stage) showing that the microbiome signature remains an independent prognostic factor for overall survival. (C) Analysis of the relationship between the microbiome signature and PD-1 and PD-L1 expression, demonstrating no significant association with these immune markers. (D) A nomogram integrating the microbiome signature for predicting 5-year overall survival, with total points indicating survival probability.

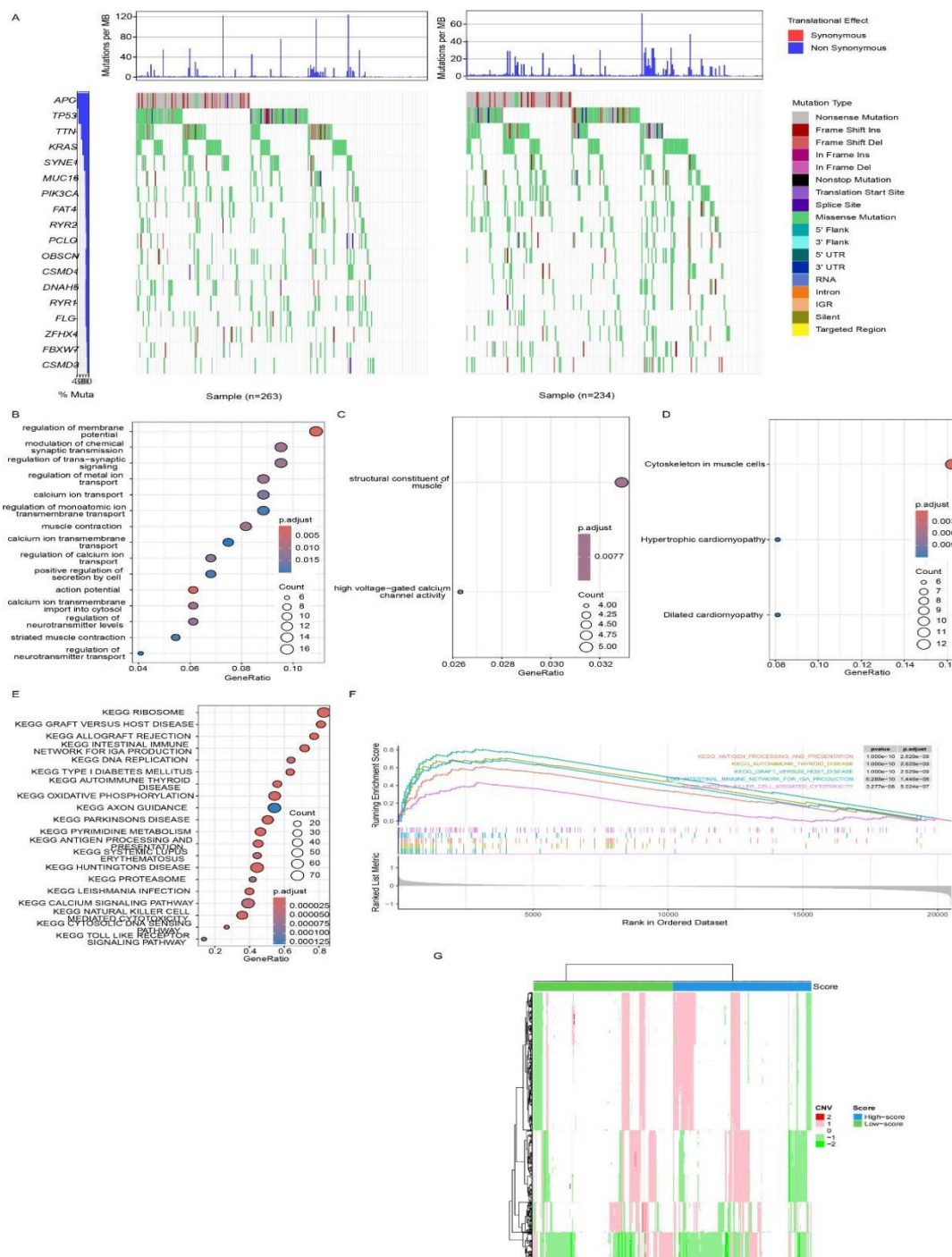


Figure 3. Genomic and transcriptomic features associated with the microbiome signature. (A) Mutation frequencies of key genes in high- and low-risk groups, with significant differences in CSMD3 mutation rates. (B-E) Gene Ontology (GO) and KEGG pathway enrichment analyses of differentially expressed genes between high- and low-risk groups. (F) Gene Set Enrichment Analysis (GSEA) linking the microbiome signature to immune-related pathways, such as antigen processing and presentation and natural killer cell-mediated cytotoxicity. (G) High-risk samples showed more amplification proportion.

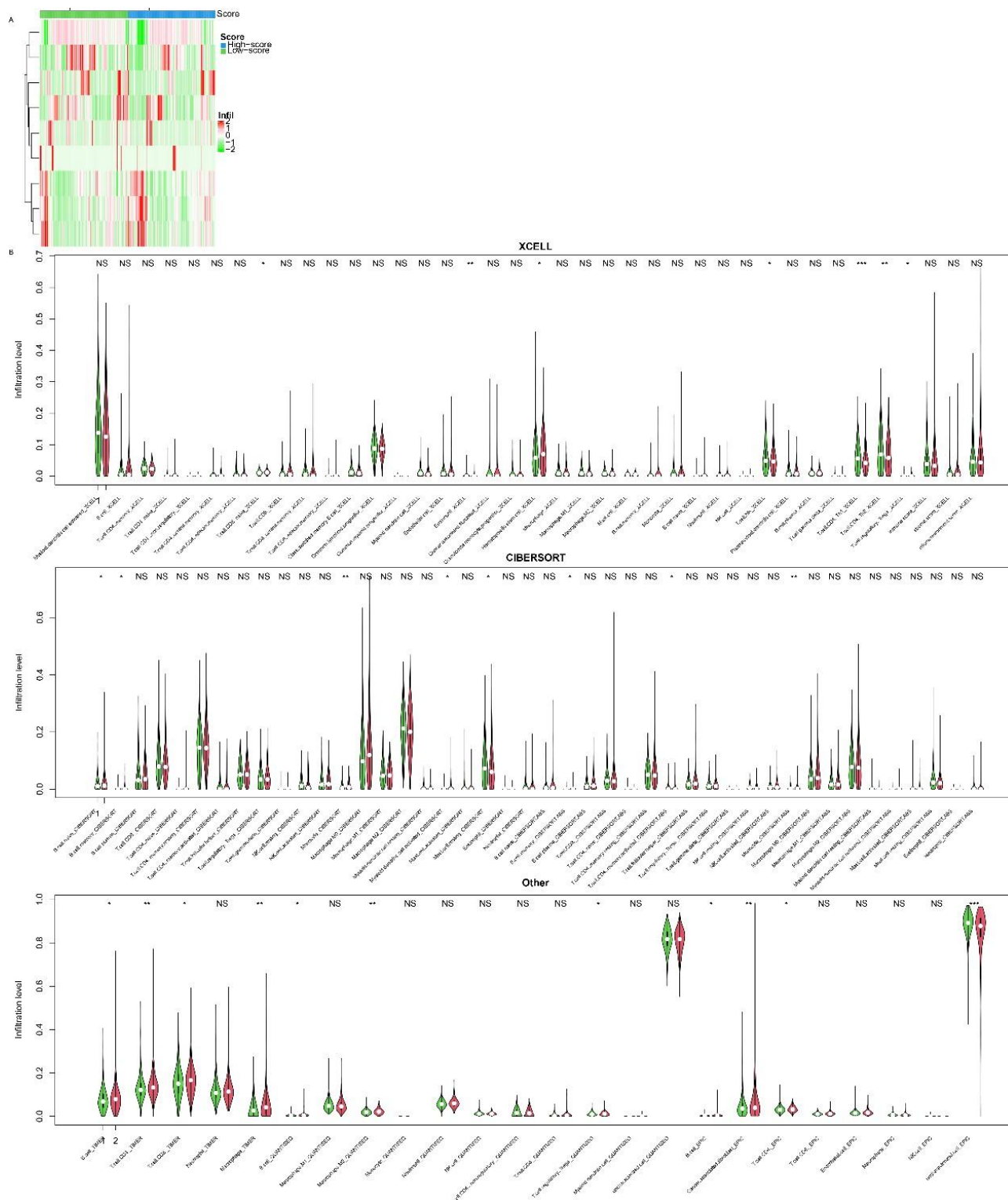


Figure 4. Immune cell infiltration differences between risk groups. (A) Heatmap showing immune cell infiltration scores estimated by multiple algorithms (XCELL, CIBERSORT, and TIMER). (B) Significant differences in immune cell populations between high- and low-risk groups, with notable variation in naïve B cells, CD4+ T cells, and CD8+ T cells infiltration levels. These findings suggest that immune infiltration plays a role in prognosis stratification by the mi-

crobiome signature.

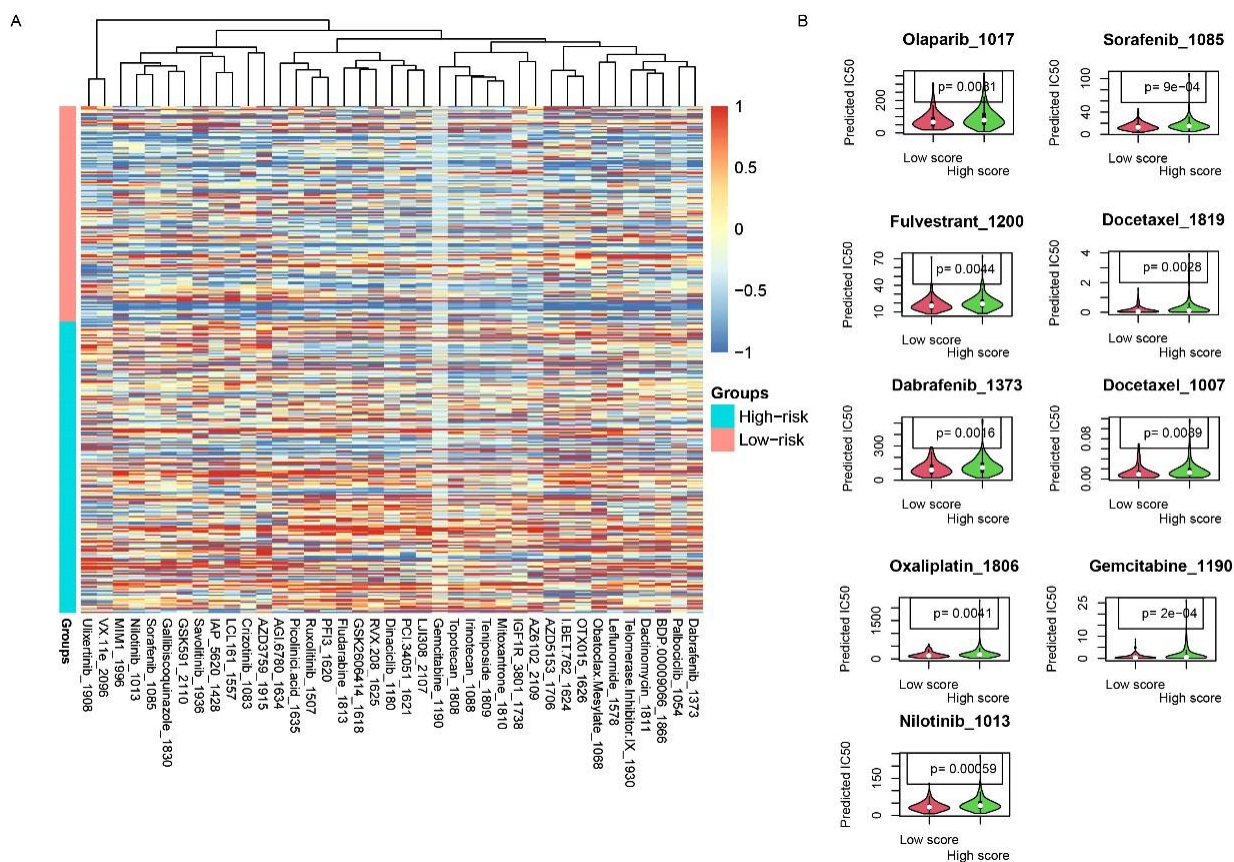


Figure 5. Drug sensitivity prediction based on microbiome signature. (A) Boxplots showing the predicted IC50 values of various drugs for high- and low-risk groups. (B) Gemcitabine, Nilotinib, and Sorafenib demonstrate significant differences in predicted drug sensitivity between the high- and low-risk groups, indicating potential treatment implications based on the microbiome signature.