npj | biofilms and microbiomes

Article

Published in partnership with Nanyang Technological University

https://doi.org/10.1038/s41522-025-00702-0

Oral microbiota signature predicts the prognosis of colorectal carcinoma

Check for updates

Shi-Hao Zhou[1,3], Yan Du[2,3], Wen-Qiong Xue[2], Min-Jun He[1], Ting Zhou[2], Zhi-Yang Zhao[1], Lu Pei[2], Yi-Wei Chen[2], Jin-Ru Xie[2], Chang-Ling Huang[1], Yong-Qiao He[2], Tong-Min Wang[2], Ying Liao[2] ✉ & Wei-Hua Jia[1,2] ✉

Emerging evidence links oral-derived gut microbes to colorectal cancer (CRC) development, but CRC prognosis-related microbial alterations in oral remain underexplored. In a retrospective study of 312 CRC patients, we examined the oral microbiota using 16S rRNA gene full-length amplicon sequencing to identify prognostic microbial biomarkers for CRC. *Neisseria oralis* and *Campylobacter gracilis* increased CRC progression risk (HR = 2.63 with $P = 0.007$, HR = 2.27 with $P = 0.001$, respectively), while *Treponema medium* showed protective effects (HR = 0.41, $P = 0.0002$). A microbial risk score (MRS) incorporating these species effectively predicted CRC progression risk (C-index = 0.68, 95% CI = 0.61–0.76). When compared to a model constructed solely from clinical factors, including tumor stage, lymphatic metastasis, and perineural invasion, the predictive accuracy significantly improved with the addition of the MRS, resulting in a C-index rising to 0.77 ($P = 2.33 \times 10^{-5}$). Our findings suggest that oral microbiota biomarkers may contribute to personalized CRC monitoring strategies, their implementation in clinical surveillance necessitates confirmatory studies.

Colorectal carcinoma (CRC) remains one of the most prevalent malignancies worldwide[1]. Despite advancements in multidisciplinary treatments, the prognosis for patients with locally advanced or metastatic CRC remains poor, largely due to high rates of distant metastasis, chemotherapy resistance, and relapse, which contribute to unsatisfactory long-term survival outcomes[2,3]. As a result, the identification of reliable prognostic factors is essential for optimizing treatment strategies and improving patient outcomes, particularly given the considerable variation in postoperative results and long-term prognosis among individuals[4].

The human oral microbiota is a complex ecosystem consisting of over 770 species-level taxa, each adapted to different niches within the oral cavity[5]. Evidence suggests that specific oral-derived species, including *Porphyromonas gingivalis*, *Fusobacterium nucleatum*, *Streptococcus anginosus*, *Peptostreptococcus stomatis,* and *Prevotella intermedia*, are associated with the occurrence of CRC[6–10]. These oral pathogens may influence CRC development by disrupting the gut microbiota via the oral-to-gut microbe translocation, or by activating the host immune inflammation[11,12]. For instance, the well-known pathogen *P. gingivalis*, upon ectopic colonization in the gut, can recruit myeloid immune cells, activate the NLRP3 inflammasome, and enhance the expression of inflammatory cytokines such as IL-1β, thereby promoting the onset and progression of colorectal cancer (CRC)[13]. While the role of oral-derived microbes in CRC development is identified, the mechanisms by which alterations in the oral microbiota are associated with CRC prognosis remain largely unknown.

The oral microbiota has been implicated in the progression of various malignancies, and certain species have been associated with cancer outcomes[14]. For example, Mohamed, et al. reported that a higher salivary carriage of the genus *Candida* was associated with a poor prognosis, while *Malassezia* was enriched in patients with favorable outcomes in oral squamous cell carcinoma[15]. Additionally, Du et al. found that reduced oral microbiota diversity was correlated with higher mortality in nasopharyngeal carcinoma, particularly among elderly patients[16]. A recent study observed that intra-tumoral infection of oral-derived *F. nucleatum* is associated with poorer disease-free survival and overall survival in patients with stage III CRC[17]. These findings underscore the critical role of the oral microbiota in influencing cancer prognosis, and further investigation of the prognostic function of oral microbiota in CRC is warranted.

In this study, we investigated the salivary microbiota of 312 CRC patients using 16S rRNA gene full-length sequencing. We identified microbial signatures associated with CRC outcomes and developed a

[1]School of Public Health, Sun Yat-sen University, Guangzhou, China. [2]State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, P.R, China. [3]These authors contributed equally: Shi-Hao Zhou, Yan Du.
✉e-mail: liaoying@sysucc.org.cn; jiawh@sysucc.org.cn

microbial risk score (MRS) to predict CRC progression. Furthermore, we integrated the MRS with key clinical factors to construct a multi-factorial prognostic model with enhanced predictive performance. While these findings require validation in multi-center cohorts, our study pioneers the application of salivary microbiota profiling as a non-invasive prognostic tool in CRC.

Results

Characteristics and microbial community structure among patients with different clinical outcomes

In this study, we collected saliva samples from 312 CRC patients who were scheduled for surgical tumor resection. The median follow-up period for patients with CRC progression was 21.5 months, whereas for survival was 25.4 months, during which 59 experienced disease progression and 25 died (Fig. 1). Demographic and clinical characteristics of the participants are presented in Table 1. There were no statistically significant differences in basic demographic characteristics, such as age, sex, marital status, body mass index (BMI), family history of tumors, and alcohol consumption history, between non-progression and progression groups, and between survival and death groups. Patients with progression or with death outcomes were more likely diagnosed at an advanced clinical stage, exhibited higher rates of lymphatic invasion, perineural invasion, and lymphatic metastasis when compared to good outcome individuals (Table 1).

We investigated the oral microbiota diversity and composition differences in several clinical factors. There were no significant differences in oral microbiota diversity with differing prognostic outcomes. Shannon diversity

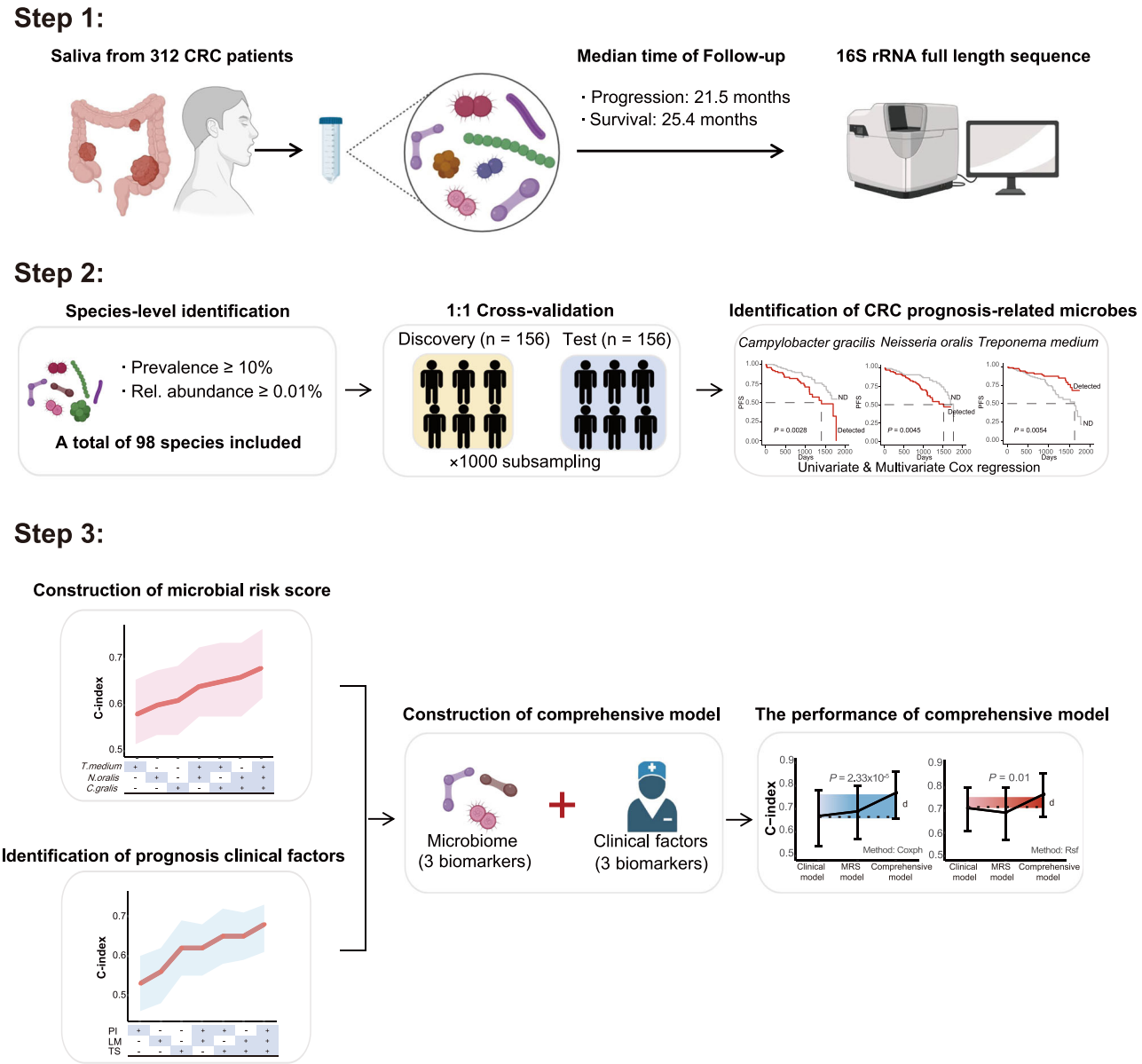

Fig. 1 | The workflow of the study. Step 1: Saliva samples were collected from 312 CRC patients scheduled for surgical tumor resection. The median follow-up time was 21.5 months for patients with disease progression and 25.4 months for survival. Salivary microbiota profiling was performed using 16S rRNA full-length sequencing. Step 2: Species-level identification was conducted with inclusion criteria of prevalence ≥10% and relative abundance ≥0.01%, resulting in 98 species. A 1:1 cross-validation strategy was applied by randomly splitting the cohort into discovery ($n = 156$) and test ($n = 156$) sets, with 1000 subsampling iterations. Prognosis-related microbes were identified using univariate and multivariate Cox regression analyses, highlighting *Campylobacter gracilis*, *Neisseria oralis*, and *Treponema medium* as significant biomarkers. Step 3: A microbial risk score (MRS) was constructed based on the identified prognostic bacterial species. Clinical prognostic factors were also evaluated, and a comprehensive model integrating microbiota biomarkers (three species) and clinical factors (three variables) was developed. The performance of the comprehensive model was validated and compared with models based solely on the clinical factors, demonstrating superior prognostic accuracy.

Table 1 | Patient demographic and clinical characteristics

Characteristics	Overall (n = 312)	Non-progressed (n = 253)	Progressed (n = 59)	P value*	Survival (n = 287)	Death (n = 25)	P value*
Age (Mean ± SD)	57.79 ± 12.07	57.97 ± 11.74	57.00 ± 11.50	0.578	57.54 ± 11.80	60.64 ± 14.80	0.219
Sex (%)							
Male	139 (44.55)	112 (44.27)	27 (45.76)	0.950	129 (44.95)	10 (40.00)	0.789
Female	173 (55.45)	141 (55.73)	32 (54.24)		158 (55.05)	15 (60.00)	
Marital status (%)							
Married	9 (2.88)	7 (2.77)	2 (3.39)	0.811	8 (2.79)	1 (4.00)	0.689
Unmarried	303 (97.12)	246 (97.23)	57 (96.61)		279 (97.21)	24 (96.00)	
BMI (%)[#]							
Normal	183 (59.22)	148 (58.96)	35 (60.34)	0.647	172 (60.35)	11 (45.83)	0.176
Obesity	13 (4.21)	12 (4.78)	1 (1.72)		12 (4.21)	1 (4.17)	
Overweight	84 (27.18)	69 (27.49)	15 (25.86)		77 (27.02)	7 (29.17)	
Underweight	29 (9.39)	22 (8.76)	7 (12.07)		24 (8.42)	5 (20.83)	
Family history of tumors (%)							
No	218 (69.87)	177 (69.96)	41 (69.49)	1	199 (69.34)	19 (76.00)	0.639
Yes	94 (30.13)	76 (30.04)	18 (30.51)		88 (30.66)	6 (24.00)	
Cigarette smoking history (%)							
No	265 (84.94)	214 (84.58)	51 (86.44)	0.875	244 (85.02)	21 (84.00)	0.778
Yes	47 (15.06)	39 (15.42)	8 (13.56)		43 (14.98)	4 (16.00)	
Alcohol consumption history (%)							
No	291 (93.27)	236 (93.28)	55 (93.22)	1	270 (94.08)	17 (80.95)	0.075
Yes	21 (6.73)	17 (6.72)	4 (6.78)		17 (5.92)	4 (19.05)	
Neoadjuvant chemoradiotherapy (%)							
No	201 (64.42)	171 (67.59)	30 (50.84)	0.023	188 (65.51)	13 (52.00)	0.256
Yes	111 (35.58)	82 (32.41)	29 (49.16)		99 (34.49)	12 (48.00)	
Tumor stage (%)							
0/1	50 (16.03)	46 (18.18)	4 (6.78)	<0.001	48 (16.72)	2 (8.00)	<0.001
2	83 (26.60)	77 (30.43)	6 (10.17)		80 (27.87)	3 (12.00)	
3	130 (41.67)	106 (41.90)	24 (40.68)		122 (42.51)	8 (32.00)	
4	41 (13.14)	16 (6.32)	25 (42.37)		29 (10.10)	12 (48.00)	
Unknown	8 (2.56)	8 (3.16)	0 (0.00)		8 (2.79)	0 (0.00)	
Histologic grade							
Low	18 (5.77)	13 (5.14)	5 (8.47)	0.397	15 (5.23)	3 (12.00)	0.293
Moderate	268 (85.90)	217 (85.77)	51 (86.44)		247 (86.06)	21 (84.00)	
High	26 (8.33)	23 (9.09)	3 (5.08)		25 (8.71)	1 (4.00)	
Lymphatic invasion (%)							
No	223 (71.47)	186 (73.52)	37 (62.71)	0.002	209 (72.82)	14 (56.00)	0.017
Yes	59 (18.91)	39 (15.42)	20 (33.90)		49 (17.07)	10 (40.00)	
Unknown	30 (9.62)	28 (11.07)	2 (3.39)		29 (10.10)	1 (4.00)	
Perineural invasion (%)							
No	230 (73.72)	198 (78.26)	32 (54.24)	<0.001	217 (75.61)	13 (52.00)	0.001
Yes	52 (16.67)	27 (10.67)	25 (42.37)		41 (14.29)	11 (44.00)	
Unknown	30 (9.62)	28 (11.07)	2 (3.39)		29 (10.10)	1 (4.00)	
Lymphatic metastasis (%)							
No	171 (54.81)	153 (60.47)	18 (30.51)	<0.001	165 (57.49)	6 (24.00)	<0.001
Yes	105 (33.65)	69 (27.27)	36 (61.02)		87 (30.31)	18 (72.00)	
Unknown	36 (11.54)	31 (12.25)	5 (8.47)		35 (12.20)	1 (4.00)	

[#]A body mass index (BMI) less than 18.5 is classified as underweight, 18.5–24 as normal, 24–28 as overweight, and greater than 28 as obese.

*The demographic and clinical characteristics of patients were described and compared among different PFS and OS groups using the Pearson χ^2 test for categorical variables and one-way ANOVA analysis for age.

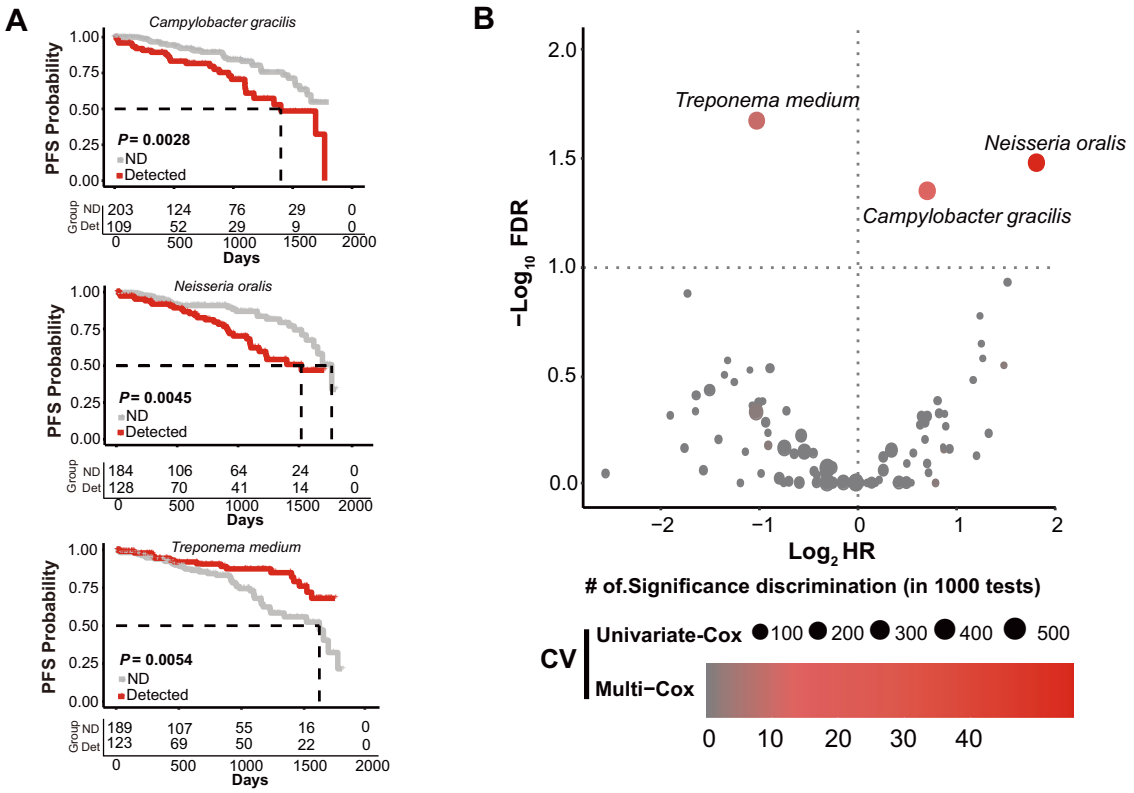

Fig. 2 | Identification of prognostic bacterial biomarkers associated with PFS in colorectal cancer patients. A Kaplan–Meier curve illustrating PFS stratified by the detection status of identified oral bacterial species. *P* values were calculated using the log-rank test. Numbers below each graph indicate the number of patients at risk at different time points. ND not detected, Det detected, PFS progression-free survival. **B** Volcano plot showing the hazard ratio (HR) and statistical significance of candidate prognostic bacterial species. The *x* axis represents $\log_2$-transformed HR, while the *y* axis shows $-\log_{10}$-transformed false discovery rates (FDR). Circle size indicates the frequency of significance in 1000 cross-validation tests using univariate Cox regression, and color intensity (gray to red) represents the frequency of significance in multivariate Cox regression analysis. The multivariate Cox regression model was adjusted for age, sex, tumor stage, and neoadjuvant chemoradiotherapy.

tended to increase with the advanced TNM stage and lower tumor histological grade. We did not observe a significant difference in sex, lymphatic metastasis status, and perineural invasion status (Supplementary Fig. 1).

The oral bacterial species are significantly associated with the prognosis of CRC patients

To determine whether the oral bacteria were associated with CRC prognosis, we conducted a two-step identification. Firstly, the univariate Cox regression model was applied to 98 oral microbial species that appeared in at least 10% of the total samples with a relative abundance of at least 0.01%, to preliminarily screen the potential species. Twelve species were found to be substantially associated with CRC progression, and five species showed substantial associations with CRC survival (Supplementary Table 1 and Supplementary Table 2). Subsequently, those significant species were incorporated into a multivariate Cox regression model, adjusting for age, sex, tumor stage, and neoadjuvant chemoradiotherapy, to assess the independent effect of these bacteria on the prognosis of CRC patients. Finally, *Campylobacter gracilis* and *Neisseria oralis* were identified as the indicators associated with an increased risk of CRC progression, while *Treponema medium* was correlated with a reduced risk of progression in the total population cohort (Hazard ratio, HR = 2.63 for *N. oralis*, 95% confidence interval (CI) = 1.50–4.63, false discovery rate (FDR) = 0.08; HR = 2.27 for *C. gracilis*, 95% CI = 1.34–3.86, FDR = 0.09; HR = 0.41 for *T. medium*, 95% CI = 0.23–0.73, FDR = 0.09; Supplementary Table 1, Fig. 2A). No oral bacteria were found to be independently associated with CRC survival (Supplementary Table 2).

To identify robust oral microbial biomarkers, Monte Carlo simulations were utilized for cross-validation across these 98 oral microbial species. The univariate and multivariate Cox regression model was conducted within both training and validation datasets by random sampling 1000 times. Consequently, *T. medium*, *C. gracilis*, and *N. oralis* exhibited the most significant statistical association in both the training and validation datasets, confirming that these three oral bacteria are robust prognostic biomarkers for CRC (Fig. 2B).

Construction of the oral MRSs as accurate predictors of CRC prognosis

To evaluate the performance of oral microbial biomarkers in predicting CRC prognosis, we developed every combination of oral MRS by integrating one, two, or all of the three above-mentioned prognostic bacteria. Using a Cox regression model, we applied these combinations of MRS to 1000 test datasets. Among all combinations, the MRS including three specific oral bacteria demonstrated a significant association with CRC prognosis, exhibiting the highest C-index (C-index = 0.68, 95% CI = 0.61–0.76). (Fig. 3A)

To better differentiate patient prognostic risks, we reclassified the MRS into three categories: 0 for MRS Low, 1 for MRS Moderate, and 2–3 for MRS-High risk. This stratification effectively distinguished progression risks among patients in addition to overall survival (Fig. 3B, C). Patients in the MRS-Low group, only 4 out of 38 patients exhibited progression (10.5%), compared to 20 out of 150 patients in the MRS-Moderate group (13.3%), and 35 out of 124 patients in the MRS-High group (28.2%) experienced disease progression post-surgery.

We also identified the clinical factors associated with the progression of CRC. Consistent with previous researches, tumor stage, perineural invasion, lymphatic metastasis were significantly associated with the progression of CRC in log-rank test, and were identified as clinical prognostic biomarkers ($P < 0.05$, Supplementary Fig. 2). Factors such as age, marital status, cigarette smoking, alcohol consumption history, histologic grade and lymphatic

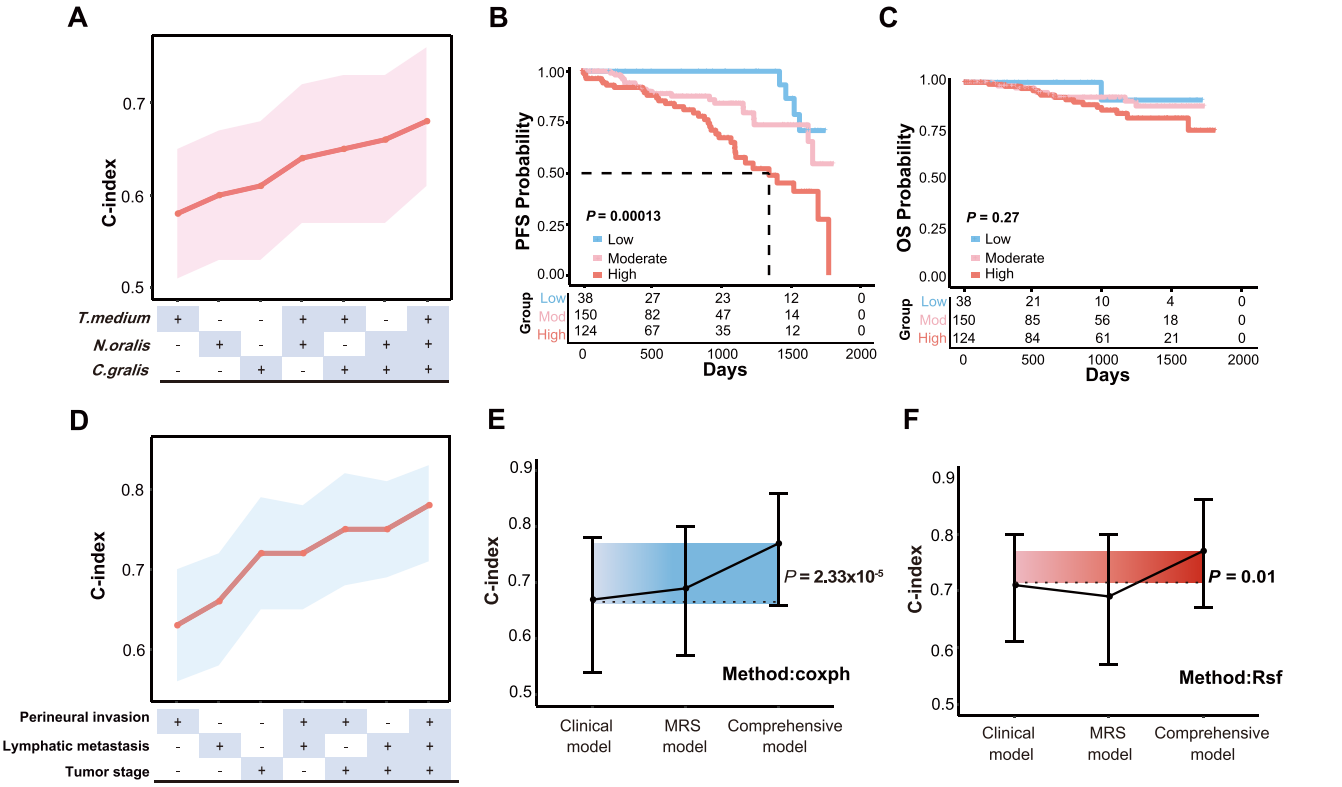

Fig. 3 | The performance of microbial risk scores, clinical model, and comprehensive model. A Comparison of the predictive performance (C-index) of all combinations of microbial risk scores derived from the three identified prognostic species in the test datasets using the Cox regression model. **B** Kaplan–Meier curves for PFS stratified by microbial risk scores (MRS) into low ($n = 38$), moderate ($n = 150$), and high-risk groups ($n = 124$). Numbers below the graph indicate patients at risk at different time points. The *P* value is calculated by the log-rank test. **C** Kaplan–Meier curves for overall survival stratified by the MRS. The *P* value is calculated by the log-rank test. **D** Comparison of the predictive performance (C-index) of all combinations of clinical factors (perineural invasion, lymphatic metastasis, and tumor stage) in the test datasets using the Cox regression model. **E** Performance comparison of the clinical model, MRS model, and comprehensive model (integrating both clinical and microbial factors) in the test datasets using the Cox regression model. The *P* value for the comparison between the clinical model and the comprehensive model was calculated using a Z score test. The bars represent the 95% confidence intervals calculated by the bootstrap method. **F** Performance comparison of the clinical model, MRS model, and comprehensive model in the test datasets using the random survival forest (Rsf) method. The *P* value for the comparison between the clinical model and the comprehensive model was calculated using a Z score test. The bars represent the 95% confidence intervals calculated by bootstrap method. *T. mediumTreponema medium,N. oralisNeisseria oralis*, *C. gralisCampylobacter gralis,* Mod moderate, Rsf random survival forest.

invasion did not show the significant association with cancer progression (Supplementary Fig. 2). We incrementally incorporated these three clinical factors into the model, and the one containing all three demonstrated a higher C-index compared to other clinical models in the 1000 test datasets (C-index: 0.78, 95% CI = 0.71 – 0.83) (Fig. 3D).

Furthermore, we integrated these clinical factors and oral MRSs to construct a combined model. We randomly partitioned the entire dataset into training and test datasets (1:1). In the training datasets, the C-index for the combined model achieved a C-index of 0.84 (95% CI: 0.75 ~ 0.91), and in the test datasets, the C-index for the combined model was 0.77 (95% CI: 0.66 ~ 0.86). The combined model exhibited a higher C-index than the clinical model 0.77 vs. 0.67, $P = 2.33 \times 10^{-5}$, with similar results observed in the train datasets and total population datasets 0.84 vs. 0.76, $P_{\text{training datasets}} = 0.02$, 0.81 vs. 0.72, $P_{\text{total population}} = 3.25 \times 10^{-5}$ (Fig. 3E, Table 2). To further assess the predictive performance of our oral MRS for CRC progression, we constructed a clinical model, an MRS model, and a combined model based on a random forest survival model. Consistently, the combined model outperformed the clinical model in both test and total population datasets, suggesting the stable predictive power of the MRS (Fig. 3F, Table 2).

Correlation between oral MRSs and predictive functional pathways

To investigate the functional roles of oral microbiota among patients with varying oral microbial risks, we used the PICRUSt2 tool to explore the potential metabolic functions of oral microbiota based on 16S rRNA gene sequences. Among the 344 identified KEGG pathways, we excluded those that appeared in fewer than 30% of patients and with an average relative abundance of below 1%, ultimately incorporating 282 metabolic pathways. Patients were categorized into two groups: MRS-low and MRS-moderate/high groups. Sixteen pathways exhibited significant differences between these two groups, with five pathways significantly increased in the MRS-moderate/high group. These pathways are predominantly associated with the promotion of cancer cell proliferation, specifically the Super-pathway of Polyamine Biosynthesis II and Polyamine Biosynthesis, which are linked to polyamine synthesis-an enhancement of cellular proliferation typically observed in cancer cells exhibiting active polyamine metabolism (Fig. 4B). Among the 11 metabolic pathways significantly enriched in the MRS-low group, the most difference in mean proportions was observed in the N-acetylglucosamine (GlcNAc) and N-Acetylgalactosamine pathways (Fig. 4B). This pathway is associated with the synthesis of GlcNAc and GalNAc, both of which can mitigate inflammatory responses, regulate glycosylation, and inhibit cancer cell proliferation. Subsequently, we analyzed the correlation between the differential metabolic pathways and identified oral microbial indicators. We found that species associated with an increased risk of CRC progression exhibited similar functional characteristics but differed markedly from those associated with a reduced risk of CRC progression. For instance, *C. gracilis* and *N. oralis* demonstrated a positive correlation with the increased KEGG pathways in the MRS-moderate/high group and a negative correlation with the depleted pathways. In contrast, the bacterium

Table 2 | The C-index values of the clinical model, the MRS model, and the comprehensive model for predicting progression-free survival probability in the train set, test set, and the overall population

	Cox proportional hazard model			Random survival forests model		
	Train set C-index (95% CI)	Test set C-index (95% CI)	Total population set C-index (95% CI)	Train set C-index (95% CI)	Test set C-index (95% CI)	Total population set C-index (95% CI)
Clinical model[#]	0.76 (0.66, 0.86)	0.67 (0.54, 0.78)	0.72 (0.64, 0.79)	0.80 (0.70, 0.88)	0.71 (0.61, 0.80)	0.76 (0.68, 0.82)
MRS model[$]	0.66 (0.58, 0.75)	0.69 (0.57, 0.80)	0.68 (0.61, 0.75)	0.66 (0.57, 0.75)	0.69 (0.57, 0.80)	0.67 (0.60, 0.74)
Comprehensive model[*]	0.84 (0.75, 0.91)	0.77 (0.66, 0.86)	0.81 (0.74, 0.86)	0.85 (0.77, 0.92)	0.77 (0.67, 0.86)	0.81 (0.74, 0.86)
P value[◆]	0.02	2.33×10^{-5}	3.25×10^{-5}	0.03	0.01	0.0007

[#]The Clinical model included Tumor stage, Perineural invasion, and Lymphatic metastasis.
[$]The MRS model included *Neisseria oralis*, *Campylobacter gracilis*, and *Treponema Medium*.
[*]The Comprehensive model integrated factors from both the clinical and MRS models.
[◆]The *P* value for the comparison between the clinical model and the comprehensive model was calculated using a Z score test.

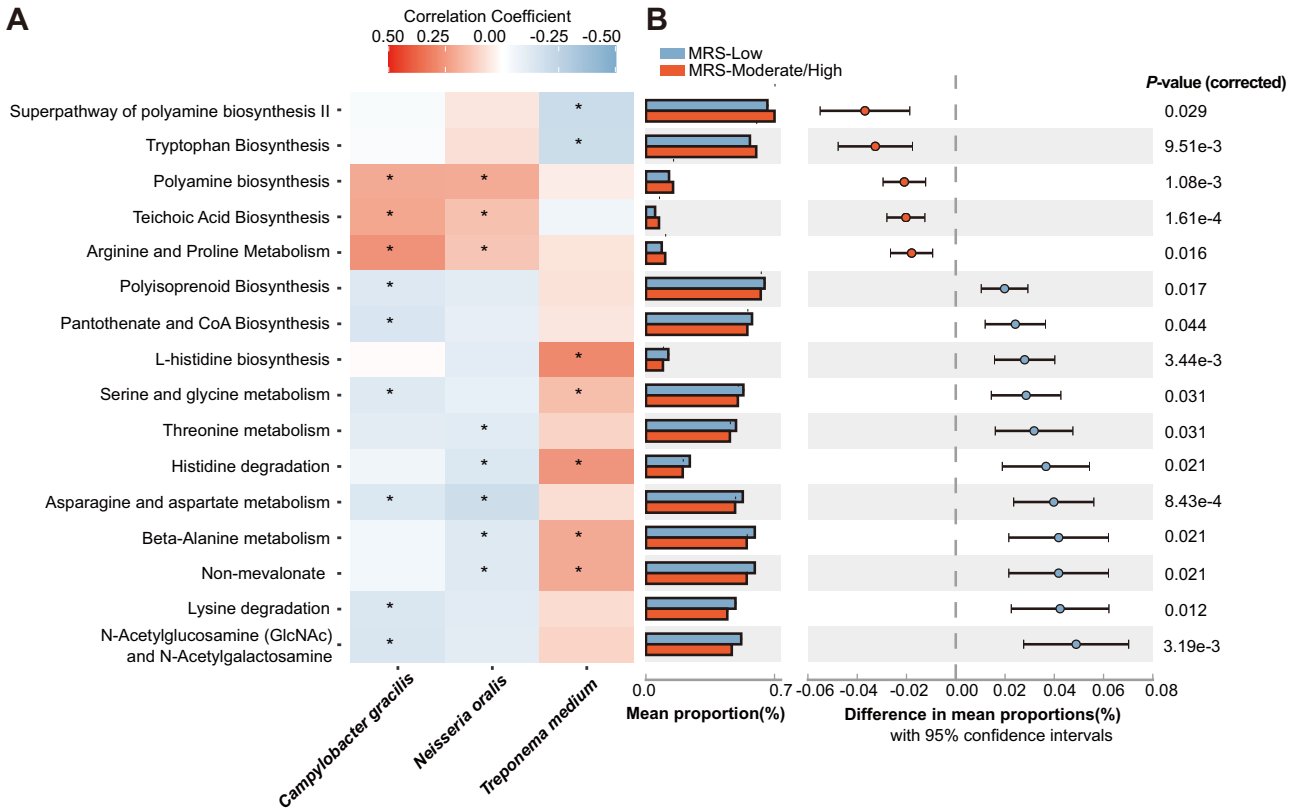

Fig. 4 | The significantly different functional pathways between MRS-Low and MRS-moderate/high patients and their associations with identified prognostic species. A Heatmap of Spearman correlation between prognostic species and differential pathways. The KEGG pathways were predicted using PICRUSt2 to infer the functional shifts in the microbiota of MRS-Low and MRS-moderate/high patients. The strength of the color depicted Spearman's correlation coefficients (negative correlation, blue; positive correlation, red). (*$P < 0.05$). **B** The differential analysis of the oral microbial community's functions between MRS-moderate/high and MRS-low groups, using STAMP software. The analysis focused on KEGG pathways with an average relative abundance exceeding 1%, and the *P* value is corrected using the Bonferroni method. The difference in mean proportion for pathways showing significant differences in abundance was shown. The 95% confidence intervals and statistical significance (*P* value corrected) were indicated as well.

T. medium, which was enriched in the MRS-low group, displayed an opposite trend (Fig. 4A).

Discussion

In this study, we systematically investigated the relationship between oral microbiota and CRC prognosis using full-length 16S rRNA sequencing data from saliva. We identified three prognostic biomarkers for CRC: *C. gracilis*, *N. oralis*, and *T. medium*. Additionally, we developed a prognostic prediction model based on the oral microbiota, which effectively predicts the prognosis of CRC patients. Given the high recurrence rates of CRC post-surgery, it is essential to identify suitable biomarkers for risk assessment[18,19].

Although many studies have explored the microbiota-based prognosis biomarkers for CRC, the majority of previous research has predominantly concentrated on gut microbiota, which is usually tested by a fecal sample[20–22]. The collection and analysis of saliva samples are simpler, more cost-effective, and demonstrate better compliance than fecal samples[8,23,24]. Prognostic models based on fecal microbiota identified the enterotypical *Prevotella* and three microbial biomarkers for predicting clinical outcomes of CRC, achieving a C-index of 0.69[25]. Microbial signatures in tumor tissues

have also demonstrated potential to refine prognostication in CRC patients[26]. Evidence of oral microbiota as prognostic indicators for CRC remains limited. In this study, we identified three oral bacteria associated with CRC prognosis based on follow-up data from patients, and constructed an oral MRS for predicting CRC prognosis. This approach enables the qualitative detection of the target species through a simple and straightforward PCR method, potentially providing advantages for monitoring the postoperative progression risk for CRC patients.

Previous epidemiological studies have established a significant association between oral diseases and CRC. A cohort study of over 700,000 individuals reported a significant correlation between clinically assessed periodontitis and CRC (HR = 1.13, 95% CI: 1.03–1.24)[27]. Similar conclusions were drawn in another study on oral health, which found that women with moderate to severe periodontal disease exhibited a modest increased risk of CRC (HR = 1.22, 95% CI: 0.91–1.63)[28]. These findings suggest that poor oral health may, to a certain extent, facilitate the onset and progression of CRC. Subsequently, numerous studies have begun to focus on the role of oral bacteria in the occurrence of CRC. Zhang S et al.[29] have evaluated the association between oral microbiota dysbiosis and CRC, and constructed a diagnostic model based on oral microbial markers, which can effectively distinguish between adenoma (AUC = 0.94) or CRC (AUC = 0.83) and healthy controls. Another study also found the value of oral microbiota as biomarkers for the detection of CRC, and a classification model containing oral microbiota can effectively distinguish CRC patients (AUC = 0.90), polyps (AUC = 0.89), and healthy controls[8]. These findings not only illuminate the connection between oral health and the risk of CRC but also underscore the potential of the oral microbiota as a diagnostic biomarker for CRC. Therefore, leveraging the oral microbiota as a biomarker for prognostic evaluation of CRC may represent an avenue for future research and clinical practice[19,30,31].

Emerging evidence suggests that oral microbiota may contribute to the development of CRC through several mechanisms. First, the oral-gut microbial translocation axis facilitates direct pathogen infiltration. The oral cavity acts as a reservoir of gut microorganisms, daily swallowing introduces substantial bacterial loads to the gastrointestinal tract. Previous studies have reported an increased oral-gut transmission in CRC patients[32]. Additionally, animal models demonstrated that buccal *F. nucleatum* can migrate to the CRC locus and impair the therapeutic efficacy and prognosis of radiotherapy[33]. Second, certain oral pathogens can enter the bloodstream. Bacteremia caused by oral microbes, such as *Peptostreptococcus* and *Gemella*, has been linked to an increased risk of CRC[34]. Furthermore, oral dysbiosis modulates gut immunoinflammatory responses via the gut-mucosal immune interconnection. Microbial dysbiosis in the oral cavity can exacerbate inflammation in the gut by regulating host immunity, specifically through the introduction of colitogenic pathobionts and pathogenic Th17 cells into the gut[35]. These multilayered interactions underscore the importance of oral microbiota balance in the development of CRC.

Our study identified *N. oralis* and *C. gracilis* as predictors of poor CRC prognosis. In a prospective cohort of 793 individuals, elevated oral abundances of *N. oralis* (OR = 1.42, 95% CI = 1.01–2.00) and *Campylobacter spp.* (OR = 1.58, 95% CI = 1.12–2.24) were significantly associated with CRC risk[36]. *C. gracilis* demonstrates oral-gut translocation capacity, with gut colonization evidenced in inflammatory bowel disease and Crohn's disease biopsies[37,38]. *C. gracilis* is also known as a pathobiont of dental diseases and periodontitis[39,40]. Periodontitis-driven systemic inflammation may contribute to CRC development (e.g., Th17 cell activation) by disrupting gut barrier integrity and amplifying microbial dysbiosis[35]. *N. oralis* typically appears in the oral cavity. One plausible mechanism is that *N. oralis* can convert ethanol into acetaldehyde in the oral cavity[41]. Acetaldehyde, a well-established carcinogen, is known for its substantial toxicity and mutagenicity[39]. These findings elucidate the mechanistic roles of these oral pathobionts in mediating unfavorable CRC prognosis through oral-gut axis translocation, systemic inflammation, and carcinogen production.

Our findings indicate that *T. medium* is associated with a reduced risk of CRC progression. *T. medium*, a member of the genus *Treponema*, is distinct from the notorious *Treponema denticola*, and research on *T. medium* remains limited. Although this bacterium has been identified in gingival plaques, there is still no consensus regarding its pathogenicity[42]. Our findings indicate that *T. medium* is associated with a reduced risk of CRC progression. The mechanisms by which *T. medium* may contribute to reducing CRC progression risk warrant further exploration in future studies.

Other oral microbes, such as *F. nucleatum*[43], *P. intermedia*[44], *Rothia*[45], and *P. gingivalis*[11], have also been identified as the higher abundance microbes enriched in the oral cavity of CRC patients. The well-known oral-derived microbes, *F. nucleatum*, have also been found in the feces and tumor tissues in CRC patients, increasing the risk of CRC by inducing inflammation, disrupting the gut microbiota, and exacerbating metabolic disorders. Its critical virulence factor, RadD, facilitates the attachment to CRC cells and plays a vital role in promoting tumorigenesis[46]. Moreover, *F. nucleatum* can effectively bind to host cells through its virulence factor, adhesin FadA. This interaction enhances bacterial adhesion, disrupts the intestinal barrier[47,48]. Besides, *F.nucleatum* can induce immune suppressive effect by promoting M2 polarization of macrophages via the TLR4/IL-6/PSTAT3/c-MYC signaling pathway[49]. These pathogen–host interactions are involved in the carcinogenesis of CRC. In this study, we did not observe the significant associations between these microbes and CRC progression, which can partly be explained by the fact that the molecular events driving carcinogenesis and the biological processes driving disease progression (e.g., metastasis, therapeutic resistance) frequently involve distinct biological mechanisms. Our current prognostic analyses are confined to CRC patient cohorts where oral microbial profiles inherently exhibit baseline homogeneity due to their shared carcinogenesis-associated signatures. Consequently, bacterial taxa already enriched in CRC populations (e.g., *F.nucleatum*) may demonstrate attenuated prognostic discriminative power.

The oral microbiota plays an important role in the biosynthesis and metabolic degradation of amino acids and carbohydrates[50]. Our study revealed that those bacteria, enriched in high oral microbial risk patients, are positively correlating with metabolic pathways, which are crucial for cellular proliferation, gene expression, and growth[51]. These bacteria had opposite functions to those associated with a reduced risk of CRC progression. Specifically, we identified *N. oralis* and *C. gracilis*, positively correlating with metabolic pathways involved in the synthesis of arginine and proline, phospholipids, and polyamines. Polyamines, such as putrescine, spermidine, and spermine, play an important role in regulating the cell cycle and apoptosis[52]. Therefore, oral microbiota may facilitate cancer growth and metastasis by increasing the biosynthesis of polyamines, providing cancer cells with essential polyamines. In contrast, the analysis of metabolic pathways revealed that *T. medium* was positively correlated with several metabolic routes, including beta-alanine metabolism, L-histidine biosynthesis and degradation, serine and glycine metabolism, and the non-mevalonate pathway. Interestingly, serine and glycine metabolism, as well as the non-mevalonate pathway, are closely associated with the synthesis of the antioxidant glutathione, while the beta-alanine pathway is considered a precursor for the synthesis of carnosine, which functions as an intracellular antioxidant and pH buffer. Oxidative stress is closely associated with cancer development, resulting in cellular damage and DNA mutations[51].

In this study, several limitations should be considered. Firstly, this study primarily utilized a single-center design and lacks an external validation cohort, which restricts the generalizability of our predictive model, necessitating future validation through multi-center studies with independent cohorts. Nevertheless, we conducted a series of sensitivity analyses to ensure the robustness of our results, including randomization of the sample division and different model-constructed approaches. Additionally, the relatively small number of progression events ($n = 59$) may affect the statistical power of our findings. Future prospective investigations incorporating multi-center designs, expanded sample sizes, and extended follow-up periods are required to reliably identify CRC prognostic biomarkers. Finally, while our analysis suggests potential microbial metabolic pathway involvement, the proposed associations between identified microbes, microbial

functional metabolic pathways, and clinical outcomes require experimental validation through in vitro and in vivo studies to elucidate the mechanisms underlying the influence of oral microbiota on CRC progression.

Besides, we acknowledge several methodological constraints of this study. First, while implementing enhanced mechanical lysis, quantitative assessment of microbial lysis efficiency was not conducted, leaving potential Gram-positive bacterial under-lysis unverified. Second, the inherent primer bias of 16S rRNA sequencing may skew microbial composition, which has the potential to be biased towards amplifying specific bacteria. As a high percentage of host genomic DNA is present in oral specimens, 16S rRNA sequencing is widely used in oral microbiota research due to its cost-effectiveness, while metagenomic sequencing data are necessary for more precise information on taxonomic composition. Finally, inferred metabolic pathways through PICRUSt2 disregard strain-level heterogeneity. Future studies should integrate shotgun metagenomics with metabolomic profiling to establish host-microbiota interactions.

In summary, our study suggests potential associations between *N. oralis*, *C. gracilis*, and *T. medium* with the risk of postoperative progression in CRC. By integrating clinical and microbial factors, we achieved improved predictive accuracy for progression risk compared to clinical factors alone. These findings collectively provide preliminary evidence for exploring oral microbiota-based biomarkers for CRC progression, the implementation of oral microbial indicators in CRC clinical surveillance required future confirmatory studies.

Methods

Participants and sample collection

From December 2018 to April 2021, we recruited 361 patients diagnosed with stage I to IV colorectal adenocarcinoma at Sun Yat-sen University Cancer Center. All patients underwent elective surgical resection, and saliva samples were collected preoperatively. Forty-nine patients were excluded as follows: (1) 42 patients were excluded due to postoperative sampling; and (2) 7 patients were excluded because their sequencing data did not meet quality control standards. Ultimately, 312 patients with colorectal adenocarcinoma were included in this analysis. Institutional review board approval was obtained from the Human Ethics Committee of Sun Yat-sen University Cancer Center (The approved number: B2024-542-01), and informed consent was obtained from all patients participating. The study data have been collected in accordance with the Declaration of Helsinki. This study followed the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline (Supplementary Table 3).

All patients were collected the clinical information, including sex, age, marital status, body mass index (BMI), cigarette smoking history, alcohol consumption history, family history of tumors, neoadjuvant chemoradiotherapy, tumor stage, histologic grade, lymphatic invasion, perineural invasion, and lymphatic metastasis. During the study registration period, the saliva samples were collected from patients who were instructed to refrain from eating and drinking for at least 30 min before the collection process. The participants were guided to open the collection tubes and let saliva flow into them effortlessly, which was collected using sterile 50 ml centrifuge tubes. The collected saliva was temporarily stored in an ice box and transferred to 2 ml centrifuge tubes within two hours, and subsequently stored at −80 °C for long-term preservation.

DNA extraction and library construction

In this study, saliva samples from patients were extracted using the DNeasy PowerSoil Pro Kit (QIAGEN, Germany). Saliva specimens were subject to mechanical lysis through bead-beating processing (four cycles of 5-minute agitation at ~2700 rpm for 50 Hz) using the Vortex-Genie 2 system (Scientific Industries) with zirconium dioxide and yttrium oxide beads. The construction of the 16S rRNA amplicon library was described previously[53,54]. Specifically, DNA extracted from each salivary sample served as the template for amplifying the full-length region of the 16S rRNA gene. We amplified the full-length 16S rRNA gene using universal primers 27 F (5'-AGR GTT YGA TYM TGG CTC AG-3') and 1492 R (5'-RGY TAC CTT GTT ACG ACT T-3'), which included 12-bp barcodes. Saliva-derived DNA was amplified using KAPA HiFi HotStart DNA Polymerase (KAPA Biosystems) for 27 cycles. The process involved denaturation at 95 °C for 30 s, annealing at 57 °C for 30 s, and extension at 72 °C for 1 min. We used Agencourt AMPure XP (Beckman Coulter) for fragment selection and purification of the PCR products. We prepared the SMRTbell libraries from the purified amplicons by ligating adapters and sequenced them using the PacBio Sequel platform (Pacific Biosciences). We obtained high-quality circular consensus sequence (CCS) reads from the raw PacBio sequencing data using SMRT Link software (v9.0.0, Pacific Biosciences) and assigned multiplexed libraries to each sample using Lima (v2.0.0) based on the barcodes.

Analysis of 16S rRNA full-length region sequencing data

We employed a customized DADA2 (v1.22.0) workflow for the PacBio full-length 16S sequencing data to conduct quality control, denoising, and identification of amplicon sequence variants (ASVs) from the demultiplexed CCS, ultimately yielding representative sequences and an abundance matrix derived from ASVs[55]. ASVs detected in fewer than five samples or with a total sequence count of fewer than ten were excluded. The "phyloseq" package in R (version 4.4.0) was employed to standardize the sequencing depth of each sample to 2500[56]. Species annotation of ASVs was performed using the pre-trained SILVA database (version 138), categorizing species information across six taxonomic levels: phylum, class, order, family, genus, and species, while calculating the relative abundance of various taxa at different classification levels[57]. Alpha diversity index and beta diversity metrics were computed using the R package "vegan"[58].

Statistical analysis

The demographic, socioeconomic, and clinical characteristics of patients were described and compared among different clinical outcomes using the Pearson $\chi 2$ test for categorical variables, and one-way ANOVA for continuous variables. Permutational multivariate analysis of variance was applied to evaluate the statistical significance of differences in microbial composition across various clinical factors[59]. The C-indexes between the clinical models and comprehensive models were compared by using the Z score test[60]. All statistical analyses and visualization procedures were performed using R software (version 4.4.0) with designated R packages.

Survival analysis

Progression-free survival (PFS) was defined as the period from surgical removal to the recurrence or progression of CRC, or death from any cause. Overall survival was defined as the period from the surgical period to death from any cause. To investigate potential prognostic factors for the CRC, we used Kaplan–Meier survival analysis and the Log-rank test to compare the relationships between various factors and prognosis, converting all variables into categorical forms, such as sex, marital status, BMI, family history of tumor, alcohol drinking history, cigarette smoking history, neoadjuvant chemoradiotherapy, tumor stage, histologic grade, lymphatic invasion, perineural invasion, lymphatic metastasis. Furthermore, we calculated the HRs and their corresponding 95% CIs for different factors influencing CRC prognosis using the Cox proportional hazards regression model. These analyses were performed using the "survival" and "survminer" R packages. Regarding the sample size for PFS analysis, the comparison between the number of progression events[59] and the number of prognostic factors used in the multivariate Cox proportional hazards model indicates that the "minimum of 10 events per predictor" rule was satisfied[61].

Identification of prognostic oral bacteria

A total of 291 bacterial species were observed, of which 98 were included in the prognosis analysis, which met criteria with a frequency exceeding 10% in the total sample and a relative abundance >0.01%. Patients were divided into two groups based on the median abundance of each species, and a univariate Cox regression model was performed. Species that were statistically significant were subsequently incorporated into a multivariate Cox regression

model, adjusted for sex, age, tumor stage, and neoadjuvant chemoradiotherapy.

Next, we applied Monte Carlo simulation cross-validation for a robust selection, by randomly partitioning the entire datasets into discovery and validation sets (1:1) and performing 1000 times randomized resampling. In the discovery set, we performed univariate and multivariate Cox regression models to identify the species associated with prognosis ($P < 0.05$). Species with significance in the discovery stage were then included in the validation set for confirmation. Ultimately, three bacterial species were identified as prognostic biomarkers, demonstrating substantial effect sizes (HR > 1.5 or HR < 0.8) and significance (corrected FDR < 0.1).

Construction of MRSs and prognostic prediction models

Three validated oral bacterial species significantly associated with CRC prognosis were included to construct an MRS. These species were *C. gracilis*, *N. oralis*, both associated with poor prognosis, and *T. medium*, which was conversely associated with good prognosis. The presence of species indicating poor prognosis or the absence of species indicating good prognosis was scored as one point each. Each patient was assigned an MRS ranging from 0 to 3. To evaluate the performance of the combinations, 50% of the samples (156 out of 312) were randomly selected 1000 times without replacement to create a pool of test datasets, and the concordance index (C-index) across the test datasets was used as an indicator of model stability. For clear stratification, a three-layered risk score (MRS), which included the above three bacteria, was categorized by assigning scores 0 for MRS low, 1 for MRS moderate, and 2–3 for MRS high risk. Furthermore, to comprehensively predict the probability of PFS in CRC patients. We constructed a clinical model based on selected clinical factors, including tumor stage, lymphatic invasion, and perineural invasion, all of which were statistically significant in a univariate Log-rank test. A comprehensive model was constructed by integrating both clinical factors and MRS. We calculated the concordance index (C-index) to evaluate the performance of the models and compared the differences in predictive performance between models incorporating and excluding MRS.

Metagenomic PICRUSt analysis

We applied the "Phylogenetic Investigation of Communities by Reconstruction of Unobserved States" (PICRUSt2) tool to infer functional variations within microbial communities[62]. Subsequently, we utilized STAMP software to evaluate significant differential metabolic pathways between moderate/high and low MRS groups, applying the Bonferroni correction with adjusted p value < 0.05 considered significant. The correlation between oral microbes and the significantly related differential pathways was performed by Spearman's rank correlation test.

Data availability

The 16S rRNA sequence data, along with the corresponding patient metadata, are currently being uploaded to the China National Center for Bioinformation (CNCB). Full access to these resources will be public prior to the publication of the study.

Code availability

The key computer codes for the analyses in this study are available on https://github.com/ZSH-AMF/Key_code.git.

Received: 6 December 2024; Accepted: 12 April 2025;
Published online: 05 May 2025

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 233–254 (2023).
3. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. Colorectal cancer. *Lancet* **394**, 1467–1480 (2019).
4. Miyata, T. et al. Predicting prognosis in colorectal cancer patients with curative resection using albumin, lymphocyte count and RAS mutations. *Sci. Rep.* **14**, 14428 (2024).
5. Asili, P. et al. The association of oral microbiome dysbiosis with gastrointestinal cancers and its diagnostic efficacy. *J. Gastrointest. Cancer* **54**, 1082–1101 (2023).
6. Kageyama, S. et al. Characteristics of the salivary microbiota in patients with various digestive tract cancers. *Front. Microbiol.* **10**, 1780 (2019).
7. Dewhirst, F. E. et al. The human oral microbiome. *J. Bacteriol.* **192**, 5002–5017 (2010).
8. Flemer, B. et al. The oral microbiota in colorectal cancer is distinctive and predictive. *Gut* **67**, 1454–1463 (2018).
9. Uchino, Y. et al. Colorectal cancer patients have four specific bacterial species in oral and gut microbiota in common-a metagenomic comparison with healthy subjects. *Cancers***13**, 3332 (2021).
10. Lopera-Maya, E. A. et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. *Nat. Genet.* **54**, 143–151 (2022).
11. Baima, G., Ribaldone, D. G., Romano, F., Aimetti, M. & Romandini, M. The Gum-gut axis: periodontitis and the risk of gastrointestinal cancers. *Cancers***15**, 4594 (2023).
12. Lourenço, T. G. B., de Oliveira, A. M., Tsute Chen, G. & Colombo, A. P. V. Oral-gut bacterial profiles discriminate between periodontal health and diseases. *J. Periodontal Res.* **57**, 1227–1237 (2022).
13. Wang, X. et al. Porphyromonas gingivalis promotes colorectal carcinoma by activating the hematopoietic NLRP3 Inflammasome. *Cancer Res.* **81**, 2745–2759 (2021).
14. Tuominen, H. & Rautava, J. Oral microbiota and cancer development. *Pathobiology* **88**, 116–126 (2021).
15. Mohamed, N. et al. Analysis of salivary mycobiome in a cohort of oral squamous cell carcinoma patients from sudan identifies higher salivary carriage of malassezia as an independent and favorable predictor of overall survival. *Front. Cell Infect. Microbiol.* **11**, 673465 (2021).
16. Du, Y. et al. Influence of pre-treatment saliva microbial diversity and composition on nasopharyngeal carcinoma prognosis. *Front. Cell Infect. Microbiol.* **12**, 831409 (2022).
17. Kim, H. S. et al. Fusobacterium nucleatum induces a tumor microenvironment with diminished adaptive immunity against colorectal cancers. *Front. Cell Infect. Microbiol.* **13**, 1101291 (2023).
18. Luo, D. et al. Clinicopathological features of stage I-III colorectal cancer recurrence over 5 years after radical surgery without receiving neoadjuvant therapy: evidence from a large sample study. *Front. Surg.* **8**, 666400 (2021).
19. Balboa-Barreiro, V. et al. Colorectal cancer recurrence and its impact on survival after curative surgery: An analysis based on multistate models. *Dig. Liver Dis.* **56**, 1229–1236 (2024).
20. Hamada, T. et al. Fusobacterium nucleatum in colorectal cancer relates to immune response differentially by tumor microsatellite instability status. *Cancer Immunol. Res.* **6**, 1327–1336 (2018).
21. Flemer, B., Herlihy, M., O'Riordain, M., Shanahan, F. & O'Toole, P. W. Tumour-associated and non-tumour-associated microbiota: addendum. *Gut Microbes* **9**, 369–373 (2018).
22. Colov, E. P., Degett, T. H., Raskov, H. & Gögenur, I. The impact of the gut microbiota on prognosis after surgery for colorectal cancer - a systematic review and meta-analysis. *APMIS* **128**, 162–176 (2020).
23. Rezasoltani, S. et al. Oral microbiota as novel biomarkers for colorectal cancer screening. *Cancers* **15**, 192 (2022).
24. Krahel, A., Hernik, A., Dmitrzak-Weglarz, M. & Paszynska, E. Saliva as diagnostic material and current methods of collection from oral cavity. *Clin. Lab.* **68**, 2072–2080 (2022).

25. Huh, J. W. et al. Enterotypical Prevotella and three novel bacterial biomarkers in preoperative stool predict the clinical outcome of colorectal cancer. *Microbiome* **10**, 203 (2022).
26. Mouradov, D. et al. Oncomicrobial community profiling identifies clinicomolecular and prognostic subtypes of colorectal cancer. *Gastroenterology* **165**, 104–120 (2023).
27. Kim, E. H. et al. Periodontal disease and cancer risk: a nationwide population-based cohort study. *Front. Oncol.* **12**, 901098 (2022).
28. Momen-Heravi, F. et al. Periodontal disease, tooth loss and colorectal cancer risk: results from the Nurses' health study. *Int. J. Cancer* **140**, 646–652 (2017).
29. Zhang, S. et al. Human oral microbiome dysbiosis as a novel non-invasive biomarker in detection of colorectal cancer. *Theranostics* **10**, 11595–11606 (2020).
30. Lee, S. A., Liu, F., Riordan, S. M., Lee, C. S. & Zhang, L. Global investigations of fusobacterium nucleatum in human colorectal cancer. *Front. Oncol.* **9**, 566 (2019).
31. Chen, Y., Chen, X., Yu, H., Zhou, H. & Xu, S. Oral microbiota as promising diagnostic biomarkers for gastrointestinal cancer: a systematic review. *Onco Targets Ther.* **12**, 11131–11144 (2019).
32. Schmidt, T. S. et al. Extensive transmission of microbes along the gastrointestinal tract. *eLife* **8**, e42693 (2019).
33. Dong, J. et al. Oral microbiota affects the efficacy and prognosis of radiotherapy for colorectal cancer in mouse models. *Cell Rep.* **37**, 109886 (2021).
34. Kwong, T. N. Y. et al. Association between bacteremia from specific microbes and subsequent diagnosis of colorectal cancer. *Gastroenterology* **155**, 383–390.e388 (2018).
35. Kitamoto, S. et al. The intermucosal connection between the mouth and gut in commensal pathobiont-driven colitis. *Cell* **182**, 447–462.e414 (2020).
36. Yang, Y. et al. Prospective study of oral microbiome and colorectal cancer risk in low-income and African American populations. *Int. J. Cancer* **144**, 2381–2389 (2019).
37. Mukhopadhya, I. et al. Detection of Campylobacter concisus and other Campylobacter species in colonic biopsies from adults with ulcerative colitis. *PloS ONE* **6**, e21490 (2011).
38. Zhang, L. et al. Detection and isolation of Campylobacter species other than C. jejuni from children with Crohn's disease. *J. Clin. Microbiol.* **47**, 453–455 (2009).
39. Seitz, H. K. & Stickel, F. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. *Genes Nutr.* **5**, 121–128 (2010).
40. Siqueira, J. F. Jr. & Rôças, I. N. Campylobacter gracilis and Campylobacter rectus in primary endodontic infections. *Int. Endod. J.* **36**, 174–180 (2003).
41. Tagaino, R. et al. Metabolic property of acetaldehyde production from ethanol and glucose by oral Streptococcus and Neisseria. *Sci. Rep.* **9**, 10446 (2019).
42. Zeng, H., Chan, Y., Gao, W., Leung, W. K. & Watt, R. M. Diversity of treponema denticola and other oral treponeme lineages in subjects with periodontitis and gingivitis. *Microbiol. Spectr.* **9**, e0070121 (2021).
43. Guven, D. C. et al. Analysis of Fusobacterium nucleatum and Streptococcus gallolyticus in saliva of colorectal cancer patients. *Biomark. Med.* **13**, 725–735 (2019).
44. Han, S. et al. Potential screening and early diagnosis method for cancer: tongue diagnosis. *Int. J. Oncol.* **48**, 2257–2264 (2016).
45. Kato, I. et al. Oral microbiome and history of smoking and colorectal cancer. *J. Epidemiol. Res.* **2**, 92–101 (2016).
46. Zhang, L. et al. The adhesin RadD enhances Fusobacterium nucleatum tumour colonization and colorectal carcinogenesis. *Nat. Microbiol.* **9**, 2292–2307 (2024).
47. Wang, N. & Fang, J. Y. Fusobacterium nucleatum, a key pathogenic factor and microbial biomarker for colorectal cancer. *Trends Microbiol.* **31**, 159–172 (2023).
48. Li, R., Shen, J. & Xu, Y. Fusobacterium nucleatum and colorectal cancer. *Infect. Drug Resist.* **15**, 1115–1120 (2022).
49. Chen, T. et al. Fusobacterium nucleatum promotes M2 polarization of macrophages in the microenvironment of colorectal tumours via a TLR4-dependent mechanism. *Cancer Immunol. Immunother.* **67**, 1635–1646 (2018).
50. Jia, Y. J. et al. Association between oral microbiota and cigarette smoking in the Chinese population. *Front. Cell Infect. Microbiol.* **11**, 658203 (2021).
51. Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked?. *Free Radic. Biol. Med.* **49**, 1603–1616 (2010).
52. Sagar, N. A., Tarafdar, S., Agarwal, S., Tarafdar, A. & Sharma, S. Polyamines: functions, metabolism, and role in human disease management. *Med. Sci.***9**, 44 (2021).
53. Gohl, D. M. et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* **34**, 942–949 (2016).
54. Liao, Y. et al. Microbes translocation from oral cavity to nasopharyngeal carcinoma in patients. *Nat. Commun.* **15**, 1645 (2024).
55. Callahan, B. J. et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* **47**, e103 (2019).
56. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
57. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
58. Okasnen J. vegan: Community Ecology Package. (2025).
59. Anderson, M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). In *Wiley StatsRef: Statistics Reference Online,* 1–15 (2017).
60. Kang, L., Chen, W., Petrick, N. A. & Gallas, B. D. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat. Med.***34**, 685–703 (2015).
61. Concato, J., Peduzzi, P., Holford, T. R. & Feinstein, A. R. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J. Clin. Epidemiol.* **48**, 1495–1501 (1995).
62. Langille, M. G. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).

Acknowledgements

We thank Hong-Ling Sun and Huan Ren from Sun Yat-sen University Cancer Center for their help in participant enrollment. This research was funded by the National Key Research and Development Program of China (2021YFC2500400), the National Natural Science Foundation of China (82404339, 82403691 and 82473703), the Guang Dong Basic and Applied Basic Research Foundation (2023A1515110442 and 2025A1515010642), the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (24ykqb002), the Science and Technology Planning Project of Guangzhou, China (2024A04J4560), and the Young Talent Support Project of Guangzhou Association for Science and Technology (QT2024-030).

Author contributions

W.H.J. and Y.L. designed the research. S.H.Z, Y.D, and Y.L had full access to all the data in the study and took responsibility for the integrity of the data and

the accuracy of the data analysis. S.H.Z. and Y.D. contributed equally as co-first authors. Y.L., S.H.Z., Y.D., W.Q.X., Y.Q.H., T.M.W., Z.Y.Z., and L.P. acquired, analyzed, and interpreted the data. S.H.Z., Y.L., and Y.D. carried out the statistical analysis; W.H.J. and Y.L. obtained funding and supervised this project. M.J.H., T.Z., C.L.H., Y.W.C., J.R.X., and Z.Y.Z. provided guidance about data analyses. All authors revised the manuscript and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41522-025-00702-0.

Correspondence and requests for materials should be addressed to Ying Liao or Wei-Hua Jia.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025