

RESEARCH

Open Access



Tissue-resident microbiota signature in nasopharyngeal carcinoma

Xi-Rong Tan^{1†}, Han Qiao^{1†}, Ying-Qing Li^{2†}, Wei Jiang^{3†}, Sheng-Yan Huang^{1†}, Sha Gong^{1†}, Wen-Fei Li⁴, Ling-Long Tang⁴, Guan-Qun Zhou⁴, Ye-Lin Liang⁴, Hui Li⁵, Qing-Mei He¹, Jie-Wen Bai¹, Ming-Liang Ye¹, Jing-Yun Wang¹, Sai-Wei Huang⁴, Jun-Yan Li⁴, Chun-Qiao Gan³, Ying-Qin Li¹, Yin Zhao¹, Ying Sun⁴, Jun Ma⁴ and Na Liu^{1*}

Abstract

Background Emerging evidence reveals that microbiota plays a crucial role in multiple cancers. Nasopharyngeal carcinoma (NPC) tissues harbour microbiota, highlighting the need to investigate the clinical implications of tissue-resident microbiota in the development of NPC. Here, we aim to clarify the specific profile of tissue-resident microbiota and its influence on NPC outcomes.

Results This retrospective study included 491 NPC patients from Sun Yat-sen University Cancer Center (Guangzhou, China) and the Affiliated Hospital of Guilin Medical College (Guilin, China). We profiled the microbial composition of 343 NPC and 36 normal nasopharyngeal tissues through sequencing of the genes encoding the 16S rRNA subunit of bacterial ribosomes. There were significant differences in microbial composition, alpha diversity (Shannon index, $P=0.007$; Simpson index, $P=0.036$), and beta diversity (Bray–Curtis distance: $R^2=0.016$, $F=5.187$, $P=0.001$; unweighted UniFrac distance: $R^2=0.017$, $F=5.373$, $P=0.001$) between NPC and normal nasopharyngeal tissues. A bacterial signature comprising four risk bacterial genera, including *Bacteroides*, *Alloprevotella*, *Parvimonas*, and *Dialister*, was constructed in the training cohort ($n=171$). Patients in the high-risk group had shorter disease-free (HR 2.80, 95% CI 1.51–5.18, $P<0.001$), distant metastasis-free (HR 4.00, 95% CI 1.77–9.01, $P<0.001$), and overall survival (HR 3.45, 95% CI 1.77–6.72, $P<0.001$) than those of patients in the low-risk group. Similar results were yielded in the internal validation ($n=172$) and external validation ($n=148$) cohorts. Integrated multi-omics analysis revealed that NPC tissues harbouring abundant risk bacteria were characterised by deficient immune infiltration, which was verified by multiplex immunohistochemistry.

Conclusions This study developed and validated the applicability of a four-bacteria signature as a prognostic tool for NPC prognostication. Integrated multi-omics analysis further uncovered that the tumour immune microenvironment was perturbed by tissue-resident microbiota, which might pave the way towards the era of microbiota-targeted precision medicine for NPC.

Keywords Nasopharyngeal carcinoma, Prognosis, Microbiota signature, Bacteria, Tumour microenvironment

[†]Xi-Rong Tan, Han Qiao, Ying-Qing Li, Wei Jiang, Sheng-Yan Huang and Sha Gong contributed equally to this work.

*Correspondence:

Na Liu

liun1@sysucc.org.cn

Full list of author information is available at the end of the article



Background

Nasopharyngeal carcinoma (NPC), a malignancy arising from the nasopharynx epithelium, prevails in South China, Southeast Asia, and North Africa [1]. Owing to high radiosensitivity and deep-seeded anatomical constraints, radiotherapy is the mainstay treatment modality for NPC, and platinum-based chemotherapy is administered to locoregionally advanced patients [2, 3]. Currently, the tumour-node-metastasis (TNM) staging system occupies a crucial position in the prediction of prognosis and the guidance of treatment decisions for NPC patients. However, prognosis varies even for patients with the same stage and analogous treatment regimens, and approximately 30% of NPC cases suffer from mortality after radical therapy because of local recurrence or distant metastasis [4], indicating the deficiency of this anatomy-based staging system for tailoring individualised therapy for NPC. As this disease is characterised by a heterogeneous nature and complex regulatory network, further refinement of advanced molecular subtyping tools for risk stratification in NPC patients is still necessary [5–8]. Herein, the search for novel efficacious biomarkers remains a major undertaking for NPC prognostication.

The evolutionary fate of humans and microbiota is intimately entwined, and microbial dysbiosis is actively involved in multiple diseases, including malignancies [9]. Almost 20% of the global cancer burden is attributed to microbial infections [10]. Recently, accumulating evidence has implied that tumour tissues, which were initially assumed to be sterile, displayed exclusive microbial signatures [11]. Investigations regarding the tumour-resident microbiota are increasing, and much effort has been made to identify unique microbial profiles and to reveal potential microbiota-tumour interactions. Employing more than 1500 primary tumours and adjacent normal tissue samples, researchers have characterised the cancer-type-specific landscape of intracellular bacteria across seven cancer types [12]. Microbial reads derived from whole-genome or transcriptome sequencing profiles in The Cancer Genome Atlas (TCGA) have been proposed for cancer diagnosis, prognosis, and prediction of therapeutic efficacy [13–15]. Bacteria has been further traced at a single-cell resolution in pancreatic tumours, and their intracellular presence is linked to cell-type-specific transcriptional shifts and clinical outcomes [16]. Additionally, a recent study suggested that bacteria such as *Staphylococcus* and *Lactobacillus* residing inside tumour cells can accelerate metastatic spread by regulating the host-cell actin network and protecting breast tumour cells against fluid shear stress in the circulatory system [17]. These data suggest that the tissue-resident microbiota is intensely associated with tumour

occurrence and development and has promising clinical prospects.

The nasopharynx serves as an essential ecological niche for the upper respiratory microbiota. The bacteria have been found to present within NPC tumour tissues, and they predominately migrate from the nasopharynx [18]. However, the specific profile of the tissue-resident microbes of NPC and its clinical implications remains unclear. In this study, we comprehensively depicted the discrepancy in microbial composition between NPC and normal nasopharyngeal tissues. Additionally, we generated and validated a four-bacteria signature, covering *Bacteroides*, *Alloprevotella*, *Parvimonas*, and *Dialister*, as an efficacious prognostic indicator for NPC. Furthermore, we gained better insights into the microbiota-host axis engaged in NPC progression by integrated functional analysis. Our findings identified a microbial shift in NPC tissues and constructed a bacterial signature for prognostic prediction, which will shed novel insights into precise individualised therapy for NPC patients.

Methods

Study cohort and design

We retrospectively collected 491 fresh-frozen biopsy tissues from patients newly diagnosed with NPC for this study. The inclusion criteria were as follows: (1) pathologically confirmed NPC; (2) aged between 18 and 70 years; (3) no prior history of malignancy; (4) without distant metastasis at initial diagnosis; (5) not received antitumour treatment before sampling; (6) with complete medical records and regular follow-up; (7) sufficient quantity and high-quality DNA for 16S rRNA gene sequencing. Among them, 343 tissues were obtained from the Sun Yat-sen University Cancer Center (SYSUCC, Guangzhou, China) between August 2010 and December 2016, and another 148 were obtained from the Affiliated Hospital of Guilin Medical College (Guilin, China) between April 2014 and May 2020. We restaged NPC patients based on the 8th edition of the American Joint Committee on Cancer staging system. All patients underwent radical radiotherapy, and 449 (91.4%) patients received additional platinum-based chemotherapy. We also collected 36 fresh-frozen normal nasopharyngeal tissues as healthy controls from SYSUCC from August 2010 to December 2016. All samples were pathologically diagnosed by two pathologists, respectively. This study was approved by the institutional ethical review boards of both hospitals for analysing anonymous data, and informed consent was obtained from each patient (B2022-788-01).

We established microbial profiles of 343 NPC and 36 normal nasopharyngeal tissues (Discovery cohort 1, SYSUCC) with 16S rRNA gene sequencing, and also acquired microbial profiles of 48 paired NPC tissues from patients

with/without posttreatment tumour relapse from our previous study (Discovery cohort 2, Public dataset) [18]. 343 NPC samples from the Discovery cohort 1 were randomly separated into a training ($n = 171$) and an internal validation ($n = 172$) cohort. Additionally, 148 samples from the Affiliated Hospital of Guilin Medical College were designated as an external validation cohort. We aimed to construct a prognostic bacterial signature from the training cohort to classify patients into high- and low-risk groups, and test its performance in two validation cohorts. We further performed RNA sequencing (RNA-seq) and multiplex immunohistochemistry (mIHC) in 32 paired NPC tissues with high- or low-risk scores, which were randomly selected from the Discovery cohort 1, to explore the potential mechanisms of the risk bacteria involved in NPC progression (Fig. 1, Fig. S1).

16S rRNA gene sequencing and data processing

Tissue samples were transferred to sterile tubes, immediately submerged in liquid nitrogen for rapid freezing, and then stored at -80°C . We extracted DNA and RNA using the AllPrep DNA/RNA Micro Kit (QIAGEN GmbH) and strictly followed the principle of aseptic operation. The 16S rRNA gene V3–V4 regions were amplified from genomic DNA using quantitative PCR with universal primers (338 F: 5'-ACTCCTACGGGAGGCAGCA- 3' and 806R: 5'-GGACTACHVGGGTWTCTAAT- 3'). Then, the libraries were constructed using the NEBNext[®] Ultra[™] II DNA Library Prep Kit for Illumina[®] (New England Biolabs, MA, USA) and sequenced on the Illumina Nova 6000 platform at Maginene Biotechnology Co., Ltd. (Guangzhou).

Raw reads were processed to obtain amplicon sequence variants (ASVs) by QIIME2 software (v2020.11; Knight and Caporaso Lab). Representative sequence datasets were used for taxonomy classification with a naive Bayes classifier according to the SILVA 16S database (v138; Ribocon GmbH). Contamination controls were set in DNA extraction and PCR amplification processes ($n = 18$) and sequencing lane ($n = 3$), and a contamination-removal (CR) procedure was established as previously reported [12, 18]. First, ASVs containing either mitochondria or chloroplasts originating from the host in taxonomic annotation were removed by KneadData bowtie2 (filter 1) [19]. Next, sequencing library batch (filter 2) was set to avoid contaminants from sequencing lane. ASVs that appear in the negative control of sequencing library batch were removed from the original data. To prevent a false-positive rate, DNA extraction batch (filter 3) and PCR amplification batch (filter 4) were set employing the nonparametric, exact, binomial test. Only taxa that achieved a P value exceeding the threshold of 0.05 in the batch comparisons were included in the subsequent

analysis. The remaining bacterial list was filtered to remove the singleton ASV, in order to avoid the random occurrence of ASVs (filter 5). Furthermore, genera whose average relative abundance did not reach 0.1% (filter 6) in at least 20% of NPC or normal tissues (filter 7) were filtered, to ensure a permanent existence [20, 21].

After contamination removal, the rarefaction curve of both the NPC and normal nasopharyngeal samples was an asymptote, supporting the adequacy of the sequence depth, and the accumulation curve of the ASV numbers flattened out, representing an appropriate sample size (Fig. S2). Identical sequencing depth per sample of 10,000 sequences was then used for the assessment of bacterial alpha and beta diversity at the ASV level using the 'phyloseq' (v1.42.0) and 'vegan' (v2.6–4) packages. Differential genera with adjusted Benjamini–Hochberg P value < 0.05 between groups were analysed with the DESeq2 package (v1.38.3).

RNA sequencing and data processing

Libraries were constructed using NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] (New England Biolabs), and sequenced on the Illumina NovaSeq6000 platform at NuoBiotech Co., Ltd (Guangzhou). Raw reads were adaptor trimmed and data quality was assessed with the fastp software (v0.23.2). Cleaned data were aligned to human reference genome (GRCh38) using STAR software (v2.7.10) in quantMode. RSEM (v1.3.3) was used for further quantification of gene expression. Differentially expressed genes were identified using the DESeq2 package. A list of Hallmark and KEGG gene sets was downloaded from the MSigDB database (v7.4) for gene set enrichment analyses using the 'clusterProfile' package (v4.6.0). Differential pathways ($P < 0.05$; q -value < 0.25 ; absolute normalised enrichment score ($|NES|$) > 1) were aggregated into four categories with a knowledge-based annotation: malignant property-related, metabolic, immunological pathways, and others. Based on the ImmPort database, we obtained differential immune genes ($P < 0.05$; q -value < 0.25 ; $|Fold\ Change| \geq 1$; Student's two-sided t -test) to construct a correlation network ($|r| > 0.30$, $P < 0.05$) with the risk bacterial genera by the Cytoscape software (v3.9). Adjustments for multiple comparisons are presented by Benjamini–Hochberg. Then, immune infiltration was assessed by the MCP-counter algorithm (v1.2.0) [22].

Multiplex immunohistochemistry

Sequential tumour sections were used for mIHC with a PANO 7-plex IHC kit (Panovue, Beijing, China) as we previously described [23, 24]. Briefly, sections were deparaffinised and rehydrated, followed by heat-induced epitope retrieval and blocking of non-specific binding

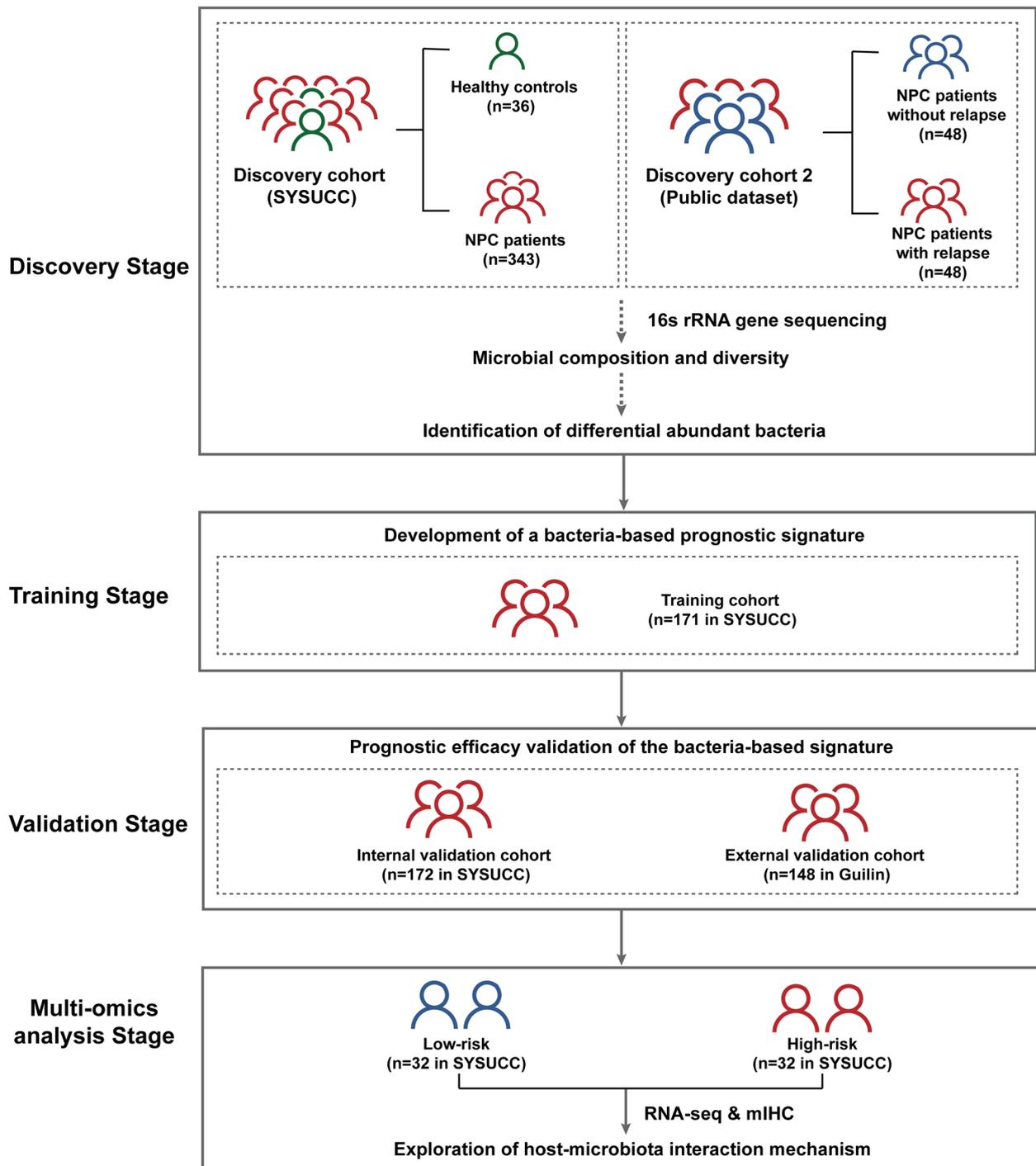


Fig. 1 Study flowchart

block. Then, sections were manually stained with the following primary antibodies: anti-CD3 (Abcam, Cambridge, UK), anti-CD8 (ZSGB-BIO, Beijing, China), anti-CD20 (Sigma-Aldrich, MO, USA), anti-CD56 (ZSGB-BIO), and anti-panCK (ZSGB-BIO), followed by

the HRP-conjugated secondary antibody, tyramide signal amplification, and OPAL fluorophore dyes (OPAL480: CD3; OPAL520: CD56; OPAL570: CD20; OPAL690: CD8; OPAL780: panCK), were subsequently stained. Whole-slide images were scanned with Vectra® Polaris®

multispectral imaging system (Akoya Biosciences, USA), and analysed in HALO[®] image software (Indica Labs, USA) as described previously [23, 24].

Statistical analysis

The primary endpoint was disease-free survival (DFS), defined as the period from the first date of treatment to tumour relapse at any site or death from any cause, whichever occurred first. The secondary endpoints were distant metastasis-free survival (DMFS), which is the interval from the first date of treatment to distant metastasis or non-cancer specific death, and overall survival (OS), which refers to death from any cause.

The Wilcoxon rank-sum test was used to compare microbial differences between groups, including NPC and normal nasopharyngeal tissues, NPC patients with different age, sex, TNM stage, and plasma EBV-DNA load. The χ^2 or Fisher's exact test was applied to compare categorical variables. We performed the least absolute shrinkage and selection operator (LASSO) on a binomial logistic regression model to identify risk bacterial genera related to DFS in the training cohort using 'glmnet' package (v4.1–6) [25]. A λ value via min (minimum error) criteria under tenfold cross-validation were used to determine how many candidate genera should be chosen for model construction. According to the λ value, genera whose beta coefficients were not zero were selected to generate a risk score for each patient. The risk scores were calculated using a formula derived from the relative abundance of risk genera weighted by their beta coefficients [26]. A receiver operating characteristic (ROC) curve was used to determine the optimal cut-off in the training cohort using 'pROC' package (v1.18.0). The threshold that produced the maximum sum of sensitivity and specificity in the ROC curve were used to divide patients into low- and high-risk groups. Survival probabilities were evaluated by Kaplan–Meier method and log-rank test, and hazard ratios (HRs) were calculated by univariable Cox regression analysis. Multivariate Cox regression analysis with backward selection was used to test the significance of independent factors. The bacterial signature, age, sex, T stage, N stage, pathological type, plasma EBV-DNA, and chemotherapy were used as covariates.

We formulated nomogram, comprising the four-bacteria signature, N stage, and plasma EBV-DNA load, using the rms package. We employed the coefficients of multivariable Cox regression to generate nomogram. Calibration curves were evaluated graphically by comparing the observed rates with the nomogram predicted probabilities, and concordance index (C-index) was calculated using survival package (v3.4). We investigated the prognostic or predictive accuracy of this classifier using

receiver operating characteristic (ROC) analysis in pROC package (v1.18.0). Comparison between immune cell infiltration was compared by Wilcoxon rank-sum test. Correlation relationships were assessed by the Spearman correlation test. All analyses were performed in R v4.2.2 and SPSS version 22.0 with two-tailed tests, and $P < 0.05$ was considered significant.

Results

Patient characteristics

We totally included 491 NPC and 36 normal nasopharyngeal samples in this study (Fig. 1, Fig. S1). In the discovery stage, we enrolled 343 NPC and 36 normal nasopharyngeal tissues from the SYSUCC (Discovery cohort 1) (Table S1), as well as analysed 48 paired NPC tissues from patients with or without relapse from a public dataset of our previous study (Discovery cohort 2) [18]. The 343 NPC samples were randomly separated into a training cohort ($n = 171$, SYSUCC) and an internal validation cohort ($n = 172$, SYSUCC). Additionally, another 148 samples from the first Affiliated Hospital of Guilin Medical University were designated as an external validation cohort ($n = 148$, Guilin). The clinical characteristics of NPC patients in the training, internal validation, and external validation cohorts are listed in Table 1. All patients were treated with radical radiotherapy, and platinum-based chemotherapy was administered to 159 (93.0%) of 171 patients in the training cohort, 152 (88.4%) of 172 in the internal validation cohort, and 138 (93.2%) of 148 in the external validation cohort. The median follow-up of the training, internal validation, and external validation cohorts was 91.7 months (IQR 66.9–107.0), 93.2 months (71.1–107.5), and 67.6 months (44.7–84.5), respectively.

The microbial profile differs between NPC and normal nasopharyngeal tissues

After the removal of contaminants using the established CR procedure (Fig. 2A, Table S2, Fig. S2–3), we first screened representative bacterial genera to perform composition analysis according to the microbiota taxon database between 343 NPC tissues and 36 normal nasopharyngeal tissues in the Discovery cohort 1 (Table S3). At the phylum level, NPC tissues exhibited significant enrichment of *Firmicutes* (16.4 vs. 5.4%, $P = 0.019$), *Actinobacteriota* (14.2 vs. 3.4%, $P < 0.001$), and *Campilobacterota* (0.6 vs. 0.1%, $P = 0.037$) and decreased abundance of *Proteobacteria* (50.5 vs. 79.7%, $P < 0.001$) compared to that of normal nasopharyngeal tissues (Fig. 2B–C, Fig. S4 A). In addition, the dominant and discrepant bacteria between NPC and normal nasopharyngeal tissues at the levels of class, order, and family are shown in Fig. S4B–D.

Table 1 Clinical characteristics of patients in the training, internal validation, and external validation cohort

	Training cohort (n = 171)			Internal validation cohort (n = 172)			External validation cohort (n = 148)		
	No of patients	Low risk (%)	High risk (%)	No of patients	Low risk (%)	High risk (%)	No of patients	Low risk (%)	High risk (%)
Sex									
Female	49	40 (28.8)	9 (28.1)	40	36 (24.2)	4 (17.4)	49	38 (36.5)	11 (25.0)
Male	122	99 (71.2)	23 (71.9)	132	113 (75.8)	19 (82.6)	99	66 (63.5)	33 (75.0)
Age/years									
< 45	91	76 (54.7)	15 (46.9)	93	86 (57.7)	7 (30.4)	48	38 (36.5)	10 (22.7)
≥ 45	80	63 (45.3)	17 (53.1)	79	63 (42.3)	16 (69.6)	100	66 (63.5)	34 (77.3)
T stage									
T1	17	16 (11.5)	1 (3.1)	16	13 (8.7)	3 (13.1)	4	2 (1.9)	2 (4.6)
T2	32	27 (19.4)	5 (15.6)	31	26 (17.5)	5 (21.7)	32	25 (24.0)	7 (15.9)
T3	88	71 (51.1)	17 (53.1)	94	86 (57.7)	8 (34.8)	69	48 (46.2)	21 (47.7)
T4	34	25 (18.0)	9 (28.1)	31	24 (16.1)	7 (30.4)	43	29 (27.9)	14 (31.8)
N stage									
N0	18	17 (12.2)	1 (3.1)	26	24 (16.1)	2 (8.7)	10	6 (5.8)	4 (9.1)
N1	83	71 (51.1)	12 (37.5)	71	64 (43.0)	7 (30.4)	59	45 (43.3)	14 (31.8)
N2	52	36 (25.9)	16 (50.0)	42	36 (24.1)	6 (26.1)	58	41 (39.4)	17 (38.6)
N3	18	15 (10.8)	3 (9.4)	33	25 (16.8)	8 (34.8)	21	12 (11.5)	9 (20.5)
TNM stage									
I	7	6 (4.3)	1 (3.1)	6	6 (4.1)	0 (0.0)	0	0 (0)	0 (0.0)
II	26	25 (18.0)	1 (3.1)	19	17 (11.4)	2 (8.7)	13	11 (10.6)	2 (4.6)
III	90	71 (51.1)	19 (59.4)	87	79 (53.0)	8 (34.8)	75	54 (51.9)	21 (47.7)
IV	48	37 (26.6)	11 (34.4)	60	47 (31.5)	13 (56.5)	60	39 (37.5)	21 (47.7)
WHO pathological type									
Undifferentiated	164	133 (95.7)	31 (96.9)	166	144 (96.6)	22 (95.6)	145	102 (98.1)	43 (97.7)
Differentiated	7	6 (4.3)	1 (3.1)	6	5 (3.4)	1 (4.4)	3	2 (1.9)	1 (2.3)
EBV DNA (copies/ml)									
< 2000	79	66 (47.5)	13 (40.6)	76	66 (44.3)	10 (43.5)	NA	NA	NA
≥ 2000	92	73 (52.5)	19 (59.4)	96	83 (55.7)	13 (56.5)	NA	NA	NA
Chemotherapy									
No	12	11 (7.9)	1 (3.1)	20	18 (12.1)	2 (8.7)	10	9 (8.7)	1 (2.3)
Yes	159	128 (92.1)	31 (96.9)	152	131 (87.9)	21 (91.3)	138	95 (91.4)	43 (97.7)
Disease progression									
No	124	107 (77.0)	17 (53.1)	125	116 (77.8)	9 (39.1)	83	66 (63.5)	17 (38.6)
Yes	47	32 (23.0)	15 (46.9)	47	33 (22.2)	14 (60.9)	65	38 (36.5)	27 (61.4)
Distant metastasis									
No	147	125 (89.9)	22 (68.7)	135	124 (83.2)	11 (47.8)	113	86 (82.7)	27 (61.4)
Yes	24	14 (10.1)	10 (31.3)	37	25 (16.8)	12 (52.2)	35	18 (17.3)	17 (38.6)
Death									
No	134	116 (83.4)	18 (56.2)	136	125 (83.9)	11 (47.8)	111	84 (80.8)	27 (61.4)
Yes	37	23 (16.6)	14 (43.8)	36	24 (16.1)	12 (52.2)	37	20 (19.2)	17 (38.6)

TNM tumour node metastasis, WHO World Health Organization, EBV Epstein-Barr virus, NA not available

To learn the association between microbial composition and clinical covariates, we also performed composition analysis according to different clinical factors (age,

sex, TNM stage, and plasma EBV-DNA load). *Fusobacterium*, *Prevotella*, and *Porphyromonas* were enriched in old NPC patients, *Glutamicibacter* was enriched in

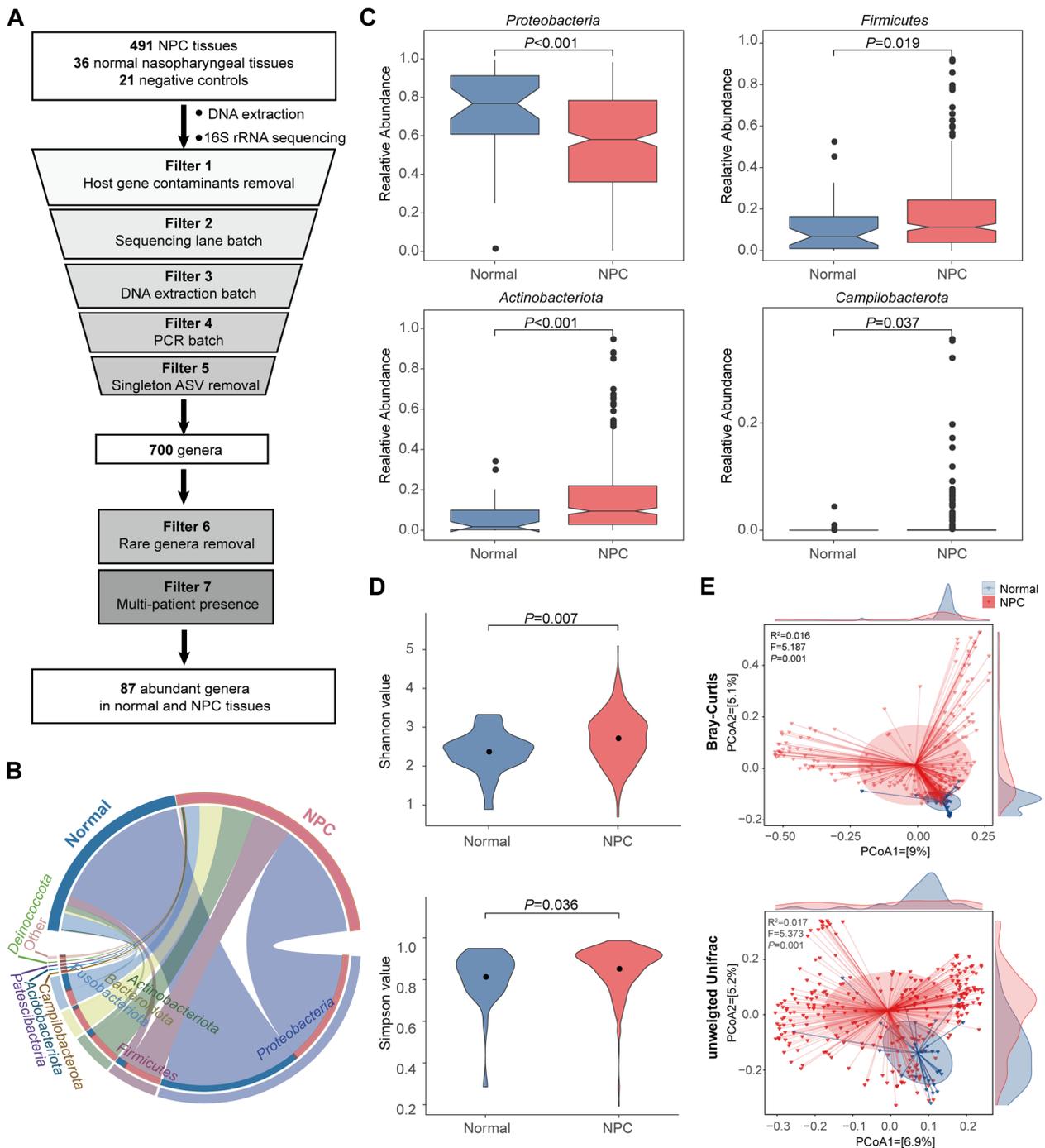


Fig. 2 The microbial profile differs between NPC and normal nasopharyngeal tissues. **A** Contamination removal procedure for 16S rRNA gene sequencing. **B** Overview of the bacterial differences in 343 NPC tumours and 36 normal nasopharyngeal tissues at the phylum level using the Circos plot. **C** Notched boxplots show the differential abundances of bacteria between NPC and normal nasopharyngeal tissues. The comparison was performed with the Wilcoxon rank-sum test. **D** Alpha diversity violin plot (Shannon and Simpson indexes) revealed differences in the community richness and evenness of the bacteria between NPC and normal nasopharyngeal tissues. The comparison was performed with the Wilcoxon rank-sum test. **E** Principal coordinate analysis (PCoA) using Bray-Curtis and unweighted UniFrac distances of beta diversity revealed the diversity distance of bacteria between NPC and normal nasopharyngeal tissues

patients with advanced-stage, and *Pseudomonas* was enriched in male patients. *Pseudomonas* was positively correlated with plasma EBV-DNA load, while *Haemophilus* was negatively correlated (Fig. S5).

Notably, compared with the normal nasopharyngeal tissues, NPC tissues displayed significantly increased microbial alpha diversity, as determined by the Shannon ($P = 0.007$) and Simpson ($P = 0.036$) indexes (Fig. 2D). In terms of beta diversity, PCoA revealed a global difference in the composition and abundance of microbiome between NPC and normal nasopharyngeal tissues (Bray–Curtis distance: $R^2 = 0.016$, $F = 5.187$, $P = 0.001$; unweighted UniFrac distance: $R^2 = 0.017$, $F = 5.373$, $P = 0.001$) (Fig. 2E).

Construction of a four-bacteria signature in the training cohort

Based on the abundant bacteria at the genus level (Table S4), we aimed to develop a bacteria-based model for NPC prognosis. Firstly, we identified 74 bacterial genera that were differentially distributed between 343 NPC and 36 normal nasopharyngeal tissues from the Discovery cohort 1 (Fig. 3A, Table S5). Secondly, we found nine differential bacterial genera between 48 paired tissues of NPC patients with or without posttreatment tumour relapse in the Discovery cohort 2 obtained from a public dataset (Fig. 3B, Table S6). After the intersection of the two comparisons, eight candidate bacterial genera were obtained for subsequent analysis. Then, we identified four risk bacterial genera, including *Bacteroides*, *Alloprevotella*, *Parvimonas*, and *Dialister*, to construct a bacterial signature for predicting DFS of NPC patients in the training cohort (Fig. 3C–D).

We generated a risk score for each patient using a formula derived from the relative abundance of the four risk bacterial genera weighted by their regression coefficients as follows: Risk score = $(3.4585 \times \text{relative abundance of } Bacteroides) + (8.5512 \times \text{relative abundance of } Alloprevotella) + (2.3116 \times \text{relative abundance of } Parvimonas) + 3.7812 \times (\text{relative abundance of } Dialister)$. Subsequently, an optimal cut-off value (0.0288) was identified for dividing patients into the low- or high-risk groups in the training cohort.

Prognostic performance of the four-bacteria signature across three cohorts

To determine the prognostic performance of the four-bacteria signature for NPC patients in the training cohort, we assigned 32 (18.7%) of 171 patients to the high-risk group and the remaining 139 (81.3%) to the low-risk group. Survival analysis demonstrated that patients in the high-risk group had significantly poorer DFS (HR 2.80, 95% CI 1.51–5.18, $P < 0.001$), DMFS (HR

4.00, 95% CI 1.77–9.01, $P < 0.001$), and OS (HR 3.45, 95% CI 1.77–6.72, $P < 0.001$) than those in the low-risk group (Fig. 4A–C).

To validate the performance of the four-bacteria signature, we included two validation cohorts. In the internal validation cohort, using the same formula and cut-off developed in the training cohort, the signature successfully classified 23 (13.4%) of 172 patients into the high-risk group and 149 (86.6%) into the low-risk group, which were obviously different in terms of DFS (HR 3.98, 95% CI 2.13–7.46, $P < 0.001$; Fig. 4D), DMFS (HR 4.37, 95% CI 2.19–8.72, $P < 0.001$; Fig. 4E), and OS (HR 4.36, 95% CI 2.18–8.74, $P < 0.001$; Fig. 4F).

In the external validation cohort, there were 44 (29.7%) patients and 104 (70.3%) patients categorised into the high- and low-risk groups, respectively. Patients in the high-risk group had markedly worse DFS (HR 2.09, 95% CI 1.28–3.43, $P = 0.003$; Fig. 4G), DMFS (HR 2.61, 95% CI 1.35–5.07, $P = 0.003$; Fig. 4H), and OS (HR 2.46, 95% CI 1.29–4.71, $P = 0.005$; Fig. 4I) than those of patients in the low-risk group.

Nomogram integrating the four-bacteria signature and clinical factors

We then performed univariate Cox regression analysis and demonstrated that the four-bacteria signature was significantly associated with DFS in all three cohorts (Fig. S6). After adjusting for the other clinicopathological characteristics, multivariate Cox regression analysis revealed that the four-bacteria signature remained a powerful independent prognostic factor for DFS in the training (HR 2.37, 95% CI 1.27–4.42, $P = 0.007$; Table S7), internal validation (HR 4.06, 95% CI 2.11–7.83, $P < 0.001$; Table S8), and external validation (HR 1.99, 95% CI 1.21–3.27, $P = 0.007$; Table S9) cohorts. Additionally, N stage and plasma EBV-DNA load were also independent prognostic factors for DFS in the three cohorts. Similar findings were acquired for DMFS and OS (Tables S7–9).

Furthermore, we generated a nomogram to predict DFS in the training cohort, comprising the four-bacteria signature, N stage, and plasma EBV-DNA load (Fig. 5A). The calibration plots for the 5-year DFS were predicted well in the training (C-index 0.713, 95% CI 0.643–0.783), internal validation (C-index 0.712, 95% CI 0.643–0.781), and external validation (C-index 0.639, 95% CI 0.568–0.710) cohorts (Fig. 5B–D). ROC analysis validated that the nomogram performed well for survival prediction in all three cohorts (Fig. 5E).

The abundance of the risk bacteria is negatively associated with immune infiltration

Given the well-defined association between commensal microbiota and host function [27], we hypothesised that

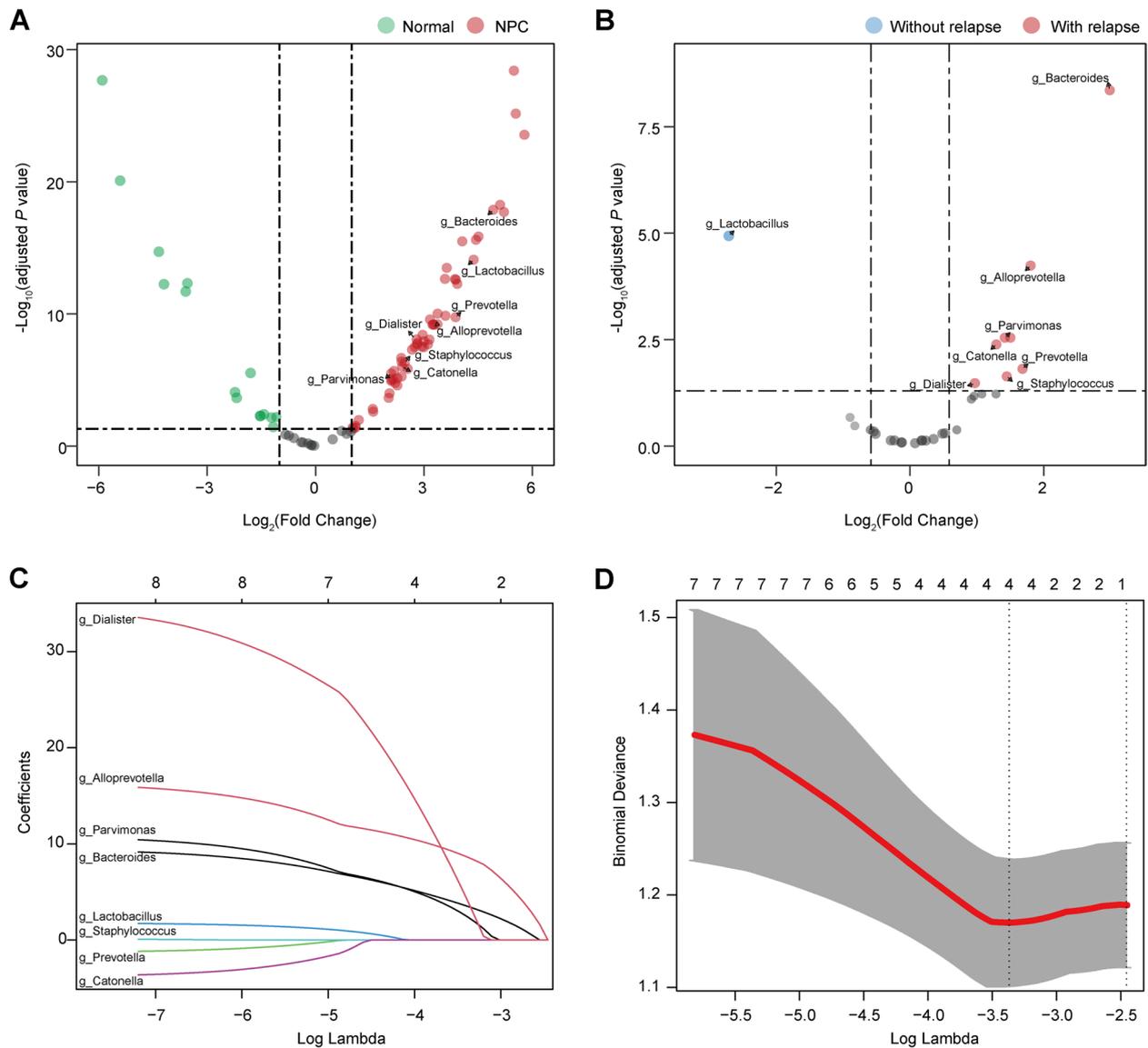


Fig. 3 Construction of a four-bacterial signature in the training cohort. **A–B** Identification of differential bacterial genera between NPC and normal nasopharyngeal tissues (adjusted P value <0.05 ; $|\text{Fold Change}| > 2$, left), as well as NPC tissues with or without relapse (adjusted P value <0.05 ; $|\text{Fold Change}| > 1.5$, right) using the DESeq2. **C** Least absolute shrinkage and selection operator (LASSO) coefficient profiles of the candidate bacterial genera. **D** Ten-time cross-validations to tune the parameter selection in the LASSO model. The dotted vertical lines are drawn at the optimal values by minimum criteria

the different abundances of risk bacteria within tumours in NPC patients may influence tumour progression by modifying signal transduction and the tumour microenvironment. We randomly selected 32 paired NPC tissues with high- or low-risk scores from the Discovery cohort 1, which were well matched by the sex, age, TNM stage, and chemotherapy, for the following transcriptome analysis (Table S10). GSEA revealed that the malignant property- and metabolic pathways were positively enriched in tumours with a higher abundance of risk bacteria,

whereas the immune-related pathways were negatively enriched (Fig. 6A, Fig. S7, Table S11). Moreover, an apparent distinction was exhibited between immune genes in tissues harbouring high or low levels of risk bacteria, predominantly those genes involved in the T-cell receptor (TCR) and B-cell receptor (BCR) signalling pathways (Fig. S8, Table S12). The correlation network showed that the four bacteria were mainly negatively correlated with immune genes, especially *Bacteroides*, which had a tight interaction with immune activity (Fig. 6B).

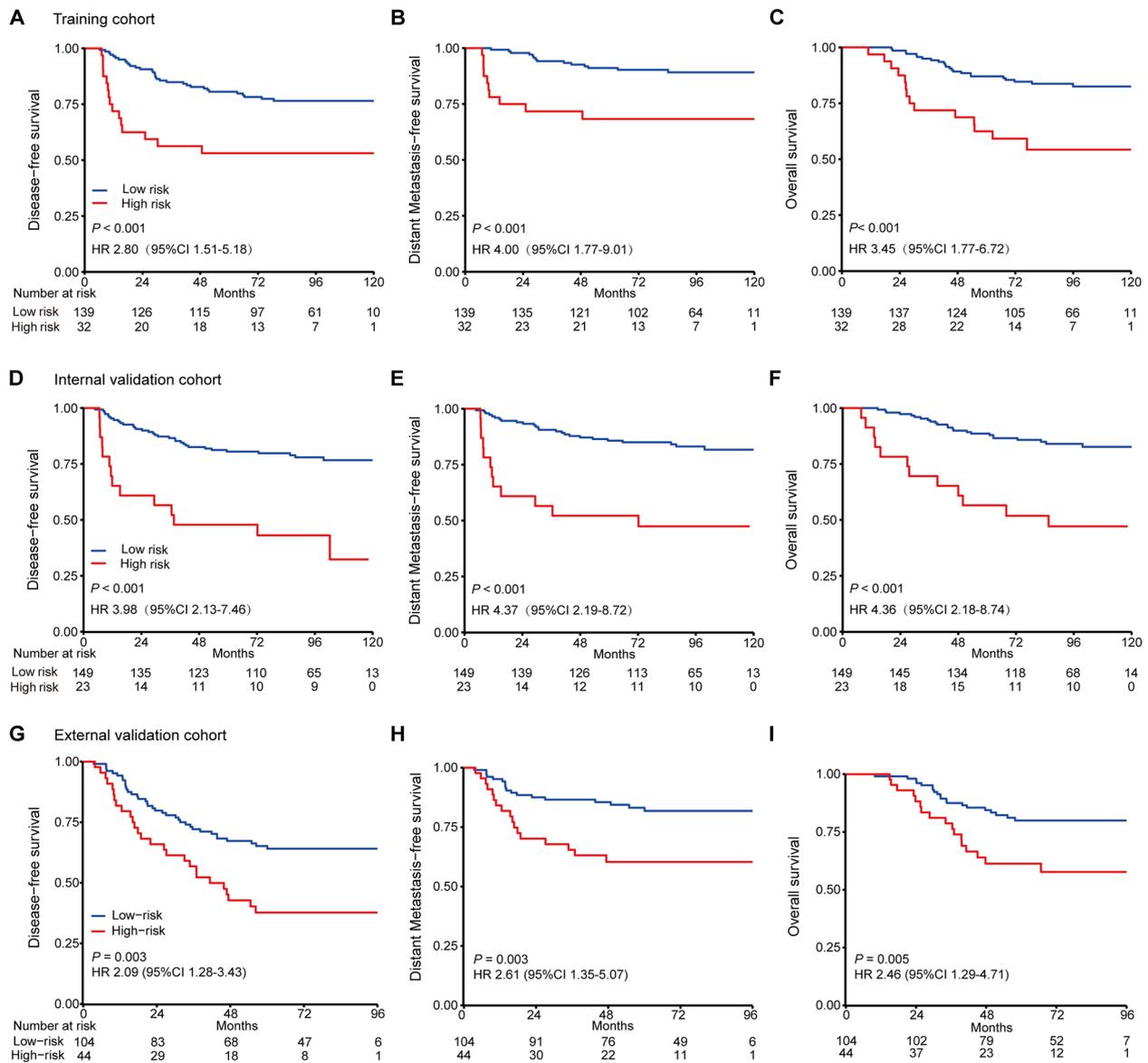


Fig. 4 Kaplan-Meier curves of survivals according to the four-bacteria signature. **A–C** Disease-free survival (DFS), distant metastasis-free survival (DMFS), and overall survival (OS) in the training cohort ($n=171$). **D–F** DFS, DMFS, and OS in the internal validation cohort ($n=172$). **G–I** DFS, DMFS, and OS in the external validation cohort ($n=148$). We calculated the P values using the unadjusted log-rank test and the hazard ratios using univariate Cox regression analysis

We further explored the immune infiltration patterns and found that tumours with a high abundance of risk bacteria had inferior infiltration of B cells, total T cells, CD8⁺ T cells, cytotoxic lymphocytes, and NK cells (Fig. 6C). The mIHC results displayed different localizations and abundances of immune cells in NPC tumours harbouring different levels of risk bacteria (Fig. 6D). Moreover, the infiltration of intratumoural and stromal CD8⁺ T cells, B cells, and NK cells was significantly lower in tumours of the high-risk group (Fig. 6E), indicating a

negative correlation between risk bacteria and immune infiltration.

Discussion

In this multicentre, retrospective cohort study, we depicted significant differences in microbial composition and diversity between NPC and normal nasopharyngeal tissues. More importantly, we developed and validated a four-bacteria signature, consisted of *Bacteroides*, *Alloprevotella*, *Parvimonas*, and *Dialister*, as an efficacious

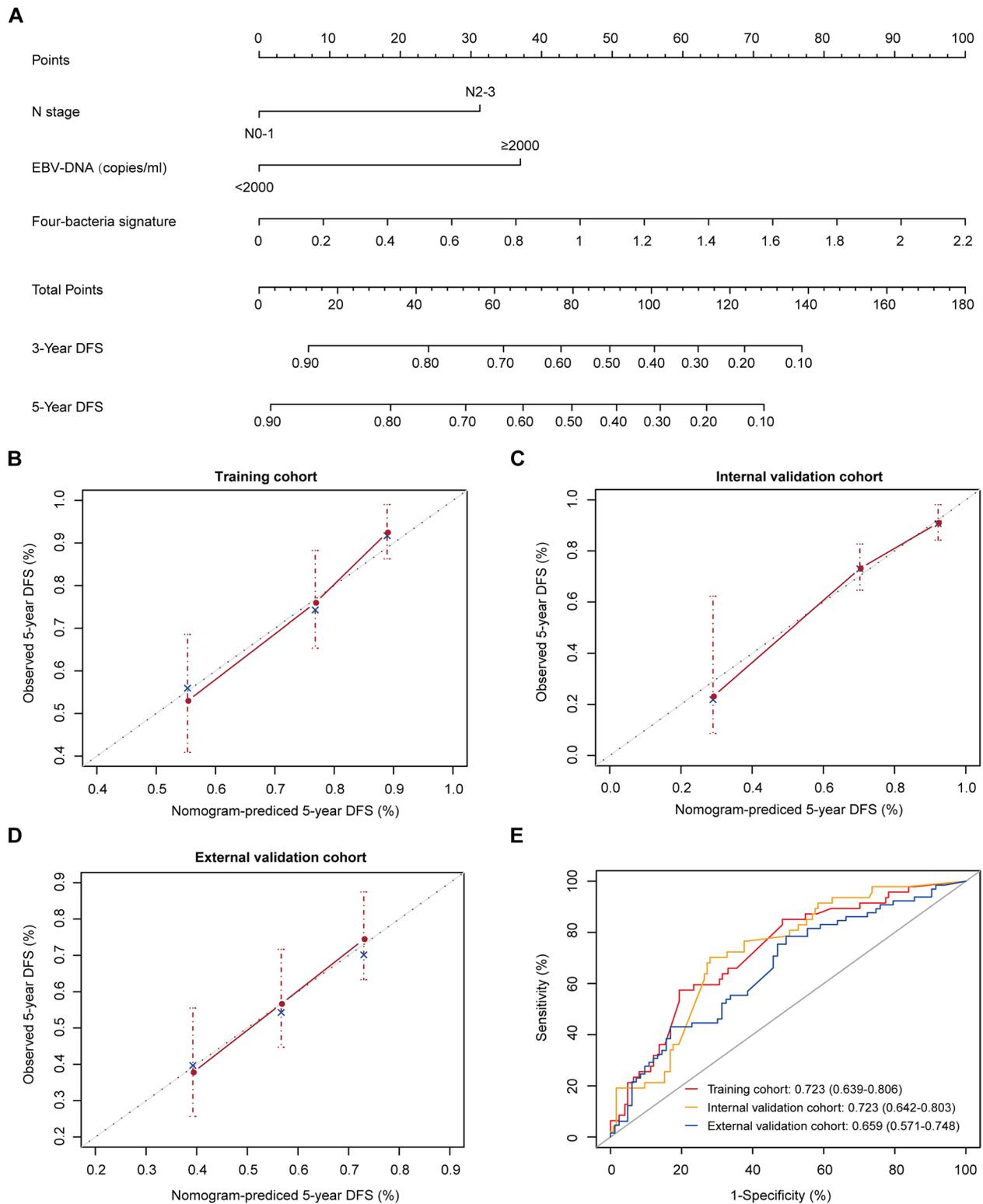


Fig. 5 Nomogram integrating the four-bacteria signature and clinical factors. **A** Establishment of a nomogram integrating the four-bacterial signature, N stage, and plasma EBV-DNA load to predict disease-free survival (DFS) in the training cohort. **B** Calibration curves of the nomogram to predict DFS at 5 years in the training cohort. **C** Calibration curves of the nomogram to predict DFS at 5 years in the internal validation cohort. **D** Calibration curves of the nomogram to predict DFS at 5 years in the external validation cohort. **E** Receiver operating characteristic (ROC) curves of the established nomogram to evaluate DFS of NPC patients in the training, internal, and external validation cohorts

prognostic indicator for NPC patients. Additionally, our results provided vital evidence that microbial shifts were associated with NPC progression, which might be mediated by alterations in malignant property-related, metabolic, and immunological pathways.

The relative abundance of the microbial composition showed that *Proteobacteria* was the most abundant phylum, which is similar to our previous study and another group's research on gastric cancer [18, 28]. We attributed the higher alpha diversity to the caustic bacterial environment in NPC tissues, and it is highly possible that the tumour microenvironment of NPC assists specific microbiota by providing favourable conditions to allow them to persist more readily than in normal tissues. It has also been reported that the intratumoural microbiota exhibits high heterogeneity in various tumours [29]. Of particular significance is the distinctive separation trend characterised by the beta diversity, noting the microbial alterations among NPC and normal nasopharyngeal tissues. Clinical characteristics, including age and sex, have an impact on bacterial communities [30], and there were no significant differences in these factors, thus avoiding their interferences in the subsequent analysis. Moreover, *Pseudomonas* and *Haemophilus* were identified to correlate with pretreatment EBV-DNA load. These results indicate that in favour of adapting to the host microenvironment, the associations among distinct bacterial genera may develop into ecological drivers of a multi-kingdom microbiota assembly [31, 32]. Therefore, an in-depth exploration of the multi-kingdom microbiota covering nonbacterial microorganisms is anticipated to provide a more comprehensive understanding of the sophisticated biological processes in NPC.

Beyond introducing the landscape of the tissue-resident microbiota in NPC patients, our team proposes the idea of a bacteria-based profile for NPC prognostic prediction. Recently, we revealed that NPC intratumoural bacteria mainly originate from the nasopharynx, and the

intratumoural bacterial load can serve as an efficacious prognostic indicator [18]. However, it is still unclear which tissue-resident microbes play critical roles in the prognosis of NPC. In this bacterial signature, all four bacteria have been demonstrated to functionally impact the host. *Bacillus fragilis*, a representative of *Bacteroides*, has been detected in various cancer types and can promote tumour progression by secreting *B. fragilis* toxin [33, 34]. The periodontal pathogen *Alloprevotella* is reported to be enriched in oral cavity squamous cell cancer [35]. *Parvimonas* is positively associated with the host gene *PARVB*, both of which are enriched in colorectal cancer [36]. *Dialister* has been confirmed to act as a central player in the ecological network of gastric cancer [37]. Based on these findings, whether the four-bacteria signature established in this study has prognostic value in multiple cancers is worthy of further investigation.

From the perspective of the host, we conducted functional enrichment analysis in NPC tumours harbouring a high or low abundance of risk bacteria as defined by the bacterial signature. Patients with abundant risk bacteria were positively correlated with malignant-related pathways involved in proliferation and metastasis. *Helicobacter pylori* infection-associated chronic inflammation is causally involved in inducing DNA methylation alterations in gastric epithelial cells to boost carcinogenesis [38]. In colorectal cancer, tumour-resident *Escherichia coli* can disrupt the gut vascular barrier and disseminate to the liver, accelerating the formation of a premetastatic niche and favouring the recruitment of metastatic cells [39]. However, the roles of these four bacteria in NPC carcinogenesis and progression warrant further exploration. Another important finding is the vigorous metabolism in tumours with abundant risk bacteria, such as the cytochrome P450 pathway. It is worth noting that tissue-resident bacteria have been reported to directly affect the activity of chemotherapy drugs [40]. *Gammaproteobacteria* can metabolise

(See figure on next page.)

Fig. 6 The abundance of the risk bacteria was negatively associated with immune infiltration. **A** Sankey plots show the significantly differential pathways according to malignant property-, metabolism-, and immune-related categories between NPC tumours with high (high-risk group) or low (low-risk group) abundances of risk bacteria based on GSEA. In the bubble plots, the size of the bubbles represents gene numbers in the labelled pathway. The colour of the bubbles represents that the labelled pathway is upregulated in the high-risk (red) or low-risk (blue) groups. The permutation-based *P* value shows the statistical significance of the NES. **B** Network plot shows the Spearman correlations ($P < 0.05$, $|r| > 0.30$) among four risk bacterial genera and significantly differentially expressed immune-related genes. **C** Violin plots show the immune infiltration in the high- or low-risk groups estimated by the MCP-counter algorithm. The comparison was performed with the Wilcoxon rank-sum test. The dots represent the mean scores, and the error bars represent the standard deviation. **D** Representative images of immune cell infiltration in NPC tissues in the high- or low-risk group determined by multiplex immunohistochemistry (mIHC). The 6-colour mIHC images (1 mm), high magnification (100 μ m) of a region of interest, and individual images of DAPI (purple), panCK (pink), CD3 (blue), CD8 (red), CD20 (yellow), and CD56 (green) are shown. **E** Box plots show the comparison of immune cell infiltration in the intratumoural (left) and stromal (right) tissues in the high- and low-risk groups. The comparison was performed with the Wilcoxon rank-sum test. The centreline represents the median, and the bounds represent the 25th and 75th percentiles. The upper whisker extends from the hinge to the largest value no further than 1.5*IQR from the hinge. The lower whisker extends from the hinge to the smallest value of at most 1.5*IQR of the hinge

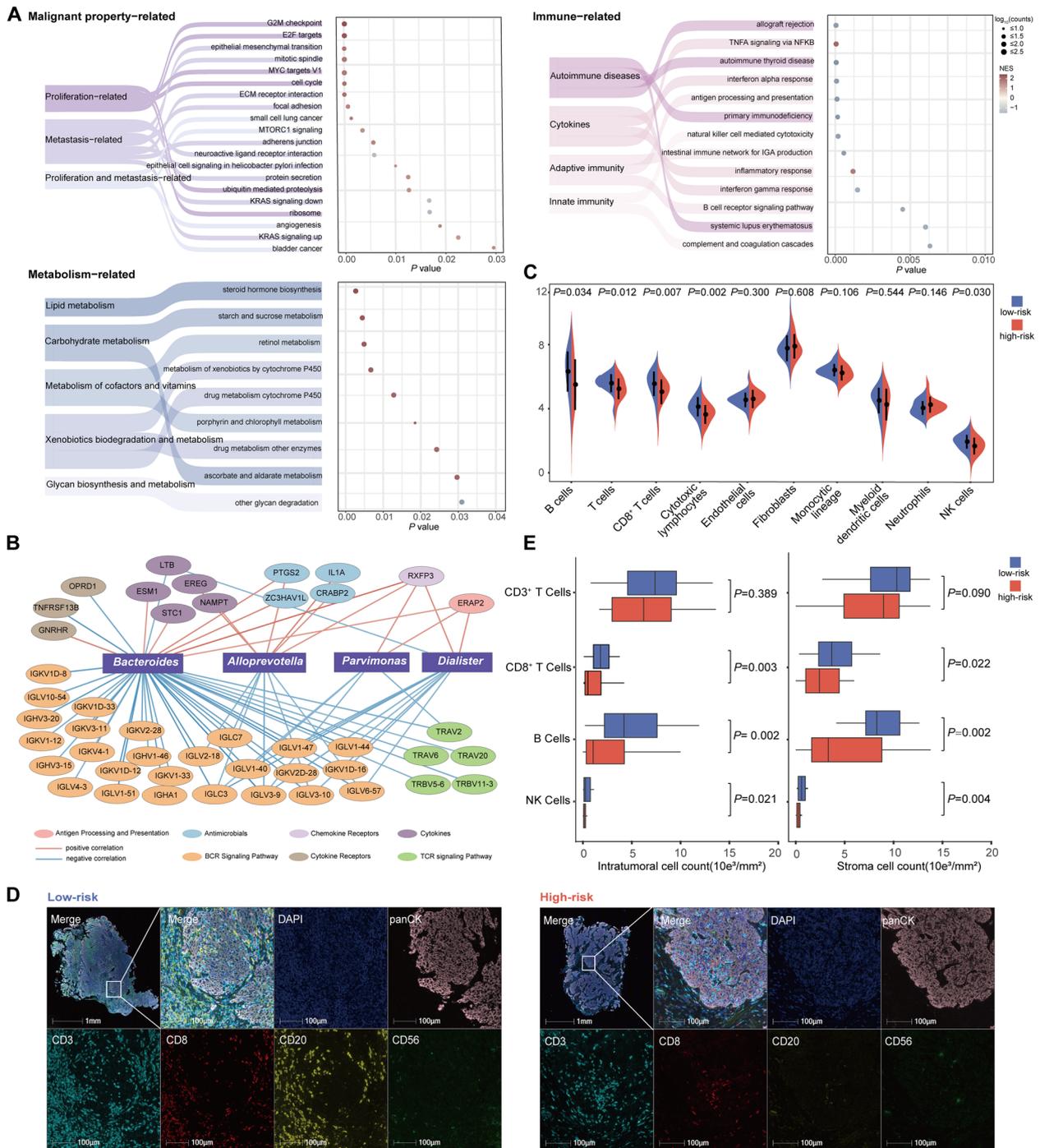


Fig. 6 (See legend on previous page.)

a chemotherapeutic drug, gemcitabine, into its inactive form by expressing a long isoform of the bacterial enzyme cytidine deaminase [41]. Besides, bacterial *preTA* operon is believed to metabolise 5-fluorouracil, intimating the main host mechanism for drug clearance

[42]. These studies suggest that tissue-resident bacteria might be directly involved in the drug metabolism of the host to influence cancer progression.

Of interest is the notion that the intratumoural microbiota are vital components of tumour microenvironment that can reprogram the immune response that ends up

impinging on tumour progression. Surprisingly, the combination of transcriptional analysis and mIHC revealed that the abundance of risk bacteria within NPC tumours was inversely associated with intratumoural and stromal immune infiltration. Consistently, a recent study showed that bacterial communities populate less vascularized and highly immuno-suppressive microniches in bacteria-positive tumour regions through GeoMx digital spatial profiling [43]. Besides, it has been reported that *Fusobacterium nucleatum* can promote tumour growth and metastasis by suppressing the infiltration of T cells in breast cancer [44]. In our study, *Bacteroides* was demonstrated to be the core genus that negatively correlated with multiple immune genes, indicating its unique mechanism involved in interacting with the host immune system, in particular with adaptive immunity. Moreover, it has been reported that antibiotics can influence the efficacy of immune checkpoint therapy in lung cancer [45], suggesting that in combination with bacteria-based therapies and other anti-tumour frontline immune regimens might have a preferable therapeutic effect.

Conclusions

To our knowledge, this is the first study with the largest sample size to uncover the tissue-resident microbial profiles in NPC and normal nasopharyngeal tissues. We developed a bacterial signature for prognostic prediction and explored the potential effect of this microbial shift on the host, which might pave a novel path towards the era of microbiota-targeted precision medicine. We acknowledge that the prognostic value of the four-bacteria signature requires prospective validation before clinical use. Advancements in sequencing technology will enable more in-depth exploration of the species-level bacterial regulatory mechanisms, which will provide novel insights into the complexity of how bacteria participate in NPC progression.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-025-02114-w>.

Additional file 1.

Acknowledgements

Not applicable.

Authors' contributions

Study design: Na Liu. Data collection: Na Liu, Xi-Rong Tan, Han Qiao, Ying-Qing Li, Wei Jiang, Sheng-Yan Huang, Sha Gong, Wen-Fei Li, Ling-Long Tang, Guan-Qun Zhou, Ye-Lin Liang, Hui Li, Qing-Mei He, Jie-Wen Bai, Ming-Liang Ye, Jing-Yun Wang, Sai-Wei Huang, Jun-Yan Li, Chun-Qiao Gan, Ying-Qin Li, Yin Zhao, Ying Sun, Jun Ma. Data analysis and interpretation: Xi-Rong Tan, Han Qiao, Ying-Qing Li, Na Liu. Statistical analysis: Xi-Rong Tan, Han Qiao, Na Liu. Writing of the manuscript: Xi-Rong Tan, Na Liu. Revision of the manuscript: All authors. All authors have reviewed the manuscript and approved the final version.

Funding

This study was supported by grants from the Guangdong Basic and Applied Basic Research Foundation (2022A1515220010); the National Natural Science Foundation of China (82372592, 82461160321); the Guangzhou Science and Technology Project (202206010013); the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (22yk1j07); Science Research and Technology Development Plan of Guilin (20230127-1); and the Young Talents Program of Sun Yat-sen University Cancer Center (YTP-SYSUCC-0010).

Data availability

Key raw data of 16S rRNA sequencing has been uploaded to the Sequence Read Archive (PRJNA962835). The raw RNA-seq data has been deposited at the Gene Expression Omnibus (GSE231572). All raw data has been uploaded to the Research Data Deposit public platform (www.researchdata.org.cn, RDDB2025567870), which is a non-profit medical research data deposit platform constructed by Sun Yat-sen University Cancer Center to ensure the authenticity.

Declarations

Ethics approval and consent to participate

The institutional ethical review boards of both hospitals approved this study for analysing anonymous data, and informed consent was obtained from each patient (B2022-788-01).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Experimental Research, State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, 651 Dongfeng Road East, Guangzhou 510060, People's Republic of China. ²Department of Out-Patient, State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China. ³Department of Radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin, People's Republic of China. ⁴Department of Radiation Oncology, State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China. ⁵Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, People's Republic of China.

Received: 28 November 2023 Accepted: 14 April 2025

Published online: 17 May 2025

References

- Chen YP, Chan ATC, Le QT, et al. Nasopharyngeal carcinoma. *Lancet*. 2019;394(10192):64–80.
- Chen YP, Ismaila N, Chua MLK, et al. Chemotherapy in combination with radiotherapy for definitive-intent treatment of stage II-IVA nasopharyngeal carcinoma: CSCO and ASCO guideline. *J Clin Oncol*. 2021;39(7):840–59.
- Tang LL, Chen YP, Chen CB, et al. The Chinese Society of Clinical Oncology (CSCO) clinical guidelines for the diagnosis and treatment of nasopharyngeal carcinoma. *Cancer Commun (Lond)*. 2021;41(11):1195–227.
- Chen YP, Liu X, Zhou Q, et al. Metronomic capecitabine as adjuvant therapy in locoregionally advanced nasopharyngeal carcinoma: a multi-centre, open-label, parallel-group, randomised, controlled, phase 3 trial. *Lancet*. 2021;398(10297):303–13.

5. Liu N, Chen NY, Cui RX, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol.* 2012;13(6):633–41.
6. Lenoci D, Resteghini C, Serafini MS, et al. Tumor molecular landscape of Epstein-Barr virus (EBV) related nasopharyngeal carcinoma in EBV-endemic and non-endemic areas: Implications for improving treatment modalities. *Transl Res.* 2024;265:1–16.
7. Tang XR, Li YQ, Liang SB, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. *Lancet Oncol.* 2018;19(3):382–93.
8. Lei Y, Li YQ, Jiang W, et al. A gene-expression predictor for efficacy of induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma. *J Natl Cancer Inst.* 2021;113(4):471–80.
9. Dai JH, Tan XR, Qiao H, et al. Emerging clinical relevance of microbiome in cancer: promising biomarkers and therapeutic targets. *Protein Cell.* 2024;15(4):239–60.
10. de Martel C, Ferlay J, Franceschi S, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol.* 2012;13(6):607–15.
11. Lu YQ, Qiao H, Tan XR, et al. Broadening oncological boundaries: the intratumoral microbiota. *Trends Microbiol.* 2024;32(8):807–22.
12. Nejman D, Livyatan I, Fuks G, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science.* 2020;368(6494):973–80.
13. Poore GD, Kopylova E, Zhu Q, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature.* 2020;579(7800):567–74.
14. Hermida LC, Gertz EM, Ruppel E. Predicting cancer prognosis and drug response from the tumor microbiome. *Nat Commun.* 2022;13(1):2896.
15. Qiao H, Li H, Wen X, et al. Multi-omics integration reveals the crucial role of fusobacterium in the inflammatory immune microenvironment in head and neck squamous cell carcinoma. *Microbiol Spectr.* 2022;10(4):e0106822.
16. Ghaddar B, Biswas A, Harris C, et al. Tumor microbiome links cellular programs and immunity in pancreatic cancer. *Cancer Cell.* 2022;40(10):1240–1253.e5.
17. Fu A, Yao B, Dong T, et al. Tumor-resident intracellular microbiota promotes metastatic colonization in breast cancer. *Cell.* 2022;185(8):1356–1372.e26.
18. Qiao H, Tan XR, Li H, et al. Association of intratumoral microbiota with prognosis in patients with nasopharyngeal carcinoma from 2 hospitals in China. *JAMA Oncol.* 2022;8(9):1301–9.
19. Pereira-Marques J, Hout A, Ferreira RM, et al. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front Microbiol.* 2019;10: 1277.
20. Burns MB, Lynch J, Starr TK, et al. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. *Genome Med.* 2015;7(1):55.
21. Wu Y, Jiao N, Zhu R, et al. Identification of microbial markers across populations in early detection of colorectal cancer. *Nat Commun.* 2021;12(1):3063.
22. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17(1):218.
23. Huang SY, Gong S, Zhao Y, et al. PJA1-mediated suppression of pyroptosis as a driver of docetaxel resistance in nasopharyngeal carcinoma. *Nat Commun.* 2024;15(1):5300.
24. Li JY, Zhao Y, Gong S, et al. TRIM21 inhibits irradiation-induced mitochondrial DNA release and impairs antitumor immunity in nasopharyngeal carcinoma tumour models. *Nat Commun.* 2023;14(1):865.
25. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med.* 1997;16(4):385–95.
26. Guo P, Luo Y, Mai G, et al. Gene expression profile based classification models of psoriasis. *Genomics.* 2014;103(1):48–55.
27. Pushalkar S, Hundeyin M, Daley D, et al. The pancreatic cancer microbiome promotes oncogenesis by induction of innate and adaptive immune suppression. *Cancer Discov.* 2018;8(4):403–16.
28. Dai D, Yang Y, Yu J, et al. Interactions between gastric microbiota and metabolites in gastric cancer. *Cell Death Dis.* 2021;12(12):1104.
29. Fu A, Yao B, Dong T, et al. Emerging roles of intratumor microbiota in cancer metastasis. *Trends Cell Biol.* 2023;33(7):583–93.
30. Vogl T, Klompus S, Leviatan S, et al. Population-wide diversity and stability of serum antibody epitope repertoires against human microbiota. *Nat Med.* 2021;27(8):1442–50.
31. Liu NN, Jiao N, Tan JC, et al. Multi-kingdom microbiota analyses identify bacterial-fungal interactions and biomarkers of colorectal cancer across cohorts. *Nat Microbiol.* 2022;7(2):238–50.
32. Rao C, Coyte KZ, Bainter W, et al. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature.* 2021;591(7851):633–8.
33. Chung L, Thiele Orberg E, Geis AL, et al. *Bacteroides fragilis* toxin coordinates a pro-carcinogenic inflammatory cascade via targeting of colonic epithelial cells. *Cell Host Microbe.* 2018;23(2):203–214.e5.
34. Parida S, Wu S, Siddharth S, et al. A procarcinogenic colon microbe promotes breast tumorigenesis and metastatic progression and concomitantly activates notch and β -catenin axes. *Cancer Discov.* 2021;11(5):1138–57.
35. Ganly I, Yang L, Giese RA, et al. Periodontal pathogens are a risk factor of oral cavity squamous cell carcinoma, independent of tobacco and alcohol and human papillomavirus. *Int J Cancer.* 2019;145(3):775–84.
36. Liu X, Tong X, Zhu J, et al. Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. *Cell Discov.* 2021;7(1):117.
37. Coker OO, Dai Z, Nie Y, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. *Gut.* 2018;67(6):1024–32.
38. Niwa T, Tsukamoto T, Toyoda T, et al. Inflammatory processes triggered by *Helicobacter pylori* infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Res.* 2010;70(4):1430–40.
39. Bertocchi A, Carloni S, Ravenda PS, et al. Gut vascular barrier impairment leads to intestinal bacteria dissemination and colorectal cancer metastasis to liver. *Cancer Cell.* 2021;39(5):708–724.e11.
40. Duan YF, Dai JH, Lu YQ, et al. Disentangling the molecular mystery of tumour-microbiota interactions: microbial metabolites. *Clin Transl Med.* 2024;14(11):e70093.
41. Geller LT, Barzily-Rokni M, Danino T, et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science.* 2017;357(6356):1156–60.
42. Spanogiannopoulos P, Kyaw TS, Guthrie BGH, et al. Host and gut bacteria share metabolic pathways for anti-cancer drug metabolism. *Nat Microbiol.* 2022;7(10):1605–20.
43. Galeano Niño JL, Wu H, LaCourse KD, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature.* 2022;611(7937):810–7.
44. Parhi L, Alon-Maimon T, Sol A, et al. Breast cancer colonization by *Fusobacterium nucleatum* accelerates tumor growth and metastatic progression. *Nat Commun.* 2020;11(1):3259.
45. Le Noci V, Guglielmetti S, Arioli S, et al. Modulation of pulmonary microbiota by antibiotic or probiotic aerosol therapy: a strategy to promote immunosurveillance against lung metastases. *Cell Rep.* 2018;24(13):3528–38.

Publisher's Note

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