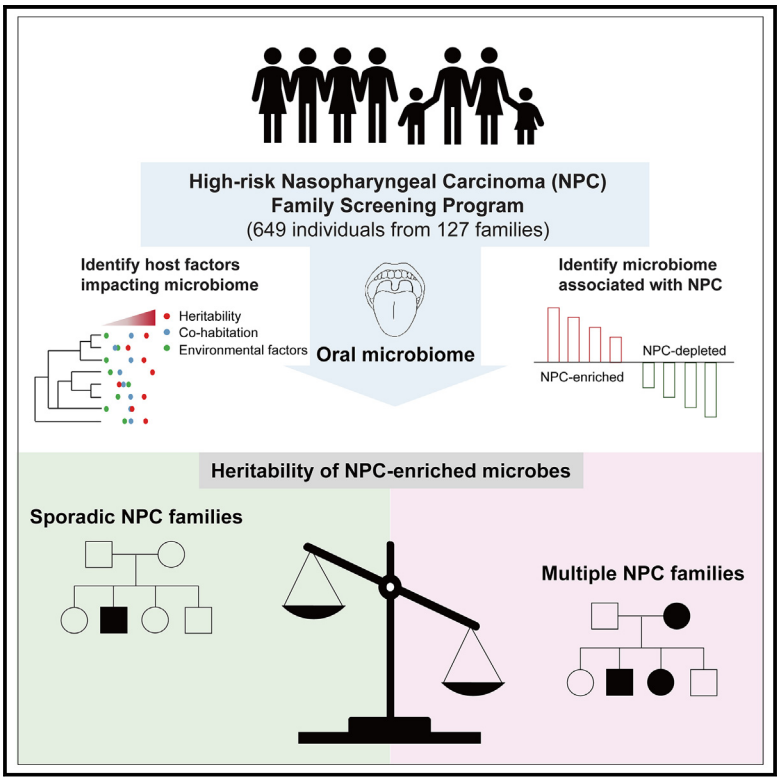
Cell Reports
Medicine

Article

Unveiling familial aggregation of nasopharyngeal carcinoma: Insights from oral microbiome dysbiosis

Graphical abstract



Authors

Ying Liao, Xia-Ting Tong, Ting Zhou, ..., Ming-Yuan Chen, Na Liu, Wei-Hua Jia

Correspondence

jiawh@sysucc.org.cn

In brief

Familial aggregation is frequent in nasopharyngeal carcinoma (NPC), but its cause remains unclear. Liao et al. investigate the oral microbiome in NPC family cohorts, revealing strong heritability effects on microbiome and identifying NPC-associated microbes with markedly higher heritability in families exhibiting increased clustering of NPC cases.

Highlights

- Family members share similar oral microbiome in NPC family cohorts
- Heritability strongly affects the microbial composition
- NPC-enriched microbes show higher heritability in NPC-clustered families
- Heritable microbes form closely interacting microbial networks

Liao et al., 2025, Cell Reports Medicine 6, 101979
March 18, 2025
https://doi.org/10.1016/j.xcrm.2025.101979

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Unveiling familial aggregation of nasopharyngeal carcinoma: Insights from oral microbiome dysbiosis

Ying Liao,[1] Xia-Ting Tong,[2] Ting Zhou,[1] Wen-Qiong Xue,[1] Tong-Min Wang,[1] Yong-Qiao He,[1] Mei-Qi Zheng,[1] Yi-Jing Jia,[2] Da-Wei Yang,[2] Yan-Xia Wu,[1] Xiao-Hui Zheng,[1] Zhi-Xiang Zuo,[1] Ming-Yuan Chen,[1,3] Na Liu,[1] and Wei-Hua Jia[1,2,4,*]

[1]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, P.R. China
[2]School of Public Health, Sun Yat-sen University, Guangzhou 510080, P.R. China
[3]Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, Guangzhou 510060, P.R. China
[4]Lead contact
*Correspondence: jiawh@sysucc.org.cn
https://doi.org/10.1016/j.xcrm.2025.101979

SUMMARY

Familial aggregation is common in nasopharyngeal carcinoma (NPC), yet the impact of oral microbiome dysbiosis on this occurrence remains largely unexplored. We recruit 127 families (649 members, 1–5 patients each) and a case-control cohort of 337 individuals, validating findings in an additional cohort of 995 individuals. Significant microbial similarity is observed among family members, with family factors contributing most to microbiome variation, followed by cigarette smoking, age, and gender. Among multi-NPC families, especially those with three or more patients, we identify three NPC-enriched taxa with notable heritability, including *Gemella* sp. (heritability, h^2 = 53.1%), *Lautropia mirabilis* (h^2 = 38.8%), and *Streptococcus* sp. (h^2 = 38.0%). Heritable bacteria present a markedly higher heritability in families with increased clustering of NPC and form closely interacting networks, suggesting their role in NPC familial aggregation. These findings open up possibilities for identifying high-risk individuals, enhancing clinical surveillance, and developing personalized prevention and treatment approaches of NPC through microbiome-based strategies.

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is an epithelial malignancy originating from the nasopharynx, with an estimated 120,000 new cases worldwide in 2022,[1] and the majority of cases were reported in Southern China and Southeastern Asia.[2] Family aggregation is common in NPC, with about 10% of cases having a family history.[3,4] Despite significant progress in uncovering genetic and environmental factors associated with familial NPC aggregation,[4–8] its etiology remains largely unknown.

The human microbiome offers insights into the occurrence of familial cancers, as microbiota can influence cancer susceptibility and progression through various mechanisms, including inflammation modulation, DNA damage induction, and metabolite production involved in oncogenesis or tumor suppression.[9–11] The family microbiomes often exhibit shared patterns probably due to the genetic backgrounds, lifestyle habits, and living environments, potentially leading to shared disease-associated microbial alterations among family members.[12,13]

Understanding of the oral microbiome, the second-largest bacterial reservoir in the human body, has advanced significantly,[14] with several oral periodontal disease-related microbes linked to an elevated risk of cancers beyond the oral cavity, including the esophagus, stomach, pancreas, and colon,[15] emphasizing its critical role in tumorigenesis. Previous investigations by our team have identified distinctive oral microbial profiles indicative of NPC[16] and uncovered the pathogenic impact of oral microbiome translocation to the nasopharynx by microbial tumor infiltration and microenvironment remodeling[17], illuminating the connections between the oral microbiome and NPC development.

In this study, we delved into a well-established multi-NPC family cohort, intensifying our focus on the oral microbiome and its interplay with NPC familial aggregation. By investigating the oral microbiome of members from families with NPC patients, we aimed to explore the familial microbiome profiles and uncover the influence of host genetics, cohabitation, and other environmental factors in shaping the oral microbiome. Additionally, by integrating data from the case-control cohorts, we sought to investigate the specific microbiome patterns shared among multiple NPC families, potentially contributing to NPC development. This study will provide insights into the intricate relationship between host-microbiome interactions in familial aggregation of NPC.

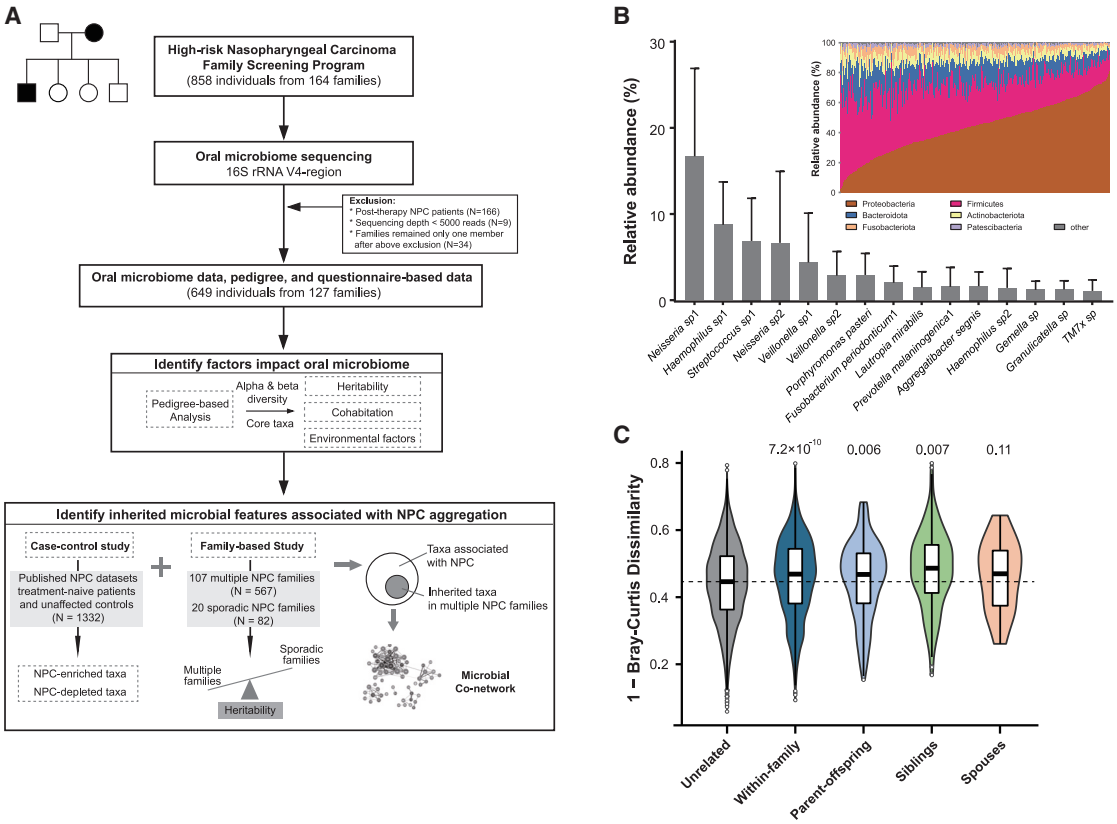


Figure 1. Microbiome similarity analysis revealed familial similarity

(A) Flow diagram outlining the study design.

(B) The stacked bar plots for the top six abundant phyla and top 15 abundant ASV taxa. Relative abundances of each taxon are shown (mean ± SEM).

(C) Violin plots showing the microbial composition similarity within family members using "1 – Bray-Curtis dissimilarity" index. There were 2,192 within-family pairs. A dashed line indicates the median of the "1 – Bray-Curtis dissimilarity" of the 6,576 matched unrelated pairs. First-pedigree pairs, including parent-offspring pairs ($N = 273$) and sibling-sibling pairs ($N = 278$), are compared, along with spouse pairs ($N = 36$). p values were evaluated by t test (two-sided). See also in Figure S1.

RESULTS

Overview of the oral microbiome data

We conducted the oral microbiome study using saliva samples from a high-risk NPC family cohort,[4,18] including 649 individuals from 127 well-characterized families (Figure 1A). Among them, 107 families (84.3%) had two or more NPC cases who met the pedigree of the first- or second-degree relatives. The mean age was 35.5 ± 17.0 years (mean ± SD). Other detailed information on demographic and environmental factors is provided in Table S1. Family pedigree information of representative families is shown in Figure S1.

An average of 24,576 ± 10,315 (mean ± SD) quality-filtered reads were obtained per sample. Using the DADA2 processing pipeline, 940 amplicon sequence variants (ASVs) were generated after filtering low-prevalence (<1%) and low-frequency reads (<20 reads). The remaining ASVs were mainly assigned to three main phyla: *Proteobacteria* (43.5% ± 17%, mean ± SD of the relative abundance), *Firmicutes* (25% ± 11.7%), and *Bacteroidetes* (15.9% ± 7.7%) (Figure 1B). Among them, 15 ASVs had an average relative abundance higher than 1%. These dominant ASVs were assigned to genera *Neisseria*, *Haemophilus*, *Streptococcus*, *Veillonella*, etc. (Figure 1B).

Family relatives shared more similar oral microbiome profiles

To assess the impact of family pedigree on the human oral microbiome, we compared the similarity of oral microbial profiles across different kinships. Our family cohort constructed a total of 232,903 pairs, including within-family pairs ($N = 2,192$) and unrelated pairs from individuals in different families ($N = 230,711$). Within the family pairs, there were 273 parent-offspring pairs, 278 sibling-sibling pairs, and 36 spouse pairs. The similarity index represented by "1 – Bray-Curtis dissimilarity" was defined as an index to represent the similarity of the microbial community, with absolutely identical communities when scored at 1 and completely different communities when scored at 0. Results showed that this index was significantly higher in within-family pairs when compared with unrelated pairs ($t = 6.18$, $p = 7.2 \times 10^{-10}$, Figure 1C). Specifically, first-relative pairs exhibited significantly higher similarity than matched unrelated pairs, observed in both parent-child and sibling-sibling pairs

(p = 0.006 and 0.007, respectively, Figure 1C). While spouse pairs showed a similar tendency, the difference was not statistically significant (p = 0.11, Figure 1C).

Host genetics considerably influences the overall composition of the oral microbiome

Family members share genetic backgrounds, cohabitation, and often similar lifestyles, and all these factors may shape the microbiome composition among individuals. To identify the factors that impacted the oral microbiome, we examined microbial differences across the 27 variables using permutational multivariate analysis of variance (PERMANOVA). Variables included the families where individuals were from and environmental factors including demographics, lifestyles, disease history, etc. (Table S1). Interestingly, the family factor was the top contributor to the variation of oral microbiome composition (Figure 2A, R^2 = 28.8% in Bray-Curtis (BC) dissimilarity, false discovery rate [FDR]-q = 0.003). Cigarette smoking ranked as the second most contributing factor (Figure 2A, R^2 = 1.7% in BC dissimilarity, FDR q = 0.003), followed by age, gender, and other 10 factors (Table S2). Consistent findings were observed using weighted and unweighted UniFrac (UF) metric (Table S2), highlighting the impact of family factors on the oral microbiome composition. Considering that excessive grouping of family factors may affect the estimation, we performed a focused analysis on four large families with $\geq$15 attendance (N = 66), reaffirming the family factor as the primary contributor (R^2 = 6.4%, p = 0.024).

To further evaluate the inheritance impact on oral microbiome profiles, we utilized a variance component model, the Sequential Oligogenic Linkage Analysis Routines (SOLAR), to estimate the heritability from family pedigrees. At the microbial composition levels, both alpha diversity indices (observed ASVs, Shannon, and Simpson index) and the top three principal coordinates (PCo) variance from beta diversity metrics (UF, weighted UniFrac [wUF], and BC dissimilarity) were evaluated. Results showed that heritability (h^2) estimates were generally higher than cohabitation (c^2) and unshared environmental factors (e^2) (Figure 2B), notably explaining significant contributions to the variances of the Shannon and Simpson alpha diversity indices (h^2 = 22%, p = 0.006; h^2 = 17%, p = 0.021), as well as nine PCo1–PCo3 from three beta diversity metrics, e.g., PCo1 and PCo2 of UF dissimilarity (h^2 = 26%, p < 0.001; h^2 = 40%, p < 0.001) and PCo2 of BC dissimilarity (h^2 = 35%, p < 0.001). Effects explained by cohabitation and environmental factors were relatively weak (Figure 2B). Detailed information on the heritability estimates is shown in Table S3.

At the ASV level, the heritability was evaluated for 275 core ASVs present in at least 10% of all individuals with a mean relative abundance higher than 0.01%. Among them, 48.3% (N = 133) of taxa reached a significant level for heritability estimates, compared to only 20.0% (N = 55) for cohabitation and 5.1% (N = 14) for non-shared environmental effect (Table S4). Consistent with the aforementioned findings, host genetic effect (h^2 = 10.2% ± 10.1%, mean ± SD) predominantly explained microbial variations, followed by cohabitant effect (c^2 = 5.0% ± 6.8%) and non-shared environmental factors (e^2 = 2.8% ± 3.2%). Moreover, 70.4% of taxa exhibited dominant effects of heritability (red in the phylogenetic tree in Figure 2C), while taxa primarily influenced by cohabitation or environmental factors were labeled blue (26.3%) or green (3.3%) in the phylogenetic tree (Figure 2C). The results of heritability estimates of different levels of taxonomies are also provided in Table S5.

Heritable oral taxa linked to NPC risk

We next ask whether specific shared familial microbiome features contribute to NPC development. To identify NPC-associated microbiome features, we combined the oral microbiome data from pre-treatment patients and matched controls in this cohort and one case-control population from our prior study.[16] Both datasets adhered to the same sampling procedures, nucleic acid processing, and sequencing methods, ensuring comparable sequencing data from positive standards. By analyzing oral microbiome datasets of 167 NPC cases and 170 healthy controls, which were well balanced for age, gender, and cigarette smoking (Table S6), we identified 34 differential abundance taxa (Figure 3A and Table S7).

Microbes associated with NPC might be attributed to various factors such as host genetics, environment, and disease progression. To further identify the heritable microbial features linked to NPC familial aggregation, the families were subdivided into multiple NPC and sporadic NPC groups based on whether the families contained two or more NPC cases meeting the pedigree of the first- or second-degree relatives. In the multiple NPC family group, the mean variance contributing to heritability (h^2) of these 34 NPC-associated ASVs was 17.2% ± 13.0% (mean ± SD), significantly higher than contributions from cohabitation and non-shared environmental factors, which averaged 5.0% ± 6.3% and 4.4% ± 3.8%, respectively (Table S7). Among the 34 NPC-associated ASVs, eight showed high heritability with h^2 > 30% in multiple NPC families (FDR q < 0.01 in SOLAR analysis, labeled "#" in Figure 3A), all enriched in NPC patients, including ASV13_*Gemella* sp. (h^2 = 39.8%), ASV19_*Fusobacterium nucleatum.1* (h^2 = 38.9%), ASV43_*Prevotella salivae* (h^2 = 38.1%), ASV5_*Veillonella* sp.*1* (h^2 = 34.6%), ASV134_*Mannheimia* sp. (h^2 = 32.7%), ASV34_*Megasphaera micronuciformis* (h^2 = 31.7%), ASV9_*Lautropia mirabilis* (h^2 = 31.1%), and ASV40_*Streptococcus* sp.*4* (h^2 = 30.1%). At the genus level, three genera, including *Gemella*, *Megasphaera*, and *Lautropia*, showed high heritability (Figure S2) and were enriched in the NPC group in the multiple NPC families. Notably, these taxa did not show a significant heritability in the sporadic NPC family group (Figure S2).

Furthermore, we performed a subgroup heritability analysis of eight NPC-risk heritable taxa across sporadic families (Num = 1), multiple families with two NPC patients (Num = 2), and multiple families with three or more NPC patients (Num = 3). The analysis revealed that seven of the taxa exhibited a higher heritability in families with three or more affected members, when compared to sporadic or two-case families (Figure 3B and Table S8). These heritable bacteria presented a markedly higher heritability in families with increased clustering of NPC, suggesting their association with NPC familial aggregation.

To further validate the association between these identified heritable taxa and NPC, we accessed published datasets comprising oral microbiome profiles from 499 NPC patients

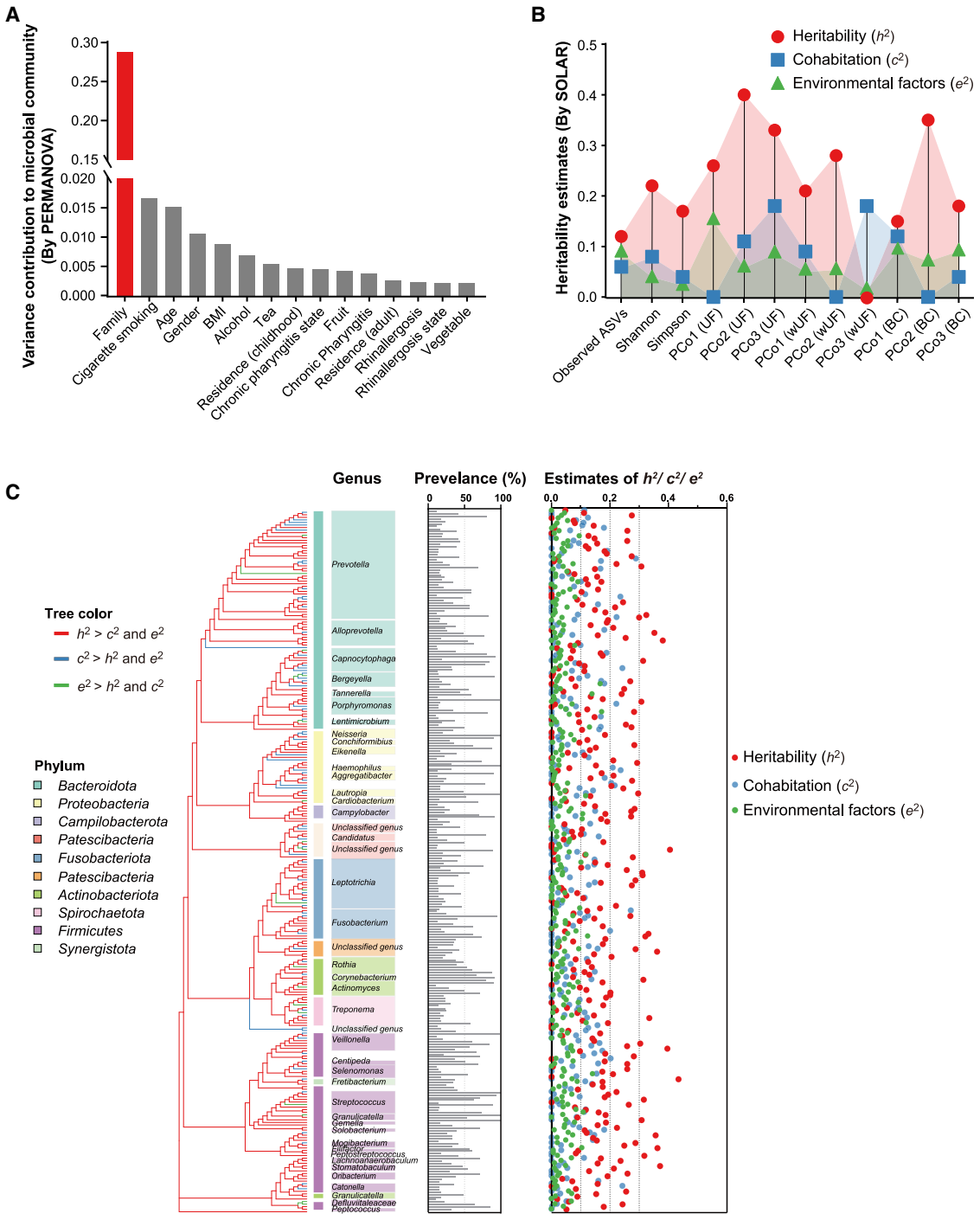


Figure 2. Oral microbiome variability and heritability in NPC families

(A) Bar plots showing variances that affected the oral microbiome community. The y axis was the R^2 estimated by PERMANOVA analysis based on Bray-Curtis metric. Only the variances with significant FDR q values ($q < 0.05$) were presented.

(B) Dot plots illustrating the proportion of microbiome diversity explained by heritability (h^2), cohabitation (c^2), and environmental factors (e^2). Alpha diversity indices (observed ASVs, Shannon, and Simpson) and the first three principal coordinates (PCo) of beta diversity dissimilarity. (UF, UniFrac; wUF, weighted UniFrac; and BC, Bray-Curtis) were evaluated.

(C) Heritability estimates of 275 core oral taxa depicted on a rooted phylogenetic tree with edges randomly split (rtree function of ape package in R). The prevalence rate, estimated heritability (h^2), cohabitation (c^2), and environmental factor (e^2) effects for each of the 275 taxa were shown. Heritability estimates were evaluated using SOLAR analysis. The branches of the phylogenetic tree were colored according to the factor with predominant effects in heritability analysis (red for heritability, blue for cohabitation, and green for environmental factors). See also in Table S1–S5.

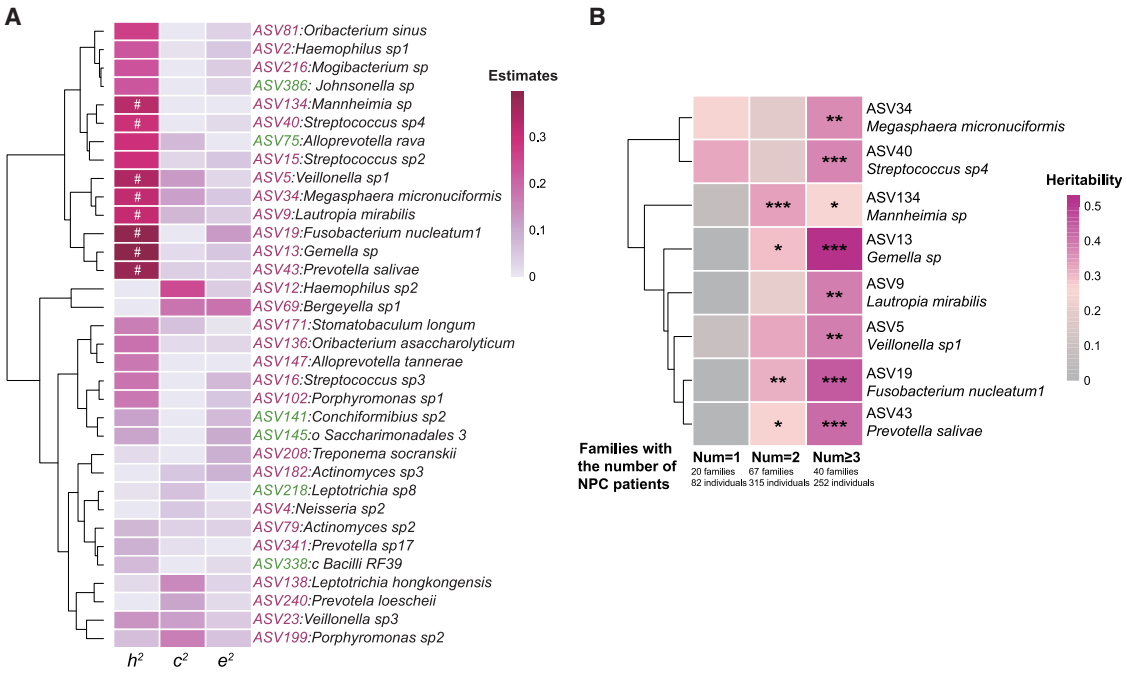


Figure 3. Identification of heritable microbiome features associated with NPC

(A) The significant differential abundant oral ASVs between NPC patients and controls, with the impact of genetics (h^2), cohabitation (c^2), and environmental factors (e^2) on these taxa in multiple NPC families. 34 NPC-associated oral ASVs identified by DESeq (differential expression analysis for sequence count data) analysis with adjusting age, gender, and cigarette smoking status ($p < 0.05$) were shown. ASVs with $h^2 > 30\%$ and FDR $q < 0.01$ in the SOLAR analysis were labeled with a "#," indicating NPC-risk heritable bacteria. NPC-enriched taxa were labeled in red, and NPC-depleted taxa were labeled in green.

(B) The heritability (h^2) of NPC-risk heritable taxa in sporadic families (Num = 1), multiple families with two NPC patients (Num = 2), and multiple families with three or more NPC patients (Num = 3). The SOLAR analysis was used for heritability (h^2) estimation and p value calculation (two-sided).

p values were obtained by t test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. See also in Figures S2–S4 and Tables S6–S9.

and 496 healthy controls.[19] In this independent cohort, we confirmed that the high-heritability bacteria identified in multiple NPC families indeed represent significant risk factors for NPC. Specifically, at the ASV level, three of the heritable ASVs were significantly enriched in NPC patients, including ASV13_*Gemella* sp. ($p = 4.0 \times 10^{-11}$), ASV9_*Lautropia mirabilis* ($p = 7.0 \times 10^{-4}$), and ASV40_*Streptococcus* sp.*4* ($p = 1.1 \times 10^{-3}$) (Table 1 and Figure S3). The heritability of ASV13_*Gemella* sp. in family groups with one, two, three or more NPC patients was 0%, 29.2%, and 53.1%, respectively; similarly, ASV9_*Lautropia mirabilis* had the heritabilities of 0.9%, 20.2%, and 38.8%, respectively (Table 1). Additionally, we identified that two more taxa, ASV193_*Treponema* sp*4* and ASV142_*Leptotrichia hofstadii*, were enriched in multiple NPC cases when compared to sporadic cases and demonstrated higher heritability in multiple families (Figure S4 and Table S9).

The co-network patterns of oral microbiome in multiple NPC families

In addition, we investigated the microbial ecological connections among individuals within multiple NPC families. The co-network included 61 taxa showing significant correlations with other taxa using SparCC, visually depicted in Figure 4 by filling the taxa nodes with colors according to the estimated values of heritability (Figure 4A), cohabitation (Figure 4B), and environmental factors (Figure 4C). Co-network showed that the taxa related to NPC tend to co-occur into hubs (node connected to other nodes) along with high heritability (Figure 4A). The heritability (h^2) values were significantly positively correlated with the degree of connectivity in the networks (Spearman's $r = 0.41$, $p < 0.01$, Figure 4D). Specifically, among eight heritable and NPC-enriched taxa, seven formed three clusters. Notably, ASV34_*Megasphaera micronuciformis*, ASV43_*Prevotella salivae*, and ASV5_*Veillonella* sp1, emerged as the top three highly connected nodes, forming hubs connecting with other 26 taxa (including five NPC-enriched taxa), represented the high correlations with each other (ASV5∼ASV34: $r = 0.66$, ASV43∼ASV5: $r = 0.60$, ASV43∼ASV34: $r = 0.60$). Additionally, ASV19_*Fusobacterium nucleatum*, a well-known periodontopathic bacteria, served as a hub of another cluster, connecting with other four periodontopathic bacteria (ASV18_*Porphyromonas gingivalis*, ASV31_*Porphyromonas endodontalis*, ASV89_*Filifactor alocis*, and ASV94_*Tannerella forsythia*). Interestingly, we did not observe the association between taxa connectivity degree and cohabitation (Figure 4E) or environmental factors (Figure 4F).

The associations between HLA genotypes and heritable microbiome

Recognizing the important roles of human leukocyte antigens (HLAs) in regulating the human microbiome, we

Table 1. Identified taxa associated with NPC familial aggregation, their relative abundance in case-control cohorts, and heritability in family cohort

Taxa	Relative abundance (%) in case-control cohort						Heritability (%) in family cohort[a]					
	Case-control cohort (N = 337)			Validation case-control cohort (N = 995)			Sporadic families Num = 1 (N = 82)			Multiple families Num = 2 (N = 315)			Multiple families Num ≥ 3 (N = 252)	
	Case group mean (SD)	Control group mean (SD)	p values[b]	Case group mean (SD)	Control group mean (SD)	p values[c]	Heritability ± SE	p values[d]	Heritability ± SE	p values[d]	Heritability ± SE	p values[d]
ASV9 *Lautropia mirabilis*	1.46 (2.36)	0.89 (1.15)	3.61×10^{-8}	0.94 (1.58)	0.65 (1.05)	6.99×10^{-4}	0.9 ± 31.0	0.488	20.2 ± 14.4	0.070	38.8 ± 16.8	4.15×10^{-3}
ASV13 *Gemella* sp	0.71 (0.71)	0.64 (0.50)	1.12×10^{-5}	2.42 (2.50)	1.54 (1.55)	4.01×10^{-11}	0.0 ± 0.0	0.500	29.2 ± 15.5	0.027	53.1 ± 16.8	2.76×10^{-4}
ASV40 *Streptococcus* sp4	0.70 (0.67)	0.65 (0.55)	1.34×10^{-3}	1.28 (1.77)	0.96 (1.42)	1.05×10^{-3}	32.7 ± 25.4	0.081	17.2 ± 14.6	0.120	38.0 ± 12.9	2.79×10^{-4}

[a]The heritability (h^2) of NPC-risk heritable taxa was evaluated in the groups of sporadic families (Num = 1), multiple families with two NPC patients (Num = 2), and multiple families with three or more NPC patients (Num = 3).
[b]p values were obtained from DESeq analysis.
[c]p values were obtained from t test.
[d]Heritability analysis was performed by SOLAR.

investigated the associations between HLA genotypes and eight NPC-enriched heritable oral microbiome taxa. This analysis incorporated three lead HLA SNPs with the strongest associations with NPC, as reported by our teams[20]: rs2860580 in the HLA-A gene, rs2894207 in the HLA-B/C gene, and rs28421666 in the HLA-DQ/DR gene. Our findings indicated potential impacts of HLA genotypes on oral microbiome composition. Specifically, the abundance of heritable taxa such as ASV34_*Megasphaera micronuciformis* and ASV43_*Prevotella salivae* showed significant correlations with the rs28421666 genotype ($p < 0.05$, Figure S5). These results suggest interactions between HLA variants and inherited oral microbiome, indicating the need for further comprehensive investigation.

DISCUSSION

This study systematically elucidated the oral microbiome profiles in NPC-affected families recruited in South China, uncovering both inheritable and environmental influences while evaluating the potential contributions of shared microbial signatures to NPC development. Our results revealed the substantial influence of host genetics on the oral microbiome among family members, identifying several taxa, including *Gemella* sp., that were significantly heritable in multi-case families and associated with elevated NPC risk. Moreover, we uncovered the co-occurrence networks of these taxa susceptible to NPC aggregation, providing new insights into the etiology of familial aggregation of NPC from a human microbiome perspective.

While multiple studies have aimed to delineate the susceptibility genes and other risk factors contributing to the familial aggregation of NPC,[4,21,22] the underlying mechanisms remain elusive. The human microbiome emerges as a potential avenue for comprehending this phenomenon. Notably, both the human oral and gut microbiomes exhibit the potential for horizontal and vertical transmission across successive family generations,[23,24] facilitating the spread of pathogenic microbes. Furthermore, shared genetic backgrounds, analogous lifestyles, and communal habitats are recognized as factors driving familial disease clustering and significantly impacting the human microbiome.[25–28] The accumulation of these disease-associated factors within families may propel the microbiome into a dysbiosis state. Additionally, given the microbiome's pivotal role in the carcinogenic process, its interaction with other familial cumulative factors may contribute to the risk of disease occurrence.

In this study, we identified three taxa that susceptibly contributed to NPC familial aggregation, including *Gemella* sp., *Lautropia mirabilis*, and *Streptococcus* sp. *Gemella* spp., commonly found in the oral cavity, has previously been reported as a disruptor in periodontal disease. Notably, our previous study revealed that *Gemella* spp., including *Gemella haemolysans* and *Gemella morbillorum*, are enriched in the nasopharynx of NPC patients and probably are attributed to the ectopic migration from the oral microbiome.[17] A recent study documented the significant enrichment of *Gemella morbillorum* in the tumor microenvironment of oral cavity squamous carcinoma.[29] Furthermore, patients experiencing bacteremia from *Gemella morbillorum* were found to have a 15-fold increased risk of colorectal carcinoma, further supporting its association with cancer.[30] *Lautropia*

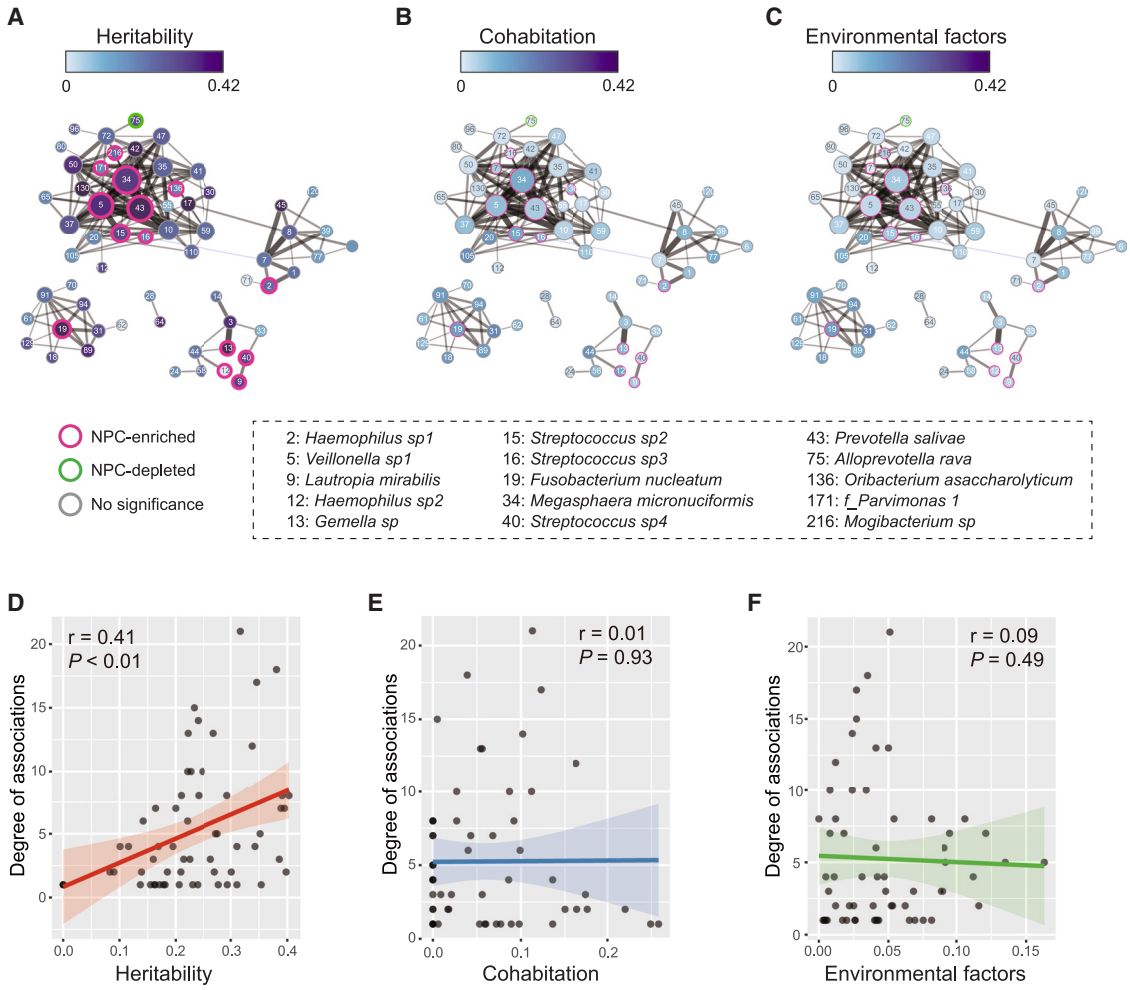


Figure 4. The microbiome occurrence patterns revealed the association between heritability and ecological connections of oral taxa

(A–C) Microbial co-networks were assessed using SparCC analysis (two-sided). Taxa meeting the criteria of $|r| > 0.35$ and p value < 0.05 were included. Each node represented one oral taxon, and each edge represented one significant association between two connected nodes. The stroke color of the node indicated the association direction with NPC phenotype, with pink representing NPC-enriched taxa, green representing NPC-depleted taxa, and gray indicating non-significance taxa. Nodes were separately colored based on the levels of heritability effects (A), cohabitation effects (B), and environmental factors effects (C).

(D–F) The scatterplots with linear correlation curves depict the degree of association and effects of heritability effects (D), cohabitation (E), and environmental factors (F). Cytoscape software was used to evaluate the degree of association in the co-networks, and Spearman analysis (two-sided) was conducted to assess the correlation effects.

mirabilis is a facultative anaerobic coccus isolated in the respiratory sites of immunodeficient patients, which was identified as the most significant abundant oral microbe that was enriched in the NPC patients in our previous study.[16] *Streptococcus* sp. was annotated as *Streptococcus sanguinis* in this study, and our previous study revealed that oral *S. sanguinis* induced the reactivation of Epstein-Barr virus (EBV) by its metabolite hydrogen peroxide and then participates in NPC tumorigenesis.[16] Although the validation of *Fusobacterium nucleatum* remains somewhat controversial, it has demonstrated high heritability in multiple families with three or more cases of NPC (h^2 = 44.7%). Our previous study also found *F. nucleatum* to be enriched in the nasopharynx of NPC patients, affecting the local microenvironment and modulating cytokine responses.[17] Moreover, a recent study reported that *F. nucleatum* promotes radioresistance in NPC by reducing PANoptosis via the SLC7A5/leucine-mTORC1 axis.[31] Given its role as an oncogenic microorganism in various cancers,[32] the pathogenic contribution of *F. nucleatum* to NPC familial aggregation deserves further investigation. These findings collectively underscore the carcinogenic potential of these heritable microbes, particularly in the context of NPC.

In this study, we observed that these heritable bacteria formed hubs in a co-occurrence network. *Streptococcus* species, as early colonizers, is one of the most abundant species in the oral biofilm.[33] Middle and late colonizers, such and *Fusobacterium nucleatum* and *Veillonella atypica*, act as "bridging species," facilitating colonization.[34] The oral biofilm is a complex multispecies community, where antagonism and mutualism interactions maintain ecological balance. The co-occurrence

of these heritable microbes might enhance oral biofilm colonization, potentially promoting microbial communication within family members and exacerbating NPC pathogenesis. Additionally, highly heritable taxa, including *Gemella* spp., *Lautropia mirabilis*, *Fusobacterium nucleatum*, and *Veillonella* spp., have been identified as the oral-to-nasopharyngeal translocation microbes, as reported in our recent work.[17] These microbes of oral origin were enriched in the nasopharyngeal mucosa of NPC patients and infiltrated into NPC tumors, resulting in an increased number of intratumoral lymphocytes, and were closely associated with EBV burden in the nasopharynx.[17] This oral-to-nasopharyngeal translocation might be one of the mechanisms by which these heritable microbes contribute to familial NPC development.

Microbiome-based interventions have advanced rapidly in recent years. This study identified several heritable microbes linked to familial aggregation of NPC, paving the way for potential treatments such as probiotics, prebiotics, or antibiotic therapies to target these key bacteria. Our findings also highlight the tight network of heritable taxa and their potential role in oral biofilm colonization. Strategies for disrupting oral biofilms, including dental scaling and biomaterial-assisted treatments, represent another promising approach to managing the oral load of these bacteria. Given the association between oral hygiene and NPC risk,[35] improving oral health education and preventive measures for relatives in multiple NPC-affected families is crucial.

Previous studies have reported that environmental factors such as wood smoke exposure and salted fish consumption during childhood are associated with an elevated risk of early-onset NPC in multiple-case NPC families.[36] These environmental factors might also contribute to microbiome dysbiosis.[37,38] However, a considerable proportion of variability in our study remains unexplained, likely due to the algorithms focusing only on documented environmental factors, thereby underestimating the effects of unknown factors.

Family-based study designs have distinct advantages in distinguishing plausibly genetic from nongenetic sources of familial correlation. In this study, we observed strong microbial similarity within the multiple NPC families, with heritability emerging as the primary source of NPC familial correlation with the oral microbiome. We identified significant associations between the rs28421666 genotype in the *HLA-DQ/DR* gene and specific heritable taxa in multiple NPC families, such as *Megasphaera micronuciformis* and *Prevotella salivae*. Previous genetic studies have pinpointed the immune-related genes, particularly those within the HLA region, as susceptible causal genes for NPC.[20,39] These genes also contribute significantly to the familial clustering of NPC.[4] The HLA system determines the specificity of T lymphocyte and natural killer cell responses to host commensal bacteria, profoundly influencing host-microbiome composition.[40,41] Recent studies have demonstrated that individuals sharing similar HLA genes tend to have similar microbiota compositions,[42] with specific HLA genotypes, such as *HLA-DQ2*, shaping early gut microbiota composition and impacting disease susceptibility.[40] Our findings suggest possible interactions between HLA variants and the inherited oral microbiome, warranting further investigation into how much these HLA genotypes contribute to microbiome heritability and their role in NPC susceptibility.

Emerging research has also illuminated the host regulation of the human microbiome through epigenetic alterations and somatic mutations.[43,44] To examine whether the heritability of identified bacteria is attributable to the predominant methylation and mutations within tumor tissues, we investigated the genetic and epigenetic landscapes using 24 Infinium HumanMethylation450 BeadChip datasets[45] and 37 whole-genome sequencing datasets[46] of NPC tissues from both familial and sporadic cases. Our analysis revealed no significant differences in methylation or somatic mutation profiles between tissues from familial and sporadic NPC patients (Figures S6 and S7). These results suggest that neither mutational status nor methylation differences in tumor tissues are the primary contributors to the high heritability of the oral microbiome observed in multiple NPC families. However, given the limited sample size, these findings should be considered preliminary, and further research is required.

In summary, our study leveraged a well-established NPC family cohort to uncover the inherited susceptibility of the oral microbiome and its potential connection to NPC familial aggregation. The identification of the specific heritable taxa linked to NPC risk provides valuable insights into the complex interplay between genetic factors, oral microbiome, and familial cancer development. While our findings offer potential avenues for risk stratification in families with multiple NPC cases, further research is needed to clarify the interactions between inherited genetic profiles and the microbiomes.

Limitations of the study

There are several limitations in our study. First, the study population focused exclusively on families recruited from South China, one of the areas with high incidence of this disease. NPC is also prevalent in some other Southeast Asian countries. Future studies that incorporate diverse geographic locations and ethnic groups will be crucial in establishing the generalizability of our findings regarding the microbiome in high-risk families for NPC. Second, our analysis was limited to a small number of SNPs within the HLA region. A more comprehensive genomic approach is necessary to reveal the intricate interactions between heritable microbiome, HLA, and other non-HLA genes in families with high risk of NPC. Additionally, this study primarily examined the relationships between oral microbiome and NPC familial aggregation, while the landscapes in the nasopharyngeal region where the cancer originates remain unexplored.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Wei-Hua Jia (jiawh@sysucc.org.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data supporting the conclusions of this article have been deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under PRJNA1062097, PRJNA721474, and PRJNA721325. The authenticity of this article has been validated by uploading the key raw data onto the Research Data Deposit platform, with the approval RDD number RDDA2024816882.

- The original code has been deposited at GitHub (https://github.com/Ying-Liao/NPC_family_microbiome) and is publicly available at Zenodo (https://doi.org/10.5281/zenodo.14214715).
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request, Wei-Hua Jia (jiawh@sysucc.org.cn).

ACKNOWLEDGMENTS

We first thank all individuals from the NPC high-risk family cohort, who participated in this study. We thank Dr. Yun-Fei Xia, Dr. Guan-Qun Zhou, and Dr. Li-Xia Lu from Sun Yat-sen University Cancer Center (SYSUCC) for their help in participants' enrollment. We thank Yu-Ying Liu and Yan-Hong Li at the Department of Cancer Prevention Center, SYSUCC, for their help with the sample collection. We thank Huan-Lin Yu, Ling-Ling Tang, Zi-Yi Wu, Dan-Hua Li, Chang-Mi Deng, Wen-Li Zhang, and all undergraduate students for their help in participants' enrollment and sample processing. This research was funded by the National Natural Science Foundation of China (grant numbers 82273705, 82404339, 82373656, and 81973131); Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501000); the Basic and Applied Basic Research Foundation of Guangdong Province, China (2021B1515420007); the National Key Research and Development Program of China (2021YFC2500400); and Fostering Program for NSFC Young Applicants (Tulip Talent Training Program) of Sun Yat-sen University Cancer Center (PT19010701).

AUTHOR CONTRIBUTIONS

W.-H.J. conceived, designed, and supervised the study. Y.L. contributed to the study design and data collection and analysis, carried out the experiments, and drafted the manuscript. X.-T.T. and T.Z. contributed to data analysis and drafted the manuscript. W.-Q.X., T.-M.W., and Y.-Q.H. recruited participants and collected samples. M.-Q.Z. and Y.-J.J. contributed to sample preparation and sequence library construction. D.-W.Y., Y.-X.W., and X.-H.Z. contributed to recruitment of participants and sample preparation. Z.-X.Z., M.-Y.C., and N.L. contributed to data collection and validation.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - DNA extraction and 16S rRNA gene amplicon sequencing
  - 16S rRNA gene sequencing analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Microbial community analyses
  - Heritability estimation by SOLAR
  - Analysis of published oral microbiome validation datasets
  - Association analysis of HLA with microbiome

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2025.101979.

Received: June 25, 2024
Revised: October 2, 2024
Accepted: January 28, 2025
Published: February 24, 2025

REFERENCES

1. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *74*, 229–263. https://doi.org/10.3322/caac.21834.
2. Chen, Y.-P., Chan, A.T.C., Le, Q.-T., Blanchard, P., Sun, Y., and Ma, J. (2019). Nasopharyngeal carcinoma. The Lancet *394*, 64–80. https://doi.org/10.1016/s0140-6736(19)30956-0.
3. Jia, W.H., Feng, B.J., Xu, Z.L., Zhang, X.S., Huang, P., Huang, L.X., Yu, X.J., Feng, Q.S., Yao, M.H., Shugart, Y.Y., and Zeng, Y.X. (2004). Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China. Cancer *101*, 363–369. https://doi.org/10.1002/cncr.20372.
4. Wang, T.M., He, Y.Q., Xue, W.Q., Zhang, J.B., Xia, Y.F., Deng, C.M., Zhang, W.L., Xiao, R.W., Liao, Y., Yang, D.W., et al. (2022). Whole-Exome Sequencing Study of Familial Nasopharyngeal Carcinoma and Its Implication for Identifying High-Risk Individuals. J. Natl. Cancer Inst. *114*, 1689–1697. https://doi.org/10.1093/jnci/djac177.
5. Feng, B.J., Huang, W., Shugart, Y.Y., Lee, M.K., Zhang, F., Xia, J.C., Wang, H.Y., Huang, T.B., Jian, S.W., Huang, P., et al. (2002). Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. Nat. Genet. *31*, 395–399. https://doi.org/10.1038/ng932.
6. Luo, X.Y., Liu, W.S., Chen, L.Z., Feng, Q.S., Zeng, Y.X., and Jia, W.H. (2009). [Comparative study on risk factors and family history of familial and sporadic nasopharyngeal carcinoma patients]. Zhonghua Yu Fang Yi Xue Za Zhi *43*, 293–298.
7. Ng, W.T., Choi, C.W., Lee, M.C.H., Chan, S.H., Yau, T.K., and Lee, A.W.M. (2009). Familial nasopharyngeal carcinoma in Hong Kong: epidemiology and implication in screening. Fam. Cancer *8*, 103–108. https://doi.org/10.1007/s10689-008-9213-9.
8. Zeng, Y.X., and Jia, W.H. (2002). Familial nasopharyngeal carcinoma. Semin. Cancer Biol. *12*, 443–450. https://doi.org/10.1016/s1044579x02000871.
9. Sepich-Poore, G.D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J.A., and Knight, R. (2021). The microbiome and human cancer. Science *371*, eabc4552. https://doi.org/10.1126/science.abc4552.
10. Irfan, M., Delgado, R.Z.R., and Frias-Lopez, J. (2020). The Oral Microbiome and Cancer. Front. Immunol. *11*, 591088. https://doi.org/10.3389/fimmu.2020.591088.
11. Bhatt, A.P., Redinbo, M.R., and Bultman, S.J. (2017). The role of the microbiome in cancer development and therapy. CA. Cancer J. Clin. *67*, 326–344. https://doi.org/10.3322/caac.21398.
12. Kunath, B.J., Hickl, O., Queirós, P., Martin-Gallausiaux, C., Lebrun, L.A., Halder, R., Laczny, C.C., Schmidt, T.S.B., Hayward, M.R., Becher, D., et al. (2022). Alterations of oral microbiota and impact on the gut microbiome in type 1 diabetes mellitus revealed by integrated multi-omic analyses. Microbiome *10*, 243. https://doi.org/10.1186/s40168-022-01435-4.
13. Monteiro, M.F., Altabtbaei, K., Kumar, P.S., Casati, M.Z., Ruiz, K.G.S., Sallum, E.A., Nociti-Junior, F.H., and Casarin, R.C.V. (2021). Parents with periodontitis impact the subgingival colonization of their offspring. Sci. Rep. *11*, 1357. https://doi.org/10.1038/s41598-020-80372-4.
14. Stasiewicz, M., and Karpiński, T.M. (2022). The oral microbiota and its role in carcinogenesis. Semin. Cancer Biol. *86*, 633–642. https://doi.org/10.1016/j.semcancer.2021.11.002.
15. Mascitti, M., Togni, L., Troiano, G., Caponio, V.C.A., Gissi, D.B., Montebugnoli, L., Procaccini, M., Lo Muzio, L., and Santarelli, A. (2019). Beyond Head and Neck Cancer: The Relationship Between Oral Microbiota and Tumour Development in Distant Organs. Front. Cell. Infect. Microbiol. *9*, 232. https://doi.org/10.3389/fcimb.2019.00232.
16. Liao, Y., Zhang, J.B., Lu, L.X., Jia, Y.J., Zheng, M.Q., Debelius, J.W., He, Y.Q., Wang, T.M., Deng, C.M., Tong, X.T., et al. (2023). Oral Microbiota

Alteration and Roles in Epstein-Barr Virus Reactivation in Nasopharyngeal Carcinoma. Microbiol. Spectr. *11*, e0344822. https://doi.org/10.1128/spectrum.03448-22.

17. Liao, Y., Wu, Y.X., Tang, M., Chen, Y.W., Xie, J.R., Du, Y., Wang, T.M., He, Y.Q., Xue, W.Q., Zheng, X.H., et al. (2024). Microbes translocation from oral cavity to nasopharyngeal carcinoma in patients. Nat. Commun. *15*, 1645. https://doi.org/10.1038/s41467-024-45518-2.
18. Zheng, M.Q., Wang, T.M., Liao, Y., Xue, W.Q., He, Y.Q., Wu, Z.Y., Yang, D.W., Li, D.H., Deng, C.M., Jia, Y.J., et al. (2020). Nasopharyngeal Epstein-Barr virus DNA loads in high-risk nasopharyngeal carcinoma families: Familial aggregation and host heritability. J. Med. Virol. *92*, 3717–3725. https://doi.org/10.1002/jmv.26198.
19. Debelius, J.W., Huang, T., Cai, Y., Ploner, A., Barrett, D., Zhou, X., Xiao, X., Li, Y., Liao, J., Zheng, Y., et al. (2020). Subspecies Niche Specialization in the Oral Microbiome Is Associated with Nasopharyngeal Carcinoma Risk. mSystems *5*, e00065-20. https://doi.org/10.1128/mSystems.00065-20.
20. Bei, J.X., Li, Y., Jia, W.H., Feng, B.J., Zhou, G., Chen, L.Z., Feng, Q.S., Low, H.Q., Zhang, H., He, F., et al. (2010). A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat. Genet. *42*, 599–603. https://doi.org/10.1038/ng.601.
21. Yu, G., Hsu, W.L., Coghill, A.E., Yu, K.J., Wang, C.P., Lou, P.J., Liu, Z., Jones, K., Vogt, A., Wang, M., et al. (2019). Whole-Exome Sequencing of Nasopharyngeal Carcinoma Families Reveals Novel Variants Potentially Involved in Nasopharyngeal Carcinoma. Sci. Rep. *9*, 9916. https://doi.org/10.1038/s41598-019-46137-4.
22. Liu, Z., Chang, E.T., Liu, Q., Cai, Y., Zhang, Z., Chen, G., Huang, Q.H., Xie, S.H., Cao, S.M., Shao, J.Y., et al. (2017). Quantification of familial risk of nasopharyngeal carcinoma in a high-incidence area. Cancer *123*, 2716–2725. https://doi.org/10.1002/cncr.30643.
23. Valles-Colomer, M., Bacigalupe, R., Vieira-Silva, S., Suzuki, S., Darzi, Y., Tito, R.Y., Yamada, T., Segata, N., Raes, J., and Falony, G. (2022). Variation and transmission of the human gut microbiota across multiple familial generations. Nat. Microbiol. *7*, 87–96. https://doi.org/10.1038/s41564-021-01021-8.
24. Valles-Colomer, M., Blanco-Míguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., Armanini, F., Cumbo, F., Huang, K.D., Manara, S., et al. (2023). The person-to-person transmission landscape of the gut and oral microbiomes. Nature *614*, 125–135. https://doi.org/10.1038/s41586-022-05620-1.
25. Gomez, A., Espinoza, J.L., Harkins, D.M., Leong, P., Saffery, R., Bockmann, M., Torralba, M., Kuelbs, C., Kodukula, R., Inman, J., et al. (2017). Host Genetic Control of the Oral Microbiome in Health and Disease. Cell Host Microbe *22*, 269–278.e3. https://doi.org/10.1016/j.chom.2017.08.013.
26. Demmitt, B.A., Corley, R.P., Huibregtse, B.M., Keller, M.C., Hewitt, J.K., McQueen, M.B., Knight, R., McDermott, I., and Krauter, K.S. (2017). Genetic influences on the human oral microbiome. BMC genomics *18*, 659. https://doi.org/10.1186/s12864-017-4008-8.
27. Shaw, L., Ribeiro, A.L.R., Levine, A.P., Pontikos, N., Balloux, F., Segal, A.W., Roberts, A.P., and Smith, A.M. (2017). The Human Salivary Microbiome Is Shaped by Shared Environment Rather than Genetics: Evidence from a Large Family of Closely Related Individuals. mBio *8*, e01237-17. https://doi.org/10.1128/mBio.01237-17.
28. Mukherjee, C., Moyer, C.O., Steinkamp, H.M., Hashmi, S.B., Beall, C.J., Guo, X., Ni, A., Leys, E.J., and Griffen, A.L. (2021). Acquisition of oral microbiota is driven by environment, not host genetics. Microbiome *9*, 54. https://doi.org/10.1186/s40168-020-00986-8.
29. Cai, L., Zhu, H., Mou, Q., Wong, P.Y., Lan, L., Ng, C.W.K., Lei, P., Cheung, M.K., Wang, D., Wong, E.W.Y., et al. (2024). Integrative analysis reveals associations between oral microbiota dysbiosis and host genetic and epigenetic aberrations in oral cavity squamous cell carcinoma. NPJ Biofilms Microbiomes *10*, 39. https://doi.org/10.1038/s41522-024-00511-x.
30. Kwong, T.N.Y., Wang, X., Nakatsu, G., Chow, T.C., Tipoe, T., Dai, R.Z.W., Tsoi, K.K.K., Wong, M.C.S., Tse, G., Chan, M.T.V., et al. (2018). Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. Gastroenterology *155*, 383–390.e8. https://doi.org/10.1053/j.gastro.2018.04.028.
31. Guo, S., Xing, S., Wu, Z., Chen, F., Pan, X., Li, Q., Liu, W., and Zhang, G. (2024). Leucine restriction ameliorates Fusobacterium nucleatum-driven malignant progression and radioresistance in nasopharyngeal carcinoma. Cell Rep. Med. *5*, 101753. https://doi.org/10.1016/j.xcrm.2024.101753.
32. Alon-Maimon, T., Mandelboim, O., and Bachrach, G. (2022). Fusobacterium nucleatum and cancer. Periodontology, 166–180. https://doi.org/10.1111/prd.12426.
33. Ge, X., Kitten, T., Chen, Z., Lee, S.P., Munro, C.L., and Xu, P. (2008). Identification of Streptococcus sanguinis genes required for biofilm formation and examination of their role in endocarditis virulence. Infect. Immun. *76*, 2551–2559. https://doi.org/10.1128/iai.00338-08.
34. Zhou, P., Li, X., Huang, I.H., and Qi, F. (2017). Veillonella Catalase Protects the Growth of Fusobacterium nucleatum in Microaerophilic and Streptococcus gordonii-Resident Environments. Appl. Environ. Microbiol. *83*, e01079-17. https://doi.org/10.1128/aem.01079-17.
35. Liu, Q.Y., Liao, Y., Wu, Y.X., Diao, H., Du, Y., Chen, Y.W., Xie, J.R., Xue, W.Q., He, Y.Q., Wang, T.M., et al. (2023). The Oral Microbiome as Mediator between Oral Hygiene and Its Impact on Nasopharyngeal Carcinoma. Microorganisms *11*, 719. https://doi.org/10.3390/microorganisms11030719.
36. Yang, X.R., Diehl, S., Pfeiffer, R., Chen, C.J., Hsu, W.L., Dosemeci, M., Cheng, Y.J., Sun, B., Goldstein, A.M., and Hildesheim, A.; Chinese and American Genetic Epidemiology of NPC Study Team (2005). Evaluation of risk factors for nasopharyngeal carcinoma in high-risk nasopharyngeal carcinoma families in Taiwan. Cancer Epidemiol. Biomarkers Prev. *14*, 900–905. https://doi.org/10.1158/1055-9965.Epi-04-0680.
37. Fitch, M.N., Phillippi, D., Zhang, Y., Lucero, J., Pandey, R.S., Liu, J., Brower, J., Allen, M.S., Campen, M.J., McDonald, J.D., and Lund, A.K. (2020). Effects of inhaled air pollution on markers of integrity, inflammation, and microbiota profiles of the intestines in Apolipoprotein E knockout mice. Environ. Res. *181*, 108913. https://doi.org/10.1016/j.envres.2019.108913.
38. Basti, A.A., Misaghi, A., Salehi, T.Z., and Kamkar, A. (2006). Bacterial pathogens in fresh, smoked and salted Iranian fish. Food Control *17*, 183–188. https://doi.org/10.1016/j.foodcont.2004.10.001.
39. Lu, S.J., Day, N.E., Degos, L., Lepage, V., Wang, P.C., Chan, S.H., Simons, M., McKnight, B., Easton, D., Zeng, Y., et al. (1990). Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. Nature *346*, 470–471. https://doi.org/10.1038/346470a0.
40. Olivares, M., Neef, A., Castillejo, G., Palma, G.D., Varea, V., Capilla, A., Palau, F., Nova, E., Marcos, A., Polanco, I., et al. (2015). The HLA-DQ2 genotype selects for early intestinal microbiota composition in infants at high risk of developing coeliac disease. Gut *64*, 406–417. https://doi.org/10.1136/gutjnl-2014-306931.
41. Davies, C.S., Worsley, S.F., Maher, K.H., Komdeur, J., Burke, T., Dugdale, H.L., and Richardson, D.S. (2022). Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. Microbiome *10*, 41. https://doi.org/10.1186/s40168-022-01233-y.
42. Andeweg, S.P., Keşmir, C., and Dutilh, B.E. (2021). Quantifying the Impact of Human Leukocyte Antigen on the Human Gut Microbiota. mSphere *6*, e0047621. https://doi.org/10.1128/mSphere.00476-21.
43. Pepke, M.L., Hansen, S.B., and Limborg, M.T. (2024). Unraveling host regulation of gut microbiota through the epigenome-microbiome axis. Trends Microbiol. *32*, 1229–1240. https://doi.org/10.1016/j.tim.2024.05.006.
44. Byrd, D.A., Fan, W., Greathouse, K.L., Wu, M.C., Xie, H., and Wang, X. (2023). The intratumor microbiome is associated with microsatellite instability. J. Natl. Cancer Inst. *115*, 989–993. https://doi.org/10.1093/jnci/djad083.
45. Jiang, W., Liu, N., Chen, X.Z., Sun, Y., Li, B., Ren, X.Y., Qin, W.F., Jiang, N., Xu, Y.F., Li, Y.Q., et al. (2015). Genome-Wide Identification of a

Methylation Gene Panel as a Prognostic Biomarker in Nasopharyngeal Carcinoma. Mol. Cancer Ther. *14*, 2864–2873. https://doi.org/10.1158/1535-7163.Mct-15-0260.

46. Lin, M., Zhang, X.L., You, R., Liu, Y.P., Cai, H.M., Liu, L.Z., Liu, X.F., Zou, X., Xie, Y.L., Zou, R.H., et al. (2023). Evolutionary route of nasopharyngeal carcinoma metastasis and its clinical significance. Nat. Commun. *14*, 610. https://doi.org/10.1038/s41467-023-35995-2.

47. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9.

48. Escapa, I.F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F.E., and Lemon, K.P. (2018). New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. mSystems *3*, e00187-18. https://doi.org/10.1128/mSystems.00187-18.

49. McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

50. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. J. Stat. Softw. *42*, 1–28. https://doi.org/10.18637/jss.v042.i08.

51. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106. https://doi.org/10.1186/gb-2010-11-10-r106.

52. Friedman, J., and Alm, E.J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. *8*, e1002687. https://doi.org/10.1371/journal.pcbi.1002687.

53. Almasy, L., and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. *62*, 1198–1211. https://doi.org/10.1086/301844.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Saliva samples as reported in the experimental model and subject details	Sun Yat-sen university cancer center	N/A
Critical commercial assays		
PowerSoil DNA isolation kit	Qiagen	Catalog No. 12888-100
Deposited data		
Raw 16S rRNA sequencing datasets	This paper	SRA: PRJNA1062097; PRJNA721474; PRJNA721325
Oral microbiome validation datasets	Debelius, J.W et al.[19]	EBI: PRJEB37445
Infinium HumanMethylation450 BeadChip datasets of NPC tissues	Jiang, W et al.[45]	GEO: GSE52068
Whole-genome sequencing datasets of NPC tissues	Lin, M et al.[46]	NGDC: HRA000034
Oligonucleotides		
515F 5′-GTGCCAGCMGCCGCGGTAA-3′	Thermo Fisher Scientific	N/A
806R 5′-GGACTACHVGGGTWTCTAAT-3′	Thermo Fisher Scientific	N/A
Software and algorithms		
R	The R Software Foundation	https://www.r-project.org
R studio	The R Software Foundation	https://www.r-project.org
QIIME2	QIIME2	https://qiime2.org/
SOLAR	SOLAR	https://hpc.nih.gov/docs/solar-8.1.1/
SparCC	SparCC	https://github.com/dlegor/SparCC
Cytoscape	Cytoscape	https://cytoscape.org/
Original code	This paper	Zenodo: https://doi.org/10.5281/zenodo.14214715

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The family cohort in this study was from the ''High-risk Nasopharyngeal Carcinoma Family Screening Program'' at the Sun Yat-sen University Cancer Center (SYSUCC, Guangzhou, China) from 2016 to 2020.[4,18] This program aimed to recruit individuals from families with two or more NPC patients, including both NPC patients and unaffected relatives. A total of 858 individuals from 164 families were recruited. On average, each family had 2.53 NPC cases and a range of 2–21 family members attending, averaging 5.11 members per family. Considering the impact of anti-cancer therapy on the oral microbiome, post-therapy NPC patients were excluded from this study. The oral microbiome data, pedigree information, and questionnaire-based data of 649 individuals from 127 families were obtained, including 564 individuals from 107 multiple NPC families, which contained two or more NPC cases who met the first- or second-degree relatives' kinship. Among them, 67 families had two NPC patients (N = 315), and 40 families had three and more NPC patients (N = 252). Besides, 82 individuals from 20 sporadic NPC families, contained only one NPC case or two NPC cases but with a third or more distant degree of kinship. Analyses were conducted by subgrouping individuals based on the families they belong to, with varying numbers of NPC patients.

All subjects were interviewed using a structured questionnaire containing 192 items, including demographics, clinical history of ear-nose-throat (ENT) and oral disease, family history of malignancy, and lifestyle information, including smoking status, alcohol consumption, *etc.* The questionnaire also included the necessary information to determine the co-household status and draw a pedigree structure for each enrolled family. Unstimulated whole saliva samples were collected as described previously.[16] All participants were asked not to eat or drink for half an hour before providing samples. Five milliliters of saliva were collected into a sterile 50mL centrifuge tube and were subsequently stored at −80°C until use.

To identify the NPC-associated oral microbiome features, we included an independent oral microbiome dataset with 150 treatment-naïve NPC cases and 153 controls,[16] as well as 17 treatment-naïve NPC patients and 17 age- and gender-matched unaffected controls in this family cohort. All case-control individuals were interviewed using a structured questionnaire, including demographics, clinical history of ear-nose-throat (ENT) and oral disease, family history of malignancy, lifestyle information, *etc*. These two datasets,

including 167 treatment-naïve NPC cases and 170 controls, followed the same sampling procedures, nucleic acid processing, and sequencing methods in the same circumstance.

The Institutional Review Board of Sun Yat-sen University Cancer Center approved this study. Informed consent was obtained from all study participants. The demographic characteristics of all the participants were provided in Table S1 and Table S6.

METHOD DETAILS

DNA extraction and 16S rRNA gene amplicon sequencing

Saliva DNA was extracted using the PowerSoil DNA isolation kit (Qiagen, Duesseldorf, Hilden, Germany) with the bead-beating method according to the manufacturer's instructions. All samples were randomized before DNA extraction to prevent batch effects. We included 31 negative controls comprising 3 sampling controls, 16 extraction controls, and 12 PCR amplification controls. Additionally, 14 positive controls were included, using the same saliva DNA samples for repeated sequencing in each sequencing library. 16S rRNA amplicon sequencing was performed using our previously established protocols.[16] DNA extracted from each sample, including quality control samples, was used as the template to amplify the V4 region of the 16S rRNA gene. Briefly, the V4 variable region of the 16S rRNA gene was first amplified using the primer pairs 515F/806R with 12 bp barcodes with 20 cycles. Products of the primary PCR were confirmed by 1.0% agarose gel electrophoresis, and a 1:100 dilution of primary PCR products was used for indexing PCR with Illumina flow cell adaptors and dual indices (6 bp). The purified amplicons were pooled in equal concentrations for sequencing using a 2 × 250 bp Illumina MiSeq system.

16S rRNA gene sequencing analysis

Amplicon sequence variants (ASVs) were generated from raw Illumina amplicon sequences using DADA2 pipeline (https://benjjneb.github.io/dada2/tutorial.html), an open-source package in R. Specifically, paired-ended fastq reads were trimmed barcodes and primers, discarding low quality reads using the function filterAndTrim with parameters of "maxN = 0, maxEE = 2, truncQ = 2, rm.phix = TRUE". Reads were performed with filtered data and learned error model using the function dada. Then, the forward and reverse reads were merged together to obtain the full denoised sequences using the function mergePairs. Chimeras were removed using the function removeBimeraDenovo with parameter "method = "consensus"". After above steps, ASVs table was constructed by function makeSequenceTable. Taxonomies were assigned to the ASVs with QIIME2[47] using the expanded Human Oral Microbiome Database[48] (eHOMD, Version 15.2) as the reference. Subsequent ASVs and sample preprocessing were conducted in phyloseq package in R. ASVs present in <1% samples and with total frequency <20 reads were removed. Samples lower than 5000 reads were discarded. An average of 24575 ± 10314 (mean ± *SD*) sequences/sample was achieved after the above filtration. The filtered ASVs table with the taxonomic assignment was generated and used in the following statistical analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Microbial community analyses

The estimate_richness and phyloseqdistance functions from the phyloseq package[49] were used in the diversity analysis of 16S rRNA data. Permutational multivariate analysis of variance (PERMANOVA; adonis function, vegan package in R) of the beta diversity metrics was used to test the influence of environmental factors on oral microbiota structure. Principal coordinate analysis (PCoA) was performed to obtain the first three principal coordinates of these four beta-diversity dissimilarity matrices (unweighted UniFrac, weighted UniFrac, Jaccard and Bray-Curtis dissimilarity).

Propensity scores were operated to randomly select unrelated pairs for different kinships with a ratio 1:3 using matchit function with parameter "method = nearest" from MatchIt[50] package. Differential gene expression analysis based on the negative binomial distribution (DESeq) function[51] of the DESeq2 package in R software was used to identify the differential abundant microbial features between NPC patients and control, and between sporadic and multiple NPC patients. Core ASVs or genera with prevalence higher than 10% and average relative abundance higher than 0.0001 were concluded into the analysis. Key confounders including age, gender and cigarette smoking were adjusted when perform DESeq analysis, other parameters were set as default. Pairwise correlation matrices were calculated from the abundance of oral taxa using SparCC,[52] and Cytoscape software (version 3.7.1) was employed to establish the networks.

Heritability estimation by SOLAR

The Sequential Oligogenic Linkage Analysis Routines (SOLAR, version 8.4.2)[53] was applied to heritability estimation based on family pedigrees. Heritability is a statistic that describes how much of the variation in a given trait can be attributed to genetic variation. Variation is teased apart to additive genetic effects (heritability, h^2), the shared/common environment (c^2), and non-shared/unique environmental factors (e^2). $h^2/c^2/e^2$ estimates range from zero to one, and their values indicate the proportion of the variability in the trait is due to genetic/cohabitant/environmental differences among people in a population. In this study, pedigree data, cohabitant status, and environmental factors were all included to construct the SOLAR model. Cohabitatation effects were estimated using household information from questionnaires, which recorded the co-housing of individuals within the families. Each participant was assigned a household identifier to indicate their cohabitation status with other family members. Variables such as age, gender, BMI,

dialect, residence (childhood or adult), disease status, lifestyle factors - including smoking, alcohol consumption, tea and herbal tea drinking – were included; dietary factors such as vegetables, fruit, soup, and salted fish consumption were also considered. These variables, detailed in Table S1, were included as non-shared environmental factors during the analysis. The SOLAR analysis was performed on the quantitative traits including the alpha-diversity index, and the first three PCos of four beta-diversity dissimilarity matrices, and the relative abundance of core ASVs in the family cohort. SOLAR analysis was also performed on the specific ASVs separately in the multiple and sporadic family subgroups. *p*-values were adjusted by the false discovery rate (FDR) method.

We analyzed twenty-four Infinium HumanMethylation450 BeadChip datasets of NPC tissues from published data,[45] including two familial NPC patients. Heatmaps were generated using 10,000 randomly selected CpG sites from the available 485,577, along with 2,173 significant methylation CpG sites identified in a previous tumor vs. normal comparison.[45] Thirty-seven whole-genome sequencing datasets of NPC tissues,[46] four of which were from familial NPC patients, were analyzed. Heatmaps were created based on reported somatic mutations in NPC tissues.[46] All heatmaps were generated using the pheatmap package in R software, clustering with "complete linkage clustering" method.

Analysis of published oral microbiome validation datasets

The NPC-association validation analysis was performed using the published datasets comprising oral microbiome profiles via 16S rRNA sequencing from 499 nasopharyngeal carcinoma patients and 496 healthy controls.[19] The bioinformatics analysis of 16S rRNA gene sequencing data followed the same methodology as described above. Since the validation dataset targeted the amplification of the V3 and V4 regions, the abundance of ASVs corresponding to the V4 region in this study was calculated by summing the abundances of ASVs with sequences that perfectly matched the V4 sequencing region.

Association analysis of HLA with microbiome

Genotypes for rs2860580 (HLA-A), rs2894207 (HLA-B/C), and rs28421666 (HLA-DQ/DR) were determined using MassARRAY technology. Based on genotype classification, individuals were categorized into three groups: homozygous wild-type, heterozygous, and homozygous mutant. Beta diversity analysis of the microbiome was then conducted across these groups. The protective allele was assigned as the reference value of 0, while heterozygotes and homozygotes mutants were assigned values of 1 and 2, respectively. Spearman correlation analyses were performed to assess the relationship between these genotype variations and the relative abundance of specific bacterial taxa.