Supplementary Methods

TCGA patient cohort

16S ribosomal RNA sequencing data and clinical profiles of HNSCC patients were obtained from the Cancer Genome Atlas (TCGA) database (http://cancergenome.nih.gov/, February 2021) and The Cancer Microbiome Atlas (TCMA) as described [1,2].

Validation HNSCC patient cohort

Patients were enrolled in this study at Peking Union Medical College Hospital (PUMCH) in 2022. All patients had pathologically confirmed, previously untreated HNSCC and underwent radical resection. The exclusion criteria were as follows: (i) antibiotic therapy in the previous month; (ii) infection with HBV, HCV, syphilis, or HIV; and (iii) a history of malignant tumors, chemotherapy, or radio therapy. The TNM stages of all participants were identified according to the National Comprehensive Cancer Network (NCCN) Guidelines in 2021. Before sampling, participants were banned from dieting, smoking, and oral hygiene prophylaxis for at least 2 hours. Cancerous tissues from the central area of the lesions were obtained during surgery. Tissue samples were placed into sterile 2 mL Eppendorf tubes (Axygen, USA) and then frozen at −80°C before further processing.

Biostatistical analysis

Alpha-diversity, based on the number of observed species, chao1 index and shannon index, and Beta-diversity, based on the pairwise bray-curtis distance among samples, were calculated with R software (version 4.0.2). Differences in the Chao1 index and the Shannon index were detected using the Wilcoxon rank sum test. Principal coordinate analysis (PCoA) was conducted using the Analysis of Phylogenetics and Evolution (APE) package in R software[3].

PERMANOVA was performed with the vegan package in R to clarify differences in microbial communities between groups. The Wilcoxon rank sum test was applied to identify significantly different taxa between groups, and a false discovery rate <0.05 was considered to be statistically significant. Kaplan–Meier curves were plotted for survival distributions with SPSS version 23.0 (IBM Corporation, Armonk, NY, USA).

Microbial DNA extraction and *Leptotrichia* quantification

Tumor tissues of HNSCC patients were divided to small pieces and digested in phosphate-buffered saline containing an enzymatic cocktail for 1 h at 37 °C [4]. A 200 μl aliquot from each tissue was processed using QIAamp DNA Mini Kit (QIAGEN) according to the manufacturer's instructions for the extraction of total genomic DNA. qPCR was performed to detect the *Leptotrichia* level by using 20 ng genomic DNA in 10 μl universal SYBR Green PCR Master Mix (Invitrogen, Thermo Fisher, Grand Island, NY, USA) with a StepOne Plus Real-Time PCR system (Applied Biosystems, Thermo Fisher, Grand Island, NY, USA). Gene ABC transporter of Leptotrichia was used in PCR. *Leptotrichia* quantitation was measured relative to the *Actb* gene. The *Leptotrichia* forward primer was 5′-GTTAAATGGGCTTCTTAAACCAA-3'; the reverse primer was 5'-AACAACTGTAAAATTGCCTT-3'.

Fluorescence in situ hybridization (FISH)

The tumor tissues of HNSCC patients were collected from frozen sections for in situ hybridization of *Leptotrichia*. The *Leptotrichia* probe sequence was 5'-CACTTCCATTCGGCCCTAATAATC-3'. According to the instructions, the 5-μm thick frozen sections were fixed with 4% (v/v) paraformaldehyde at room temperature for 20min. After fixation, the sections were permeabilized with proteinase K at 37 °C for 20 min and 0.1 M fresh glycine solution for 1 min to stop the proteinase K. After rinsing in PBS buffer, the cells were prehybridized

in hybridization buffer at 65 °C for 1 h in a hybridization box. The prehybridization buffer was replaced with hybridization solution (containing 100 μm Cy3-labeled *Leptotrichia* nucleotide probe), and the sections were incubated at 65 °C for 48 h. The sections were then washed once with 0.2× SSC for 1 min at room temperature, washed three times with formamide plus 4× SSC for 20 min at 65 °C, and five times with PBS for 1 min at room temperature. Finally, the sections were treated with DAPI (Thermo Fisher Scientific) for 5 min. Image acquisition was performed using a Zeiss confocal microscope system (Thornwood). Fluorescence was semiquantitatively assessed on the basis of the mean fluorescence intensity (MFI) of each image.

References:

1. Dohlman AB, Arguijo Mendoza D, Ding S, et al. The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. *Cell Host Microbe.* 2021;29(2):281-298 e285.
2. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2(5):401-404.
3. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics.* 2004;20(2):289-290.
4. Long X, Wong CC, Tong L, et al. Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. *Nat Microbiol.* 2019;4(12):2319-2330.