From Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

DIETARY HABITS, COMMENSAL MICROBIOME, AND NASOPHARYNGEAL CARCINOMA

Tingting Huang
黄婷婷

Karolinska Institutet

Stockholm 2020

Front cover illustration: 壹日创意 and Tingting Huang
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2020

ISBN 978-91-8016-002-5

Dietary Habits, Commensal Microbiome, and Nasopharyngeal Carcinoma

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Tingting Huang

December 17th, 2020. 09:00 a.m.
Lecture hall Atrium, Nobels väg 12B, Karolinska Institutet, Solna

Principal Supervisor:
Professor Weimin Ye
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Co-supervisor(s):
Professor Eva Munck af Rosenschöld Wikland
Karolinska Institutet
Department of Clinical Science, Intervention and Technology (CLINTEC)
Division of ENT diseases
Karolinska University Hospital
Medical Unit Head Neck, Lung and Skin Cancer

Assistant Professor Zongli Zheng
Karolinska Institutet
Ming Wai Lau Centre for Reparative Medicine

Dr. Justine W. Debelius
Karolinska Institutet
Department of Microbiology, Tumor, and Cell Biology
Centre for Translational Microbiome Research

Opponent:
Professor Maria Li Lung
The University of Hong Kong
LKS Faculty of Medicine
Department of Clinical Oncology

Examination Board:
Docent Johan Dicksved
Swedish University of Agricultural Sciences
Department of Animal Nutrition and Management

Professor Per Hall
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Professor Agneta Åkesson
Karolinska Institutet
Institute of Environmental Medicine
Unit of Cardiovascular and Nutritional Epidemiology

To my parents for always believing in me

致我最爱的爸爸妈妈 ❤

& all the efforts to improve human health

ABSTRACT

Nasopharyngeal carcinoma (NPC) is a malignant disease characterized by unique geographic distribution endemic to southern China, Southeast Asia, and the Middle East/North Africa. It has been widely accepted that the interaction of Epstein-Barr Virus infection, environmental and lifestyle factors, and genetic susceptibility, contributes to NPC carcinogenesis. In the past two decades, extensive application of intensity-modulated radiotherapy and widespread introduction of chemotherapy have significantly contributed to a desirable prognosis: better survival with fewer toxicities. However, there are still numerous knowledge gaps and unsolved questions about NPC risk and prognosis. In this thesis, we investigated whether dietary habits (Study I), and oral commensal microbiome (Study II) were associated with NPC risk, using a population-based case-control study in endemic southern China. We also conducted a longitudinal hospital-based NPC cohort study (Study III) to deliver proof-of-concept data on the commensal microbiome patterns in patients' nasopharynx during radiotherapy and their role in NPC prognosis.

In Study I/Paper I, we analyzed a total of 4398 study participants (2174 NPC cases and 2224 controls) with data about adolescent diet and 4832 participants (2387 NPC cases and 2445 controls) with data about adulthood diet. We demonstrated a strong positive association between higher consumption of the "animal-foods-based diet" and NPC risk and a strong negative association with higher intake of the "plant-based diet". Following mutual adjustment for adolescence and adulthood dietary patterns, risk estimates for the former were attenuated and no longer statistically significant, whereas associations with adulthood dietary patterns remained virtually unchanged.

In Study II/Paper II, we explored the relationship between NPC status and the oral microbiome using 16S rRNA sequencing in a study of 994 participants (499 NPC cases, 494 controls). We observed a significant reduction in community richness in NPC cases compared to that in controls. We also identified a pair of *Granulicatella adiacens* amplicon sequence variants (ASVs; Gran-7770 and Gran-5a37), which were strongly associated with NPC status and differed by a single nucleotide. We further revealed that Gran-7770 and Gran-5a37 each formed co-occurring nodes with a dozen ASVs, which were exclusive. These results suggest differences in the oral microbiomes between NPC patients and controls, which may be associated with both a loss of microbial diversity and niche specialization among closely related microorganisms.

In Study III/Paper III, we analyzed 445 nasopharyngeal samples longitudinally collected from 39 NPC patients during radiotherapy-based treatment. We addressed stable, temporal changes in the nasopharyngeal microbial community structure among NPC patients during treatment. These changes were associated with patients' short-term clinical outcomes measured three months after the completion of radiotherapy. We also identified 23 out of 73 abundant ASVs that showed statistically significant changes in the ratio of abundance between early and late

responders throughout treatment. These results provided evidence of an association between nasopharyngeal commensal microbiome and NPC patients' short-term clinical outcome.

By studying a range of topics, this thesis provides more insights into NPC in terms of risk and prognosis: plant-based and animal-foods-based diets are differentially associated with NPC risk in endemic southern China, suggesting a possibility of primary prevention of NPC through dietary intervention; oral microbiome is associated with NPC risk; the niche specialization among closely related commensals associated with NPC status calls for future culture-based investigations; moreover, stable, temporal changes of the nasopharyngeal microbiome are associated with NPC patients' short-term clinical outcome, which calls for more extensive longitudinal studies with long-term follow-up for verification and serves as a base of generating new hypotheses for future studies.

LIST OF SCIENTIFIC PAPERS

* Equal contribution.

I. **Huang T**, Ploner A, Chang ET, Liu Q, Cai Y, Zhang Z, Chen G, Huang QH, Xie SH, Cao SM, Jia WH, Zheng Y, Liao J, Chen Y, Lin L, Ernberg I, Huang G, Zeng Y, Zeng YX, Adami HO, Ye W. Dietary Patterns and Risk of Nasopharyngeal Carcinoma: A Population-Based Case-Control Study in Southern China. (Submitted)

II. Debelius JW*, **Huang T***, Cai Y*, Ploner A, Barrett D, Zhou X, Xiao X, Li Y, Liao J, Zheng Y, Huang G, Adami HO, Zeng Y, Zhang Z, Ye W. Subspecies Niche Specialization in the Oral Microbiome Is Associated with Nasopharyngeal Carcinoma Risk. *mSystems*, 2020 Jul
DOI: 10.1128/mSystems.00065-20

III. **Huang T**, Debelius JW, Ploner A, Xiao XL, Zhang T, Hu K, Zhang Z, Wang R, Ye W. Radiation Therapy-Induced Changes of the Nasopharyngeal Commensal Microbiome in Nasopharyngeal Carcinoma Patients. *International Journal of Radiation Oncology Biology Physics*. 2020 Aug 28;S0360-3016(20)34201-2.
DOI: 10.1016/j.ijrobp.2020.08.054

RELATED WORK

(Not included in the thesis)

- Jiang L, Bo L, Zhang Y, Ma S, Liu C, Liang F, Wei Z, **Huang T***, Wang R*. Influence of pelvic intensity-modulated radiation therapy with concurrent cisplatin-based chemotherapy of cervical cancer on the vaginal microbiome. (Submitted; * Corresponding authors)

- Feng R, Chang ET, Liu Q, Cai Y, Zhang Z, Chen G, Huang QH, Xie SH, Cao SM, Zhang Y, Yun JP, Jia WH, Zheng Y, Liao J, Chen Y, **Huang T**, Lin L, Ernberg I, Huang G, Zeng YX, Adami HO, Ye W. Intake of alcohol and tea and risk of nasopharyngeal carcinoma: a population- based case-control study in southern China. (Accepted by *Cancer Epidemiology, Biomarkers & Prevention*, 2020, Nov)

- Zhu S; Xu K; Jiang Y; Suo C; Cui M; Wang Y; Yuan Z; Xue J; Wang J; Zhang T; Zhao G; Ye W; **Huang T**; Lu M; Tian W; Li J; Chen X. The gut microbiome in subclinical atherosclerosis: a population-based multi-phenotype analysis. (Submitted; DOI: 10.21203/rs.2.17233/v1)

- Liang T; Liu F; Ma L; Zhang Z; Liu L; **Huang T**; Li, J; Dong W; Zhang H; Li Y; Jiang Y; Ye W; Bai S; Kang L. Effect of Migration on Intestinal Microbiota of Tibetans with a Distinct Genetic Background. (Manuscript)

- Chen Y, Xu Y, Zhao W, Xiao X, Zhou X, Lin L, **Huang T**, Liao J, Li Y, Zeng X, Huang G, Ye W, Zhang Z. Lack of association between cigarette smoking and Epstein Barr virus reactivation in the nasopharynx in people with elevated EBV IgA antibody titres. *BMC Cancer*. 2018 Feb 14;18(1):190. DOI: 10.1186/s12885-018-4110-6.

- Zhou X, Xiao X, **Huang T**, Du C, Wang S, Mo Y, Ma N, Murata M, Li B, Wen W, Huang G, Zeng X, Zhang Z. Epigenetic inactivation of follistatin-like 1 mediates tumor immune evasion in nasopharyngeal carcinoma. *Oncotarget*. 2016 Mar 29; 7(13): 16433–16444. DOI: 10.18632/oncotarget.7654

- Zhou X, Wei J, Chen F, Xiao X, **Huang T**, He Q, Wang S, Du C, Mo Y, Lin L, Xie Y, Watanabe H, Murata M, Huang G, Ernberg I, Matskova L, Zhang Z. Epigenetic downregulation of the ISG15-conjugating enzyme UbcH8 impairs lipolysis and correlates with poor prognosis in nasopharyngeal carcinoma. *Oncotarget*. 2015 Dec 1;6(38):41077-91. DOI: 10.18632/oncotarget.6218

CONTENTS

LIST OF ABBREVIATIONS

16S rRNA	16S ribosomal ribonucleic acid
^{18}F-FDG-PET/CT	^{18}F-fluorodeoxyglucose-positron emission tomography/computed tomography
AIC	Akaike information criterion
AJCC	American Joint Committee on Cancer
ANCOM	Analysis of composition of microbiome
ASIR	Age-standardized incidence rate
ASV	Amplicon sequence variant
BLAST	Basic local alignment search tool
CA	Cluster analysis
CI	Confidence interval
CR	Complete response
CT	Computed tomography
CUP	Continuous Update Project
DNA	Deoxyribonucleic acid
DSS	Disease-specific survival
EBV	Epstein-Barr Virus
FA	Factor analysis
FDR	False discovery rate
FFQ	Food frequency questionnaire
HLA	Human leukocyte antigen
IARC	International Agency for Research on Cancer
IMRT	Intensity-modulated radiotherapy
LME	Linear mixed-effects model
MRI	Magnetic resonance imaging
NCCN	National Comprehensive Cancer Network
NMIT	Non-parametric microbial interdependence test
NPC	Nasopharyngeal carcinoma
NPCGEE	Gene-environment Epstein-Barr Virus Interactions in the Etiology of Nasopharyngeal Carcinoma

OR	Odds ratio
OS	Overall survival
PC	Principle coordinate
PCA	Principal component analysis
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction
PD-L1	Programmed death-ligand 1
PERMANOVA	Permutational multivariate analysis of variance
PR	Partial response
RECIST	Response evaluation criteria in solid tumors
RRR	Reduced rank regression
RT	Radiation therapy
SCNIC	Sparse co-occurrence network investigation for compositional data
SD	Stable disease
SS-ANOVA	Smoothing-spline analysis of variance
TEI	Total energy intake
TNM	Tumor-node-metastasis
UICC	Union for International Cancer Control

1 INTRODUCTION

Nasopharyngeal carcinoma (NPC) is one type of malignant head and neck tumors, characterized by unique geographic distribution endemic to southern China, Southeast Asia, and the Middle East/North Africa. It is widely accepted that the interaction of Epstein-Barr Virus (EBV) infection, environmental and lifestyle factors, and genetic susceptibility, contributes to NPC carcinogenesis. In the past two decades, extensive application of intensity-modulated radiotherapy (IMRT) and widespread introduction of chemotherapy with various strategies (concurrent, adjuvant, induction) have significantly contributed to a desirable prognosis: better survival with less and/or lower level of toxicities.

However, there are still numerous knowledge gaps in relation to NPC risk and prognosis. The ubiquitous EBV infection (mainly persists latently) in over 90% of the world population is insufficient to explain the distinct disease burden and makes the primary intervention difficult. Monitoring and/or modifying environmental and lifestyle risk factors are reasonable measures in disease control and intervention. A progressively declined NPC incidence trend has been observed worldwide, partly reflecting a better understanding of the disease, rapid economic development, and gradual changes in environment and lifestyle. Nevertheless, epidemiological studies in China, especially in southern regions, showed generally stable incidence trends from the 1980s to 2010s.

Currently, the 5-year survival rates of non-metastasis NPC patients are ranging from 60-90%. However, around 30% of them will relapse with either recurrent disease or metastasis. There is a clear need to identify patients with diverse prognosis and provide optimal local and systematic treatments in clinical practice. Beyond the tumor stage, EBV-related markers are currently recommended for prognostic risk stratification and therapeutic monitoring in NPC, although the evidence in clinical practice is less conclusive. Researches on various biomarkers concerning treatment response and disease prognosis are warranted.

The purpose of this thesis is to study whether dietary habits, oral commensal microbiome are associated with NPC risk in southern China, and to explore the longitudinal pattern of nasopharyngeal commensal microbiome among NPC patient during radiotherapy-based treatment.

2 BACKGROUND

Nasopharyngeal carcinoma (NPC) occurs in the epithelial lining of the nasopharynx, which is part of the pharynx and connects the nasal cavity and oropharynx. The nasopharynx is a box-like area, under and in front of the base of skull and pharyngeal tonsils, above and behind the soft palate, with an opening of the eustachian tube (called pharyngeal recess or Rosenmüller's fossa) on each side (Figure 2.1). NPC frequently rises in the pharyngeal recess [1]. NPC is a nonlymphomatous, squamous-cell carcinoma, histologically classified into keratinizing squamous cell carcinoma, non-keratinizing carcinoma (undifferentiated and differentiated), and basaloid squamous cell carcinoma by World Health Organization [2-4]. The following topics regarding NPC will be discussed: descriptive epidemiology, risk factors, clinical characteristics, treatment and prognosis, as well as prognosis factors.

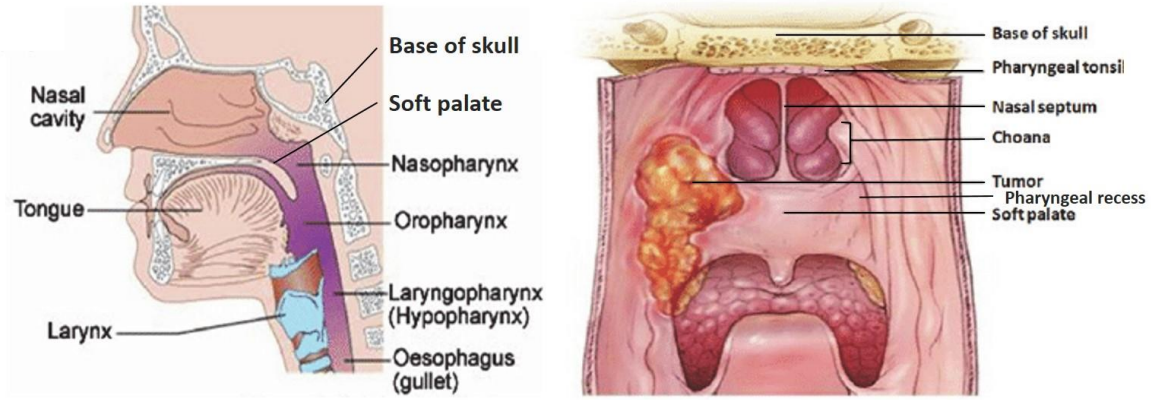


Figure 2.1 Anatomy of the nasopharynx (adapted from Hong LL et al. [5]).

2.1 DESCRIPTIVE EPIDEMIOLOGY

Worldwide, NPC is an uncommon cancer with approximated 129 000 new cases (0.7% of all cancers combined) and 73 000 deaths (0.9% of all cancer-related deaths combined) in 2018 [6]. It is rare in most parts of the world with an age-standardized incidence rate (ASIR) of less than 1 per 100 000 person-years (Figure 2.2). However, it is endemic in the East and Southeast Asia, South-Central Asia, as well as North and East Africa [6-13]. Around half of the estimated new cases were reported in China, with an ASIR (world) of 3.0 per 100 000 person-years [6]. Notably, the NPC incidence increases from northern China to southern China with up to 30 per 100,000 person-years in southern China (e.g., Hong Kong, Sihui in Guangdong province), while Tianjin in northern China has less than 2 per 100,000 person-years (Figure 2.3) [10]. Males have two to three times higher NPC incidence than females in different populations [6-12]. Like most cancers, NPC incidence increases with age in the low-risk population. On the contrary, the incidence among high-risk population peaks at ages 50 - 64 years following a decline afterward [12-15]. Migrant studies have suggested that Chinese immigrants have higher NPC incidence than other ethnicities in Malaysia and the United States [16, 17]. Genetic, anthropological, and linguistically studies have proposed that NPC might have originated in

the Bai-Yue peoples and was transmitted to the Han Chinese in southern China via intermarriage [18, 19].

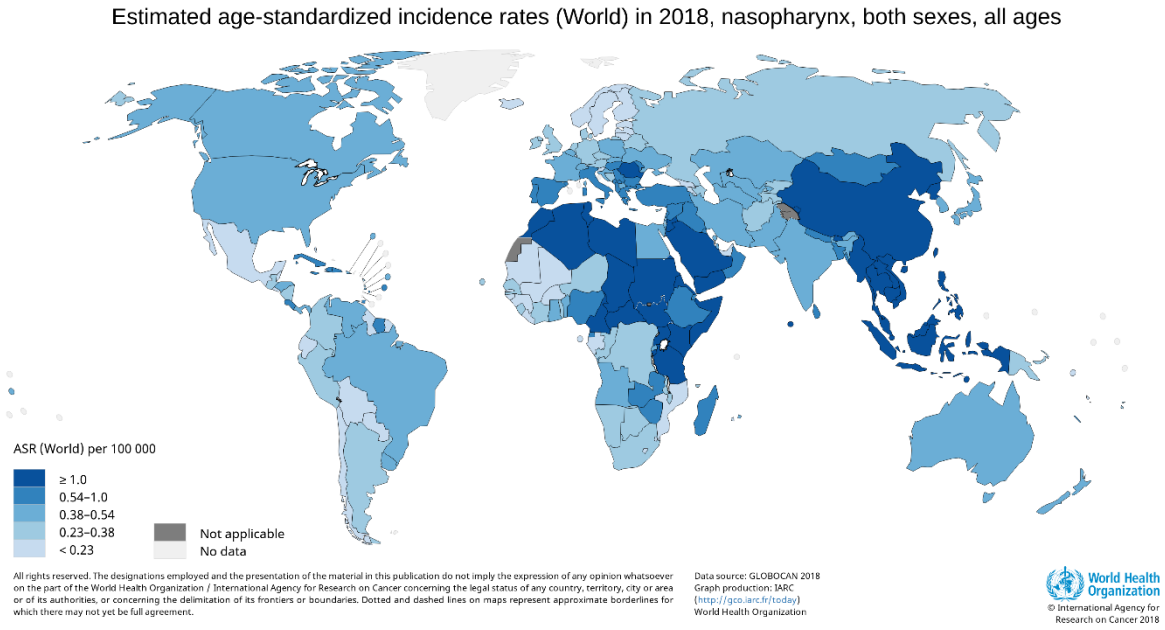
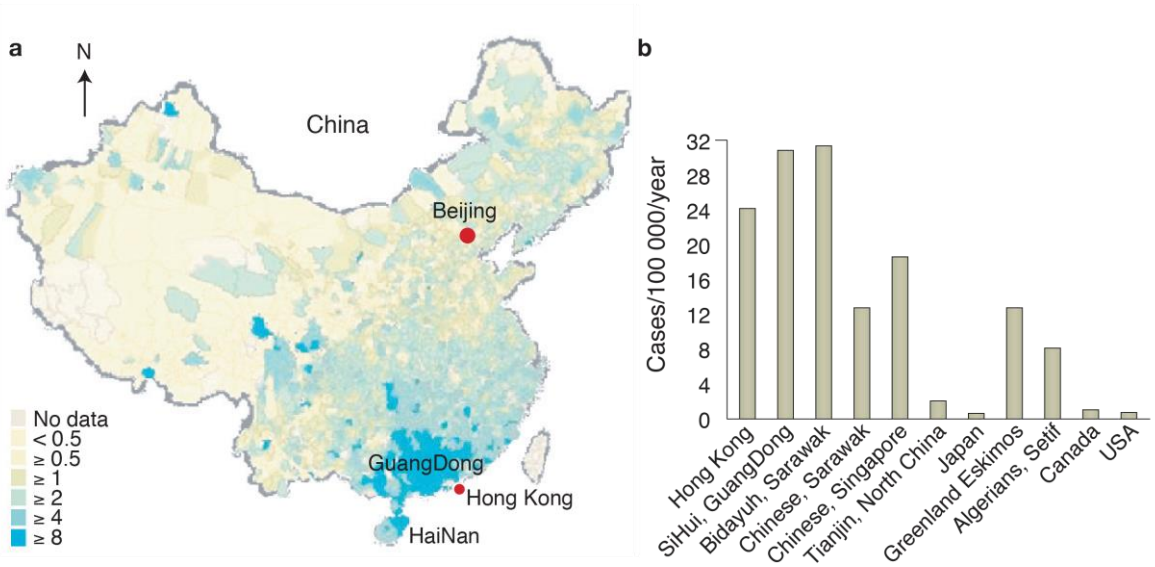


Figure 2.2 Global nasopharyngeal carcinoma incidence: estimated age-standardized incidence rate per 100,000 person-years (both sexes, all age). Data and graph production: GLOBOCAN 2018, IARC [20].



The geographical distribution of nasopharyngeal carcinoma

Expert Reviews in Molecular Medicine © 2007 Cambridge University Press

Figure 2.3 The geographical distribution of nasopharyngeal carcinoma in China (adopted from Tao Q et al. [10]). (a) The mortality rate of NPC among males (world standardized rate; per 100,000 person-years) in China, 1970s (Data source: http://cancernet.cicams.ac.cn). It is in line with incidence of NPC within China, increasing from north to south. (b) The incidence rate (per 100,000 person-years) of NPC in males, from selected areas.

Gradually declining trends of NPC incidence have been observed in Southern and Eastern Asia, North America, and Nordic countries [17, 21-24], which might partly be explained by a better understanding of the disease, rapid economic development, and gradual changes in environment and lifestyle (e.g., declined exposure to traditional preserved-foods and tobacco control) in the last few decades [9, 10, 25]. However, some researchers considered that these results need to be viewed cautiously in the context of population shifting due to constant immigration, especially in the high-risk regions that underwent dramatic economic development in the past 40 years, such as Hong Kong, Singapore, and Guangzhou [10]. Notably, some poorly developed areas (e.g., Sihui in Guangdong, and Cangwu in Guangxi; Figure 2.4) showed generally stable trends from 1980s to 2010s [10, 26, 27]. Nationwide studies analyzing data from more than 250 local population-based cancer registries in China reported that NPC incidence was at a high level in 2013 compared to other countries, and no apparent increasing or decreasing trend was witnessed [11, 12].

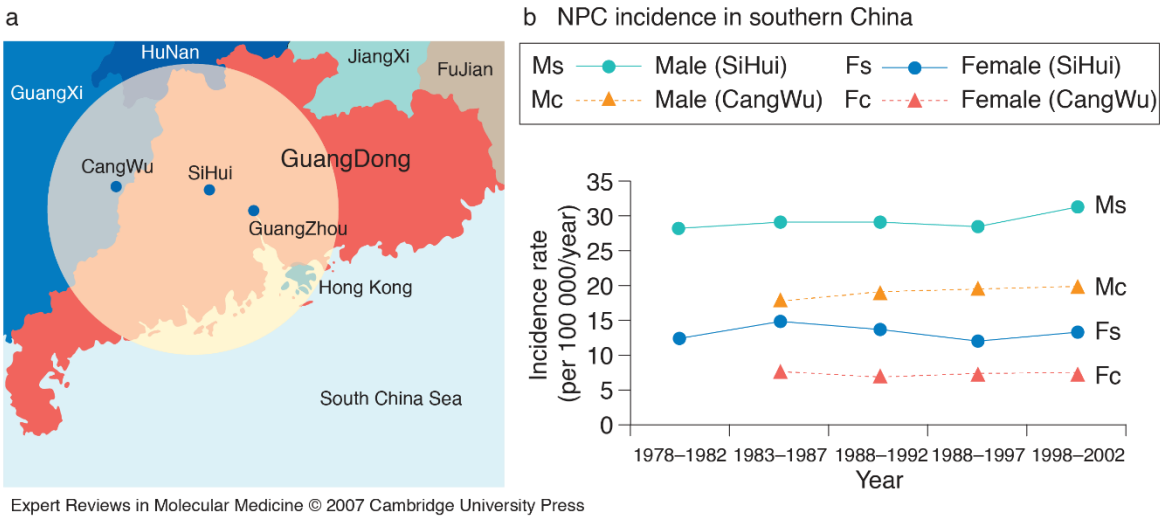


Figure 2.4 The incidence of nasopharyngeal carcinoma in southern China from 1978 to 2002 (figure adopted from Tao Q et al. [10]).

2.2 RISK FACTORS

Around 95% of NPC cases in endemic regions, such as southern China, are non-keratinizing type, which has different pathogenesis and risk factors from the keratinizing type [28-30]. This section will mainly focus on the risk factors of non-keratinizing NPC, including EBV infection, environmental and lifestyle factors, as well as genetic susceptibility.

2.2.1 EBV infection

Epstein-Barr virus, also known as human herpesvirus 4, is a linear double-stranded deoxyribonucleic acid (DNA) virus and the first human tumor virus discovered [31]. It is the most widely spread and persistent virus infection known in humans, with over 90% world population sustaining an asymptomatic life-long infection [32]. EBV primarily infects human

oropharyngeal epithelium via saliva (lytic infection) very early in life with no or minor symptoms. In contrast, a small proportion of people with delayed primary infection may result in infectious mononucleosis [33-35]. In any case, EBV is programmed to be a life-long latent infection in the human memory B-cell pool afterward [36-40].

EBV, including but not limited to its genome, compositions, antigens, is found virtually in all non-keratinizing NPC [41-47] but hardly detected in the normal nasopharyngeal epithelium [47, 48]. EBV-related NPC constitutes more than 95% of the cases in endemic regions, but less than one-third of NPC in non-endemic regions [37, 49-51]. However, the question about how a ubiquitous virus causes endemic cancer only in certain parts of the world has not yet been answered.

The latent infection of EBV in epithelial cells is abnormal [33, 34, 52]. It remains unclear how EBV enters the nasopharyngeal epithelium and establishes a latent infection, as well as how latent EBV contributes to the NPC carcinogenesis. Currently, several key receptors have been found involved in the process of EBV entering (neuropilin 1 [53], epidermal growth factor [54], integrins and non-muscle myosin heavy chain IIA [55], ephrin receptor A2 [56, 57]); two high-risk EBV variants associated with endemic NPC in southern China have also been identified [58]. In addition, researchers have established immortalized nasopharyngeal epithelial cells with type II EBV latency, which may be useful models for the study of EBV infection and its role in NPC development [59].

2.2.2 Environmental and lifestyle factors

2.2.2.1 Dietary factors

The World Cancer Research Fund concludes in their Expert Reports of the Continuous Update Project (CUP) that the consumption of Cantonese-style salted fish is a probable cause of NPC [60, 61]; and the International Agency for Research on Cancer (IARC) has also classified Cantonese-style salted fish as a carcinogen [62, 63]. In the 1960s, researchers first reported an association between Chinese-style salted fish and NPC risk among Tankas (boat dwelling Cantonese) in Hong Kong [64]. Afterward, various epidemiological studies among different populations consistently demonstrated similar associations (in Hong Kong [65]; Guangzhou [66]; Cangwu [67]; Shanghai [68]; Tianjin [69]; Guangdong [70]; Guangxi [71]; Taiwan [72]; southern China [73]; and Malaysian Chinese [74]). These studies also reported stronger associations with NPC risk among people who consumed salted fish during childhood, especially during weaning (salted fish pre-chewed by parents has been used as a traditional weaning food in south China), than for consumption during older ages. Similarly, the CUP Reports indicated strong dose-response associations across different age periods (risk ratio [RR]: 1.28 [1.13-1.44] in adulthood; 1.32 [1.14-1.60] at age 10; 1.42 [1.11-1.81] at ages 0-3) [60, 61]. Cantonese-style salted fish is produced using less salt and a higher degree of fermentation than other types of preserved fish, and is favorable to bacterial growth, such as *Staphylococci*, facilitated by the high outdoor temperatures and air humidity in southern China

[75-77]. It has been found to be rich in N-nitroso compounds that have been associated with cancer development in experimental studies [77-79].

The CUP reports also provide evidence that intake of preserved non-starchy vegetables, processed meat, and red meat are associated with increased risk of NPC, while increased consumption of fresh non-starchy vegetables and fruits is associated with decreased risk of NPC [61, 80-82]. In addition, several studies have suggested that milk consumption may be a protective factor for NPC [83-85], although, conflicting results exist [86, 87].

It is noticeable that, along with the rapid economic development in the endemic areas since the second half of the 20th century, dietary habits have significantly and progressively changed over time: decreased consumption of preserved foods (especially Cantonese-style salted fish), increased intake of fresh foods, and the fact that the habit of weaning with salted fish has become rare [22, 88]. However, this is not reflected in the relatively stable NPC incidence in southern China over the past decades [12, 26, 27]. Theoretically, this may imply that some of the dietary risk factors primarily in early life affect disease development, and an expected decrease in NPC incidence might emerge in about 4 - 5 decades afterwards. Data on NPC incidence from the endemic areas, especially southern China, in the coming two decades is therefore crucial.

On the other hand, given the substitution effects in dietary behavior, high consumption of some food items is generally related to low intake of other food items [89]. Focusing on single food items is insufficient for understanding the potential impact of dietary habits, namely how people choose to combine their foods, on the disease risk. Nonetheless, to our knowledge, the potential association between dietary habits and NPC risk is unclear.

2.2.2.2 Tobacco smoking and alcohol intake

IARC classifies tobacco as a group I carcinogen [63]. Tobacco smoking has been suspected as a risk factor for NPC and investigated by numerous epidemiological studies in both endemic and non-endemic areas [73, 90-96]. However, the results have been contradictory. Recently, two meta-analyses indicated an increased NPC risk among smokers compared to non-smokers, despite substantial heterogeneity among the included studies [97, 98]. Alcohol consumption has also been studied for years, and it seems not to be associated with NPC risk. Although some studies observed a J-shape association [99-101], most observational studies found null results [9, 95, 102-104]. Therefore, meta-analysis using individual participant data from observational studies are warranted to better understand the association of tobacco smoking and alcohol intake with NPC risk.

2.2.2.3 Oral hygiene

Poor oral hygiene has long been implicated as a risk factor for various cancers, such as oral cancer, esophageal cancer, and gastric cancer [105-108]. Poor oral hygiene indicates dysbiosis in the oral microbiome, which has been linked to head and neck cancers, colorectal cancer, and pancreatic cancer [109-112]. A recent study in southern China has shown a positive association

between NPC risk and the number of repaired teeth and NPC risk, while daily frequency of teeth brushing was inversely related with NPC risk [113]. Although little is known about the underlying mechanism, it is suspected that dysbiosis in the oral microbiome indicated by poor oral health might facilitate EBV reactivation via oncogenic metabolites (e.g., butyric acid, nitrosamines), and/or conduce partly to the heavy infiltration of non-malignant lymphocytes around malignant epithelial cells [109, 114-119]. It is noteworthy that several potential NPC risk factors mentioned above, such as tobacco smoking and diet, may affect oral hygiene and oral microbiome [120-124]. Thus, a well-designed, exploratory study is needed to study the association between the oral microbiome and NPC risk, with careful handling of potential confounding.

2.2.2.4 Other factors

Occupational exposure to fumes, smokes, wood/cotton dust, asbestos, and formaldehyde has been linked to increased risk of NPC [67, 73, 125-129], though no consensus has been established. Additionally, living conditions, residential history, use of wood fuel (smoke), education level, and socioeconomic status have also been investigated; however, results vary across studies. One concern is that those factors seem to correlate with each other tightly and are hard to measure properly.

2.2.3 Genetic susceptibility

Genetic susceptibility is another essential factor in NPC etiology, supported by a series of scientific researches addressing NPC's familial clustering since the 1970s. Later on, this phenomenon has been well documented across different populations [130-135]. A recent case-control study in southern China reported a more than fourfold higher NPC risk among individuals with first-degree relatives diagnosed with NPC [136]. Meanwhile, researchers have committed to reveal the basis of familiar clustering by demonstrating the connection between NPC susceptibility and genomic polymorphisms [30, 137, 138]. Numerous studies have reported that the most influential factors were located in the human leukocyte antigen (HLA) region on chromosome 6p21 [139-141]. Recent genome-wide association studies with large sample sizes mainly from the endemic areas have identified more loci (HLA-A, HLA-B, HLA-C, HLA-DQA1, HLA-DQB1, HLA-DRB1, HLADPB1, and HLA-F) strongly associated with NPC risk [140, 142-147]. Benefitting from the substantial development of genomic analysis, several studies from southern China have extensively demonstrated more candidate genes other than HLA loci associated with NPC risk (including *MECOM*, *TNFSFR19*, *CDKN2A/B*, *TERT* [143, 148-150], human macrophage-stimulating 1 receptor (MST1R) [151], and *ITGA9* [152]). To date, the search for determinants of NPC susceptibility and the investigation of the biological mechanism are ongoing to understand NPC's genetic susceptibility fully.

2.2.4 Early-life exposure

The familial aggregation of NPC might result from inherited genetic susceptibility and shared environmental and lifestyle factors (e.g., living conditions, diet, oral health, and socioeconomic status), respectively, or jointly. As previously described, researchers have consistently

observed a stronger association of NPC risk with salted-fish intake during an earlier lifetime. Some have suggested that the weaning habit might be a critical mechanism for EBV's initial infection in China [153]. Moreover, exposure to passive smoking during childhood has been reported to be independently associated with increased NPC risk in never-smoking men and women [96]. Accordingly, it is reasonable to hypothesize that a person carrying high-risk HLA allele(s), infected with high-risk EBV strains (most likely caught it from family member) in early life, harboring a certain pattern of oral/nasopharyngeal microbiome, exposed to specific carcinogen/chemical (such as N-nitrosamine rich in salted fish or other preserved foods), may have a substantially higher risk of non-keratinizing NPC in endemic regions.

However, numerous open questions remain: is a person carrying high-risk HLA allele(s) especially susceptible to high-risk EBV strains? Does the oral/nasopharyngeal microbiome contribute to EBV genotype selection pressures? Does the oral/nasopharyngeal microbiome or specific chemicals contribute to EBV entry into epithelial cells and the establishment of latent infection? Are tobacco smoking, preserved food consumption, and dietary habits involved in NPC risk or general cancer risk?

Nevertheless, to test the hypothesis of early-life exposure and to address all these questions, a well-designed, large-scale birth cohort with more than 50 - 60 years of follow-up in endemic areas is needed, which is generally difficult, if not impossible. Besides, the lack of a feasible and reliable non-keratinizing NPC model is another difficulty and hamper.

2.3 CLINICAL CHARACTERISTICS

Because of the deep-seated anatomical location, most NPC patients present non-specific symptoms when the disease is initiated. NPC’s clinical presentation is correlated with the anatomical structures being invaded as the primary tumor growth and varies accordingly (Figure 2.1) [5, 25]. NPC usually occurs in the pharyngeal recess and invades surrounding structures. Thus, patients can develop epistaxis (post-nasal drip with blood), nasal obstruction, and discharge when the tumor presents in the nasopharynx and/or anteriorly spreads to the nasal cavity, pterygoid fossa, and maxillary sinuses. Dysfunction of the eustachian tube (e.g., tinnitus, secretory otitis media, and deafness) indicates the paranasopharyngeal space invasion. Palsy of cranial nerves (the III, V, VI, XII cranial nerves being most affected), presenting as diplopia, facial pain and numbness, and headache, implies an advanced disease locally in which the tumor superiorly and posteriorly erodes the base of the skull, clivus, and intracranial structures. Neck masses, usually appearing in the upper neck, point to a regional spread of the tumor. In the clinic, painless neck mass is the most common symptom among newly diagnosed NPC patients (76%), followed by nasal dysfunction (73%), aural dysfunction (62%), headaches (35%), and diplopia (11%) [154]. Around 80% of the patients present with multiple symptoms, while 20% report a single symptom or no symptom (1%).

Clinical assessments shall be administrated to patients with symptoms. Nasopharyngoscopy is routine for detecting a suspected nasopharyngeal mass; meanwhile, an endoscopic biopsy is necessary for definitive histopathological diagnosis of NPC. Imaging modalities, such as

magnetic resonance imaging (MRI), computed tomography (CT), and ^{18}F-fluorodeoxyglucose-positron emission tomography/computed tomography (^{18}F-FDG-PET/CT), are commonly used for assessing the tumor size and location and evaluating the stage of NPC.

The Union for International Cancer Control /American Joint Committee on Cancer (UICC/AJCC) TNM (tumor-node-metastasis) staging system, 8th edition (in 2016), is the most important and widely used staging guideline for NPC [155-157]. Given the deep-seated anatomical location and aggressive behavior, more than 75% of NPC patients have advanced disease with either local or regional spread (stage III and IV) and around 10% with distant metastasis at diagnosis [13, 29, 30, 154, 158], making effective treatment challenging.

2.4 MANAGEMENT AND PROGNOSIS

Primary NPC is radiosensitive. Radiation therapy (RT) is the mainstay of curative-intent treatment for non-disseminated NPC, instead of surgery and chemotherapy [13]. At present, most cancer centers have achieved five-year overall survival (OS) rates around 60%−80% for non-metastatic patients with radiotherapy +/- chemotherapy. However, around 30% of NPC patients will suffer from distant metastasis and resistant/recurrent disease, resulting in a poor prognosis with 5-year OS rate lower than 40%.

2.4.1 Treatment for non-metastatic NPC

Currently, intensity-modulated radiotherapy (IMRT) is the most widely used radiation technique [13]. IMRT provides full coverage of the complex-shaped gross tumor and subclinical diseases with precise dose delivery while minimizing the dose to critical organs at risk by controlling the intensity of the radiation beam. IMRT is superior to conventional radiotherapies (e.g., two-dimensional radiotherapy, three-dimensional conformal radiotherapy) with significant improvement in survival and reduction of radiation-related toxicities among non-metastatic NPC patients [159-169]. At present, there is little doubt that IMRT alone is the standard of care for stage I NPC (Table 2.1), with expected 5-year disease-specific survival (DSS) and OS over 95% [170].

The use of chemotherapy as part of the management for advanced locoregionally diseases, especially concurrent chemotherapy during radiotherapy, is another crucial progress. Compelling and consistent evidence from numerous clinical trials and meta-analyses has been demonstrated the survival benefit of concurrent chemoradiotherapy among NPC patients with stage III to IVb disease, with absolute survival benefits of around 6% at 5 years [171-179]. However, the benefits of adjuvant and induction chemotherapy as a combination treatment for stage II to IVb disease, beyond the concurrent chemoradiotherapy, are still being discussed [25, 30, 178-180]. Current viewpoints propose that adjuvant chemotherapy cannot lead to better survival alone and may not achieve further survival benefits incorporating concurrent chemoradiotherapy, while induction chemotherapy, on top of concurrent chemoradiotherapy, may contribute to improving distance control and survival in the IMRT era [13]. Nonetheless, the latest version of the National Comprehensive Cancer Network (NCCN) Guidelines (Version 2.2020; Jun 9, 2020) recommends concurrent chemoradiotherapy ± sequential

chemotherapy for a broad range of NPC diseases: from stage II to IVb (Table 2.1) [181]. It means that the decision-making for patients with stage II to IVb diseases depends primarily on experienced experts or the clinical trials they participated in (if available). The ambiguous recommendations for this group of NPC patients remain to be improved by a better prognostic risk stratification model and therapeutic monitoring for providing more accurate clinical practice guidelines.

Table 2.1 Treatment strategies for NPC with different stages

Stage (8th ed. [a])	**NCCN (V2.2020 [b])**
Stage I	Radiotherapy alone
Stage II Stage III Stage IVA - B	Radiotherapy plus concurrent chemotherapy + adjuvant chemotherapy (2A [c]) or induction chemotherapy + concurrent chemotherapy (2A) or concurrent chemotherapy (2B)
Stage IVC	Chemotherapy alone or radiotherapy + chemotherapy

[a] Stage system: American Joint Committee on Cancer (AJCC) – TNM Staging System (8th ed. 2016)

[b] National Comprehensive Cancer Network (NCCN) Guidelines Version 2.2020

[c] NCCN Categories (1-3) of Evidence and Consensus

2.4.2 Management of residual/recurrent and metastatic NPC

Residual/recurrent and metastatic diseases in NPC are not uncommon and present poor outcomes [164-169, 182-185]. Thus, post-treatment surveillance is essential and critical for NPC patients. The first clinical check-up for evaluating a patient's treatment response usually takes place 10-12 weeks after the completion of radiotherapy, including physical examination, nasopharyngoscopy, and radiation imaging tests (MRI/CT/PET-CT, etc.) [30, 186, 187]. NCCN guidelines provide detailed follow-up strategies.

After primary treatment, around 10-15% of NPC patients will experience residual/recurrent disease, and up to 20% will develop distant failure [165, 167, 168, 182-185]. It has been reported that up to 40% of the T4 disease experienced treatment failure [165, 188-190]. Salvage surgery is recommended for residual neck mass and a resectable tumor which relapses within one-year after primary radiotherapy [191, 192]. Tumors recurring more than one year after initial treatment are considered for re-irradiation [193]. Taking systematic therapy (cisplatin/gemcitabine are first-line regimens) or participating in clinical trials may otherwise be regarded [181]. Current studies report a similar 5-year OS rate of 40% among NPC patients treated with neck dissection or re-irradiation [192, 193]. Nonetheless, aggressive salvage treatments involve inevitably acute and late toxicities, impacting the survivors' quality of life significantly. On the other hand, it is recommended for patients with metastatic disease to participate in feasible clinical trials, as preferred. Palliative chemotherapy is alternatively

recommended with cisplatin plus gemcitabine as first-line regimens. Prognosis of metastatic NPC varies greatly with an OS of approximately 12 to 20 months and a 5-year OS rate lower than 40% [194-197].

Tumor immunotherapy is a rising star in cancer treatment. Theoretically speaking, NPC patients may be eligible for immune checkpoint blockade therapies due to high programmed death-ligand 1 (PD-L1) expression and abundant infiltration of non-malignant lymphocytes around the primary tumor [198-201]. Several single-arm trials evaluating immunotherapy targeting the PD1/PD-L1 pathway in recurrent/metastatic NPC patients have shown promising outcomes [202-204]. Some ongoing phase 2/3 trials, investigating the immunotherapies targeting EBV and/or PD1/PD-L1 on managing locoregional, recurrent, and metastatic NPC, are anticipated to provide useful evidence [13].

2.4.3 Prognostic factors

The anatomical TNM stage is considered the pivotal factor for guiding treatment strategy and prognostic stratification. Besides, compelling evidence has shown that a pretreatment EBV-DNA test can provide further prognostic information and assist the assessment of treatment response; meanwhile, a post-treatment EBV-DNA involves in the surveillance of locoregional failure and distant metastasis in a high-risk population [205-211]. Currently, the NCCN guideline (Version 2.2020) recommends considering plasma EBV-DNA for therapeutic monitoring. However, several vital concerns remain: EBV-DNA is not 100% detectable among NPC patients [212]; large variabilities in the detection of plasma EBV-DNA have been reported across skilled laboratories [213] and a harmonized quantitative protocol/assay for EBV-DNA detection is not yet available; plasma EBV-DNA levels may be affected by EBV-infected B cells [30]. There is a clear need to search for reliable and effective prognostic markers in NPC management.

Recently, the human commensal microbiome has been reported to be shaped by anticancer treatment and to modulate the host's clinical response to treatment (e.g., CpG-oligonucleotide, gemcitabine, and immune checkpoint inhibitors) [214-218]. Two recent studies demonstrated severe compositional and functional imbalance in the gut and oral microbial communities after radiotherapy among cancer patients, which may contribute to the pathogenesis of therapy-induced gastrointestinal and oral mucositis [219, 220]. Similar to the intestinal tract and oral cavity, the human nasopharynx harbors diverse microbes. It is reasonable to suspect that aggressive radiotherapy might induce a dysbiosis of the nasopharyngeal microbiome among NPC patients. Yet, no data is available, and little is understood whether the commensal microbiome is associated with the host's response to radiotherapy [221].

In the modern IMRT era, a promising prognosis can be achieved by incorporating chemotherapy among the majority of NPC patients, leaving 30% of them suffering from poor outcomes due to treatment failure. Therefore, it is essential to build an effective strategy based on crucial prognostic stratification and therapeutic monitoring in the initial treatment to improve therapeutic outcomes and avoid unnecessary salvage.

3 AIMS AND RESEARCH QUESTIONS

By studying a range of topics concerning nasopharyngeal carcinoma in the endemic area, this thesis aimed to provide more insights into the disease in terms of risk factor and prognosis. Specifically, we investigated whether dietary habits were associated with NPC risk in southern China, using a population-based case-control study. Among a subset of the case-control study, we studied whether oral microbiome was associated with the risk of NPC. We also conducted a longitudinal hospital-based NPC cohort study in southern China to deliver a proof-of-concept data on the patterns of the commensal microbiome in patients' nasopharynx during radiotherapy and their role in NPC prognosis (Figure 3.1).

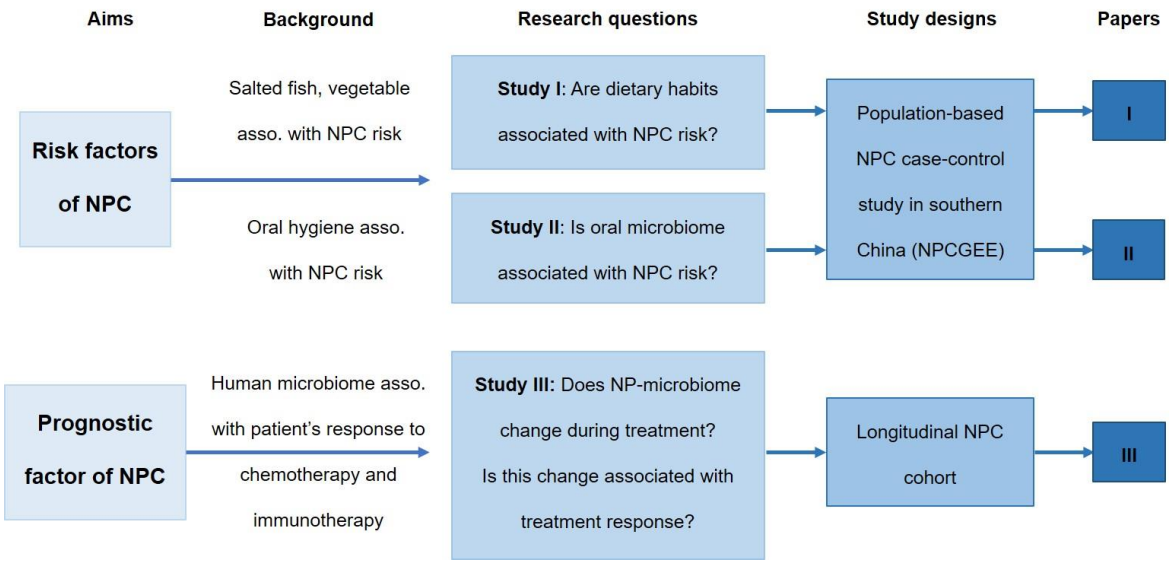


Figure 3.1 Overview of three studies in this thesis. (*NPC, nasopharyngeal carcinoma; NP, nasopharyngeal; asso, be associated*)

4 METHODS & MATERIALS

Table 4.1 shows an overview of the study designs, study population, main measurements, and statistical methods used for each paper in the thesis.

Study	Study design	Study population	Main measures	Statistical methods
Study I/ Paper I	Population-based case-control study	NPCGEE project: People (aged 20-74, lived in Wuzhou, Guiping/pingnan, and Zhaoqing in Southern China) diagnosed with NPC during 2010-2013, and their frequency matched controls (by age, sex, and residential region)	Diet intakes in: 10 years before interview; aged 16-18 yrs	Principal component analysis Unconditional logistic regression
Study II/ Paper II		Wuzhou subset derived from the NPCGEE project	Oral microbiome (saliva collected during interview)	Alpha/beta diversities PCoA, PERMANOVA Poisson regression, Phylofactor, ANCOM Network analysis
Study III/ Paper III	Prospective cohort study with repeated measurement	62 newly diagnosed NPC patients who were treated with RT in GXMU during 2014-2015	Nasopharyngeal microbiome (swabs successively collected over the course of RT)	Alpha/beta diversities PCoA, Volatility analysis Linear mixed effects model NMIT, PERMANOVA SS-ANOVA, ANCOM

Table 4.1 Overview of methods and materials of each paper in this thesis. (*NPCGEE: Gene-environment Epstein-Barr Virus Interactions in the Etiology of nasopharyngeal carcinoma; NPC: nasopharyngeal carcinoma; PCoA; principal coordinates analysis; PERMANOVA: permutational multivariate analysis of variance; ANCOM: analysis of composition of microbiome; GXMU: Guangxi Medical University; RT: radiation therapy; NMIT: non-parametric microbial interdependence test; SS-ANOVA: smoothing-spline analysis of variance*).

4.1 STUDY DESIGN, STUDY POPULATION, AND DATA COLLECTION

4.1.1 Population-based NPC case-control study in southern China

A population-based NPC case-control study, entitled "Gene-environment Epstein-Barr Virus Interactions in the Etiology of nasopharyngeal carcinoma (NPCGEE) project," was launched in southern China, 2010 [222]. People officially residing in 13 cities/counties in Wuzhou, Guiping, and Pingnan areas in Guangxi Autonomous Region, and Zhaoqing area in Guangdong Province, were defined as study area for the NPCGEE project (Figure 4.2). From 2010 to 2013, 3,047 newly diagnosed NPC patients aged 20-74 were identified and invited; 83.8% of them

(2,554) participated. The enrollment was closed by November of 2014, when 2,648 (82.7%) controls consented to participate, among 3,202 who were frequency-matched with the NPC cases on age (in five-year groups), sex, and residential region.

A team of trained interviewers performed face-to-face interviews with all NPC cases and 95% population controls (5% interviewed by phone) using a computerized questionnaire [222]. During the structured interview, information on demographics and known potential NPC risk factors, including family history of cancer [136], residential history, occupational history, smoking habits [96], alcohol drinking, herbal tea and soup consumption [223], and dietary habits was collected [224]. Biosamples (including blood and saliva) were collected at the interview (not available for the participants interviewed by phone). Specifically, participants were required not to eat or chew gum 30 minutes before saliva collection. Saliva samples (2ml) were collected into 50ml tubes with a Tris-EDTA buffer, transferred by cold-chain (-20 ºC) within 3 days, and stored at -80 ºC until further use.

In total, 2,554 NPC cases and 2,648 controls were included in Study I, while a subset from Wuzhou area (Wuzhou subset: 532 cases and 534 controls) was included in Study II.

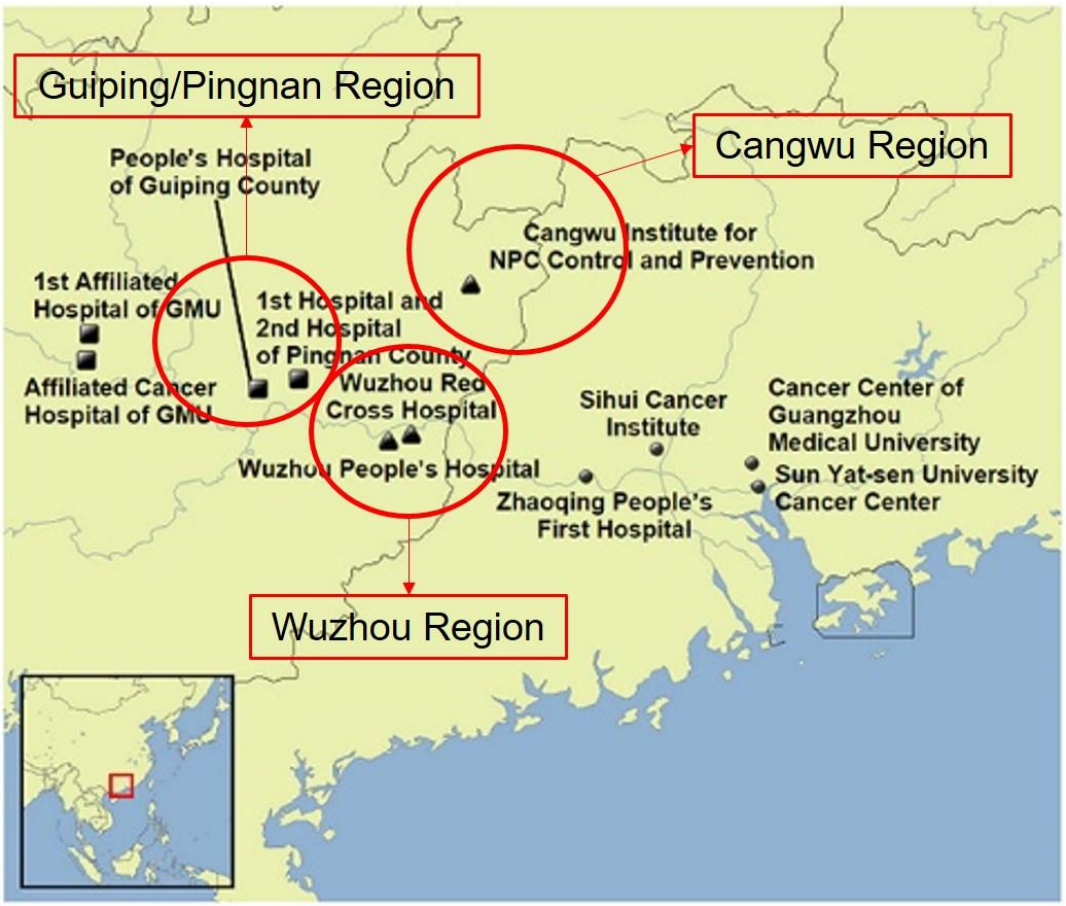


Figure 4.2 Three study regions of the NPCGEE project (adapted from Ye W et al. [222]).

4.1.2 Prospective NPC patient cohort

A hospital-based prospective cohort of NPC patients was recruited in the First Affiliated Hospital of Guangxi Medical University, Guangxi Autonomous Region, China. Between 2014 and 2015, 62 newly diagnosed, treatment naïve, non-metastatic NPC patients were enrolled in Study III (Figure 4.3). All of them were treated according to the standard of care (radiotherapy ± concurrent chemotherapy ± adjuvant/induction chemotherapy ± other treatments; according to NCCN guidelines). During radiotherapy (7 to 8 weeks), successive nasopharyngeal swabs were repeatedly collected, specifically, starting from treatment-naïve (S0), twice a week afterwards until the completion of radiotherapy, and at the completion of radiotherapy (SE). The sampling time points correspond to an accumulated dose of radiation.

The first clinical check-up evaluating patients' treatment response (based on the response evaluation criteria in solid tumors, RECIST) [225, 226] was performed three months after the completion of radiotherapy.

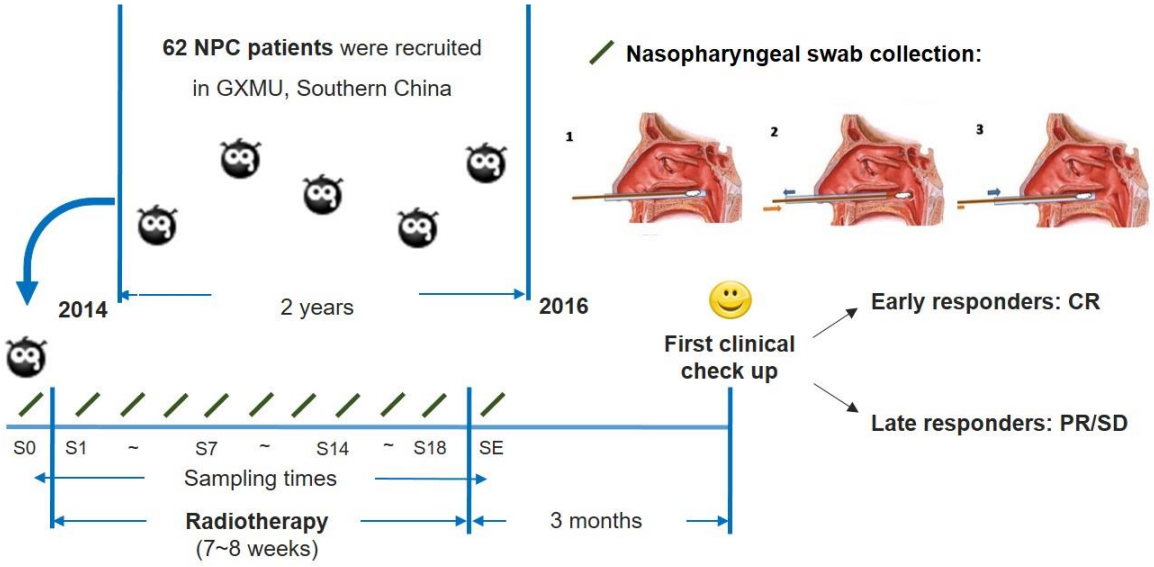


Figure 4.3 Study design and sample collection in the prospective cohort of nasopharyngeal carcinoma patients. (*GXMU, Guangxi Medical University; CR, complete response; PR, partial response; SD, stable disease*)

4.2 MAIN MEASUREMENTS

4.2.1 Dietary intakes

The association between dietary intakes from two life periods and NPC risk was investigated in Study I. During a structured interview, dietary intake information of all participants in the NPCGEE project was assessed by a food frequency questionnaire (FFQ) including 77 of the most commonly consumed foods in the study area (estimated to cover ca. 80% of total energy intake, TEI) [224]. Study participants were asked to provide the frequency of each food item (daily, weekly, monthly, yearly, or never) and its estimated portion size (converted into gram

later) in two life periods: 10 years before the interview (i.e., adulthood) and at ages 16-18 years (i.e., adolescence, for subjects aged above 35 years at the interview). Interviewers used an illustrated booklet with pictures of serving sizes of various food items to assist participants in estimating the portion size correctly. Individual alcohol consumption and household cooking oil intake were collected only 10 years ago, not in adolescence. Therefore, the daily TEI in adulthood was calculated from the 77 food items, cooking oil (on household average), and alcohol intake if applicable. Daily TEI in adolescence was solely considered energy from the 77 food items. Chinese food composition tables were used as the base of energy estimation [227].

4.2.2 Oral microbiome

A total of 1066 saliva samples derived from the Wuzhou subset of the NPCGEE project were used for evaluating the oral microbiome among NPC cases and controls in Study II.

We used a conserved marker gene, namely 16S ribosomal ribonucleic acid (rRNA) gene, which contains one or more highly variable regions as "fingerprint identification" to profile the oral microbial communities [228]. Lab processing included DNA extraction and targeted 16S rRNA amplicon library preparation, which was implemented in collaboration with Guangxi Medical University (China). Salivary DNA was extracted using a protocol with two steps: 1) processed by lysozyme and bead beating, 2) applied to the TIANamp blood DNA kit (Tiangen, Beijing, China). The 16S rRNA library was prepared using a two-step polymerase chain reaction (PCR) strategy: 341F/805R primers (341F-*CCTACGGGNGGCWGCAG*, 805R-GACTACHVGGGTATCTAATCC) targeting the variable region 3 to 4 were used for PCR-round 1, and the sample-specific barcode was added during PCR-round 2 [229, 230]. The library cleanup was performed using an Agencourt AMPure XP purification kit. Negative controls (water control) and mock communities (known bacterial DNA) were used in DNA extraction and 16S library preparation. The quality and quantity of cleaned libraries were measured on an Agilent 2100 Bioanalyzer system and real-time PCR. In total, 1080 sample-libraries were submitted to the Beijing Genome Institute (BGI) using a 2x300-bp paired-end strategy on an Illumina MiSeq.

Afterward, the raw amplicon-sequences from 1066 samples were processed to classify the oral microbiome into a feature table, taxonomy, and phylogeny (Figure 4.4). In brief, we used Deblur to resolve sequencing data into exact sequence feature (amplicon sequence variants, ASVs) based on sequencing error profiles [231]. For taxonomy (a taxonomic name is assigned to an ASV), we used a naïve Bayesian classifier trained against the August 2013 Greengenes database (q2-feature-classifier in QIIME 2, 2018 November release) [232-234]. Furthermore, the same reference was used to build a phylogenetic tree using fragment insertion (99% identity tree backbone with q2-fragmentinsertion) [235].

4.2.3 Nasopharyngeal microbiome

Eight hundred seventy nasopharyngeal swabs were collected from 62 NPC patients for quantifying the longitudinal patterns in the nasopharyngeal microbiome during radiotherapy-

based treatment in Study III. The same lab procedure and bioinformatic workflow for sample preparation and feature extraction as in Study II was used (Figure 4.4).

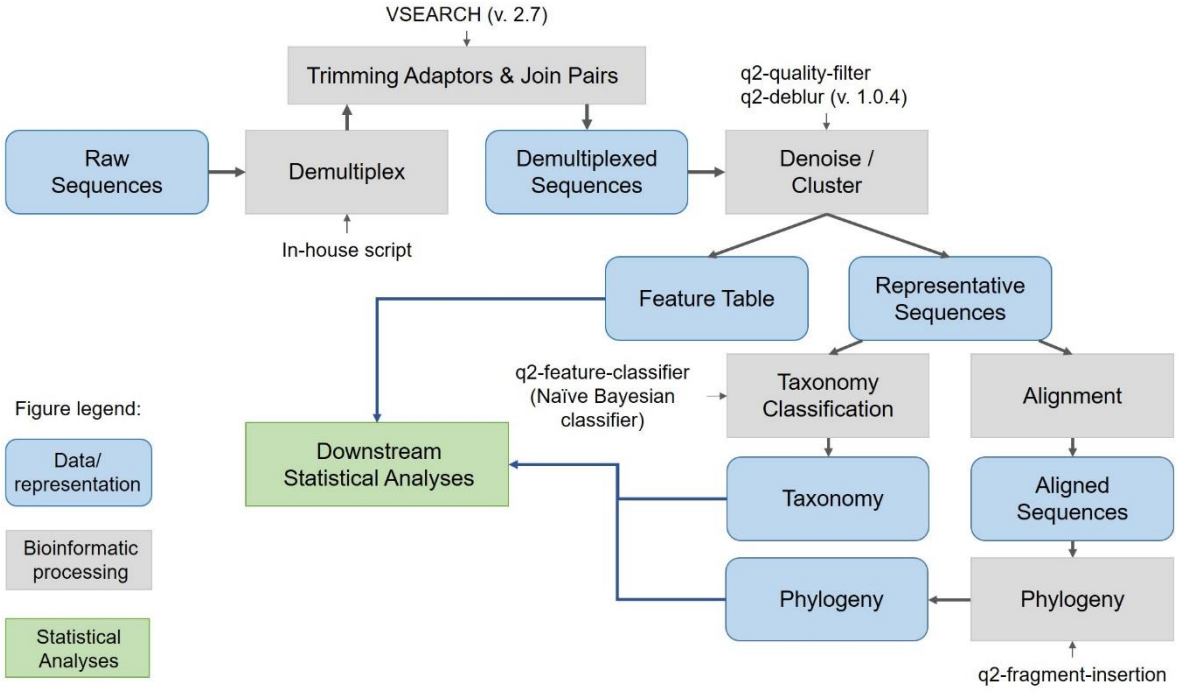


Figure 4.4 Overview of the bioinformatics workflow in Study II and Study III.

4.2.4 Treatment outcome of NPC patients

In Study III, all NPC patients received radiotherapy, with or without sequential chemotherapy. All patients were recommended for a clinical check-up at three months after the completion of radiotherapy to assess their response status using RECIST criteria (by three oncologists). A short-term clinical outcome in this study was defined as: a patient who achieved a CR at the check-up was an early responder to radiotherapy-based treatment, otherwise a late responder.

4.3 STATISTICAL ANALYSES

4.3.1 Dietary patterns and NPC risk

In Study I, principal component analysis (PCA) was implemented to identify underlying dietary patterns. First, the 77 individual food items were classified into 18 food groups based on food group characteristics, similarity of nutrient contents, and prior knowledge [236, 237]. The sum of the absolute intake of each food group was calculated as the intake of the corresponding food items (in gram). In the PCA procedure, principal components were selected based on a scree plot of the eigenvalues. Component scores were computed for each study subject for downstream analysis. Dietary components were identified separately for the two study periods (adolescence and adulthood). Iterated principal factor analysis was applied as sensitivity analysis.

The associations between the identified dietary patterns and NPC risk were estimated by odds ratios (ORs) and the corresponding 95% confidence intervals (CIs) derived from unconditional logistic regression models, in Study I. Each component score was transformed to quartiles before being entered into the regression models; all identified components were included in the models simultaneously. Variables of age (in 10 years group), sex, and residential region were considered in the minimally adjusted models; additionally, multiple covariates known or suspected to affect both diet and NPC risk based on prior knowledge or findings from the NPCGEE project were also included in the fully adjusted models. The ORs and 95% CIs were estimated separately for the two study periods.

A subset of 4215 subjects who provided dietary information in both adolescence and adulthood were further investigated in a joint analysis. All dietary components classified in both periods were included in the models simultaneously. The same sets of covariates were used as before. Moreover, a variable representing birth before or after year 1963 (according to the median age of the study population) was used to evaluate potential effect modification by age. This variable was included in regression models together with an interaction term.

4.3.2 Oral microbiome and NPC risk

The statistical analyses used to study the association between oral microbiome and NPC status in Study II are illustrated in Figure 4.5.

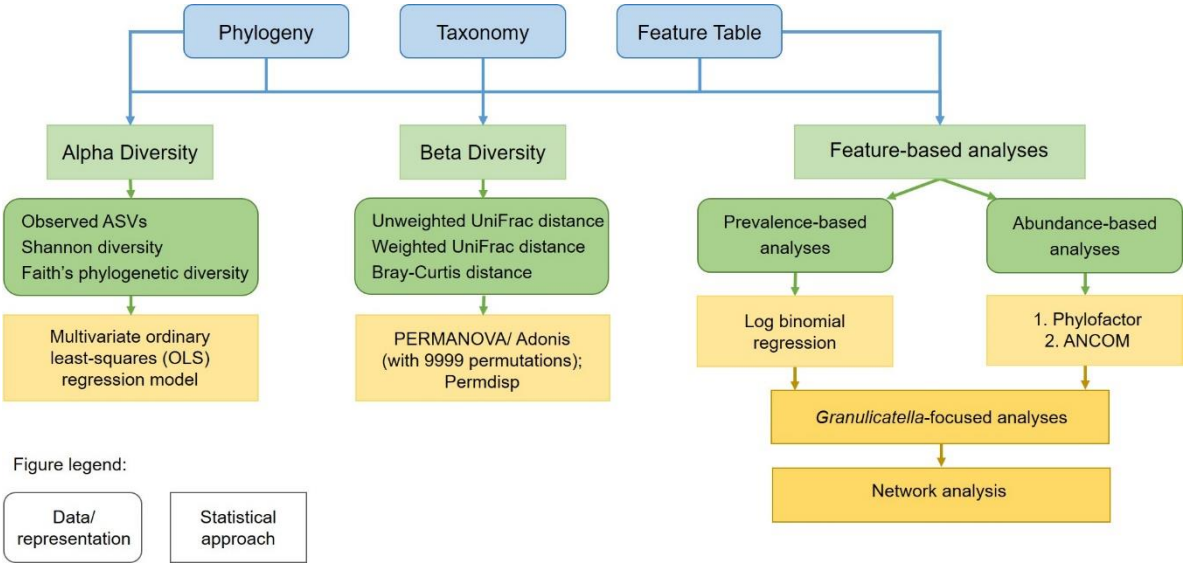


Figure 4.5 Biostatistics workflow in Study II.

4.3.3 Longitudinal patterns of the nasopharyngeal microbiome among NPC patients during radiotherapy

Figure 4.6 shows the statistical analyses used for profiling the change of nasopharyngeal microbiome among NPC patients during radiotherapy-based treatment and for estimating the association between the change and patients' short-term outcomes in Study III.

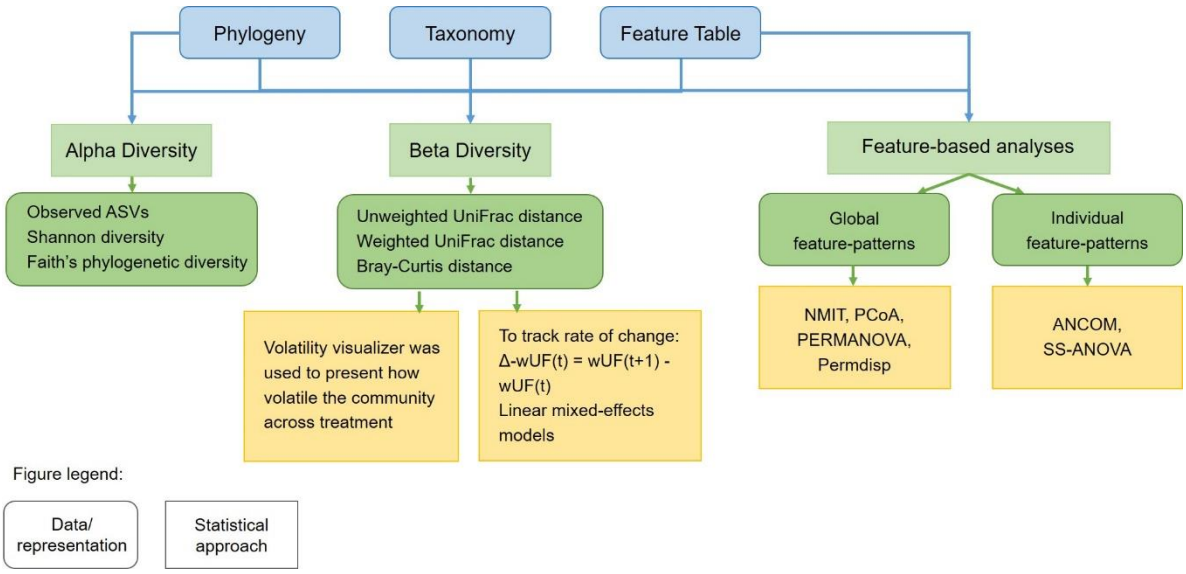


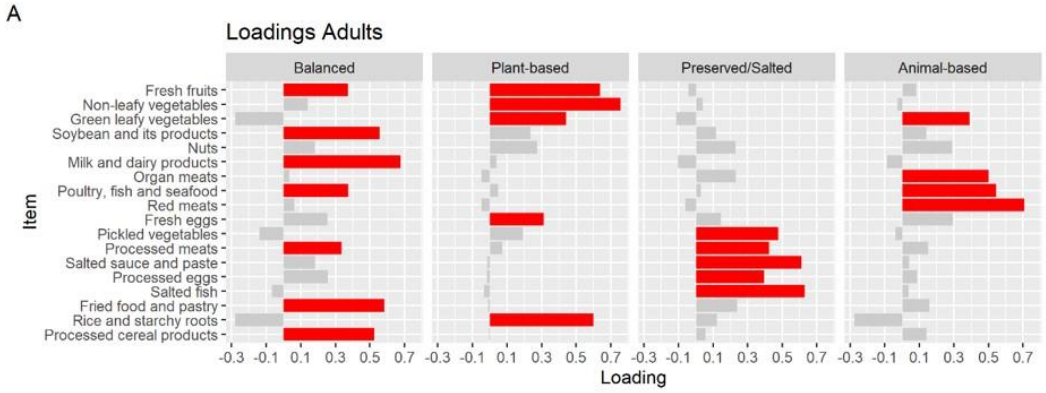
Figure 4.6 Biostatistics workflow in Study III.

5 MAIN FINDINGS

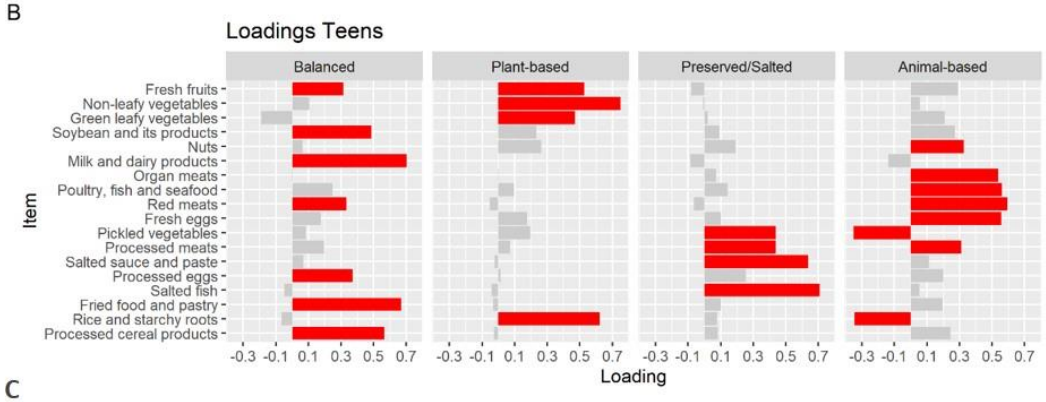
5.1 DIETARY PATTERNS ARE ASSOCIATED WITH NPC RISK

We analyzed a total of 4398 study participants (2174 NPC cases and 2224 controls) with adolescence dietary data and 4832 participants (2387 NPC cases and 2445 controls) with adulthood dietary data in Paper I. In both analytical periods, we characterized four dietary components among controls and summarized them as four dietary patterns: “balanced diet”, “plant-based diet”, “preserved/salted diet”, as well as “animal-foods-based diet” (Figure 5.1).

A



B



C

Dietary Pattern	High Intake	Low Intake
Balanced diet	Milk and dairy products; fried food and pas-tries; soybean and its products; processed cereal products; poultry, fish, and seafood; fresh fruits; processed meats	Rice and starchy roots; green leafy vegetables
Plant-based diet	Non-leafy vegetables; fresh fruits; Rice and starchy roots; green leafy vegetables	-
Preserved/salted diet	Salted fish; salty sauce and pastes; pickled vegetables; processed meats; processed eggs	-
Animal-foods-based diet	Red meats; poultry, fish, and seafood; organ meats; fresh eggs; nuts	Rice and starchy roots

Figure 5.1 Component loadings of dietary patterns during adolescence and adulthood

In Table 5.2, the ORs of four dietary patterns derived from fully adjusted models in two life periods are listed. Among participants who had the highest quartiles of the “plant-based diet”,

a 34% decreased NPC risk in adolescence ($OR_{q4\ vs.\ q1}$ = 0.66, 95% CI = 0.52 – 0.84, P_{trend} = 0.0067) and a 50% decreased risk in adulthood ($OR_{q4\ vs.\ q1}$ = 0.50, 95% CI = 0.41 – 0.63, P_{trend} < 0.0001) were observed. On the contrary, increased NPC risk was found along with increased quartiles of the "animal-foods-based diet" in both periods, although the effect in adolescence was smaller than in adulthood (adolescent: $OR_{q4\ vs.\ q1}$ = 1.39, 95% CI = 1.13 – 1.70, P_{trend} = 0.0002; adulthood: $OR_{q4\ vs.\ q1}$ = 2.26, 95% CI = 1.86 – 2.75, P_{trend} < 0.0001). Weak associations were observed between NPC risk and the "preserved/salted diet", while no evident association was found between the NPC risk and the "balanced diet". Besides, we studied 4215 participants with diet information in both period in a joint analysis which included eight age-specific dietary patterns simultaneously. Remarkably, the OR estimates of adulthood diet in association with NPC risk remained largely unchanged after adjusting for adolescent diet (for "plant-based diet": $OR_{q4\ vs.\ q1}$ = 0.49, 95% CI = 0.37 – 0.65; for "animal-foods-based diet": $OR_{q4\ vs.\ q1}$ = 2.20, 95% CI = 1.75 – 2.77). In contrast, following adjustment for adulthood diet, the previous associations of NPC risk and adolescent dietary patterns were moderated.

Table 5.2 Odds ratios of nasopharyngeal carcinoma by quartiles intake of four dietary patterns in adolescence and adulthood

	Adolescence [a] (*N*=4398)		**Adulthood [b] (*N*=4832)**	
Dietary pattern	**OR (95% CIs)**	**P_{trend}**	**OR (95% CIs)**	**P_{trend}**
Balanced diet				
Quartile 1 (ref.)	1.0 (ref.)	0.3245	1.0 (ref.)	0.9901
Quartile 2	1.03 (0.86-1.25)		0.95 (0.80-1.13)	
Quartile 3	1.06 (0.88-1.28)		1.01 (0.84-1.21)	
Quartile 4	0.84 (0.68-1.03)		0.98 (0.80-1.19)	
Plant-based diet				
Quartile 1 (ref.)	1.0 (ref.)	0.0067	1.0 (ref.)	<.0001
Quartile 2	0.77 (0.64-0.93)		0.72 (0.61-0.86)	
Quartile 3	0.81 (0.66-0.99)		0.63 (0.53-0.77)	
Quartile 4	0.66 (0.52-0.84)		0.50 (0.41-0.63)	
Preserved/salted diet				
Quartile 1 (ref.)	1.0 (ref.)	0.0011	1.0 (ref.)	0.0207
Quartile 2	0.95 (0.79-1.14)		0.76 (0.64-0.90)	
Quartile 3	1.17 (0.97-1.42)		1.01 (0.85-1.19)	
Quartile 4	1.31 (1.08-1.59)		1.13 (0.94-1.34)	
Animal-foods-based diet				
Quartile 1 (ref.)	1.0 (ref.)	0.0002	1.0 (ref.)	<.0001
Quartile 2	0.98 (0.81-1.19)		1.16 (0.97-1.39)	
Quartile 3	1.33 (1.10-1.61)		1.37 (1.14-1.65)	
Quartile 4	1.39 (1.13-1.70)		2.26 (1.86-2.75)	

Abbreviations: OR, odds ratio; CIs, confidence intervals; ref., reference.

[a] Adolescence analysis: estimates from multivariate logistic regression models were adjusted for age (10-years groups), sex, residential area, body mass index, total energy intake, education level, current housing type, current occupation, smoking status, NPC history among first-degree relatives, number of filled tooth, frequency of tooth-brushing, frequency tea consumption, weekly frequency of soup consumption, and herbal tea consumption, and quartiles of the daily total energy intakes in adolescence. Results referred to the composite model fitting all the four dietary components simultaneously.

[b] Adulthood analysis: estimates from multivariate logistic regression models were adjusted for the same set of co-variates as in adolescence, the quartiles of adulthood TEI was used instead of adolescent TEI.

5.2 ORAL MICROBIOME IS ASSOCIATED WITH NPC RISK

We analyzed 994 participants (499 NPC cases, 494 controls) enrolled in the Wuzhou region of the NPCGEE project in Paper II [238].

In the global comparisons of alpha diversities regarding the mean species diversity in a community, we observed significantly fewer amplicon sequence variants (ASVs) (*P*-value < $1x10^{-12}$, Figure 5.2 a), as well as reduced Faith' phylogenetic diversity and reduced Shannon diversity among NPC cases. We also found significant differences in beta diversities (scale the level of pairwise dissimilarity) between NPC cases and controls, based on unweighted UniFrac (Figure 5.2 d), weighted UniFrac, and Bray-Curtis distance (PERMANOVA, false discovery rate [FDR] adjusted *P*-value < 0.001, 9,999 permutations). We noticed that NPC status was accounted for the largest variation explained in the unweighted UniFrac distance (Figure 5.2 b), while it was the second largest in weighted UniFrac distance following smoking (Figure 5.2c).

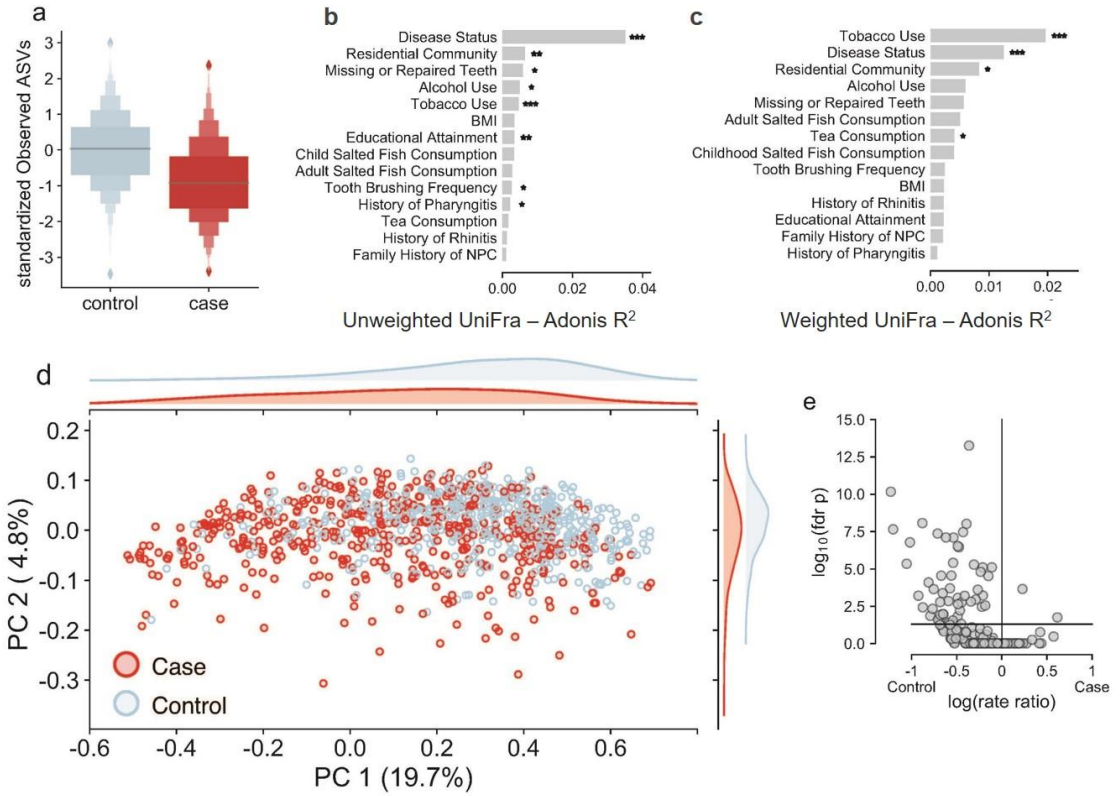


Figure 5.2 The oral microbiome differs between nasopharyngeal carcinoma cases and controls.

On the other hand, we identified 51 ASVs out of 245 abundant ASVs (relative abundance larger than 0.02% and present in more than 10% of 994 samples) which presented more in controls (log-binomial model adjusted for age, sex, sequencing run, residential community, and oral health, FDR *P*-value < 0.05; Figure 5.2 e). In contrast, two ASVs among 245 abundant ASVs showed more prevalent in NPC patients: a member of the genus *Granulicatella* ASV (Gran-7770) and a *Lactobacillus* ASV (Lact-eca9). We also observed that the Gran-7770 was 3.4 times more abundant among NPC cases than in controls (95% CI = 2.4 - 4.9), and a second *Granulicatella* ASV (Gran-5a37) was less abundant in cases. In addition, we found that all participants carried at least one or more *Granulicatella* species. Among them, 972 (97.8%) only carried three *Granulicatella* ASVs: Gran-5a37 and Gran-7770 (both mapped to species *Granulicatella adiacens*), and Gran-6959 (*Granulicatella elegans*) (Figure 5.3 a). Based on multinomial logistic regression, the variants of *Granulicatella adiacens* a participant carried were significantly associated with their NPC status (case or control). NPC cases presented higher odds of having both ASVs, and even higher odds of carrying Gran-7770 alone, compared to carrying Gran-5a37 alone (Figure 5.3 b). We found clear separations in principal coordinates analysis (PCoA) space in weighted and unweighted UniFrac and Bray-Curtis distances comparing samples carrying Gran-7770, both, or Gran-5a37, which was statistically highly significant (PERMANOVA, *P*-value < 0.001, 999 permutations; Figure 5.3 c).

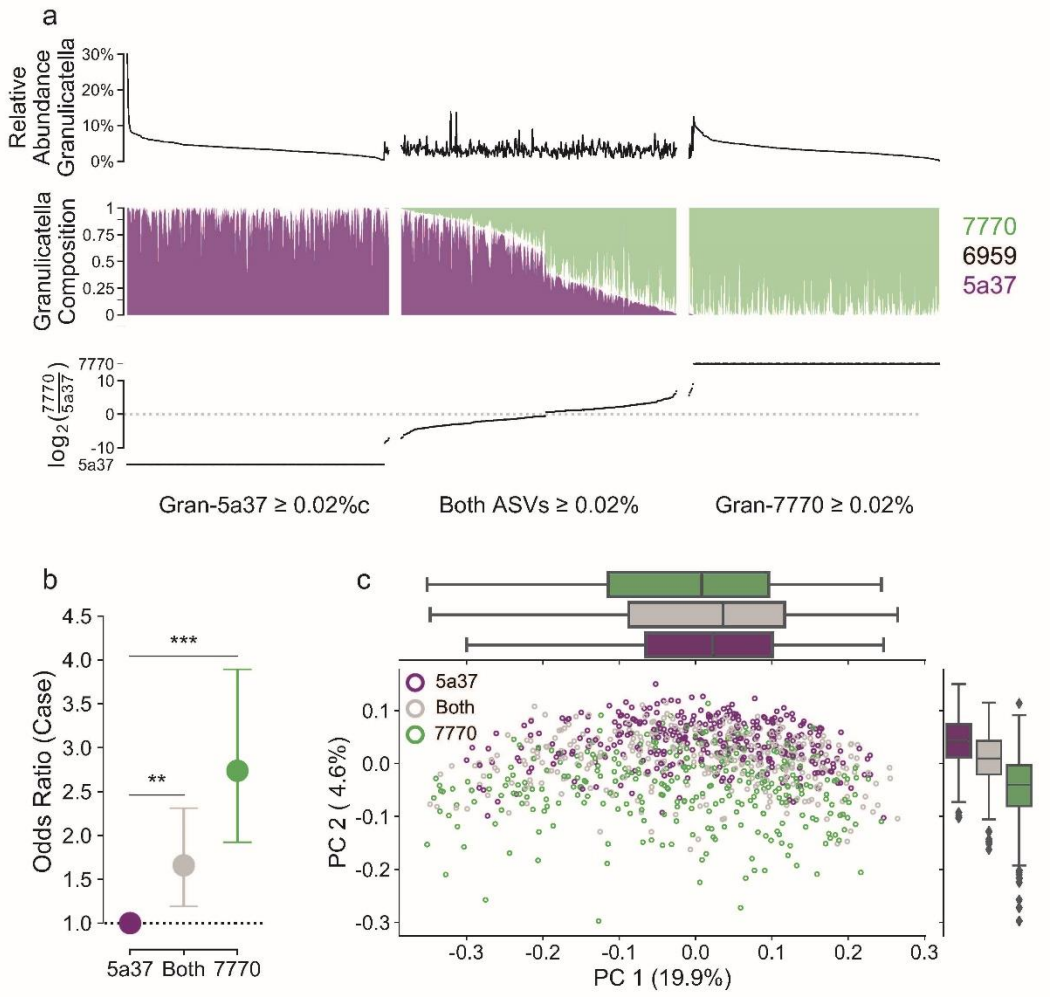


Figure 5.3 The *Granulicatella adiacens* variant predicts community structure.

Furthermore, we distinguished five networks using SparCC-based network analysis. Among them, a large network with 29 co-occurring and co-excluding ASVs caught our attention (Figure 5.4). In this network, Gran-7770 and Gran-5a37 each formed co-occurring nodes with a dozen ASVs, which were exclusive. By BLAST (basic local alignment search tool) searching against the Human Oral Microbiome Database, we found another two pairs of ASVs which were exclusive between nodes but assigned to the same clones: Prev-b7f2 and Prev-71e7 (*Prevotella melaninogenica*) and Stre-900d and Stre-0531 (*Streptococcus parasanguinis*). These results imply a possibility of niche specialization that is associated with disease status.

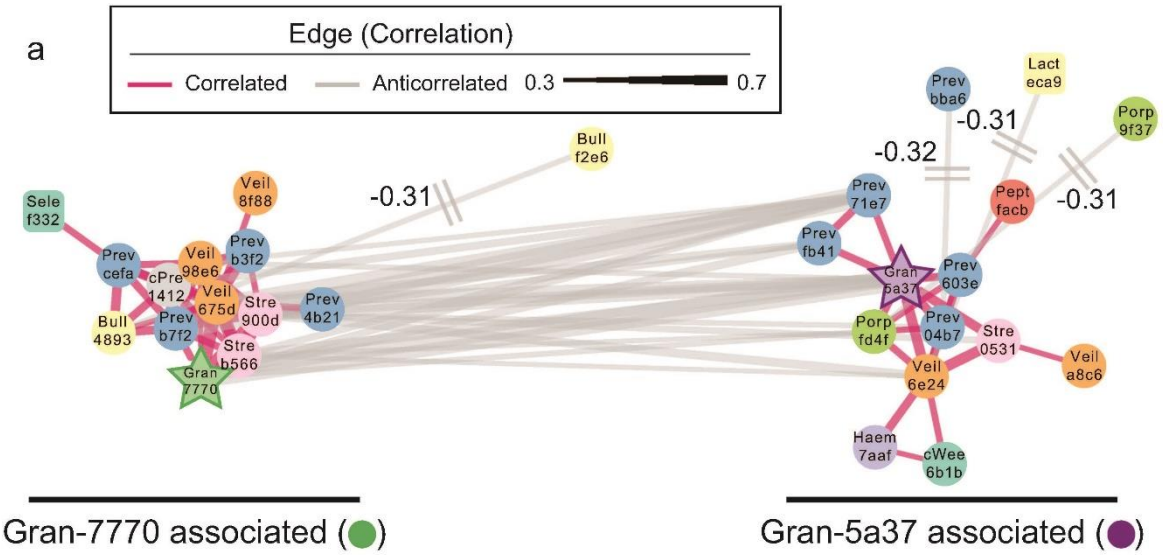


Figure 5.4 *Granulicatella adiacens* variants set at the center of a network of closely related co-occurring organisms.

5.3 CHANGES IN NASOPHARYNGEAL MICROBIOME DURING RADIOTHERAPY

We analyzed a total of 445 nasopharyngeal samples collected from 39 NPC patients in Paper III [239]. Among these, 27 (69.2%) patients achieved a complete response at the clinical check-up three months after completion of radiotherapy and were classified as early responders; the remaining 12 patients completely responded within 24 months and were classified as late responders.

We observed that the global structure of the nasopharyngeal microbiome among NPC patients changed over time in a constant manner, along with the treatment progress (Figure 5. 5, A and B). Volatility plots illustrated a consistent change in the first principle coordinate axis (PC1) along with sampling time; we also noticed that the changes in PC2 differed by patients ' response to treatment. Results from the linear mixed-effects models (LMEs) estimating the change in weighted UniFrac distance support the volatility plots that the global nasopharyngeal microbiome changed significantly during treatment (slope, *P-value* = 0.0005; Figure 5.5 C).

Moreover, we found larger magnitude of the longitudinal changes among early responders than late responders, indicating significantly different trajectories between the two groups (*P-value* = 0.0005; Figure 5.5 D).

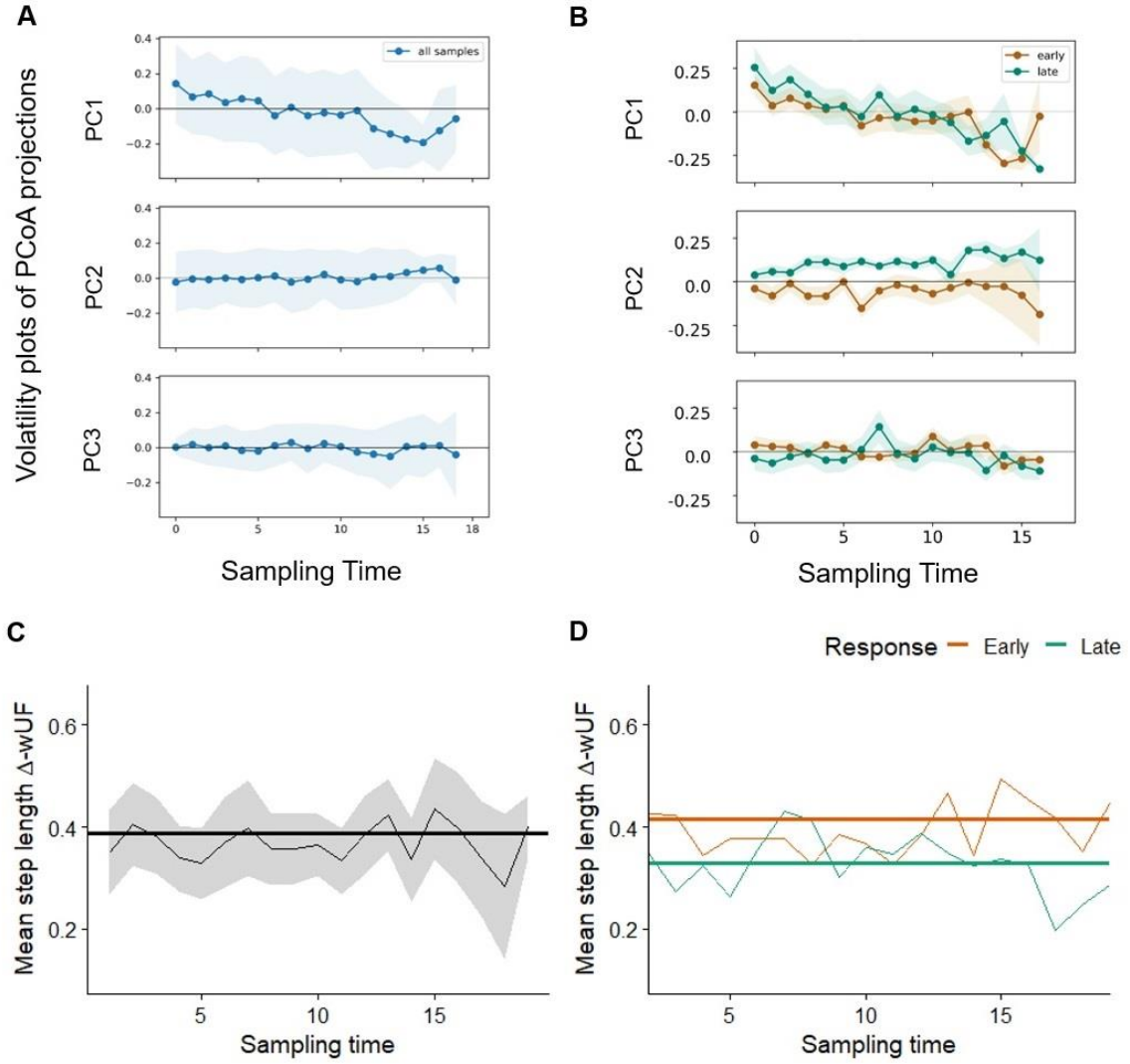


Figure 5.5 Global change of nasopharyngeal microbiome over treatment course.

We quantified the longitudinal differences in abundance of 73 abundant ASVs between the early and late responders (log-2 transformed). By applying smoothing-spline analysis of variance (SS-ANOVA), we identified 23 abundant ASVs with significant differences between groups (FDR < 0.05). Based on the similar pattern of changes, the estimated ratios between abundance in the early and late responders of 73 ASVs were clustered and visualized as a heatmap in Figure 5.6. In cluster C, half of the ASVs (gTherm.b26c, gTherm.087b, gTherm.94d, fCaulo.a200, gAcine.4f58, gRalst.edc7, gAcine.dfae, and gAcine.4c95) differed between groups since radiotherapy and consistently to the end. Most of these ASVs were assigned to genera *Ralstonia* and *Thermus*. Notably, some ASVs were assigned to the same genera, but showed different dynamic trajectories (increased, decreased, and relatively stable) across the treatment, such as members of the genus *Acinetobacter* (gAcine.32b4 and

gAcine.efe8 in Cluster A, compared to gAcine.4f58, gAcine.4c95, and gAcine.dfae in Cluster C), which might reflect niche specialization within the same genus.

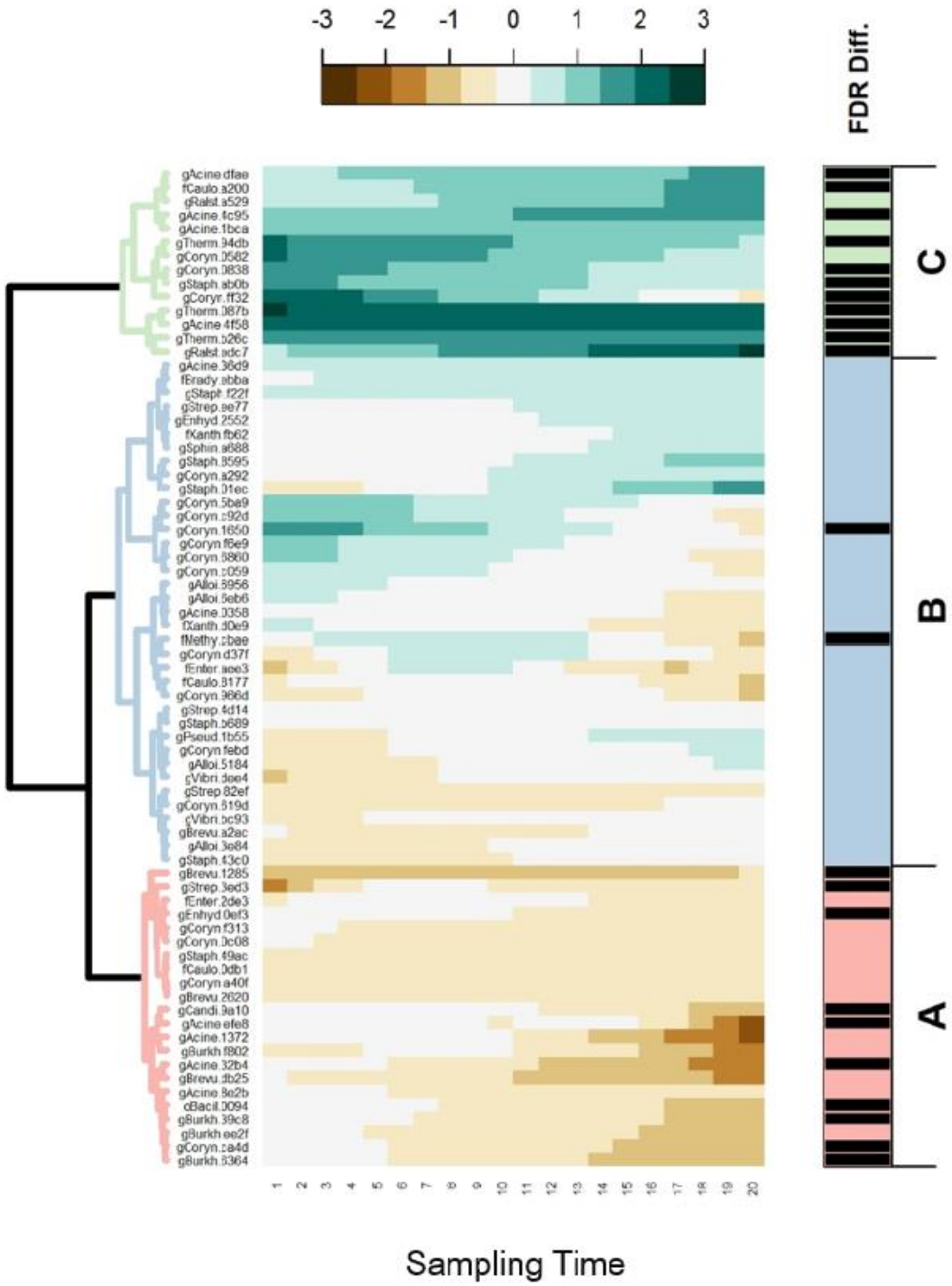


Figure 5.6 Longitudinal changes among the abundant ASVs differed by clinical outcomes.

6 DISCUSSION

6.1 INTERPRETATION AND IMPLICATIONS

6.1.1 Dietary patterns are associated with NPC risk

Using food frequency data from the NPCGEE project in southern China, we identified similar dietary patterns in both adolescence and adulthood that were associated with NPC risk in endemic areas, in Paper I. We demonstrated a clear negative association between higher consumption of the "plant-based diet" and NPC risk, as well as a strong positive association with higher intake of the "animal-foods-based diet." After mutual adjustment for adolescent and adulthood patterns, we observed that the OR estimates of adulthood patterns remained mostly unchanged, while associations with adolescent patterns were attenuated and no longer evident.

These findings suggest that adult dietary patterns are better in predicting NPC risk in southern China. It could be explained partly by underlying cumulative effects of diets rich in plants or high in animal foods along time, while it could also be affected by recall bias (in that people tend to provide more accurate information for recent than for distant diet). The CUP Reports suggest foods from plant sources, such as grains, fresh non-starchy vegetables, and fruits, decrease risks of colorectal cancer and many aerodigestive cancers, with strong evidence [61]. Meanwhile, they report that foods from animal sources, such as processed meat and red meat, increase colorectal cancer risk with substantial evidence. We cannot eliminate the possibility that our observed associations capture the underlying effects of foods from plants and animals on cancer risk in general, rather than NPC-specific effects. Therefore, our findings not only provide evidence for the possibility of primary prevention of NPC through dietary intervention but also support the public health recommendations of "eat a diet rich in wholegrains, vegetables, fruits and beans" as well as "limit consumption of red and processed meat" [61].

Notably, we found a positive association between higher intake "preserved/salted diet" during adolescence in the joint analysis, although it did not meet the statistical significance. In line with previous evidence showing the earlier exposure to Cantonese-style salted fish, the higher NPC risk, our finding may support an earlier susceptible exposure window for a preserved-foods diet concerning NPC risk. Although conducting a birth-cohort study on NPC would be difficult, if not impossible, our findings, together with numerous previous evidence, call for a further investigation focusing on early-life exposure and NPC risk.

Besides, studies in Malaysian Chinese and Italy have reported that a diet rich in fresh vegetables and fruits was associated with a decreased NPC risk, and the Italian study suggested a positive association between an "animal unsaturated fatty acids" diet and NPC risk [236, 240]. Therefore, our findings may be partially generalized to other NPC endemic areas outside China (e.g., Southeast Asia) and non-endemic area (e.g., Italy).

6.1.2 Oral microbiome is associated with NPC risk

We address the association between the oral microbiome and NPC risk, which may be explained partly by reduced microbial diversity and niche specialization among closely related organisms.

Our results suggest that the oral microbiome is significantly associated with NPC status (i.e., case/control). These results are robust because they consider known confounders (e.g., residential community, smoking, and oral health) and less likely to be confounded by unmeasured factors because of the strong signal. It is also implausible that these results are related to cancer treatment since NPC patients were recruited after diagnosis but before taking radiotherapy/chemotherapy. However, we cannot rule out the effect of antibiotics (self-medication) on the oral microbiome. Because the history of antibiotic use was unavailable (prior medical use was asked for both cases and controls, but the resulting information was not detailed enough to classify antibiotic use reliably), the current use of antibiotics was not an exclusion criterion. Although antibiotics could induce declined diversity, the oral microbiome has been suggested to be a more stable community than those at other body sites with regard to antibiotic treatment [241]. Besides, the significant drop in richness among NPC cases would require extensive use of antibiotics, which is not in line with interview data.

Remarkably, our results reveal a significant association between NPC status and a pair of closely related ASVs, Gran-7770 and Gran-5a37, mapped to species *Granulicatella adiacens*. The presence/absence status and relative abundance of *G. adiacens* variants were significantly associated with disease status, and the variant carried by a participant predicted the community structure. These findings suggest that the *G. adiacens* variants might be a path through which NPC status shapes the oral microbiome. Furthermore, the co-occurring and co-excluding network centered on the *G. adiacens* variants may reflect partial niche specialization and link to metabolic changes. In the context of NPC in southern China, we hypothesize that *G. adiacens* and co-occurring organisms are involved in producing oncogenic metabolites (e.g., nitrate and nitrite production) and/or immune regulation [109, 242].

However, given our cross-sectional study design, a disease status may either be a reaction to bacterial strain-specific variation or a factor driving this variation. It is also challenging, if not impossible, to evaluate how these organisms may function in concert without culture-based experiments. Besides, the limited resolution of operational taxonomic unit-based clustering in existing database challenges *in-silico* validation. Therefore, our findings call for future efforts to explore the temporality and the underlying mechanism.

6.1.3 Changes in nasopharyngeal microbiome during radiotherapy are associated with NPC patients' treatment outcome

In Paper III, we demonstrated for the first time stable, temporal changes of the nasopharyngeal microbiome among NPC patients during radiotherapy-based treatment. We observed that the nasopharyngeal microbiome showed temporal shifts throughout treatment (i.e., UniFrac distance-based volatility plots); seven abundant ASVs (assigned as *Corynebacterium*) showed

a significant and consistent loss. This proof-of-principle data provides compelling evidence for exploring possible interactions between the host, radiotherapy, and commensal microbiome.

Additionally, we reported a significant association between the microbial changes among NPC patients and their short-term outcomes. First, we found distinct and consistent separations in volatility plots regarding patients' short-term outcomes (i.e., early responders vs. late responders). Results from LMEs modelling the magnitude of change (UniFrac distance-based) suggested a statistically significant difference between the early responders and late responders, which implied a possibility of feature-based differences. Thus, with the focus on 73 abundant ASVs, the non-parametric microbial interdependence test (NMIT)-based procedure showed that the microbial network's temporal changes over treatment among early responders differed from the changes among late responders. These findings challenge Anna Karenina's principle, which would suggest that a favorable prognosis is associated with community stability [243]: in contrast, our results may implicate that a specific structure or trajectory of the nasopharyngeal microbiome reflecting less change during treatment indicates delayed response to radiotherapy.

Furthermore, we identified 23 abundant ASVs significantly different between response groups during radiotherapy-based treatment. Five ASVs assigned to genera *Ralstonia* and *Thermus* showed a consistent difference between groups throughout treatment, from beginning to end; these organisms were at very low abundance or absent in early responders but present in late responders. Members of these extremophile genera are known to be radiation- or ROS-resistant, suggesting a potentially plausible biological mechanism for their inclusion in these communities [244]. These findings may implicate bacterial resilience to radiotherapy as a feature of delayed response. We hypothesize that a community might be protective against both radiation-induced perturbations on the microbiome and radiation damage of the host cells.

We acknowledge this as an exploratory observational study with a small number of NPC patients. However, it is a relatively large study in terms of the number of longitudinal samples collected and evaluated for each patient. Few studies have a larger sample size in the same research area due to the difficulty of repeated sampling in clinical practice. There is no comparable study that could be used as an external validation set for our findings. Although the short-term outcome was assessed at the first clinical check-up with 21 months of follow-up, it might be insufficient in the context of direct applicability of the microbes as prognostic markers for NPC patients.

We argue that several genera showed distinct and stable differences between response groups from treatment-naïve until the end of the full course of radiotherapy, which should be further validated for monitoring the initial treatment to improve the therapeutic outcome and moderate radiation-related toxicities. By focusing on NPC, for which radiotherapy is the most common treatment, we hypothesize that the relationship between the commensal microbiome and radiotherapy response may hold for other diseases and anatomical locations. Therefore, our findings call for more extensive longitudinal studies with long-term follow-up to verify the proposed hypotheses and serve as a base for generating new hypotheses in the coming future.

6.2 METHODOLOGICAL CONSIDERATIONS

6.2.1 Observational studies

Internal and external validities are crucial considerations during all stages of an observational study: internal validity captures to what degree a study has measured what it aimed to; external validity, also known as generalizability, describes how specific study results can be generalized to other settings [245]. Regarding internal validity, selection bias, information bias, and confounding are present to some extent in all observational studies, and need to be identified carefully and avoided or controlled as much as possible. We have attempted to address these biases and potential confounding in the presented papers, in line with the different study designs. The generalizability has been discussion in each paper previously.

6.2.1.1 Case-control studies

A case-control study is considered inexpensive, easy to conduct, and relatively quick compared to other study designs like cohort studies or randomized controlled trials, especially when investigating rare diseases or preliminary studies where little is known about the association between the exposure of interest and disease. Therefore, a population-based case-control design is suitable for the research questions in Study I and Study II.

However, a case-control study is retrospective in nature: it starts with an outcome and traces back to the exposure, making it more sensitive to various biases and more challenging to establish the timeline of exposure-outcome association.

Accordingly, great efforts have been made in the NPCGEE project to reduce selection bias and information bias, which are inevitable in case-control studies [222]. We equally applied the eligibility criteria of cases to controls, assuring that cases and controls came from the same study population. We also randomly selected controls from the total population registries in study area. We trained and monitored interviewers for the proper conduct of an interview in an identical manner for cases and controls (due to the lack of regional cancer registers in the study area, we identified NPC cases via a network of physicians who diagnosed and/or treated NPC at hospitals in the study area and informed the study interviewers afterward; thus the interviewers were not able to be blinded to subject's disease status). Besides, we interviewed around 5% of controls by telephone after several unsuccessful attempts of a face-to-face interview in order to capture as many controls as possible. Furthermore, we used a structured electronic questionnaire to reduce logic errors and missing values and designed an illustrated booklet of various foods in different serving portions to facilitate participants in estimating dietary intake properly. Despite relatively high participation rates among cases (85.8%) and controls (82.7%), we acknowledge that a possible selection bias cannot be ruled out.

To establish diet as a possible exposure for NPC development, we collected information on dietary intakes 10 years ahead of the interview for all participants, as well as 16-18 years of age for those above age 35. This, however, relies on remote memories of the participants and introduces a pervasive and inescapable recall bias. We excuse that recall bias for dietary intake

is likely non-differential between cases and controls, tending to obscure real differences. In Study II, we can only study the association between the oral microbiome and NPC status, without temporal ordering (cross-sectional case-control study).

Confounding exists when a factor is associated with both exposure and outcome, but not an intermediator lying in the causal pathway between exposure and outcome. Complete elimination of unmeasured and unknown confounding is virtually impossible in an observational study. We used frequency matching on age, sex, and resident region in the NPCGEE project. Furthermore, we applied multivariate regression models to control for potential confounding based on prior knowledge and identified risk factors in the study population. Although socioeconomic status was not available directly, we argue that it correlates well with the occupation, housing type, and educational level, which were considered in the multivariate logistic regression models. We also examined whether age (born before or after 1963) modified the effect of dietary patterns as risk factors for NPC risk in Paper I.

Before accepting the results and further interpreting them, one should be aware of a chance finding. *P*-values are not the arbiters of validity but a measure of chance, which advises us of the likelihood of a false-positive conclusion [245].

6.2.1.2 Cohort study

There is no prior knowledge or evidence about how the nasopharyngeal microbiome changes during radiotherapy. We conducted a hospital-based prospective NPC patient cohort with the repeated collection of nasopharyngeal swabs along with radiotherapy-based treatment to provide longitudinal data. The repeated measurement design effectively controls for within-individual variation, but limits the number of possible participants in this study. However, the longitudinal characteristic provides a comprehensive view of microbial diversity, an effective way for assessing the temporal microbial correlations, and an opportunity to reveal the degree of change of abundant features during the observation period [246].

In Paper III, all 39 NPC patients received the first clinical check-up at three months after completion of radiotherapy following the standard of care. Their short-term treatment outcomes were assessed and determined by the same three experienced oncologists based on RECIST criteria. Information bias is less likely to exist. Among 62 initially enrolled patients, 23 were excluded due to various reasons [239]. We found no statistically significant difference in the distributions of age, sex, TNM stage, and treatment strategies between included patients and excluded patients. However, this study is limited by a lack of a comparative cohort as a validation set.

6.2.2 Principal component analysis

Dietary habits describe what foods people habitually choose to eat in their daily life. Unlike individual food items, habitual dietary intake is a multi-dimensional exposure covering numerous possible combinations of food items [89] and requires a comprehensive characterization. In observational studies, two strategies have been widely used for descriptive

purposes and to summarize dietary patterns for subsequent modelling and hypothesis testing: one is to describe a number of different food items using an established score, such as estimating adherence to the Mediterranean-diet score [247]; another is to efficiently summarize a large set of food items into a smaller group of variables (also known as dietary patterns) using a data reduction approach. As there was no eating score/index for the Chinese diet available when we planned Study I, we considered data reduction techniques. There are several commonly used approaches, including principal component analysis (PCA) and factor analysis (FA), reduced rank regression (RRR), and cluster analysis (CA) [248]. PCA and FA are two similar techniques chosen when a researcher is willing to summarize a broad set of food intakes as several dietary variables (i.e., components or factors) and to include these new variables in a regression model without prior theory/knowledge [249, 250]. In comparison, a RRR is a useful approach if a researcher has sufficient knowledge to determine response variables (such as body mass index, income level, etc.) as a set of intermediate factors associated with outcomes [251]. Unlike previous methods, groups/clusters generated from a CA are mutually exclusive, and there is no standard, well-validated procedure suitable across all settings [248]. Given the aims of (a) describing dietary habits in two study periods among participants from the NPCGEE project and (b) evaluating their associations with NPC risk, we decided to use PCA as the primary analysis, with an iterated principal FA as a sensitivity analysis.

However, approaches like PCA/FA are not free from limitations:

1) Researchers decide how many components/factors should be included in the downstream analysis, which is subjective.

2) In many cases, only a low to a medium proportion of dietary habits variation is explained by selected components/factors.

3) It can be challenging to interpret results, such as which food items/groups characterize the component/factor.

4) Related to #3, naming the components/factors is also often subjective and challenging in practice.

6.2.3 Biostatistical analyses in microbial studies

Unlike traditional observational studies, there is no standard analytical workflow for human microbiome studies yet [228]. The choice of analytical strategies and statistical approaches differs due to the inherent characteristics of 16S rRNA amplicon sequencing data. For instance, the total number of bacteria in a biosample is not recoverable from the bacterial sequence counts due to numerous biases introduced by the DNA-PCR-sequencing procedures [252]. Therefore, relative instead of absolute abundances must be used, i.e., the data to be *compositional*: the relative abundances of all sequences in a sample are constrained to sum to one [228]. It means that when the relative abundance of a sequence increases, the relative abundances of the other sequences must decrease correspondingly, which is critical when interpreting results from a naïve analysis biologically. Additionally, a feature table representing

the number of times each feature (e.g., amplicon sequence variant [ASV], operational taxonomic unit [OTU]) is observed in each sample is sparse, often containing around 90% of zero counts; the data therefore generally does not follow a standard normal distribution even approximately, or more common discrete distribution like Poisson or negative binomial distribution. These characteristics lead to erroneous results and spurious correlations when naively applying traditional statistical procedures [252]. Consequently, many concepts and analytic approaches in microbial studies have been adopted from the field of ecology, and differ from methods commonly used in conventional epidemiological research [253].

Here, we outline our experiences and considerations across the microbial methodological workflow in Paper II (Figure 6.1) and Paper III (Figure 6.2), starting from three key representations generated from amplicon sequencing data:

1) The *feature* table is the fundamental data matrix of all the downstream analyses, containing the count (abundance) for all features (as rows) across all samples (as columns). In the literature, this feature table may be referred to as an ASV or OTU table, depending on the denoising technique employed.

2) The *taxonomy* contains information on each feature's bacterial taxonomic classification in the context of the established hierarchy of known bacterial species.

3) The *phylogeny* (phylogenetic tree) contains the evolutionary relationships between individual features reconstructed from the data at hand and acts as a backbone for many diversity analyses like, e.g., Faith's phylogenetic diversity or UniFrac distance [254].

6.2.3.1 Oral microbiome

With these three essential representations in place, we can address three scientific questions with regard to our NPC case-control Study II:

1) How does the number of species/features in salivary samples differ by NPC status?

2) How similar are NPC cases and controls in terms of microbiome composition?

3) What features specifically are associated with differences between NPC cases and controls?

The first and second questions can be answered by calculating alpha- and beta-diversity measures, central concepts in ecology, from our key data. Before calculating these diversity measures, however, normalization (standardizing the number of sequences across samples) must be performed to allow for valid comparison across samples, since the full range of species is rarely saturated, i.e., the higher the sequence yield for a sample, the higher the observed species count. A rarefaction curve is a useful guide for choosing a suitable normalization threshold to ensure that sufficient diversity can be observed in all samples [255]. However, rarefaction is a balance between maximizing the number of sequences per sample and keeping as many samples as possible. Thus, the choice of threshold is highly dependent on the studied

data. For the oral data, we chose a threshold of 6,500 sequences based on rarefaction curves (C-1 in Figure 6.1).

Alpha-diversity as a concept quantifies within-sample diversity based on the observed number of species in a sample. Practically, many, often closely related measures for alpha-diversity have been proposed, of which we chose the Shannon index, which is based on relative species abundances, and Faith's phylogenetic diversity, which takes into account phylogenetic relationships between species [256, 257]. Alpha-diversities were modelled as outcomes in multivariate ordinary least-squares regression to quantify the relationship between diversity and environment, adjusted for age, sex, and sequencing run number (C-2). Model fit for a range of different models, including different covariates, was assessed via the Akaike information criterion (AIC). The model with the lowest AIC score was selected as the best fitting. A "leave-one-out" approach was used to estimate each covariate's relative contribution to the AIC metric. To visualize alpha-diversity in answering the first question, we used boxplots by NPC cases and controls.

Beta-diversity conceptually captures pairwise dissimilarity in microbial composition between two samples, with the two most commonly employed measures being the Bray-Curtis dissimilarity and UniFrac distance. The Bray-Curtis dissimilarity quantifies the compositional dissimilarity between two samples based on species counts. The UniFrac distance calculates the distance between species present in two samples based on the underlying phylogenetic tree, as a fraction of non-shared branch lengths: *weighted UniFrac* is a quantitative measure considering the actual species counts, while *unweighted UniFrac* is only based on the presence or absence of species [258, 259]. Beta-diversity is typically presented as a sample-by-sample matrix. The distance or dissimilarity between two matrices can be approximated and visualized via principal coordinate analysis (PCoA). In a PCoA ordination, the *n* samples are placed in *n*-dimensional space of synthetic variables (the principal coordinates, PCs) so that the resulting Euclidean distances between the samples based on the principal coordinates approximate the original distance or dissimilarity matrix as well as possible. PCs are constructed in such a manner that the first PC (PC1) explains as much variation in distance between samples as possible, with subsequent PCs explaining less and less variation. Generally, the first two or three PCs are used for projection as 2-dimension plots or 3-dimension plots to visualize beta-diversity.

As the visualization of beta-diversities via PCoA in Study II showed differences between NPC cases and controls, we decided to continue with formal statistical inference. A combination of permutational multivariate analysis of variance (PERMANOVA/Adonis) and permutational analysis of multivariate dispersions (Permdisp) was used (C-3) [260]. PERMANOVA is a test for dissimilarity (here: beta-diversity) between groups, with the null hypothesis that the centroids and dispersions of the groups are all equal, while Permdisp tests for homogeneity of dispersions within each group, with the null hypothesis that there is no difference in dispersion across groups. In Study II, a permutation *P-value* < 0.001 from PERMANOVA based on UniFrac distances implies that NPC cases and controls have statistically significantly different

beta-diversity structures. However, this does not distinguish whether the difference is primarily due to variation in within-group dispersion, or primarily due to shifts in centroids between groups; a complementary Permdisp test with a non-significant *P*-value (> 0.55) indicates the latter case.

The Adonis-implementation of PERMANOVA also reports an R^2 measure of variance explained that describes how variation in community structure can be attributed to the main exposure (here: NPC status) or to other covariates included in the model. Base on this R^2 measure, we found that NPC disease status was the strongest factor associated with differences in microbial community structures when considering the presence/absence of species (unweighted UniFrac-based), and the second strongest (after smoking) element when also considering the relative abundances of species (weighted UniFrac-based).

In the next step of the analysis, we addressed the question of differences in feature-based patterns between NPC cases and controls: specifically, but not limited to the question of which features are more prevalent in cases (controls) and which features are more abundant in cases (control)? Our approach is based on the same three key representations, but processing them in a different manner.

To begin with, we filtered the feature table by defining presence as a relative abundance greater than 0.02% (corresponding to the shallowest sequencing depth for the abundant counts) in at least 10% of samples in order to eliminate noise and unreliable measurements and to avoid spurious correlations. This resulted in 245 abundant ASVs (C-4). In addition to age, sex, and sequencing run, we also chose to include covariates with an Adonis R2 at least 60% of the value associated with NPC (weighted or unweighted UniFrac) in the preceding global analysis for full adjustment in the feature-based analyses, to account for potential confounding; this led to the inclusion of smoking, residential community, and oral health.

As NPC disease status was strongly associated with the microbial community structure considering the presence/absence of species, we first aimed to identify feature patterns (focused on the presence/absence of species) differed between NPC cases and controls. Thus, we fitted log-binomial regression models (approximated via a Poisson regression with robust standard errors) for presence/absence as a binary outcome for all 245 abundant ASVs, adjusting for age, sex, sequencing run, smoking, residential community, and oral health (C-5). A Benjamini-Hochberg FDR corrected *P*-value of 0.05 was considered as statistically significant. At this significant level, we identified only two (out of 245) ASVs, namely Gran-7770 and Lact-eca9, which were more prevalent in cases, compared to 51 ASVs which were more prevalent among controls. This is in line with our previous findings that controls had higher alpha-diversity compared to cases.

As we found strong associations between disease status and beta-diversity as measured by weighted UniFrac distance and Bray-Curtis dissimilarity, which incorporate relative abundances of species, we also wanted to study feature patterns of abundant ASVs based on relative abundances (C-6). We chose Phylofactor, a compositionally and phylogenetically

aware generalization of factor analysis to microbiome data, to evaluate how disease status corresponds to microbial abundance patterns (given the strong signal we saw in weighted UniFrac) [261]. Phylofactor is based on a specialized algorithm, "phylofactorization", that aims to identify the most significant phylogenetic clades influencing variation in the data via their associations with different covariates. It uses isometric log transforms to model differences in the feature data with multivariate adjustment. Visualizations using phylofactorization can also map back to the phylogeny to provide taxonomic microbial annotation for each identified clade, which facilitates biological interpretation. The results from Phylofactor gave a summary of microbial abundance patterns associated with NPC status, sorted by explained variability. We found that the results highlight that individual ASVs like Gran-7770 and Gran-5a37, were significantly associated with NPC status, with Gran-7770 more abundant in cases and Gran-5a37 more abundant in controls.

As a complement to the Phylofactor, we also used a modified multivariate analysis of composition of microbiomes (ANCOM) with the abundant ASVs. ANCOM is also a compositionally aware statistical framework that compares log ratios of the abundance for each ASV to the abundance of all remaining ASVs one at a time; it controls the resulting risk of false discoveries in detecting differentially abundant ASVs with this approach while still maintaining high statistical power [262]. In the ANCOM model, we adjusted for age, sex, sequencing run, smoking, residential community, and oral health. Results from ANCOM corresponded well with Phylofactor and provided us microbial abundance patterns of individual features.

After addressing our three original questions, we decided to investigate the surprising finding that two ASVs, Gran-7770 and Gran-5a37, repeatedly appeared in almost all feature-based analyses (C-7). First, we conducted a "background investigation" for these ASVs in the literature, microbiome databases, and our data. We found that these ASVs represent closely related organisms, or variants, of *Granulicatella adiacens*, and almost every participant in our study carried at least one variant. Second, a multinomial logistic regression was used to verify the associations between disease status and the variant carried by participants. The modelling results indicate that NPC cases were more likely to carry Gran-7770, while controls were more likely to carry Gran-5a37. We also tried to validate our findings using the publicly available data: however, as almost all the available oral microbiome data is at OTU-based resolution, which is not comparable with ASV-based data, with resolution down to single-nucleotides. We were, however, able to investigate the presence of other species that *Granulicatella adiacens* might interact within our material (C-8): we used the sparse co-occurrence network investigation for compositional data (SCNIC) to generate and analyze co-occurrence networks among the abundant ASVs [263]. Again, surprisingly and amazingly, we identified a co-occurring and co-excluding network centered on the *G. adiacens* variants; this may reflect partial niche specialization and links to metabolic changes [264].

Nevertheless, given our cross-sectional design, a disease status may either equally be a driver or a consequence of bacterial strain-specific variation. Besides, it is difficult to evaluate further

how these organisms may function in concert without culture-based experiments, and we decided to end the oral microbial adventure here.

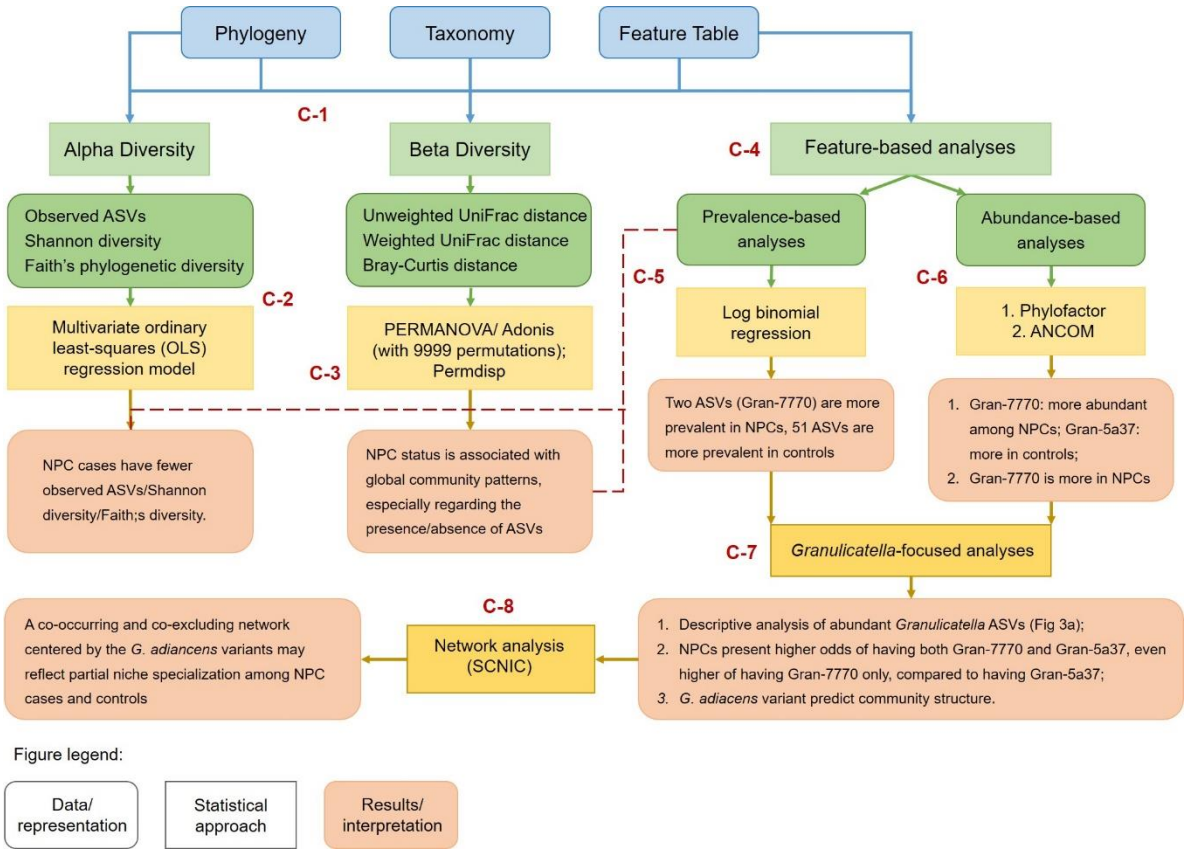


Figure 6.1 Methodological consideration about biostatistical analyses in Paper II. *C-n: considerations during the workflow.*

6.2.3.2 Nasopharyngeal microbiome

For investigating the nasopharyngeal microbiome during radiotherapy in Study III, we started our workflow with the same three essential representations generated from 16S-amplicon sequencing data: feature table, taxonomy, and phylogeny. However, given the different anatomical location and prospective cohort design with repeated sampling, the methodological considerations and research questions are inevitably different from those in Study II:

1) How many and which species/ASVs can be observed in nasopharyngeal samples?

2) How does the nasopharyngeal microbiome change along with radiotherapy-based treatment?

3) How similar is the nasopharyngeal microbiome between the early and late responders?

Based on rarefaction curves, we rarefied feature table to 1,500 sequences. Alpha- and beta-diversities were computed afterward. To test for differences in alpha-diversity between sampling time points and response groups, we used the pairwise Kruskal-Wallis test; we found

no statistically significant evidence for such differences. This is reasonable because samples collected longitudinally from the same individual are more likely to be similar to each other, which means that they are not independent (C-2).

Thus, we wanted to explore how the microbiome changes over sampling time, in parallel with treatment progress. We created a volatility visualization, which shows interactive line plots to represent how the nasopharyngeal microbiome changes across time in one or more groups (by response, by individual, etc.). Both alpha- and beta-diversities (represented by PCs) were visualized. We observed a clear shift in the nasopharyngeal microbiome as a whole in PC1, as well as a separation between early and late responders in PC2 over treatment, based on weighted UniFrac distances (i.e., reflecting abundance). However, we found no such patterns for unweighted UniFrac distance-based and Shannon index-based analyses. Together, these results suggest that the change of the nasopharyngeal microbiome and the different patterns between groups may be associated with the relative abundance of species in the community rather than their presence/absence (C-3).

Consequently, we attempted to measure the magnitude of change in the microbiome in terms of weighted UniFrac distance across treatment by defining the variable Δ-wUF(t), corresponding to the length of the (t-1)-th step of a subject's trajectory through microbial space (C-4). By transforming the distance matrices of successive time points into values of the magnitude of change, we were able to fit a series of linear mixed-effects models (LMEs), allowing for varying degrees of change over treatment course as well as an examination of between-subject variation: based on the model comparison, we could assess whether there was significant change over treatment, whether the change differed by treatment outcome, and whether the change differed among individuals. Differences between nested models were tested via likelihood ratio tests. AIC was used to compare different models. To address sensitivity towards possible violations of model assumptions, standard LME inference was complemented with the parametric bootstrap *P*-values. Based on this modelling approach, we concluded that our data presented evidence for a). a stable, temporal change in the nasopharyngeal microbiome among NPC patients during radiotherapy-based treatment; b). different patterns of change between early and late responders.

In order to examine patterns of change for individual features, we started as for Study II by filtering the feature table, in this case with a threshold of relative abundance greater than 0.1% (corresponding to the shallowest sequencing depth for the abundant counts) in at least 10% of samples, resulting in 73 abundant ASVs (C-5).

As mentioned, samples collected from the same individual are not independent. Thus, we would like to explore the subject-specific time dynamics of the abundant features as a whole during treatment, and examine whether there is a difference between response groups (C-6). Therefore, we employed the non-parametric microbial interdependence test (NMIT) permutational testing framework [265]. Briefly, for each subject, a within-subject correlation matrix between the logarithmized relative abundances across sampling time points was calculated as a measure of subject-specific time dynamics; and the differences between subjects

were captured by the Frobenius norm (i.e., the sum of squared differences) of their relative within-subject correlation matrices, resulting in a subject-by-subject matrix of coefficients quantifying the dissimilarity of time dynamics between subjects. Differences between the groups were visualized using a PCoA approximation. Based on the results of the PERMANOVA (*P*-value = 0.014) and Permdisp tests (*P*-value = 0.315), we suggest that the subject-specific time dynamics differed between the early and late responders. The difference is more likely due to between-group variation rather than a difference in within-group variation.

Meanwhile, we also wanted to investigate whether individual features changed over treatment (C-7). We modelled the trajectory of observed counts for each ASVs via smoothing-spline ANOVA (SS-ANOVA) [266], which fits smoothing-spline curves to the longitudinal trajectories of log-2 transformed differences in abundance between two response groups, using a penalized least-squares approach. These predicted trajectories summarize how an ASV's relative abundance between early and late responders shifts over sampling time. The jackknife-based penalty term in this model is explicitly designed as a regularization parameter to avoid overfitting the spline terms. For ensuring statistical robustness, we also increased the number of random re-samples to k = 10,000 to allow for multiple testing adjustment, and a FDR < 0.05 was considered as statistically significant. We identified 23 out of 73 abundant ASVs that statistically significantly differed between early and late responders over time (FDR < 0.05). Among significant ASVs, 5 ASVs assigned to genera *Ralstonia* and *Thermus* showed a consistent difference between early and late responders from the beginning to the end of treatment; these organisms were at very low abundance or absent in early responders but present in late responders. These results may implicate bacterial resilience to radiotherapy as a feature of delayed response.

However, given the relatively small sample size, observational design, and limited follow-up time in Study III, we were limited to further investigate whether the change is associated with patient's long-term outcomes (e.g. 5-year overall survival rate, 5-year locoregional control rate, etc.) and verify our results in a validation cohort. Nonetheless, we believe that these findings are essential as a proof of principle and hope that our work will raise researchers' interest in radiation-related changes in the human commensal microbiome and lead to more extensive longitudinal investigations. There are some plans for continuously studying this topic, which will be presented in the Future Perspective.

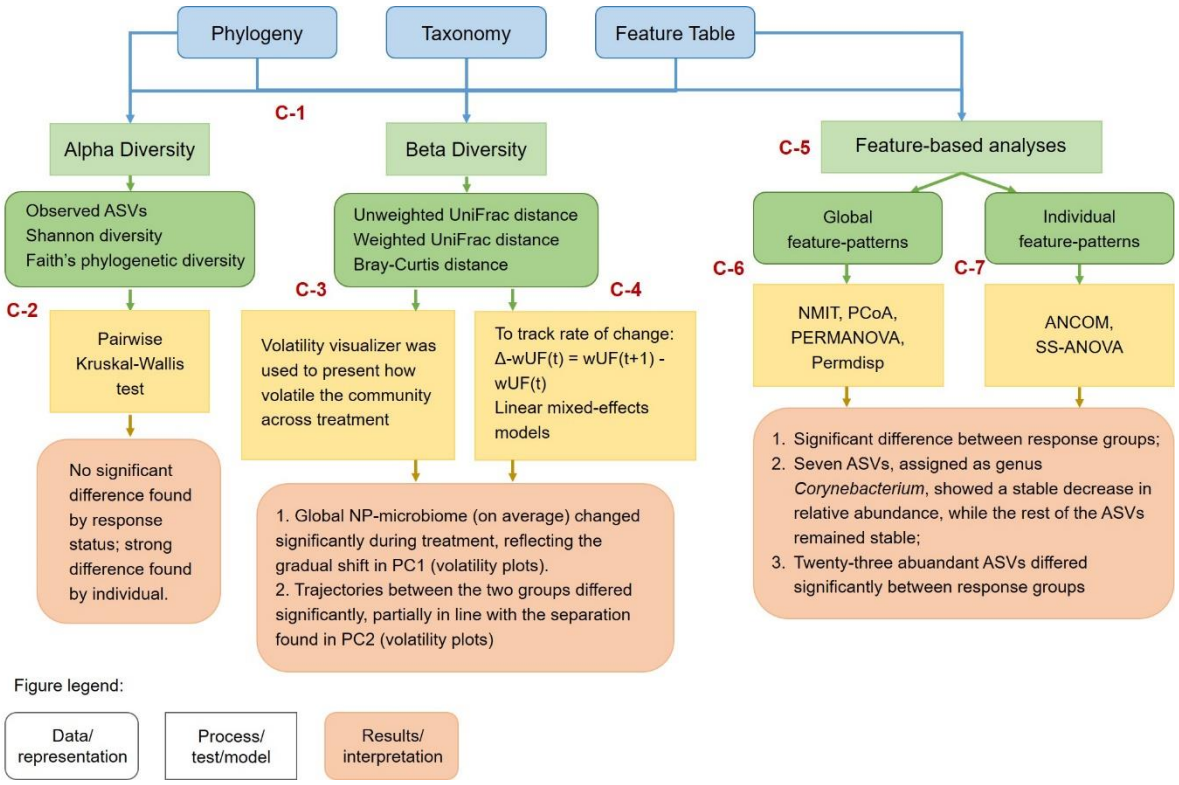


Figure 6.2 Methodological consideration about biostatistical analyses in **Paper III** [265-267]. *C-n: considerations during the workflow*.

6.3 ETHICAL CONSIDERATIONS

The data and saliva samples used in Study I and Study II were derived from the NPCGEE project, which was approved by the institutional ethical review board of Harvard T.H. Chan School of Public Health (Boston, United States), the Regional Ethical Review Board in Stockholm (Sweden), the Institute for Viral Disease Control and Prevention of the Chinese Center for Disease Control and Prevention (Beijing, China), Sun Yat-sen University Cancer Center (Guangzhou, China), and Guangxi Medical University (Nanning, China). During the interview, all participants were informed and provided consent of the future utility of questionnaire data and biosamples. The data and nasopharyngeal samples used in Study III were collected from the prospective NPC patients cohort in the First Affiliated Hospital of GXMU (Nanning, China), which was approved by the Ethical Review Committee of the First Affiliated Hospital. All participants were informed and provided consent of the future utility of questionnaire data and biosamples.

The questionnaire data used in Study I and II is stored on the server at the Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden. Data manager of the NPCGEE project replaced the identification number of questionnaire for each participant by an auto identification number of electronic questionnaire, which is used for linkages between sections of questionnaire data by sub-users. The sub-users cannot get access to the original questionnaire data. Saliva samples used in Study II are stored at Guangxi Medical University,

and the subsequent 16S-amlicon sequences have been made publicly available and stored at the European Nucleotide Archive [268]. The only metadata can be accessed is sex, disease status (case/control), and sequencing plate number. In Study III, the clinical data and nasopharyngeal samples are stored at Guangxi Medical University. The clinical data of NPC patients, including age, sex, cancer stage, treatment strategy, sampling time, and response status, as well as 16S-amplicon sequence is stored on the servers at Karolinska Institutet, which was approved by the Regional Ethical Review Board in Stockholm (Sweden).

Tingting Huang, and the corresponding authors of Study I - III are the guarantors of these works and, as such, have full access to all the data and samples, and take responsibility for the integrity of the data and the accuracy of the data analysis.

Ethical approvals from Regional Ethical Review Board in Stockholm related to this thesis work include: Dnr 2009/1293-31/3 and Dnr 2015/175/39 (for Study I and II); Dnr 2017/1393-31 (Study III).

7 CONCLUSIONS

- Plant-based and animal-foods-based diets are associated with nasopharyngeal carcinoma risk (in differential direction), which supports the possibility of primary prevention of nasopharyngeal carcinoma through dietary intervention.

- Oral microbiome is associated with nasopharyngeal carcinoma risk. The niche specialization among closely related commensals associated with nasopharyngeal carcinoma status, calls for future culture-based investigations.

- Stable, temporal changes of the nasopharyngeal microbiome among nasopharyngeal carcinoma patients during radiotherapy-based treatment are observed. These changes are associated with patients' short-term clinical outcomes measured three months after the completion of radiotherapy. Our findings call for more extensive longitudinal studies with long-term follow-up for verification, and serve as a base of generating new hypotheses for future studies.

8 FUTURE PERSPECTIVE

As concluded, this thesis has demonstrated the associations of dietary habits and suggested a potential diet intervention, addressed that subspecies niche specialization of oral microbiome is associated with NPC risk, and provided proof-of-concept evidence on the possibility of nasopharyngeal microbiome contributing to NPC treatment. Meanwhile, our findings trigger new thoughts and hypotheses, which call for future efforts.

8.1 DIET AND NPC

The take-home messages from Paper I are:

1) Having more wholegrains, vegetables, fruits, and beans, and limiting the intakes of red meat and processed meat in daily diet shall be widely recommended not only because it may reduce the risk of NPC, and more importantly, it may improve the overall health and prevent many chronic diseases not limited to cancers.

2) Although a birth-cohort study in an endemic area is necessary and sufficient for studying NPC etiology given that early-life exposure is deemed to be essential, unluckily, it is almost impossible in reality. However, well-designed case-control studies and cohort studies with high quality are warranted in further understanding the disease, supplemented by individual participant data meta-analysis.

8.2 ORAL MICROBIOME AND NPC

Based on the same study setting of Paper II, we shall further investigate whether oral microbiome is associated with EBV infection. For example, whether oral microbiome is associated with EBV load/EBV strains among the study participants, what are the correlation patterns among NPC cases and controls, whether oral microbiome is associated with NPC cases' clinical characteristics (e.g., cancer stage, treatment outcome, and survival)?

On the other hand, we hypothesize that *G. adiacens* and co-occurring microorganisms are involved in the generation of oncogenic metabolites (e.g., nitrate and nitrite production) and/or immune regulation. Culture-based experiments are required. Furthermore, the established immortalized nasopharyngeal epithelial cells with type II EBV latency may be an ideal working model. This will request a multidisciplinary, international approach, involving experts from microbiology, virology, biochemistry, and cancer biology.

8.3 COMMENSAL MICROBIOME DURING RADIOTHERAPY

In Paper III, we provide compelling data on the possibility that commensal microbiome/microbes might be biomarkers for treatment outcome prediction and monitoring. Next step, we plan to validate the results and evaluate the performance of commensal microbiome as a biomarker in a larger NPC-patient-cohort (with less sampling time points) using different detection measures (16S-based sequencing, quantitative PCR, etc.). We will also look for more collaboration with other cancer centers.

8.4 NP-MICROBIOME AMONG HIGH-RISK POPULATION OF NPC

We are conducting a study on evaluating the association between the nasopharyngeal microbiome (nasopharyngeal swabs) and EBV status (serum EBV titers, EBV load in the nasopharynx, etc.) based on an NPC cohort in southern China with around 900 participants who were tested EBV positive (serum titers) at the baseline survey.

8.5 COMMENSAL MICROBIOME AND OTHER CANCERS

Besides NPC, we found a significant difference in the vaginal microbial community between patients with cervical cancer and healthy controls. We identified several microorganisms with increased relative abundance along with radiotherapy among patients treated with pelvic IMRT-based treatment (unpublished data). We will investigate the potential association between the human commensal microbiome and various types of cancers in the coming future.

9 ACKNOWLEDGEMENTS

It is a challenging but rewarding adventure for my Ph.D. Here, at the very end, I would like to express my sincere gratefulness to every past and current person who has accompanied me through this journey.

First of all, I would like to express my sincere gratitude to the financial support from the China Scholarship Council, the Swedish Cancer Society (Cancerfonden), the Swedish Research Council (Vetenskapsrådet), the National Cancer Institute at the U.S. National Institutes of Health, and the National Nature Science Foundation of China, as well as support from Karolinska Institutet. Further, I would like to acknowledge all participants in the NPCGEE project and the hospital-based NPC patient cohort; and to thank all the fieldworkers, interviewers, doctors, and nurses who helped us with data and sample collection. None of the studies would have been done without their contributions.

I would like to thank **Weimin Ye**, my brilliant principal supervisor, for offering me the opportunity to pursue a Ph.D. at Karolinska Institutet and work with world-class researchers. Thank you for guiding me to the fields of cancer epidemiology and human microbiome, sharing your profound knowledge in epidemiology and biostatistics, and encouraging me to be an independent researcher. Your expertise, passion, dedication to science, and rigorous scholarship, will light the way of my upcoming journey.

Eva Munck-Wikland, my amiable co-supervisor, thank you for sharing your knowledge and experience as a brilliant and skillful surgery and oncologist. I remember that you brought me the Swedish guidelines in the clinical practice of head and neck cancer in our second meeting. Thank you for taking care of me when I need help.

Zongli Zheng, my talented co-supervisor, thank you for sharing your valuable knowledge and skills in cutting-edge molecular epidemiology and precision medicine. Thank you for all the financial and technical support, as well as inspiring discussions.

To my reliable co-supervisor and friend, **Justine Debelius**. Thank you for walking me through the microbial adventures patiently and for all the support to my Ph.D. I sincerely appreciate that you always stimulate me to think independently. I truly enjoyed our talks and discussions about science, life, foods, and everything. You are an excellent researcher and teacher, as well as a precious friend (良师益友)!

To the excellent statistician and amiable teacher, **Alexander Ploner**. Thank you for making this journey as bright as it can be! It is not only because of all the propound knowledge, brilliant programming skills, and rigorous logical thinking I learned from you, but also the fact that you would be my side whenever I needed it. You are the superhero of my Ph.D.!

Fang Fang, I am sincerely grateful for your presence and considerate care during these years. Both made me go further professionally and personally. Many thanks for your generous and warm support when I needed the most. I genuinely enjoy working with you and hope to build

up our collaboration in the future. You are my excellent role model in science, and I am a big fan of yours!

I would like to thank my previous supervisors and lifetime mentors, **Guangwu Huang** and **Zhe Zhang**, for guiding me from the clinic to scientific research and supporting my Ph.D. studies as always. What I have learned from you will benefit the rest of my life. I am also grateful to **Rensheng Wang**, my mentor in Radiation Oncology. Thank you for introducing me to radiation oncology, sharing your broad knowledge and precious experience, supporting my choice of pursuing a Ph.D. I am looking forward to working with you soon.

I want to thank all of my co-authors for these studies. Special thanks to all the help and input from my amazing co-authors: **Hans-Olov Adami**, **Ellen T. Chang**, **Ingemar Ernberg**, whose expertise was invaluable in formulating the research questions and interpreting the results. Your insightful feedback brought my work to a higher level. I also want to thank **Yixin Zeng**, **Yi Zeng**, **Yonglin Cai**, **Miao Xu**, **Weihua Jia**, **Sumei Cao**, and **Mingfang Ji** for your help and advice. What I have learned from all of you will benefit the rest of my life.

Many thanks to **Yudi Pawitan** and **Paul Lichtenstein** for thoughtful consideration of my Ph.D. and for providing me helpful advice. I would like to thank **Per Hall**, **Kamila Czene**, and **Pär Sparén** for the inspiring discussions we had at MEB. I am grateful for having **Yang Cao** as my Ph.D. mentor.

A big thanks to previous and current colleagues and friends from *Weimin*'s group: **Huan Song**, **Jiaqi Huang**, **Zhiwei Liu**, **Donal Barrett**, **Haiyun Wang**, **Xingdong Chen**, **Yuan Zhang**, **Ping Hu**, **Jingru Yu**, **Isabella Ekheden**, **Nelson Ndegwa Gichora**, **Nikolaos Tertipis**, **Tracy Peters**, **Alessandra Grotta**, **Ulrika Zagai**, **Amelie Plymoth**, **Anna Berglund**, **Weiyao Yin**, **Yawen Sun**, **Lijie Ding**, **Xiaorong Yang**, **Joar Franzen**, **Ji Zhang**, **Yufeng Chen**, **Fei He**, **Song Lin**, **Ping Yuan**, **Tao Tao**, **Weiwei Bian**, **Yun Du**, **Amir Sohrabi**, and **Yvonne Nartey**, for being always supportive. *Huan*, thank you for all the considerate care and joyful time! *Jingru*, thank you for always being by my side, listening, sharing, and accompanying. It means a lot to me! I wish you all the best in the rest part of your Ph.D. *Donal*, I am happy to know you and lovely **Masha Anufrieva**, and I enjoyed all the conversations we had about science, cultures, and life. *Zhiwei*, thank you for sharing epi knowledge and SAS skills. *Xingdong* and *Yuan*, I enjoyed our meeting every time. *Ping Hu*, thank you for sharing knowledge and experience in microbial research. *Isabella*, thank you for all the thoughtful talks and advice. *Nelson*, thank you for sharing the Ph.D. experience and the help with my thesis.

I am profoundly grateful for the excellent working environment at MEB because of all **MEBers**! Many thanks to the outstanding administrators, IT group, data managers, and Biobankers, in particular **Camilla Ahlqvist**, **Alessandra Nanni**, **Mariam Lashkariani**, **Gunilla Sonnebring**, **Frank Pettersson**, **William Day**, and **Mark Divers**. I want to thank all the past and current friends and colleagues for being helpful and supportive, including but not limited to: **Shadi Azam**, **Qing Shen**, **Jie Song**, **Haomin Yang**, **Xu Chen**, **Jiayao Lei**, **Wei He**, **Elisa Longinetti**, **Shuyang Yao**, **Tong Gong**, **Bojing Liu**, **Ruyue Zhang**, **Donghao Lu**,

Jianwei Zhu, **Ruoqing Chen** & **Yiqiang Zhan**, **Daniela Mariosa, Jiangrong Wang** & **Xiaoyuan Ren, Ci Song** & **Zheng Chang, Yunzhang Wang** & **Yinxi Wang**, **Mwenya Mubanga**, **Andreas Yangmovez**, **Emilio Ugalde-Morales**, **Pui San Tan**, **Felix Grassmann**, **Xinhe Mao**, **Erwei Zeng**, **Jiangwei Sun** & **Lin Li**, **Zheng Ning**, **Hong Xu**, **Xia Li**, **Shihua Sun**, **Wenjiang Deng**, **Fei Yang**, **Chen Suo**, **Qi Chen**, **Le Zhang**, **Can Cui**, **Kejia Hu**, **Mei Wang**, **Xiaoying Kang**, **Yuanjun Ma**, **Qi Wang**, **Qian Yang**, **Yingying Yang**, and **Chen Wang**. I also want to acknowledge *Shadi*, my lovely officemate; I enjoyed all the time we're together, inside the office, outside the office, and by phone. Thank you for all the support, sharing, and care for this journey! Special thanks to *Qing* & **Tiansheng Shi**, *Haomin*, *Jingru*, *Wei*, *Xu*, *Jiayao* & **Karl Dahlgren**, for all the joyful and colorful gatherings and activities. *Qing*, many thanks for all the sharing and accompanying along the journey! *Haomin*, thank you for all the helpful discussions and generous sharing!

I am glad to have *Jie* & **Yanpeng Yang**, **Jiangning Gao** & **Xinhai Zhang**. Thank you for accompanying and caring, which made my life brighter. Many thanks to *Tong*, **Shixing Zheng** & **Dongli Liu**, **Yang Xuan**, and **Dong Yang** for the exciting trips, films, and gatherings. I would also like to thank **Yang Lei** for your friendship and support in these years; **Qing Cheng**, for taking care of me from the first day I arrived in Stockholm and always being supportive; **Liudmila Matskova** & **Gösta Winberg**, for inspiring talks and considerate care; **Mo Chen**, for all the inspiring discussions and interesting talks; **Shuping Que** & **Zhengtao Liu**, for your friendship; **Min Guo**, **Rong Yu**, **Dongmei Tong**, **Abdul Aleem**, **Siyu Bao**, **Quan Tang, Xia Hao**, **Muyi Yang**, **Yue Hu**, and **Baoyi Zhou**, for all the interesting talks when we met.

I would like to thank all of my Chinese colleagues in *Guangwu Huang* and *Zhe Zhang*'s research group, in *Rensheng Wang*'s research group, as well as in the dept. of Radiation Oncology, the First Affiliated Hospital of Guangxi Medical University (China). Special thanks to **Xiaoying Zhou**, **Xue Xiao**, **Shumin Wang**, **Yingxi Mo**, **Li Jiang**, **Kai Hu**, and **Tingting Zhang**, for all the supports on my studies.

To my friends in China, included but not limited to: **Xinghe Jiang** & **Yi Su**, **Shan Huang** & **Lie Huang**, **Qiujie Chen** & **Weicheng Tang**, **Chunjie Liao** & **Zijian Ou-yang**, **Wen He**, **Tao Li**, **Liang Li**, **Chuan Yun**, and **Jun Jian**, for our precious friendship! Thank you for taking care of my family and me; for all the lovely time we spent together!

老爸和老妈，这是我第一次对你们说：感谢你们对我的"惯"，让我勇敢地千里求学；感谢你们对我的爱，让我勇敢地完成征程！

感谢我的所有亲人们对我的爱和支持！

"Life isn't like in the movies. Life… is much harder. If you don't walk out, you will think that is the whole world." — Thank me for walking out bravely, for making unremitting efforts, for being better.

10 REFERENCES

1. Sham, J.S., et al., *Detection of subclinical nasopharyngeal carcinoma by fibreoptic endoscopy and multiple biopsy.* Lancet, 1990. **335**(8686): p. 371-4.
2. Vaughan, T.L., et al., *Nasopharyngeal cancer in a low-risk population: defining risk factors by histological type.* Cancer Epidemiol Biomarkers Prev, 1996. **5**(8): p. 587-93.
3. El-Naggar, A.K., et al., *WHO Classification of Head and Neck Tumours*. 2017: International Agency for Research on Cancer.
4. Stelow, E.B. and B.M. Wenig, *Update From The 4th Edition of the World Health Organization Classification of Head and Neck Tumours: Nasopharynx.* Head Neck Pathol, 2017. **11**(1): p. 16-22.
5. Hong Lok Lung, A.K.L.C., Josephine Mun Yee Ko, Yue Cheng and Maria Li Lung, *Identification of Tumor Suppressor Genes via Cell Fusion and Chromosomal Transfer, Tumor Suppressor Genes*, ed. Y. Cheng. 2012: IntechOpen.
6. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
7. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.
8. Jemal, A., et al., *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.
9. Chang, E.T. and H.O. Adami, *The enigmatic epidemiology of nasopharyngeal carcinoma.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(10): p. 1765-77.
10. Tao, Q. and A.T. Chan, *Nasopharyngeal carcinoma: molecular pathogenesis and therapeutic developments.* Expert Rev Mol Med, 2007. **9**(12): p. 1-24.
11. Wei, K.R., et al., *Nasopharyngeal carcinoma incidence and mortality in China in 2010.* Chin J Cancer, 2014. **33**(8): p. 381-7.
12. Wei, K.R., et al., *Nasopharyngeal carcinoma incidence and mortality in China, 2013.* Chin J Cancer, 2017. **36**(1): p. 90.
13. Chen, Y.P., et al., *Nasopharyngeal carcinoma.* Lancet, 2019. **394**(10192): p. 64-80.
14. Burt, R.D., T.L. Vaughan, and B. McKnight, *Descriptive epidemiology and survival analysis of nasopharyngeal carcinoma in the United States.* Int J Cancer, 1992. **52**(4): p. 549-56.
15. Lee, A.W., et al., *Changing epidemiology of nasopharyngeal carcinoma in Hong Kong over a 20-year period (1980-99): an encouraging reduction in both incidence and mortality.* Int J Cancer, 2003. **103**(5): p. 680-5.
16. Armstrong, R.W., et al., *Incidence of nasopharyngeal carcinoma in Malaysia, 1968--1977.* Br J Cancer, 1979. **40**(4): p. 557-67.
17. Sun, L.M., et al., *Trends in the incidence rates of nasopharyngeal carcinoma among Chinese Americans living in Los Angeles County and the San Francisco metropolitan area, 1992-2002.* Am J Epidemiol, 2005. **162**(12): p. 1174-8.

18. Li, C.C., M.C. Yu, and B.E. Henderson, *Some epidemiologic observations of nasopharyngeal carcinoma in Guangdong, People's Republic of China.* Natl Cancer Inst Monogr, 1985. **69**: p. 49-52.

19. Wee, J.T., et al., *Is nasopharyngeal cancer really a "Cantonese cancer"?* Chin J Cancer, 2010. **29**(5): p. 517-26.

20. Ferlay J, E.M., Lam F, et al., *Global Cance Observatory: cancer today. International Agency for Research on Cancer 2018. https://gco.iarc.fr/today (accessed Oct 8, 2020).* Lyon, France.

21. Yu, M.C. and J.M. Yuan, *Epidemiology of nasopharyngeal carcinoma.* Semin Cancer Biol, 2002. **12**(6): p. 421-9.

22. Hsu, C., et al., *Difference in the incidence trend of nasopharyngeal and oropharyngeal carcinomas in Taiwan: implication from age-period-cohort analysis.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 856-61.

23. Luo, J., et al., *Secular trends of nasopharyngeal carcinoma incidence in Singapore, Hong Kong and Los Angeles Chinese populations, 1973-1997.* Eur J Epidemiol, 2007. **22**(8): p. 513-21.

24. Tang, L.L., et al., *Global trends in incidence and mortality of nasopharyngeal carcinoma.* Cancer Lett, 2016. **374**(1): p. 22-30.

25. Chua, M.L.K., et al., *Nasopharyngeal carcinoma.* Lancet, 2016. **387**(10022): p. 1012-1024.

26. Jia, W.H., et al., *Trends in incidence and mortality of nasopharyngeal carcinoma over a 20-25 year period (1978/1983-2002) in Sihui and Cangwu counties in southern China.* BMC Cancer, 2006. **6**: p. 178.

27. Zhang, L.F., et al., *Incidence trend of nasopharyngeal carcinoma from 1987 to 2011 in Sihui County, Guangdong Province, South China: an age-period-cohort analysis.* Chin J Cancer, 2015. **34**(8): p. 350-7.

28. Pathmanathan, R., et al., *Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx. Variants of Epstein-Barr virus-infected neoplasia.* Am J Pathol, 1995. **146**(6): p. 1355-67.

29. Wei, W.I. and J.S. Sham, *Nasopharyngeal carcinoma.* Lancet, 2005. **365**(9476): p. 2041-54.

30. *Nasopharyngeal Carcinoma: From Etiology to Clinical Practice*, ed. M.L.L.a.W.T.N. Anne W.M. Lee. 2019: Elsevier Inc.

31. Epstein, M.A., B.G. Achong, and Y.M. Barr, *Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma.* Lancet, 1964. **1**(7335): p. 702-3.

32. Lieberman, P.M., *Virology. Epstein-Barr virus turns 50.* Science, 2014. **343**(6177): p. 1323-5.

33. Young, L.S., L.F. Yap, and P.G. Murray, *Epstein-Barr virus: more than 50 years old and still providing surprises.* Nat Rev Cancer, 2016. **16**(12): p. 789-802.

34. Niller, H.H., et al., *Epigenetic Alterations in Epstein-Barr Virus-Associated Diseases.* Adv Exp Med Biol, 2016. **879**: p. 39-69.

35. Evans, A.S., *Clinical syndromes associated with EB virus infection.* Adv Intern Med, 1972. **18**: p. 77-93.

36. Babcock, G.J., et al., *EBV persistence in memory B cells in vivo.* Immunity, 1998. **9**(3): p. 395-404.

37. Young, L.S. and C.W. Dawson, *Epstein-Barr virus and nasopharyngeal carcinoma.* Chin J Cancer, 2014. **33**(12): p. 581-90.

38. Young, L.S. and A.B. Rickinson, *Epstein-Barr virus: 40 years on.* Nat Rev Cancer, 2004. **4**(10): p. 757-68.

39. Thorley-Lawson, D.A. and A. Gross, *Persistence of the Epstein-Barr virus and the origins of associated lymphomas.* N Engl J Med, 2004. **350**(13): p. 1328-37.

40. Thompson, M.P. and R. Kurzrock, *Epstein-Barr virus and cancer.* Clin Cancer Res, 2004. **10**(3): p. 803-21.

41. Gunven, P., et al., *Epstein-Barr virus in Burkitt's lymphoma and nasopharyngeal carcinoma. Antibodies to EBV associated membrane and viral capsid antigens in Burkitt lymphoma patients.* Nature, 1970. **228**(5276): p. 1053-6.

42. zur Hausen, H., et al., *EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx.* Nature, 1970. **228**(5276): p. 1056-8.

43. Wolf, H., H. zur Hausen, and V. Becker, *EB viral genomes in epithelial nasopharyngeal carcinoma cells.* Nat New Biol, 1973. **244**(138): p. 245-7.

44. Pathmanathan, R., et al., *Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesions related to nasopharyngeal carcinoma.* N Engl J Med, 1995. **333**(11): p. 693-8.

45. Brooks, L., et al., *Epstein-Barr virus latent gene transcription in nasopharyngeal carcinoma cells: coexpression of EBNA1, LMP1, and LMP2 transcripts.* J Virol, 1992. **66**(5): p. 2689-97.

46. Young, L.S., et al., *Epstein-Barr virus gene expression in nasopharyngeal carcinoma.* J Gen Virol, 1988. **69 (Pt 5)**: p. 1051-65.

47. Sam, C.K., et al., *Analysis of Epstein-Barr virus infection in nasopharyngeal biopsies from a group at high risk of nasopharyngeal carcinoma.* Int J Cancer, 1993. **53**(6): p. 957-62.

48. Tao, Q., et al., *Evidence for lytic infection by Epstein-Barr virus in mucosal lymphocytes instead of nasopharyngeal epithelial cells in normal individuals.* J Med Virol, 1995. **45**(1): p. 71-7.

49. Chan JKC, B.F., McCarron P, W. Foo, Lee AWM, Yip T, Kuo TT, Pilch BZ, Wenig BM, Huang D, Lo KW, Zeng YX, Jia WH, *Pathology and genetics of head and neck tumours*. Nasopharyngeal carcinoma, ed. E.J. Barnes L, Reichart P, Sidransky D. 2005, Lyon: IARC Press, 2005: 85–97.: WHO classifi cation of tumours.

50. Niedobitek, G., *Epstein-Barr virus infection in the pathogenesis of nasopharyngeal carcinoma.* Mol Pathol, 2000. **53**(5): p. 248-54.

51. Tsao, S.W., C.M. Tsang, and K.W. Lo, *Epstein-Barr virus infection and nasopharyngeal carcinoma.* Philos Trans R Soc Lond B Biol Sci, 2017. **372**(1732).

52. Tsang, C.M., et al., *Epstein-Barr virus infection and persistence in nasopharyngeal epithelial cells.* Chin J Cancer, 2014. **33**(11): p. 549-55.

53. Wang, H.B., et al., *Neuropilin 1 is an entry factor that promotes EBV infection of nasopharyngeal epithelial cells.* Nat Commun, 2015. **6**: p. 6240.

54. Chesnokova, L.S. and L.M. Hutt-Fletcher, *Fusion of Epstein-Barr virus with epithelial cells can be triggered by alphavbeta5 in addition to alphavbeta6 and alphavbeta8, and integrin binding triggers a conformational change in glycoproteins gHgL.* J Virol, 2011. **85**(24): p. 13214-23.

55. Xiong, D., et al., *Nonmuscle myosin heavy chain IIA mediates Epstein-Barr virus infection of nasopharyngeal epithelial cells.* Proc Natl Acad Sci U S A, 2015. **112**(35): p. 11036-41.

56. Zhang, H., et al., *Ephrin receptor A2 is an epithelial cell receptor for Epstein-Barr virus entry.* Nat Microbiol, 2018. **3**(2): p. 1-8.

57. Chen, J., et al., *Ephrin receptor A2 is a functional entry receptor for Epstein-Barr virus.* Nat Microbiol, 2018. **3**(2): p. 172-180.

58. Xu, M., et al., *Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma.* Nat Genet, 2019. **51**(7): p. 1131-1136.

59. Tsang, C.M., et al., *Cyclin D1 overexpression supports stable EBV infection in nasopharyngeal epithelial cells.* Proc Natl Acad Sci U S A, 2012. **109**(50): p. E3473-82.

60. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective*, in *The Second Expert Report*. 2007, World Cancer Research Fund/American Institute for Cancer Research.

61. *Diet, Nutrition, Physical Activity and Cancer: a Global Perspective*, in *The Third Expert Report*. 2018, World Cancer Research Fund/American Institute for Cancer Research.

62. Secretan, B., et al., *A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish.* Lancet Oncol, 2009. **10**(11): p. 1033-4.

63. *Personal habits and indoor combustions. A review of human carcinogens*. IARC monographs on the evaluation of carcinogenic risks to humans. Vol. 100 E. 2012: International Agency for Research on Cancer.

64. Ho, J.H.C., *Nasopharyngeal carcinoma in Hong Kong.* In *Cancer of the Nasopharynx. UICC Monograph Series*, 1967: p. 58-63.

65. Yu, M.C., et al., *Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong.* Cancer Res, 1986. **46**(2): p. 956-61.

66. Yu, M.C., T.B. Huang, and B.E. Henderson, *Diet and nasopharyngeal carcinoma: a case-control study in Guangzhou, China.* Int J Cancer, 1989. **43**(6): p. 1077-82.

67. Zheng, Y.M., et al., *Environmental and dietary risk factors for nasopharyngeal carcinoma: a case-control study in Zangwu County, Guangxi, China.* Br J Cancer, 1994. **69**(3): p. 508-14.

68. Yuan, J.M., et al., *Preserved foods in relation to risk of nasopharyngeal carcinoma in Shanghai, China.* Int J Cancer, 2000. **85**(3): p. 358-63.

69. Ning, J.P., et al., *Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China.* J Natl Cancer Inst, 1990. **82**(4): p. 291-6.

70. Jia, W.H., et al., *Traditional Cantonese diet and nasopharyngeal carcinoma risk: a large-scale case-control study in Guangdong, China.* BMC Cancer, 2010. **10**: p. 446.

71. Yu, M.C., et al., *Preserved foods and nasopharyngeal carcinoma: a case-control study in Guangxi, China.* Cancer Res, 1988. **48**(7): p. 1954-9.

72. Ward, M.H., et al., *Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan.* Int J Cancer, 2000. **86**(5): p. 603-9.

73. Guo, X., et al., *Evaluation of nonviral risk factors for nasopharyngeal carcinoma in a high-risk population of Southern China.* Int J Cancer, 2009. **124**(12): p. 2942-7.

74. Armstrong, R.W., et al., *Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese.* Cancer Res, 1983. **43**(6): p. 2967-70.

75. Armstrong, R.W. and A.C. Eng, *Salted fish and nasopharyngeal carcinoma in Malaysia.* Soc Sci Med, 1983. **17**(20): p. 1559-67.

76. Cancer, W.H.O.I.A.f.R.o., *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans* Some naturally occurring substances: food items and constituents, heterocyclic aromatic amines and mycotoxins. Vol. Vol.56 pp. 1993: World Health Organization.

77. Zou, X.N., S.H. Lu, and B. Liu, *Volatile N-nitrosamines and their precursors in Chinese salted fish--a possible etological factor for NPC in china.* Int J Cancer, 1994. **59**(2): p. 155-8.

78. Mirvish, S.S., *Role of N-nitroso compounds (NOC) and N-nitrosation in etiology of gastric, esophageal, nasopharyngeal and bladder cancer and contribution to cancer of known exposures to NOC.* Cancer Lett, 1995. **93**(1): p. 17-48.

79. Dodd, L.E., et al., *Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(11): p. 2216-25.

80. Jin, J., Z. Ouyang, and Z. Wang, *Association of fruit and vegetables with the risk of nasopharyngeal cancer: evidence from a meta-analysis.* Sci Rep, 2014. **4**: p. 5229.

81. Li, F., et al., *Red Meat and Processed Meat Consumption and Nasopharyngeal Carcinoma Risk: A Dose-response Meta-analysis of Observational Studies.* Nutr Cancer, 2016. **68**(6): p. 1034-43.

82. Gallicchio, L., et al., *Adulthood consumption of preserved and nonpreserved vegetables and the risk of nasopharyngeal carcinoma: a systematic review.* Int J Cancer, 2006. **119**(5): p. 1125-35.

83. Mai, Z.M., et al., *Milk consumption in relation to incidence of nasopharyngeal carcinoma in 48 countries/regions.* BMC Cancer, 2015. **15**: p. 994.

84. Hsu, W.L., et al., *Lowered risk of nasopharyngeal carcinoma and intake of plant vitamin, fresh fish, green tea and coffee: a case-control study in Taiwan.* PLoS One, 2012. **7**(7): p. e41779.

85. Mai, Z.M., et al., *Milk Consumption Across Life Periods in Relation to Lower Risk of Nasopharyngeal Carcinoma: A Multicentre Case-Control Study.* Front Oncol, 2019. **9**: p. 253.

86. Polesel, J., et al., *Consumption of fruit, vegetables, and other food groups and the risk of nasopharyngeal carcinoma.* Cancer Causes Control, 2013. **24**(6): p. 1157-65.

87. Feng, B.J., et al., *Dietary risk factors for nasopharyngeal carcinoma in Maghrebian countries.* Int J Cancer, 2007. **121**(7): p. 1550-5.

88. Hjalgrim, H., J. Friborg, and M. Melbye, *The epidemiology of EBV and its association with malignant disease*, in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, A. Arvin, et al., Editors. 2007: Cambridge.

89. Schulze, M.B., et al., *Food based dietary patterns and chronic disease prevention.* BMJ, 2018. **361**: p. k2396.

90. Chow, W.H., et al., *Tobacco use and nasopharyngeal carcinoma in a cohort of US veterans.* Int J Cancer, 1993. **55**(4): p. 538-40.

91. Doll, R., et al., *Mortality from cancer in relation to smoking: 50 years observations on British doctors.* Br J Cancer, 2005. **92**(3): p. 426-9.

92. Friborg, J.T., et al., *A prospective study of tobacco and alcohol use as risk factors for pharyngeal carcinomas in Singapore Chinese.* Cancer, 2007. **109**(6): p. 1183-91.

93. Hsu, W.L., et al., *Independent effect of EBV and cigarette smoking on nasopharyngeal carcinoma: a 20-year follow-up study on 9,622 males without family history in Taiwan.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(4): p. 1218-26.

94. Xu, F.H., et al., *An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein-Barr virus activation.* J Natl Cancer Inst, 2012. **104**(18): p. 1396-410.

95. Fachiroh, J., et al., *Tobacco consumption and genetic susceptibility to nasopharyngeal carcinoma (NPC) in Thailand.* Cancer Causes Control, 2012. **23**(12): p. 1995-2002.

96. Chang, E.T., et al., *Active and Passive Smoking and Risk of Nasopharyngeal Carcinoma: A Population-Based Case-Control Study in Southern China.* Am J Epidemiol, 2017. **185**(12): p. 1272-1280.

97. Xue, W.Q., et al., *Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011.* Am J Epidemiol, 2013. **178**(3): p. 325-38.

98. Long, M., et al., *Cigarette smoking and the risk of nasopharyngeal carcinoma: a meta-analysis of epidemiological studies.* BMJ Open, 2017. **7**(10): p. e016582.

99. Chen, L., et al., *Alcohol consumption and the risk of nasopharyngeal carcinoma: a systematic review.* Nutr Cancer, 2009. **61**(1): p. 1-15.

100. Ruan, H.L., et al., *Alcohol and tea consumption in relation to the risk of nasopharyngeal carcinoma in Guangdong, China.* Front Med China, 2010. **4**(4): p. 448-56.

101. Du, T., et al., *Association Between Alcohol Consumption and Risk of Nasopharyngeal Carcinoma: A Comprehensive Meta-Analysis of Epidemiological Studies.* Alcohol Clin Exp Res, 2019. **43**(11): p. 2262-2273.

102. Ji, X., et al., *Nasopharyngeal carcinoma risk by histologic type in central China: impact of smoking, alcohol and family history.* Int J Cancer, 2011. **129**(3): p. 724-32.

103. Polesel, J., et al., *Tobacco smoking, alcohol drinking, and the risk of different histological types of nasopharyngeal cancer in a low-risk population.* Oral Oncol, 2011. **47**(6): p. 541-5.

104. Yong, S.K., et al., *Associations of lifestyle and diet with the risk of nasopharyngeal carcinoma in Singapore: a case-control study.* Chin J Cancer, 2017. **36**(1): p. 3.

105. Stolzenberg-Solomon, R.Z., et al., *Tooth loss, pancreatic cancer, and Helicobacter pylori.* Am J Clin Nutr, 2003. **78**(1): p. 176-81.

106. Dar, N.A., et al., *Poor oral hygiene and risk of esophageal squamous cell carcinoma in Kashmir.* Br J Cancer, 2013. **109**(5): p. 1367-72.

107. Michaud, D.S., et al., *Periodontal Disease, Tooth Loss, and Cancer Risk.* Epidemiol Rev, 2017. **39**(1): p. 49-58.

108. Ndegwa, N., et al., *Association between poor oral health and gastric cancer: A prospective cohort study.* Int J Cancer, 2018. **143**(9): p. 2281-2288.

109. Gholizadeh, P., et al., *Role of oral microbiome on oral cancers, a review.* Biomed Pharmacother, 2016. **84**: p. 552-558.

110. Hayes, R.B., et al., *Association of Oral Microbiome With Risk for Incident Head and Neck Squamous Cell Cancer.* JAMA Oncol, 2018. **4**(3): p. 358-365.

111. Flemer, B., et al., *The oral microbiota in colorectal cancer is distinctive and predictive.* Gut, 2018. **67**(8): p. 1454-1463.

112. Fan, X., et al., *Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study.* Gut, 2018. **67**(1): p. 120-127.

113. Liu, Z., et al., *Oral Hygiene and Risk of Nasopharyngeal Carcinoma-A Population-Based Case-Control Study in China.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(8): p. 1201-7.

114. Saygun, I., et al., *Periodontitis lesions are a source of salivary cytomegalovirus and Epstein-Barr virus.* J Periodontal Res, 2005. **40**(2): p. 187-91.

115. Santangelo, R., et al., *Bacterial and viral DNA in periodontal disease: a study using multiplex PCR.* New Microbiol, 2004. **27**(2): p. 133-7.

116. Hirayama, T. and Y. Ito, *A new view of the etiology of nasopharyngeal carcinoma.* Prev Med, 1981. **10**(5): p. 614-22.

117. Zeng, Y., et al., *Screening of Epstein-Barr virus early antigen expression inducers from Chinese medicinal herbs and plants.* Biomed Environ Sci, 1994. **7**(1): p. 50-5.

118. Gourzones, C., C. Barjon, and P. Busson, *Host-tumor interactions in nasopharyngeal carcinomas.* Semin Cancer Biol, 2012. **22**(2): p. 127-36.

119. Man, S.M., *Inflammasomes in the gastrointestinal tract: infection, cancer and gut microbiota homeostasis.* Nat Rev Gastroenterol Hepatol, 2018. **15**(12): p. 721-737.

120. Nociti, F.H., Jr., M.Z. Casati, and P.M. Duarte, *Current perspective of the impact of smoking on the progression and treatment of periodontitis.* Periodontol 2000, 2015. **67**(1): p. 187-210.

121. Wu, J., et al., *Cigarette smoking and the oral microbiome in a large study of American adults.* ISME J, 2016. **10**(10): p. 2435-46.

122. Kato, I., et al., *Nutritional Correlates of Human Oral Microbiome.* J Am Coll Nutr, 2017. **36**(2): p. 88-98.

123. Kilian, M., et al., *The oral microbiome - an update for oral healthcare professionals.* Br Dent J, 2016. **221**(10): p. 657-666.

124. Baker, J.L. and A. Edlund, *Exploiting the Oral Microbiome to Prevent Tooth Decay: Has Evolution Already Provided the Best Tools?* Front Microbiol, 2018. **9**: p. 3323.

125. Wang, H.J., et al., *[A molecular phylogeny of Shennongjia white bear based on mitochondrial cytochrome b gene sequence]*. Yi Chuan, 2006. **28**(10): p. 1237-41.

126. Hildesheim, A., et al., *Occupational exposure to wood, formaldehyde, and solvents and risk of nasopharyngeal carcinoma.* Cancer Epidemiol Biomarkers Prev, 2001. **10**(11): p. 1145-53.

127. Li, W., et al., *Occupational risk factors for nasopharyngeal cancer among female textile workers in Shanghai, China.* Occup Environ Med, 2006. **63**(1): p. 39-44.

128. Yu, M.C., et al., *Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China.* Int J Cancer, 1990. **45**(6): p. 1033-9.

129. Armstrong, R.W., et al., *Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles, formaldehyde and heat.* Int J Epidemiol, 2000. **29**(6): p. 991-8.

130. Albeck, H., et al., *Familial clusters of nasopharyngeal carcinoma and salivary gland carcinomas in Greenland natives.* Cancer, 1993. **72**(1): p. 196-200.

131. Ko, J.Y., et al., *Familial clustering of nasopharyngeal carcinoma.* Otolaryngol Head Neck Surg, 1998. **118**(5): p. 736-7.

132. Zhang, F. and J. Zhang, *Clinical hereditary characteristics in nasopharyngeal carcinoma through Ye-Liang's family cluster.* Chin Med J (Engl), 1999. **112**(2): p. 185-7.

133. Jia, W.H., et al., *Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China.* Cancer, 2004. **101**(2): p. 363-9.

134. Friborg, J., et al., *Cancer susceptibility in nasopharyngeal carcinoma families--a population-based cohort study.* Cancer Res, 2005. **65**(18): p. 8567-72.

135. Loh, K.S., et al., *Familial nasopharyngeal carcinoma in a cohort of 200 patients.* Arch Otolaryngol Head Neck Surg, 2006. **132**(1): p. 82-5.

136. Liu, Z., et al., *Quantification of familial risk of nasopharyngeal carcinoma in a high-incidence area.* Cancer, 2017. **123**(14): p. 2716-2725.

137. Simons, M.J., et al., *Probable identification of an HL-A second-locus antigen associated with a high risk of nasopharyngeal carcinoma.* Lancet, 1975. **1**(7899): p. 142-3.

138. Simons, M.J., et al., *Immunogenetic aspects of nasopharyngeal carcinoma. IV. Increased risk in Chinese of nasopharyngeal carcinoma associated with a Chinese-related HLA profile (A2, Singapore 2).* J Natl Cancer Inst, 1976. **57**(5): p. 977-80.

139. Lu, S.J., et al., *Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region.* Nature, 1990. **346**(6283): p. 470-1.

140. Yu, K.J., et al., *Association of human leukocyte antigens with nasopharyngeal carcinoma in high-risk multiplex families in Taiwan.* Hum Immunol, 2009. **70**(11): p. 910-4.

141. Tang, M., et al., *Haplotype-dependent HLA susceptibility to nasopharyngeal carcinoma in a Southern Chinese population.* Genes Immun, 2010. **11**(4): p. 334-42.

142. Tse, K.P., et al., *Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3.* Am J Hum Genet, 2009. **85**(2): p. 194-203.

143. Bei, J.X., et al., *A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci.* Nat Genet, 2010. **42**(7): p. 599-603.

144. Zhao, M., et al., *Further evidence for the existence of major susceptibility of nasopharyngeal carcinoma in the region near HLA-A locus in Southern Chinese.* J Transl Med, 2012. **10**: p. 57.

145. Tang, M., et al., *The principal genetic determinants for nasopharyngeal carcinoma in China involve the HLA class I antigen recognition groove.* PLoS Genet, 2012. **8**(11): p. e1003103.

146. Chin, Y.M., et al., *HLA-A SNPs and amino acid variants are associated with nasopharyngeal carcinoma in Malaysian Chinese.* Int J Cancer, 2015. **136**(3): p. 678-87.

147. Mokni-Baizig, N., et al., *HLA-A*26-A*30 and HLA-DRB1*10 could be predictors of nasopharyngeal carcinoma risk in high-risk Tunisian families.* J Oral Sci, 2017. **59**(2): p. 289-296.

148. Yee Ko, J.M., et al., *Multigene pathway-based analyses identify nasopharyngeal carcinoma risk associations for cumulative adverse effects of TERT-CLPTM1L and DNA double-strand breaks repair.* Int J Cancer, 2014. **135**(7): p. 1634-45.

149. Bei, J.X., et al., *A GWAS Meta-analysis and Replication Study Identifies a Novel Locus within CLPTM1L/TERT Associated with Nasopharyngeal Carcinoma in Individuals of Chinese Ancestry.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(1): p. 188-192.

150. Cui, Q., et al., *An extended genome-wide association study identifies novel susceptibility loci for nasopharyngeal carcinoma.* Hum Mol Genet, 2016. **25**(16): p. 3626-3634.

151. Dai, W., et al., *Whole-exome sequencing identifies MST1R as a genetic susceptibility gene in nasopharyngeal carcinoma.* Proc Natl Acad Sci U S A, 2016. **113**(12): p. 3317-22.

152. Ng, C.C., et al., *A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma.* J Hum Genet, 2009. **54**(7): p. 392-7.

153. Xiong, G., et al., *Epstein-Barr virus (EBV) infection in Chinese children: a retrospective study of age-specific prevalence.* PLoS One, 2014. **9**(6): p. e99857.

154. Lee, A.W., et al., *Nasopharyngeal carcinoma: presenting symptoms and duration before diagnosis.* Hong Kong Med J, 1997. **3**(4): p. 355-361.

155. (AJCC), T.A.J.C.o.C., *AJCC Cancer Staging Manual*. 8 ed. 2016: Springer International Publishing.

156. Amin, M.B., et al., *The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging.* CA Cancer J Clin, 2017. **67**(2): p. 93-99.

157. Tang, L.L., et al., *Validation of the 8th Edition of the UICC/AJCC Staging System for Nasopharyngeal Carcinoma From Endemic Areas in the Intensity-Modulated Radiotherapy Era.* J Natl Compr Canc Netw, 2017. **15**(7): p. 913-919.

158. Lee, A.W.M., et al., *The strength/weakness of the AJCC/UICC staging system (7th edition) for nasopharyngeal cancer and suggestions for future improvement.* Oral Oncol, 2012. **48**(10): p. 1007-1013.

159. Kam, M.K., et al., *Prospective randomized study of intensity-modulated radiotherapy on salivary gland function in early-stage nasopharyngeal carcinoma patients.* J Clin Oncol, 2007. **25**(31): p. 4873-9.

160. Pow, E.H., et al., *Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: initial report on a randomized controlled clinical trial.* Int J Radiat Oncol Biol Phys, 2006. **66**(4): p. 981-91.

161. Fang, F.M., et al., *Quality of life and survival outcome for patients with nasopharyngeal carcinoma receiving three-dimensional conformal radiotherapy vs. intensity-modulated radiotherapy-a longitudinal study.* Int J Radiat Oncol Biol Phys, 2008. **72**(2): p. 356-64.

162. Su, S.F., et al., *Long-term outcomes of early-stage nasopharyngeal carcinoma patients treated with intensity-modulated radiotherapy alone.* Int J Radiat Oncol Biol Phys, 2012. **82**(1): p. 327-33.

163. Peng, G., et al., *A prospective, randomized study comparing outcomes and toxicities of intensity-modulated radiotherapy vs. conventional two-dimensional radiotherapy for the treatment of nasopharyngeal carcinoma.* Radiother Oncol, 2012. **104**(3): p. 286-93.

164. Lee, A.W., et al., *Evolution of treatment for nasopharyngeal cancer--success and setback in the intensity-modulated radiotherapy era.* Radiother Oncol, 2014. **110**(3): p. 377-84.

165. Sun, X., et al., *Long-term outcomes of intensity-modulated radiotherapy for 868 patients with nasopharyngeal carcinoma: an analysis of survival and treatment toxicities.* Radiother Oncol, 2014. **110**(3): p. 398-403.

166. Zhang, B., et al., *Intensity-modulated radiation therapy versus 2D-RT or 3D-CRT for the treatment of nasopharyngeal carcinoma: A systematic review and meta-analysis.* Oral Oncol, 2015. **51**(11): p. 1041-1046.

167. Mao, Y.P., et al., *Prognostic factors and failure patterns in non-metastatic nasopharyngeal carcinoma after intensity-modulated radiotherapy.* Chin J Cancer, 2016. **35**(1): p. 103.

168. Lee, A.W., et al., *Management of Nasopharyngeal Carcinoma: Current Practice and Future Perspective.* J Clin Oncol, 2015. **33**(29): p. 3356-64.

169. Zhang, M.X., et al., *Intensity-modulated radiotherapy prolongs the survival of patients with nasopharyngeal carcinoma compared with conventional two-dimensional radiotherapy: A 10-year experience with a large cohort and long follow-up.* Eur J Cancer, 2015. **51**(17): p. 2587-95.

170. Zong, J., et al., *Impact of intensity-modulated radiotherapy on nasopharyngeal carcinoma: Validation of the 7th edition AJCC staging system.* Oral Oncol, 2015. **51**(3): p. 254-9.

171. Lin, J.C., et al., *Phase III study of concurrent chemoradiotherapy versus radiotherapy alone for advanced nasopharyngeal carcinoma: positive effect on overall and progression-free survival.* J Clin Oncol, 2003. **21**(4): p. 631-7.

172. Chan, A.T., et al., *Overall survival after concurrent cisplatin-radiotherapy compared with radiotherapy alone in locoregionally advanced nasopharyngeal carcinoma.* J Natl Cancer Inst, 2005. **97**(7): p. 536-9.

173. Baujat, B., et al., *Chemotherapy in locally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis of eight randomized trials and 1753 patients.* Int J Radiat Oncol Biol Phys, 2006. **64**(1): p. 47-56.

174. Wu, X., et al., *Long-term follow-up of a phase III study comparing radiotherapy with or without weekly oxaliplatin for locoregionally advanced nasopharyngeal carcinoma.* Ann Oncol, 2013. **24**(8): p. 2131-6.

175. Wee, J., et al., *Randomized trial of radiotherapy versus concurrent chemoradiotherapy followed by adjuvant chemotherapy in patients with American Joint Committee on Cancer/International Union against cancer stage III and IV nasopharyngeal cancer of the endemic variety.* J Clin Oncol, 2005. **23**(27): p. 6730-8.

176. Lee, A.W., et al., *Randomized trial of radiotherapy plus concurrent-adjuvant chemotherapy vs radiotherapy alone for regionally advanced nasopharyngeal carcinoma.* J Natl Cancer Inst, 2010. **102**(15): p. 1188-98.

177. Chen, Y., et al., *Progress report of a randomized trial comparing long-term survival and late toxicity of concurrent chemoradiotherapy with adjuvant chemotherapy versus radiotherapy alone in patients with stage III to IVB nasopharyngeal carcinoma from endemic regions of China.* Cancer, 2013. **119**(12): p. 2230-8.

178. Blanchard, P., et al., *Chemotherapy and radiotherapy in nasopharyngeal carcinoma: an update of the MAC-NPC meta-analysis.* Lancet Oncol, 2015. **16**(6): p. 645-55.

179. Chen, Y.P., et al., *A Bayesian network meta-analysis comparing concurrent chemoradiotherapy followed by adjuvant chemotherapy, concurrent chemoradiotherapy alone and radiotherapy alone in patients with locoregionally advanced nasopharyngeal carcinoma.* Ann Oncol, 2015. **26**(1): p. 205-11.

180. Ribassin-Majed, L., et al., *What Is the Best Treatment of Locally Advanced Nasopharyngeal Carcinoma? An Individual Patient Data Network Meta-Analysis.* J Clin Oncol, 2017. **35**(5): p. 498-505.

181. Pfister, D.G., et al., *Head and Neck Cancers, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology.* J Natl Compr Canc Netw, 2020. **18**(7): p. 873-898.

182. Jiang, F., et al., *Long-term outcomes and failure patterns of patients with nasopharyngeal carcinoma staged by magnetic resonance imaging in intensity-modulated radiotherapy era: The Zhejiang Cancer Hospital's experience.* J Cancer Res Ther, 2015. **11 Suppl 2**: p. C179-84.

183. Ou, X., et al., *Treatment outcomes and late toxicities of 869 patients with nasopharyngeal carcinoma treated with definitive intensity modulated radiation therapy: new insight into the value of total dose of cisplatin and radiation boost.* Oncotarget, 2015. **6**(35): p. 38381-97.

184. Setton, J., et al., *Long-term patterns of relapse and survival following definitive intensity-modulated radiotherapy for non-endemic nasopharyngeal carcinoma.* Oral Oncol, 2016. **53**: p. 67-73.

185. Au, K.H., et al., *Treatment outcomes of nasopharyngeal carcinoma in modern era after intensity modulated radiotherapy (IMRT) in Hong Kong: A report of 3328 patients (HKNPCSG 1301 study).* Oral Oncol, 2018. **77**: p. 16-21.

186. Kwong, D.L., et al., *The time course of histologic remission after treatment of patients with nasopharyngeal carcinoma.* Cancer, 1999. **85**(7): p. 1446-53.

187. Lin, G.W., et al., *The use of MR imaging to detect residual versus recurrent nasopharyngeal carcinoma following treatment with radiation therapy.* Eur J Radiol, 2013. **82**(12): p. 2240-6.

188. Cao, C.N., et al., *Clinical outcomes and patterns of failure after intensity-modulated radiotherapy for T4 nasopharyngeal carcinoma.* Oral Oncol, 2013. **49**(2): p. 175-81.

189. Kong, F.F., et al., *Effectiveness and toxicities of intensity-modulated radiation therapy for patients with T4 nasopharyngeal carcinoma.* PLoS One, 2014. **9**(3): p. e91362.

190. Cao, C.N., et al., *Update report of T4 classification nasopharyngeal carcinoma after intensity-modulated radiotherapy: an analysis of survival and treatment toxicities.* Oral Oncol, 2015. **51**(2): p. 190-4.

191. Tsang, R.K. and W.I. Wei, *Salvage surgery for nasopharyngeal cancer.* World J Otorhinolaryngol Head Neck Surg, 2015. **1**(1): p. 34-43.

192. Liu, Y.P., et al., *Surgery for isolated regional failure in nasopharyngeal carcinoma after radiation: Selective or comprehensive neck dissection.* Laryngoscope, 2019. **129**(2): p. 387-395.

193. Leong, Y.H., et al., *Long-term outcomes after reirradiation in nasopharyngeal carcinoma with intensity-modulated radiotherapy: A meta-analysis.* Head Neck, 2018. **40**(3): p. 622-631.

194. Chan, O.S. and R.K. Ngan, *Individualized treatment in stage IVC nasopharyngeal carcinoma.* Oral Oncol, 2014. **50**(9): p. 791-7.

195. Shen, L.J., et al., *Subdivision of M category for nasopharyngeal carcinoma with synchronous metastasis: time to expand the M categorization system.* Chin J Cancer, 2015. **34**(10): p. 450-8.

196. Zou, X., et al., *Establishment and validation of M1 stage subdivisions for de novo metastatic nasopharyngeal carcinoma to better predict prognosis and guide treatment.* Eur J Cancer, 2017. **77**: p. 117-126.

197. Prawira, A., et al., *Systemic therapies for recurrent or metastatic nasopharyngeal carcinoma: a systematic review.* Br J Cancer, 2017. **117**(12): p. 1743-1752.

198. Lee, V.H., et al., *Correlation of PD-L1 Expression of Tumor Cells with Survival Outcomes after Radical Intensity-Modulated Radiation Therapy for Non-Metastatic Nasopharyngeal Carcinoma.* PLoS One, 2016. **11**(6): p. e0157969.

199. Zhu, Q., et al., *Tumor cells PD-L1 expression as a favorable prognosis factor in nasopharyngeal carcinoma patients with pre-existing intratumor-infiltrating lymphocytes.* Oncoimmunology, 2017. **6**(5): p. e1312240.

200. Wang, Y.Q., et al., *Prognostic significance of tumor-infiltrating lymphocytes in nondisseminated nasopharyngeal carcinoma: A large-scale cohort study.* Int J Cancer, 2018. **142**(12): p. 2558-2566.

201. Larbcharoensub, N., et al., *Characterization of PD-L1 and PD-1 Expression and CD8+ Tumor-infiltrating Lymphocyte in Epstein-Barr Virus-associated Nasopharyngeal Carcinoma.* Am J Clin Oncol, 2018. **41**(12): p. 1204-1210.

202. Hsu, C., et al., *Safety and Antitumor Activity of Pembrolizumab in Patients With Programmed Death-Ligand 1-Positive Nasopharyngeal Carcinoma: Results of the KEYNOTE-028 Study.* J Clin Oncol, 2017. **35**(36): p. 4050-4056.

203. Ma, B.B.Y., et al., *Antitumor Activity of Nivolumab in Recurrent and Metastatic Nasopharyngeal Carcinoma: An International, Multicenter Study of the Mayo Clinic Phase 2 Consortium (NCI-9742).* J Clin Oncol, 2018. **36**(14): p. 1412-1418.

204. Fang, W., et al., *Camrelizumab (SHR-1210) alone or in combination with gemcitabine plus cisplatin for nasopharyngeal carcinoma: results from two single-arm, phase 1 trials.* Lancet Oncol, 2018. **19**(10): p. 1338-1350.

205. Lin, J.C., et al., *Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma.* N Engl J Med, 2004. **350**(24): p. 2461-70.

206. Leung, S.F., et al., *Plasma Epstein-Barr viral deoxyribonucleic acid quantitation complements tumor-node-metastasis staging prognostication in nasopharyngeal carcinoma.* J Clin Oncol, 2006. **24**(34): p. 5414-8.

207. Chan, J.Y. and S.T. Wong, *The role of plasma Epstein-Barr virus DNA in the management of recurrent nasopharyngeal carcinoma.* Laryngoscope, 2014. **124**(1): p. 126-30.

208. Leung, S.F., et al., *Plasma Epstein-Barr viral DNA load at midpoint of radiotherapy course predicts outcome in advanced-stage nasopharyngeal carcinoma.* Ann Oncol, 2014. **25**(6): p. 1204-8.

209. Twu, C.W., et al., *Metronomic adjuvant chemotherapy improves treatment outcome in nasopharyngeal carcinoma patients with postradiation persistently detectable plasma Epstein-Barr virus deoxyribonucleic acid.* Int J Radiat Oncol Biol Phys, 2014. **89**(1): p. 21-9.

210. Chan, A.T.C., et al., *Analysis of Plasma Epstein-Barr Virus DNA in Nasopharyngeal Cancer After Chemoradiation to Identify High-Risk Patients for Adjuvant Chemotherapy: A Randomized Controlled Trial.* J Clin Oncol, 2018: p. JCO2018777847.

211. Sun, X.S., et al., *Identifying optimal candidates for local treatment of the primary tumor among patients with de novo metastatic nasopharyngeal carcinoma: a retrospective cohort study based on Epstein-Barr virus DNA level and tumor response to palliative chemotherapy.* BMC Cancer, 2019. **19**(1): p. 92.

212. Kim, K.Y., et al., *Current State of PCR-Based Epstein-Barr Virus DNA Testing for Nasopharyngeal Cancer.* J Natl Cancer Inst, 2017. **109**(4).

213. Le, Q.T., et al., *An international collaboration to harmonize the quantitative plasma Epstein-Barr virus DNA assay for future biomarker-guided trials in nasopharyngeal carcinoma.* Clin Cancer Res, 2013. **19**(8): p. 2208-15.

214. Iida, N., et al., *Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment.* Science, 2013. **342**(6161): p. 967-70.

215. Geller, L.T., et al., *Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine.* Science, 2017. **357**(6356): p. 1156-1160.

216. Routy, B., et al., *Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors.* Science, 2018. **359**(6371): p. 91-97.

217. Roy, S. and G. Trinchieri, *Microbiota: a key orchestrator of cancer therapy.* Nat Rev Cancer, 2017. **17**(5): p. 271-285.

218. Bhatt, A.P., M.R. Redinbo, and S.J. Bultman, *The role of the microbiome in cancer development and therapy.* CA Cancer J Clin, 2017. **67**(4): p. 326-344.

219. Ferreira, M.R., et al., *Microbiota and radiation-induced bowel toxicity: lessons from inflammatory bowel disease for the radiation oncologist.* Lancet Oncol, 2014. **15**(3): p. e139-47.

220. Schuurhuis, J.M., et al., *Head and neck intensity modulated radiation therapy leads to an increase of opportunistic oral pathogens.* Oral Oncol, 2016. **58**: p. 32-40.

221. McQuade, J.L., et al., *Modulating the microbiome to improve therapeutic response in cancer.* Lancet Oncol, 2019. **20**(2): p. e77-e91.

222. Ye, W., et al., *Development of a population-based cancer case-control study in southern china.* Oncotarget, 2017. **8**(50): p. 87073-87085.

223. Lin, C., et al., *Chinese nonmedicinal herbal diet and risk of nasopharyngeal carcinoma: A population-based case-control study.* Cancer, 2019. **125**(24): p. 4462-4470.

224. Barrett, D., et al., *Past and Recent Salted Fish and Preserved Food Intakes Are Weakly Associated with Nasopharyngeal Carcinoma Risk in Adults in Southern China.* J Nutr, 2019.

225. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.

226. Schwartz, L.H., et al., *RECIST 1.1 - Standardisation and disease-specific adaptations: Perspectives from the RECIST Working Group.* Eur J Cancer, 2016. **62**: p. 138-45.

227. Yang Y, W.G., Pan X., *China food composition*. 2009, Beijing, China: Peking University Medical Press.

228. Knight, R., et al., *Best practices for analysing microbiomes.* Nat Rev Microbiol, 2018. **16**(7): p. 410-422.

229. Herlemann, D.P., et al., *Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea.* ISME J, 2011. **5**(10): p. 1571-9.

230. Hugerth, L.W., et al., *DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies.* Appl Environ Microbiol, 2014. **80**(16): p. 5116-23.

231. Amir, A., et al., *Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns.* mSystems, 2017. **2**(2).

232. Bolyen, E., et al., *Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2.* Nat Biotechnol, 2019. **37**(8): p. 852-857.

233. Wang, Q., et al., *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.* Appl Environ Microbiol, 2007. **73**(16): p. 5261-7.

234. McDonald, D., et al., *An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea.* ISME J, 2012. **6**(3): p. 610-8.

235. Janssen, S., et al., *Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information.* mSystems, 2018. **3**(3).

236. Armstrong, R.W., et al., *Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures.* Int J Cancer, 1998. **77**(2): p. 228-35.

237. Lo, Y.L., et al., *Partial Least Square Discriminant Analysis Discovered a Dietary Pattern Inversely Associated with Nasopharyngeal Carcinoma Risk.* PLoS One, 2016. **11**(6): p. e0155892.

238. Debelius, J.W., et al., *Subspecies Niche Specialization in the Oral Microbiome Is Associated with Nasopharyngeal Carcinoma Risk.* mSystems, 2020. **5**(4).

239. Huang, T., et al., *Radiation Therapy-Induced Changes of the Nasopharyngeal Commensal Microbiome in Nasopharyngeal Carcinoma Patients.* Int J Radiat Oncol Biol Phys, 2020.

240. Edefonti, V., et al., *Nutrient-based dietary patterns and nasopharyngeal cancer: evidence from an exploratory factor analysis.* Br J Cancer, 2015. **112**(3): p. 446-54.

241. Zaura, E., et al., *Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces.* mBio, 2015. **6**(6): p. e01693-15.

242. Hyde, E.R., et al., *Metagenomic analysis of nitrate-reducing bacteria in the oral cavity: implications for nitric oxide homeostasis.* PLoS One, 2014. **9**(3): p. e88645.

243. Zaneveld, J.R., R. McMinds, and R. Vega Thurber, *Stress and stability: applying the Anna Karenina principle to animal microbiomes.* Nat Microbiol, 2017. **2**: p. 17121.

244. Battista, J.R., A.M. Earl, and M.J. Park, *Why is Deinococcus radiodurans so resistant to ionizing radiation?* Trends Microbiol, 1999. **7**(9): p. 362-5.

245. Grimes, D.A. and K.F. Schulz, *Bias and causal associations in observational research.* Lancet, 2002. **359**(9302): p. 248-52.

246. Goodrich, J.K., et al., *Conducting a microbiome study.* Cell, 2014. **158**(2): p. 250-262.

247. Trichopoulou, A., et al., *Adherence to a Mediterranean diet and survival in a Greek population.* N Engl J Med, 2003. **348**(26): p. 2599-608.

248. Gleason, P.M., et al., *Publishing nutrition research: a review of multivariate techniques--part 3: data reduction methods.* J Acad Nutr Diet, 2015. **115**(7): p. 1072-82.

249. Nettleton, J.A., et al., *Associations between markers of subclinical atherosclerosis and dietary patterns derived by principal components analysis and reduced rank regression in the Multi-Ethnic Study of Atherosclerosis (MESA).* Am J Clin Nutr, 2007. **85**(6): p. 1615-25.

250. Reedy, J., et al., *Comparing 3 dietary pattern methods--cluster analysis, factor analysis, and index analysis--With colorectal cancer risk: The NIH-AARP Diet and Health Study.* Am J Epidemiol, 2010. **171**(4): p. 479-87.

251. Hoffmann, K., et al., *Application of a new statistical method to derive dietary patterns in nutritional epidemiology.* Am J Epidemiol, 2004. **159**(10): p. 935-44.

252. Tsilimigras, M.C. and A.A. Fodor, *Compositional data analysis of the microbiome: fundamentals, tools, and challenges.* Ann Epidemiol, 2016. **26**(5): p. 330-5.

253. Gloor, G.B., et al., *It's all relative: analyzing microbiome data as compositions.* Ann Epidemiol, 2016. **26**(5): p. 322-9.

254. *QIIME 2*. 2020; Available from: https://docs.qiime2.org/2020.8/tutorials/.

255. Gotelli NJ, C.R., *Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness.* Ecol Lett, 2001. **4: 379-391**.

256. CE., S., *A mathematical theory of communication.* 2001. **GetMobile 5:3–55**.

257. Faith, D.P. and A.M. Baker, *Phylogenetic Diversity (PD) and Biodiversity Conservation: Some Bioinformatics Challenges.* Evolutionary Bioinformatics, 2006. **2**: p. 117693430600200007.

258. Lozupone, C. and R. Knight, *UniFrac: a new phylogenetic method for comparing microbial communities.* Appl Environ Microbiol, 2005. **71**(12): p. 8228-35.

259. Lozupone, C.A., et al., *Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities.* Appl Environ Microbiol, 2007. **73**(5): p. 1576-85.

260. Anderson, M.J. and D.C.I. Walsh, *PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing?* Ecological Monographs, 2013. **83**(4): p. 557-574.

261. Washburne, A.D., et al., *Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets.* PeerJ, 2017. **5**: p. e2969.

262. Kaul, A., et al., *Analysis of Microbiome Data in the Presence of Excess Zeros.* Front Microbiol, 2017. **8**: p. 2114.

263. Friedman, J. and E.J. Alm, *Inferring correlation networks from genomic survey data.* PLoS Comput Biol, 2012. **8**(9): p. e1002687.

264. Mark Welch, J.L., F.E. Dewhirst, and G.G. Borisy, *Biogeography of the Oral Microbiome: The Site-Specialist Hypothesis.* Annu Rev Microbiol, 2019. **73**: p. 335-358.

265. Zhang, Y., et al., *A multivariate distance-based analytic framework for microbial interdependence association test in longitudinal study*. Genet Epidemiol, 2017. **41**(8): p. 769-778.

266. Joseph N. Paulson, H.T., Hector Corrada Bravo, *Longitudinal differential abundance analysis of microbial marker-gene surveys using smoothing splines.* bioRxiv, 2017.

267. Mandal, S., et al., *Analysis of composition of microbiomes: a novel method for studying microbial composition.* Microb Ecol Health Dis, 2015. **26**: p. 27663.

268. *European Nucleotide Archive* 2020; Available from: https://www.ebi.ac.uk/ena/browser/view/PRJEB37445.