From DEPARTMENT OF MEDICAL EPIDEMIOLOGY AND BIOSTATISTICS
Karolinska Institutet, Stockholm, Sweden

ETIOLOGY AND EARLY DETECTION OF NASOPHARYNGEAL CARCINOMA – AN EPIDEMIOLOGICAL APPROACH

Zhiwei Liu

Karolinska Institutet

Stockholm 2016

Cover figure: Designed and drawn by Qiushi Cheng.
Published by Karolinska Institutet.
Printed by E-print AB 2016

ISBN 978-91-7676-371-1

Institutionen för Medicinsk Epidemiologi och Biostatistik

Etiology and early detection of nasopharyngeal carcinoma – an epidemiological approach

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i hörsal Atrium, Nobels väg 12B.

Onsdag den 31 augusti 2016, kl 09.00

av

Zhiwei Liu

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

Bihandledare:
Professor Hans-Olov Adami
Harvard School of Public Health
Department of Epidemiology

Professor Ingemar Ernberg
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Professor Su-Mei Cao
Sun Yat-sen University Cancer Center, Guangzhou, China
Cancer Prevention Center
Department of Epidemiology

Fakultetsopponent:
Professor Tai-hing Lam
The University of Hong Kong
Division of Community Medicine and Public Health Practice

Betygsnämnd:
Docent Marie Eriksson
Umeå University
Umeå School of Business and Economics

Docent Nele Brusselaers
Karolinska Institutet
Department of Molecular Medicine and Surgery

Docent Elena Kashuba
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Stockholm 2016

To my beloved family

ABSTRACT

Nasopharyngeal carcinoma (NPC) is a rare malignancy worldwide, but it is endemic in a few areas including southern China, Southeast Asia, North Africa, and the Arctic. The underlying mechanisms behind this remarkable geographic distribution remain unclear. Although Epstein-Barr virus (EBV) infection has been suggested as a necessary cause of undifferentiated NPC, EBV itself is not sufficient to cause this malignancy. Other co-factors, such as environmental risk factors, and/or genetic susceptibility, may interact with EBV to play a role in the carcinogenesis of NPC.

Survival rates differ significantly between NPC patients in early stages and late stages. Due to the close associations between EBV infection and NPC risk, EBV-related biomarkers have been used for early detection and screening for NPC in a few high-incidence areas. However, the cost-effectiveness of this approach has not been demonstrated. Pilot efforts have highlighted that identification of additional biomarkers is needed to improve the early detection rate and reduce the mortality rate of NPC in high-incidence populations. The aim of this thesis is to study the association of early-life social environment, oral hygiene, and family history with the risk of NPC, and to utilize an affinity proteomic approach to identify biomarkers that can potentially facilitate the early detection.

In Study I, we investigated associations between childhood family structure, in terms of sibship size and number of older/younger siblings, and the risk of NPC or infectious mononucleosis (IM, another EBV-associated disease) in Sweden, an NPC low-incidence area. For each outcome, a nested case-control study was conducted within the Swedish national health and population registers, including 251 NPC cases, 11,314 IM cases, and five population controls per case matched by sex and year of birth. Clearly contrasting findings were observed between NPC and IM risk. We detected a monotonically increased risk of NPC with a larger sibship size ($P_{trend} = 0.006$), especially with an increasing number of older siblings. For example, the odds ratio (OR) of NPC for those with three or more siblings compared with no siblings was 2.03 (95% confidence interval [CI]: 1.23, 3.35). In contrast, we detected a lower risk of IM among subjects with a larger sibship size, and more older/younger siblings. We concluded that early-life social environment could contribute to NPC pathogenesis in non-endemic areas. Earlier infection with EBV might be associated with an elevated risk of NPC, which is further supported by the clearly contrasting findings between NPC and IM.

In Study II, we estimated the associations between oral hygiene and NPC risk in a high-incidence area. We conducted a population-based case-control study in southern China between 2010 and 2014, with a total of 2528 newly diagnosed NPC cases aged 20-74, and 2596 randomly selected controls. Controls were frequency matched to the age and sex distribution of the cases by geographic region. Based on questionnaire information, we found an increased risk of NPC with a higher number of filled teeth. Compared with those having no filled teeth, subjects having 1 to 3 and more than 3 filled teeth had adjusted ORs of 1.25 (95% CI: 1.06, 1.49) and 1.55 (95% CI: 1.13, 2.12), respectively ($P_{trend} = 0.002$). In contrast, more frequent tooth brushing was inversely associated with the risk of NPC. We concluded that poor oral health could be associated with an elevated risk of NPC. Prospective cohort studies with a comprehensive measurement on oral health condition are needed to confirm our findings and explore the underlying mechanisms.

Absolute NPC risks in the general population with and without a family history of NPC in NPC-endemic geographic regions, where the great majority of NPC cases occur worldwide, are largely unknown. In Study III, we utilized family data from the aforementioned population-based case-control study in southern China and found that subjects with a first-

degree family history of NPC were at a greater than 4-fold higher risk for NPC, compared to those without such a history, whereas a family history of other malignancies did not confer an increased risk of NPC. The excess risk was higher for a maternal than a paternal history and slightly stronger for a sibling than a parental history, and for a sororal than a fraternal history. Among first-degree relatives of cases, the cumulative risk of NPC up to age 74 years was 3.7%, whereas that among relatives of controls was 0.9%. Cumulative risk was higher in siblings than in parents among relatives of cases, whereas no such difference was noted among relatives of controls.

The affinity proteomic approach is valuable for biomarker validation and identification. In Study IV, we utilized plasma samples from 174 NPC cases and 175 community-based controls from Taiwan to identify biomarkers that could potentially facilitate early detection of NPC. We established a panel of eight such biomarkers, including proteins encoded by the genes *CCNB1, KDR, PDGFB, LGALS1, HAS1, LY6K, IL2RA*, and *CXCL10*. The combination of these eight markers showed promising value to distinguish early-stage NPC patients from controls (areas under the receiver operating characteristic curves = 0.808, 95% CI: 0.745, 0.871). Additional studies with prospectively collected biospecimens are warranted to examine the use of these plasma biomarkers in the early detection of NPC.

LIST OF SCIENTIFIC PAPERS

I. Liu Z, Fang F, Chang ET, Adami HO, Ye W. **Sibship size, birth order and risk of nasopharyngeal carcinoma and infectious mononucleosis: a nationwide study in Sweden.** *Int J Epidemiol*. 2015 Apr 28. [Epub ahead of print]

II. Liu Z, Chang ET, Liu Q, Cai Y, Zhang Z, Chen G, Xie, SH, Cao SM, Shao JY, Jia WH, Liao J, Chen Y, Ernberg I, Vaughan TL, Adami HO, Huang G, Zeng Y, Zeng YX, Ye W. **Oral hygiene and risk of nasopharyngeal carcinoma - a population-based case-control study in China.** *Cancer Epidemiol Biomarkers Prev*. 2016 May 19. [Epub ahead of print]

III. Liu Z, Chang ET, Liu Q, Cai Y, Zhang Z, Chen G, Huang QH, Xie, SH, Cao SM, Shao JY, Jia WH, Liao J, Chen Y, Lin L, Liang L, Ernberg I, Vaughan TL, Adami HO, Huang G, Zeng Y, Zeng YX, Ye W. **Quantification of familial risk for nasopharyngeal carcinoma in a high-incidence area.** *(Manusript)*

IV. Liu Z, Byström S, Ploner A, Chen CJ, Hildesheim A, Jochen S, Ye W. **Search for biomarkers for early detection of nasopharyngeal carcinoma by antibody suspension bead array assays.** (*Manuscript)*

CONTENTS

LIST OF ABBREVIATIONS

AUC	Area under the curve
BL	Burkitt's lymphoma
CCNB1	Cyclin B1
CI	Confidence interval
CXCL10	Chemokine (C-X-C motif) ligand 10
EA	Early antigen
EBNA1	Epstein-Barr nuclear antigen 1
EBV	Epstein-Barr virus
ELISA	Enzyme-linked immunosorbent assay
HAS1	Hyaluronan synthase 1
HL	Hodgkin lymphoma
HLA	Human leukocyte antigen
HR	Hazard ratio
ICC	Intraclass correlation coefficient
ICD	International Classification of Disease
IL2RA	Interleukin 2 receptor subunit alpha
IM	Infectious mononucleosis
KDR	Kinase insert domain receptor
LGALS1	Lectin galactoside-binding soluble 1
LMP1	Latent membrane protein 1
LY6K	Lymphocyte antigen 6 complex locus K
MA	Multi-normalization
MFI	Median fluorescence intensity
NPC	Nasopharyngeal carcinoma
NPCGEE	NPC Genes, Environment, and EBV
NPV	Negative predictive value
OR	Odds ratio
PDGFB	Platelet-derived growth factor beta polypeptide
PPV	Positive predictive value
PQN	Probabilistic quotient normalization
ROC	Receiver operating characteristic
VCA	Capsid antigen
WHO	World Health Organization

1 INTRODUCTION

Nasopharyngeal carcinoma (NPC) is one of the Epstein-Barr virus (EBV)-associated malignancies, characterized by a remarkable geographical distribution. It is a major cause of morbidity and mortality in southern China, where it is one of the most common cancers. Despite the heavy public-health burden of NPC in southern China and other endemic areas, relatively little is known about the etiology and prevention of NPC. Although certain environmental exposures, including high consumption of salt-preserved fish and other preserved foods, low consumption of fresh fruits and vegetables, and tobacco smoking, are generally well accepted as NPC risk factors, to date there has been no rigorous population-based case-control study of NPC in southern China. Evidence accumulated so far indicates a probable causal role of EBV in the pathogenesis of undifferentiated NPC (the most common histological subtype of NPC). However, despite establishing lifelong latency in the majority of humans, only a small proportion of individuals infected with EBV develop cancer. This indicates that EBV alone is not a sufficient cause for this malignancy. Environmental exposures and/or genetic risk factors likely also play a role in the pathogenesis of this tumor.

Despite the unknown etiology, using antibodies against EBV for early diagnosis and screening for NPC has been conducted in a few high-incidence areas in southern China since the 1970s. Traditional markers include IgA antibodies against EBV capsid antigen (VCA/IgA) and early antigen (EA/IgA) measured by immunofluorescence assays. Recent studies demonstrated that a combination of IgA antibodies against the Epstein-Barr nuclear antigen1 (EBNA1/IgA) and VCA/IgA measured by enzyme-linked immunosorbent assay (ELISA) has higher sensitivity, specificity, and positive predictive value compared with the traditional method. Individuals identified as being at high risk of NPC based on EBV serological markers can be offered fiberoptic endoscopy/biopsy and close medical surveillance to enable early diagnosis of NPC and, ideally, reduced mortality. However, the cost-effectiveness of this labour-intensive strategy has yet not been proved, and new biomarkers are needed more specifically identify the high-risk population, in order to provide screening for NPC in the general population.

2 BACKGROUND

2.1 DESCRIPTIVE EPIDEMIOLOGY

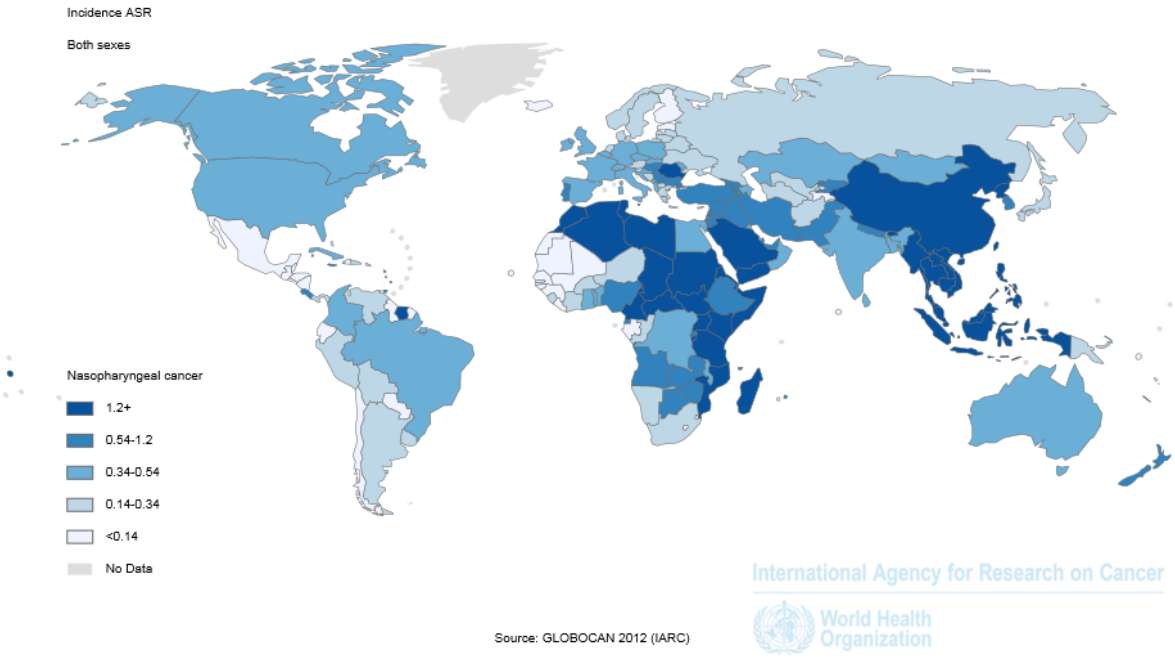

Figure 1 Global nasopharyngeal carcinoma incidence: estimated age-standardized incidence rate per 100,000, both sexes, all ages. Source: GLOBOCAN 2012 (IARC)

Worldwide, NPC is a rare malignancy, with an incidence rate of generally <1 per 100,000 person-years, but areas of the high incidence are in regions of southern China, Southeast Asia, North Africa, and the Arctic [1] (**Figure 1**). In 2012, ~ 86,000 incident cases of NPC were diagnosed worldwide and the estimated number of deaths exceeded 50,000, making it the 24th most common new cancer in the world; in contrast, NPC was the 7th most common new malignancy in Southeast Asia [1]. Other areas with incidence rates ranging from 5 to 15 per 100,000 person-years, exist in northern China, the Middle East and Mediterranean, and North Africa, and among the Inuit of Alaska and Greenland [1]. In some parts of southern China, where NPC is endemic, NPC was the 3rd most common new malignancy in males in cities of Guangzhou, Zhongshan, and Sihui [2]. The overall age-standardized incidence rates in these areas are >20 per 100,000 person-years among males, which are more than100-fold higher than the rest of the world.

The World Health Organization (WHO) classification divides NPC into four histopathological types: keratinizing squamous cell carcinoma (type I), differentiated nonkeratinizing carcinoma (type II), undifferentiated nonkeratinizing carcinoma (type III), and basaloid squamous cell carcinoma [3, 4]. In NPC endemic areas, over 90% of NPC cases in endemic areas are type III and most of the remaining 10% is type II [5]; On the other hand, ~ 50% of cases in non-endemic areas are type I [6, 7]. Type I appears to have different risk factors and pathogenic processes compared to types II/III NPC [8]. For example, gene products of Epstein-Barr virus (EBV) are detected in the tumor cells of all type III NPC [9, 10], whereas the role of EBV infection on the etiology of type I remains uncertain. Cigarette smoking and alcohol drinking account for >50% of type I NPC in non-endemic areas, whereas they are much more weakly associated with types II and III NPC [7, 11]. Basaloid squamous cell carcinoma, a new histopathological subtype introduced in 2005, is very rare in high-incidence areas and morphologically is the same as the same tumor elsewhere in the head and neck region [4].

Despite the decreasing incidences of NPC in some areas of Southeast Asia, such as Taiwan [12], Hong Kong [13], Singapore[13], and Sweden (**Figure 2A**), the secular trend was stable in high-risk areas of southern China over the last 20 years (**Figure 2B)** [14, 15] . Of note, the decreased incidence was found only for type I NPC whereas the incidence of type II/III NPC remains stable in Hong Kong [16]. The overall decline in incidence of type I NPC may be attributed to the consequences of economic development in Hong Kong (e.g. decline in smoking), whereas the relatively steady in incidence of type II/III NPC may be attributed to the interplay of genetics, virus infection, and/or stable environmental exposures.

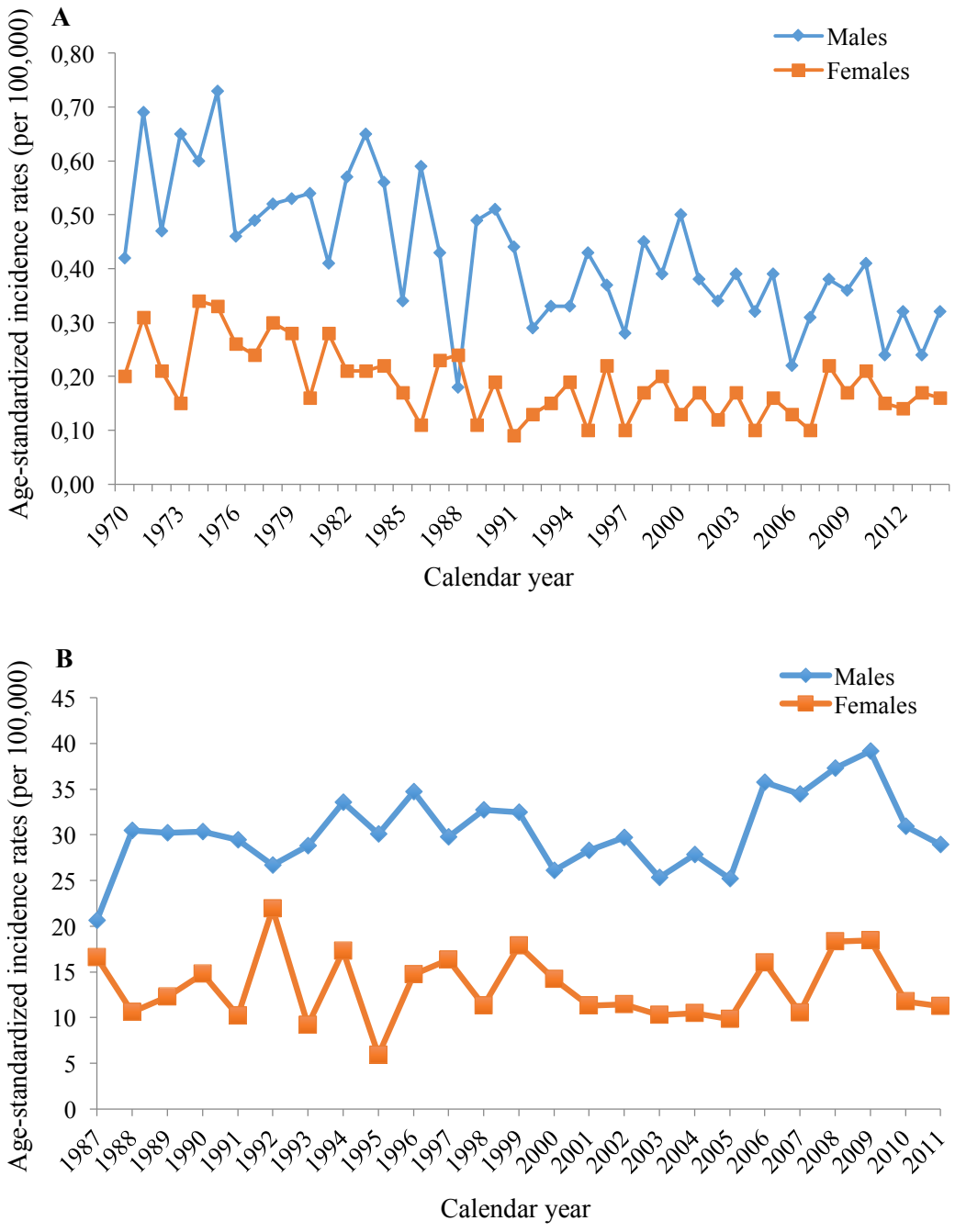

Figure 2 Age-standardized incidence rates (world) of nasopharyngeal carcinoma among males and females.
A. Sweden (1970-2015); data was downloaded from Socialstyrelsen: http://www.socialstyrelsen.se/statistics/statisticaldatabase/cancer
B. Sihui, Guangdong Province, China (1987-2011); reproduced from [15]

In both NPC-endemic and non-endemic areas, age-standardized incidence rates are consistently 2- to 3-fold higher in males than in females [17]. A bimodal pattern of age distributions in low-incidence populations has been observed, with an initial peak in early adulthood, followed by a second peak later in life [18]. In contrast, the peak of incidence is around ages 45 to 59 years in high-incidence populations [14], apparently highlighting the importance of early exposure to carcinogens, such as childhood environmental factors and genetic predisposition, is essential in the etiology of type II/III NPC [14, 17].

2.2 CLINICAL FEATURES

The symptoms and signs at presentation of NPC include neck masses, epistaxis, nasal obstruction and discharge, headache, and other nonspecific indicators. Furthermore, because the cancer is located in a silent anatomic site (**Figure 3**), and NPC exhibits a higher metastatic rate compared to other head and neck cancers [19], NPC tends to present at advanced stages (clinical stages III and IV) when diagnosed. It has been shown that >70% of patients were at advanced stage when diagnosed in clinics [20]. A 10-year survival rate for NPC patients can reach 98% for stage I and 60% for stage II [21]. In contrast, median survival is 3 years for patients at advanced stages [22], highlighting that to improvements in diagnosis rate could help to reduce NPC mortality.

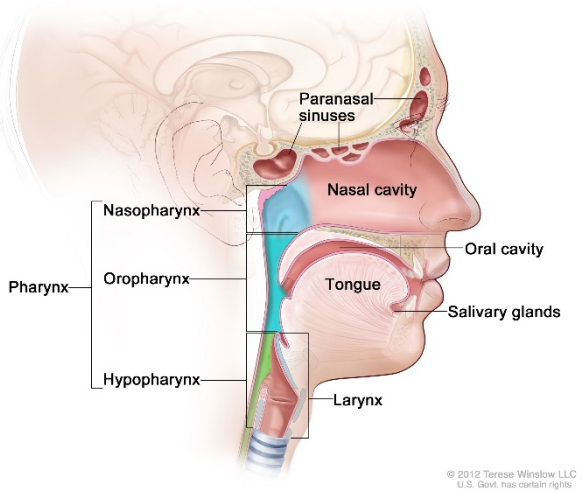

Figure 3 Anatomy of the head and neck
(Source: http://www.cancer.gov/types/head-and-neck)

Especially in high-incidence area, patients with symptoms should be clinically assessed for physical signs of the disease. The examination of the nasopharynx is firstly made by an indirect nasopharyngoscope, followed by a direct nasopharyngoscope (fiberoptic endoscope). A biopsy should be performed if a suspicious growth of in the nasopharynx (e.g. an elevated, nodular, rough, or ulcerative nasopharyngeal surface) is detected. Cross-sectional imaging by computed tomography (CT) scan or magnetic resonance imaging (MRI) should be undertaken even if a suspected tumor is not visualized with endoscopic examination.

2.3 RISK FACTORS

Since the malignancy was firstly reported in 1901 [23], the etiology of NPC has remained a puzzle for more than a century. Migrant studies show that when southern Chinese settle in other countries, their incidence of NPC is 10 to 30 times higher than that of other races – a rare pattern among malignancies suggesting a strong genetic component of NPC risk [24, 25].

A higher incidence of NPC is also observed among North African immigrants in Israel and Sweden, when compared to the native Israelis and native Swedes [24, 26]. Incidence of NPC among Chinese born in western countries is still higher than that among Caucasians, although it is about half that of those living in China or migrating within Southeast Asia [27-29]. In addition, compared with those born in southern France, men of French origin born in North Africa also had a higher incidence of NPC [30]. The latter findings indicate that in addition to genetics, environmental factors also play an important role in NPC.

To date, established risk factors for type III NPC include Cantonese ethnicity [31], male sex [31, 32], EBV infection [10], a family history of NPC [33-36], high consumption of salt-preserved fish [37], low intake of fresh vegetables and fruits intake [38, 39], smoking [40], and some human leukocyte antigen (HLA) class I alleles [41-43]. On the other hand, other HLA genotypes [44] and a history of infectious mononucleosis (IM) [7, 45, 46] may be associated with a decreased risk. Further potential risk factors include high consumption of other preserved foods [38, 47, 48], a history of chronic respiratory tract conditions [49, 50], and genetic polymorphisms in cytochrome P450 2E1 (*CYP2E1*), *CYP2A6*, glutathione S-transferase M1 (*GSTM1*), and *GSTT1* [51]. Less established risk factors include consumption of herbal medicine, occupational exposures to dust and formaldehyde, and nickel exposure [17, 52]. This thesis focuses on two unexplored risk factors (i.e. early family structure and oral hygiene), and quantification of familial risk of NPC in a high-incidence area of southern China.

2.3.1 Epstein-Barr Virus (EBV)

EBV, a gamma-herpesvirus that infects lymphocytes and epithelial cells, establishes lifelong latency in more than 90% of adults globally [53-55]. Primary infection with EBV usually occurs early in life and transmission is mainly through saliva. In high NPC incidence areas, such as Hong Kong, Taiwan, and mainland China, ~60% of children have been infected by age 2, ~80% by age 6, and almost 100% by age 10 [56-58]. In contrast, age at primary infection is relatively late in children from western countries, such as US, Denmark, and Sweden [55, 59, 60]. Although primary infection with EBV is usually asymptomatic, it is associated with certain diseases, including IM and Burkitt's lymphoma (BL), ~30% of HL, certain subtypes of non-Hodgkin lymphoma, type III NPC, and a subset (~10%) of gastric carcinoma [61]. It is estimated that ~143,000 deaths worldwide in 2010 could be attributed to EBV-associated malignancies [62].

The link between EBV and NPC was first proposed in 1966 when NPC patients were reported to have higher titers of antibodies against an antigen later demonstrated as a product of EBV [63]. Since then, extensive evidence suggests that EBV is a potential cause of NPC, especially type III [10]. First, monoclonal EBV genome and viral gene products are detected in virtually all tumors in NPC-endemic areas [63-67], indicating that the tumors result from clonal proliferations of a single cell that is initially infected with EBV. Second, elevated IgA antibodies against EBV antigens are highly specific markers for subsequent NPC in high-incidence areas [68, 69], while elevated EBV-neutralizing antibodies blocking B-cell infection and anti-gp350 antibodies are inversely associated with NPC risk [70]. Third, the expression of viral proteins, such as latent membrane protein 1 (LMP1), LMP2, Epstein-Barr nuclear antigen1 (EBNA1) and EBNA2, has been demonstrated to drive tumor progression in invasive epithelial cancers [10]. Nevertheless, EBV has never been detected in non-cancerous epithelial cells of the nasopharynx [71], and epithelial infection is much less efficient *in vitro* than B-lymphocyte infection [72]. The viral target, complement receptor type 2 (CR2), which is presented on B cells and attaches to EBV envelop, gp 350/220, is expressed at low levels on epithelial cells (**Figure 4**) [73, 74]. Therefore, other mechanisms of viral entry into epithelial cells have been postulated [75], including attachment to two additional

glycoproteins, gHgL and gB (**Figure 5**) [76], cell-to-cell contact [77, 78], or IgA/secretory component (SC) protein mediation [79].

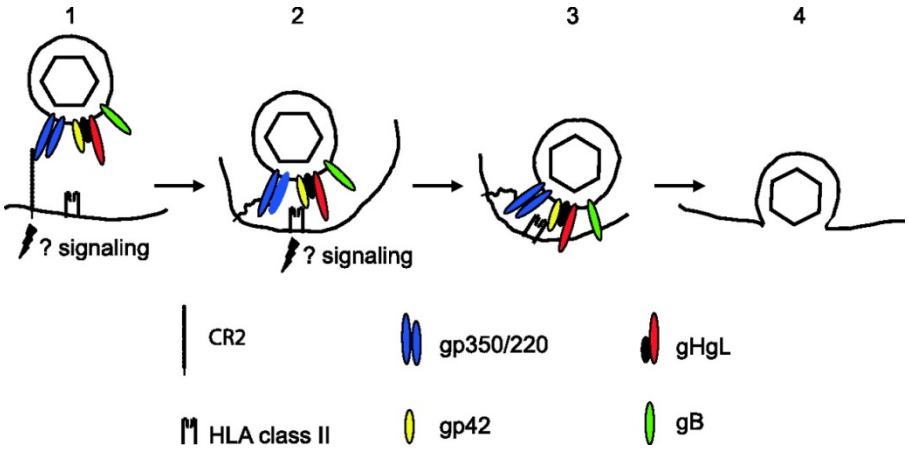

Figure 4 A putative model of the steps involved in entry of EBV into a B lymphocyte. (Adapted from Hutt-Fletcher LM [74])

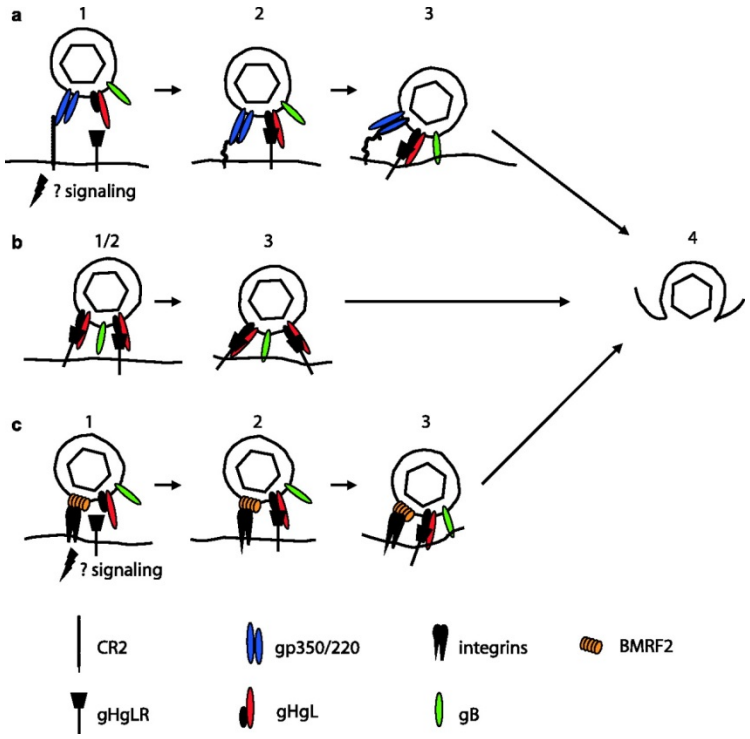

Figure 5 Putative models of the steps involved in entry of EBV into epithelial cells after attachment via different cell proteins. (Adapted from Hutt-Fletcher LM [74])

2.3.2 Early-Life Risk Factors

Factors that could potentially alter the oncogenicity of EBV include the age and immune response at the time of primary EBV infection [46, 80]. For example, when primary infection is delayed until adolescence, the EBV-related immune response is robust and can lead to symptoms of IM, which is linked to risk of EBV-related HL in adulthood [81-83]. In NPC-endemic areas, IM and HL are not prevalent, leading us to hypothesize that timing at infection may play a role in the development of NPC. Accumulated evidence shows that childhood exposure to certain environmental factors confer a higher risk of NPC [84, 85]. Nevertheless, the associations of factors influencing the timing of common childhood infections with the risk of NPC have not been studied. There are a few studies showing that a

history of IM is associated with a lower risk of NPC [7, 45, 46], although commonly based on small number of cases. In high-risk populations, perhaps due to the fact that late infection is rare, it is difficult to estimate the association between a history of IM and NPC risk [46]. In NPC non-endemic areas, however, the rarity of NPC makes the evaluation of this hypothesis a big challenge.

The household environment during childhood, when primary EBV infection is most probable, including number of siblings and population density of the household, could be important predictors for the immunological control of EBV and eventual EBV-related disease risk [86]. Childhood family structure may serve as an indirect indicator of early infection with common childhood pathogens. For example, birth order has been linked to risk of other EBV-related malignancies, such as HL [87, 88], and also to hepatocellular carcinoma [89, 90]. Hence, we hypothesized that very early exposure to EBV and other carcinogens may play a role in NPC pathogenesis. Studies on early childhood family structure may lend insights into whether timing of primary infection with EBV is associated with a subsequent risk of NPC.

2.3.3 Oral Hygiene

Poor oral health, as a modifiable risk factor that is common among the elderly [91-94], has been linked to cancers of the pancreas, esophagus, stomach, and head and neck [95-104]. In the case of NPC, periodontitis might increase inflammation [105, 106] and thus might increase the risk for NPC, given that inflammatory response could be one pathway of carcinogenesis promotion [61, 107, 108]. In addition, bacterial load increases with a greater number of teeth lost [109], and some of the bacteria have been implicated in the production of nitrosamines, which are known carcinogens for NPC development [110, 111]. Poor oral health could also increase the risk of NPC by stimulating EBV replication, as indicated by higher viral load among individuals with periodontal disease than those without [112-116]. In rats, *n*-butyrate produced by anaerobic bacteria found in the nasopharynx, combined with phorbol ester (TPA)-like plant and soil microelements, can enhance EBV-mediated B-cell transformation and promote NPC development [117]. The proposed underlying mechanistic pathways are demonstrated in **Figure 6**.

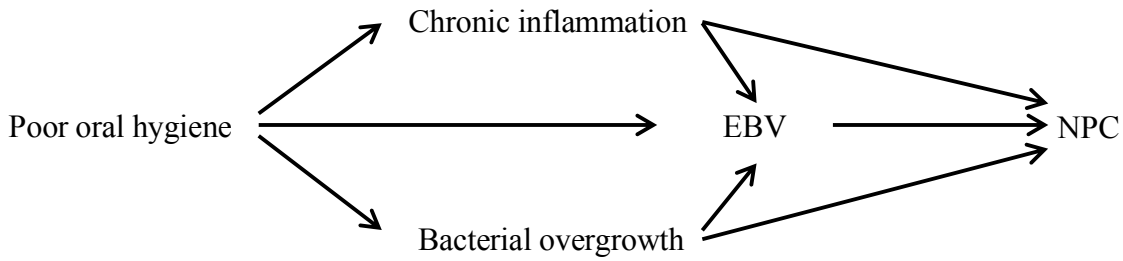

Figure 6 Hypothetical underlying mechanistic pathways linking poor oral hygiene with NPC risk

Few epidemiological studies have tried to address this research question of whether poor oral hygiene is related to NPC risk. One hospital-based case-control study in Turkey showed that infrequent tooth brushing and an increasing number of decayed teeth were associated with a higher NPC risk [118]. Compared with those who brushed teeth daily, those who teeth brushed rarely had an odds ratio (OR) of 6.17 (95% CI: 3.60, 10.55). However, when examining poor oral health as a risk factor for cancer in general, any positive associations could be due to residual confounding by smoking, low socioeconomic status, diet, and/or medical history. Detailed risk factor information in a large, population-based case-control study could help to facilitate the rigorous evaluation of oral health as a risk factor for NPC.

2.3.4 Familial Aggregation

Familial clustering has been consistently reported in NPC high-incidence [119-122], intermediate-incidence [34, 123, 124], and low-incidence [125, 126] areas. In southern China, where NPC is endemic, more than 5% of incident cases reported a positive family history of NPC among the first-degree relatives [122, 127, 128]. Previous case-control studies in different populations showed that ORs ranged from 2 to 20 in individuals who reported a first-degree family history of NPC compared with those with no such history [48, 122, 123, 127, 129-135]. This magnitude of association is among the highest of any malignancy [136], suggesting that environmental factors themselves cannot fully explain the observed association. A cohort study in Taiwan showed that the risk of NPC among males in a multiplex family cohort was 6.8-fold higher than that in a community cohort [35]. Genes and environmental exposures likely play a combined role in the etiology of NPC [17]. An inheritance pattern that cannot be explained by activation of a single major susceptibility gene is supported by results from a complex segregation analysis of familial NPC showing that the etiology of NPC involves interaction of multiple genetic and environmental factors [137].

Whether familial NPC cases differ substantially from sporadic cases in terms of clinical features (i.e. histology, stage, and prognosis), ethnicity, sex, age at diagnosis, environment risk factors, EBV serology, and/or genetic risk factors is still controversial. A few studies showed that familial cases did not have characteristics notably distinct from sporadic cases [119, 129, 130, 138]. On the other hand, others found that familial NPC cases tend to be younger [139-141], and have better survival than sporadic cases [141, 142]. Three previous studies reported significant modification of the association with family history of NPC by smoking [133], wood fuel use [143], and salt-preserved fish consumption [122], whereas a prospective study did not find an interaction between smoking and family history of NPC [35]. The small number of controls with a positive first-degree family history of NPC and the low power of the statistical test of heterogeneity make it difficult to draw firm conclusions about the joint effects of family history and environmental risk factors. Pooled studies with larger numbers of subjects will enable more powerful tests of such interactions.

To date, previous epidemiologic studies of NPC have been limited in number, size, scope, and rigorousness of study design. Few studies have investigated the relative-specific risk among families with affected members. Because most studies have not ascertained all first-degree relatives and are not population-based, absolute NPC risks in the general population with and without a family history in NPC-endemic geographic regions, where the great majority of NPC cases occur worldwide [2], are largely unknown. The lack of evidences precludes cost-effectiveness modeling of screening for NPC among families with a positive history of NPC.

2.3.5 Genetic Susceptibility

As reviewed by Hildesheim and Wang [51], since 2000, a total of 83 published papers have investigated genetic risk factors for NPC. Only one genome-wide association study (GWAS) utilized samples from more than 1000 cases and controls [144]; other GWAS were limited to a few hundred cases and controls. A number of genetic polymorphisms associated with NPC risk have been identified, including classical *HLA* class I/II genes, non-classical *HLA* genes, phase I/II metabolic activation/detoxification and DNA repair genes, and other genes such as those involved in cell cycle control, cell adhesion/migration, angiogenesis, and DNA methylation.

Many investigators have focused on the possible pathogenetic role of HLA molecules, which are required for the presentation of foreign antigens, including viral peptides, to the immune system for targeted lysis. In Chinese and other high-risk Asian populations, *HLA-A2-B46*, and *B17* are associated with a 2- to 3-fold increase in NPC risk, whereas an increased risk is associated with *HLA-B5* in Caucasians. One-third to one-half lower risk is found in association with *HLA-A11* across all races, B13 in Chinese and Tunisians, and A2 in non-Chinese. Several other HLA associations have been reported, but must be interpreted with caution due to multiple-testing considerations [44]. Genetic polymorphisms other than *HLA* are also reported. However, most genetic association studies are based on small sample sizes, and the lack of replication precludes a full understanding of genetic influences on NPC development.

2.4 SCREENING

2.4.1 Screening in High-incidence Areas

The notion that testing for antibodies against EBV could a useful screening tool to facilitate the early detection of NPC is supported by several lines of evidence. First, EBV infection is an early event during the tumor progression [67], and the EBV genome and gene products can be detected in virtually all tumors of type III NPC [63-67]. Second, VCA/IgA [35, 68, 69], neutralizing antibodies against EBV DNAse [35, 68], and EA/IgA [145-149], can be detected in serum even years prior to clinical evidence of the cancer, making them the basis for successful NPC screening tests in high-incidence areas. A few pilot efforts have been made to conduct NPC mass screening in high-incidence counties in southern China since the 1970s [145], using the two biomarkers of VCA/IgA and EA/IgA measured by immunofluorescence assays. More recently, studies in southern China demonstrated that a combination of EBNA1/IgA and VCA/IgA measured by enzyme-linked immunosorbent assay (ELISA) had a higher diagnostic accuracy (i.e. high sensitivity, specificity, and positive predictive value [PPV]) in both the general population [150] and families with at least two affected relatives [148].

Although the value of using antibodies against EBV to facilitate NPC diagnosis is generally accepted, there are a few barriers to the implementation of screening for NPC by testing these antibodies in high-incidence populations. First, only a fraction of the ~2% of individuals with elevated titers of VCA/IgA in high-risk areas develop NPC [68, 69]. As shown by one of our large cohort studies in Sihui, southern China, after 20 years of follow-up, only 42 NPC cases were found among 1318 individuals who were positive for VCA/IgA at baseline [69]. Our results from an initial round of screening showed that anti-EBV antibodies can increase the early diagnosis rate to ~70% but with low PPV (~3%) [128]. Second, serologic evidence of EBV reactivation from latency, as indicated by elevated antibody titers against viral lytic antigens, can also be detected in normal individuals, particularly during periods of psychological or physical stress [151, 152], thereby decreasing their specificity. Third, previous efforts were not carefully controlled (i.e. no randomized controlled trial has yet been conducted) and do not permit accurate quantification of the impact of EBV-based screening on detection rates of early-stage NPC and on NPC mortality. These results are required to support evidence-based decisions regarding the efficacy and cost-effectiveness of such screening strategies.

2.4.2 Search for Novel Biomarkers for Early Detection

Other than biomarkers related to EBV, biomarkers related to the human proteome may also exhibit great potential for early diagnosis of NPC. A handful of studies have used proteomics to investigate potential biomarkers for early diagnosis of NPC [153]. Although a number of

biomarkers have been identified, few have been replicated in other independent studies. Most of them may have biological implications rather than diagnostic capacity. Limited sample size, inappropriate study design, heterogeneity of NPC, and different technologies used across different studies may contribute to the inconsistency of findings.

To date, the technology most commonly used for human protein biomarker discovery is mass spectrometry (MS), which is limited to the analysis of a relatively small number of samples in parallel [154]. Recently, plasma antibody profiling technologies, such as antibody suspension bead array assays [155, 156], have been developed for multiplex screening of a large number of proteins in patient cohorts (**Figure 7**). This advance might bring a hope to identify and validate biomarkers for early diagnosis of NPC.

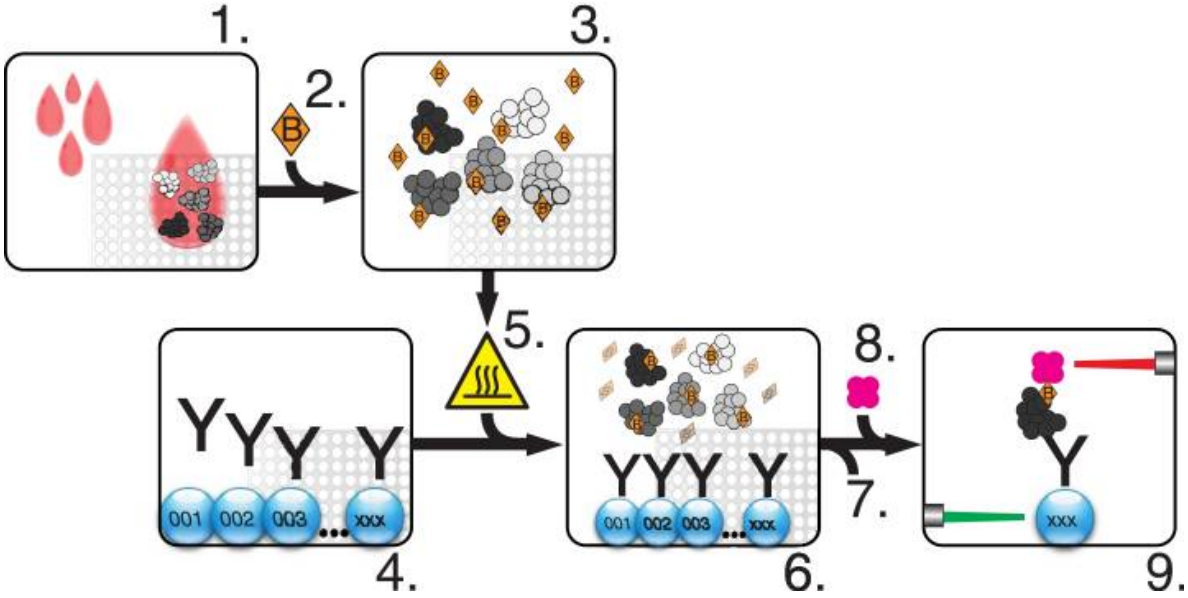

Figure 7 Schematic view of bead array assay (adapted from Schwenk JM, *et al.* [156])

3 AIMS

The overall aim of this thesis was towards identification of risk factors and early diagnostic biomarkers for NPC

The specific study aims were:

- To examine the association between early childhood family structure and the risk of NPC in a low-incidence area.

- To study the association of oral hygiene with risk of NPC in a high-incidence area based on a large, population-based case-control study in southern China.

- To explore familial risks of NPC in detail and to quantify the lifetime risk of familial NPC in a high-incidence population.

- To identify potential biomarkers for early diagnosis of NPC by using affinity antibody suspension bead array assays.

4 MATERIALS AND METHODS

4.1 DATA SOURCES AND COLLECTION

4.1.1 Swedish Multi-Generation Register

The familial data from the Swedish Multi-Generation Register includes all residents born in 1932 or later in Sweden, together with their parents (biological or adoptive) [157, 158]. Such information was available for 60% of individuals who died between 1968 and 1990, and for more than 95% of residents in Sweden alive in 1991 or deceased before 1968. Although information on individuals who died before 1991 were not available, almost 100% of individuals deceased before 1968 and ~60% of individuals deceased between 1968 and 1990 can be identified from Statistics Sweden's register of births and other personal records. All supplementary information is now included in the Register.

4.1.2 Swedish Cancer Register

In 1958, the Swedish Cancer Register was established and is now considered approaching 98% complete [159]. In Sweden, it is mandatory for each health care provider to report newly diagnosed cancer cases from different resources, including clinical examinations, morphological analyses, laboratory tests, and autopsies. Information includes demographics (i.e. personal identification number, sex, age, and location of residence) and medical and follow-up data. Important medical information includes diagnosis (as indicated by the 7th edition of International Classification of Disease [ICD-7] through the whole period), date of diagnosis, histological type (old histology code – WHO/HS/CANC/24.1 – is available through the whole period), and stage (since 2004). The ICD-7 code "146" was used for NPC diagnosis in this Register.

4.1.3 National Patient Register

In the mid-1960s, the Swedish Patient Register was established by the National Board of Health and Welfare. It included only inpatient information from public hospitals in a few counties at the beginning. It has had nearly complete nation-wide coverage of somatic and psychiatric hospital inpatient discharges since 1987, and has also covered nation-wide outpatient care since 2001 with approximately 80% coverage [160]. Diagnoses are reported to the National Patient Register using the ICD-7 between 1961 and 1968, ICD-8 between 1969 and 1986, ICD-9 between 1987 and 1996, and ICD-10 since 1997 [160].

4.1.4 Other Registers

In 1960, 1970, 1980 and 1990, Swedish nationwide censuses were conducted respectively [161]. This population with cross-linkage to the Swedish Multi-Generation Register was our study population for **Study I**. Parental region of residence and occupation were assembled from these Censuses. Emigration and death information for censoring was collected from the Emigration and Immigration Register and Cause of Death Register.

4.1.5 Population Based Case-Control Study in Southern China

In 2010, we launched a population based case-control study in southern China, where NPC is endemic. The collaborative NPC Genes, Environment, and EBV study (NPCGEE) was conducted in the Zhaoqing area of Guangdong Province and the Wuzhou and Guiping/Pingnan areas of Guangxi Autonomous Region. We defined the study base as persons living in 13 cities/counties (Deqing, Fengkai, Gaoyao, Huaiji, Sihui, Zhaoqing, Guangning, Wuzhou, Cenxi, Cangwu, Tengxian, Pingnan, and Guiping) between March

2010 and December 2013 for cases, and between November 2010 and November 2014 for controls. These 13 cities/counties were chosen because of their high incidence of NPC, their geographic contiguity, their existing opportunities for collaboration with local investigators, and their relatively stable population base compared with other, more urban areas in this geographic region. The total population of the 13 study cities/counties is approximately 8 million and the estimated total number of incident NPC cases is approximately 850 per year, based on information from local cancer registries located in Sihui and Cangwu Counties [14]. To ensure adequate statistical power to detect interactive risk factors, we initially aimed to recruit 2,600 cases and 2,600 controls.

Patients with incident cases of histologically confirmed NPC were recruited. We developed a rapid case ascertainment system involving a network of physicians who diagnosed and/or treated NPC at hospitals in the study area (**Figure 8**). Contact persons notified the study personnel as soon as a new NPC case was histopathologically confirmed, after which physician permission was sought to contact each patient, as long as the physician deemed the patient physically and mentally able to participate in the study. This rapid case ascertainment system was backed up by two population-based cancer registries that have collected cancer incidence data in Sihui County since 1977 and Cangwu County since 1982 [14]. However, these registries do not cover the entire population of the study area and are often not immediately notified about new cancer diagnoses, making them insufficient on their own to fulfill the aims of our study. Therefore, ad-hoc case ascertainment networks based on local hospitals are required for population-based case identification in most of parts of the areas.

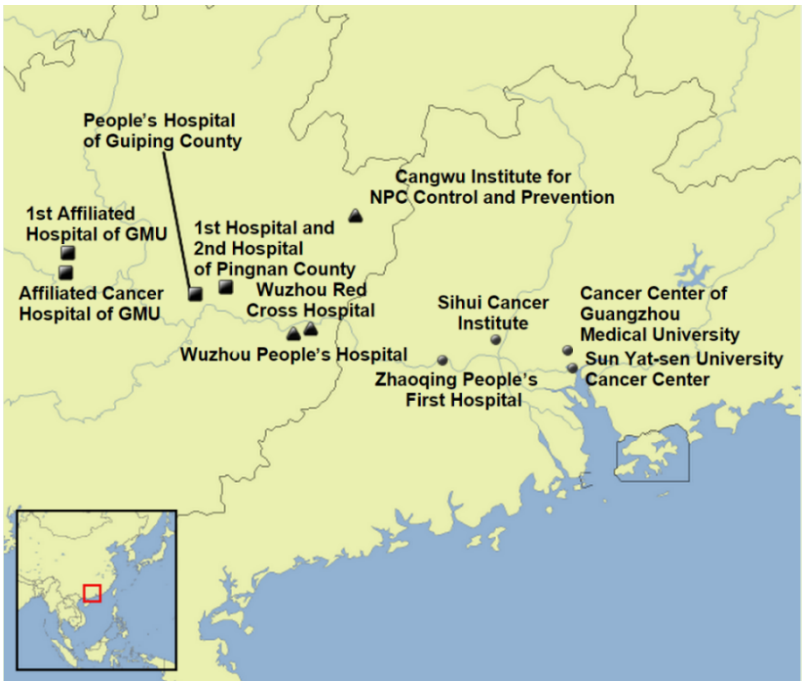

Figure 8 Participating hospitals and cancer research institutions

Controls were randomly selected every 6–12 months from total population registries located in Guangdong Province and Guangxi Autonomous Region, with frequency matching to the expected 5-year age and sex distribution of the cases by residential areas. We first attempted to contact controls by village doctors in rural areas or community committee members in urban areas. Interviews were conducted at the subject's home or a nearby hospital. Potential controls with outdated contact information or a history of working outside of the study area for more than 10 years, as identified with the help of the local government in each town or community, were replaced.

Structured, computerized questionnaire was administered to subjects in person by trained interviewers in the hospital or at home, as appropriate. Interviews were audiotaped for quality control, and logic checks and automatic skip patterns were built into the questionnaire. Preliminary questionnaire data were analyzed periodically to ensure that results were within expectation. The questionnaire covers potential NPC risk factors, including socioeconomic background, race/ethnicity, residential history, occupational history, smoking habits, alcohol consumption, herbal medicine use, medical history, reproductive history, family history of cancer, and past diet (i.e. 10 years ago at interview, adolescent, and childhood). After the interviews, biospecimens including blood, saliva, hair and nails were also collected.

4.1.6 Community Based Case-Control Study in Taiwan

Between July 15, 1991 and December 31, 1994, a community based case-control study was conducted in Taipei City, Taiwan, an NPC intermediate-incidence area. Incident cases of histologically confirmed NPC were recruited from the National Taiwan University and MacKay Memorial Hospitals, two large referral hospitals in Taipei. Eligible patients were less than aged 75 years, had had no prior diagnosis or treatment for NPC, and had resided in Taipei city or county for 6 months or longer. For each case, one control was selected individually matching on sex, 5-year age, and residence area (same district or township). Controls were those without a history of NPC before selection in the study. Ineligible controls were replaced, but no replacement was attempted for eligible control subjects who refused participation. Of 378 eligible cases and 372 controls, 369 cases (98%) and 320 controls (86%) provided questionnaire data and blood samples. A subset of samples was used for testing proteomics.

4.2 STUDY DESIGN

4.2.1 Nested Case-control Study in Sweden (Study I)

To investigate whether early family structure, including maternal age, paternal age, number of siblings, number of older siblings, or number of younger siblings, is associated with the risks for NPC and IM, we have conducted two nested case-control studies, one for each health outcome.

Regarding NPC, we included residents with available biological maternal information, alive and without a recorded diagnosis of NPC on January 1, 1961, or born after 1961 as our study population. We only focused residents born as singletons in Sweden in 1936 or later and up to 1992 [162]. Eligible NPC cases were those who were first diagnosed with NPC (ICD-7 code: 146) between 1961 and 2009. Follow-up started on 1 January 1961 and was censored at the date of first NPC diagnosis, diagnosis of any other cancer, death, emigration out of Sweden or 31 December 2009, whichever occurred first.

We identified 301 NPC cases during follow-up, of whom 251 aged ≥ 20 years at the date of diagnosis. We randomly selected 1255 control subjects (five controls per case) individually matched to NPC cases on year of birth and sex using a method of incidence density sampling [163]. Eligible controls were those who had not yet died, emigrated out of Sweden, or been diagnosed with any cancers at the identification of index case's diagnosis.

Regarding IM, we included individuals who were not diagnosed with IM (ICD-10: B27) before 2001 as our study population, because outpatient data are only available from 2001 onward and most IM patients are not hospitalized. Eligible cases were those with a recorded diagnosis of IM between 2001 and 2009. Follow-up started on January 1, 2001, and was censored at the date of first diagnosis of IM, death, emigration out of Sweden, or December 31, 2009, whichever occurred first.

We included 11,314 IM cases during follow-up. 56,570 control subjects (five controls per case) individually matched to IM case patients on year of birth and sex were randomly selected using incidence density sampling.

4.2.2 Population Based Case-Control Study in Southern China (Studies II, III)

Of 3027 histopathologically confirmed, first incident NPC cases identified from the study base in southern China, we have recruited a total of 2554 patients (84%). We have selected a total of 3202 controls; of these, 2648 (83%) agreed to participate the study. After excluding subjects with missing, ineligible, indeterminate, or poor-quality data, 2528 cases and 2596 controls were analyzed in **Study II**, and 2499 cases and 2576 in **Study III**.

We used data from questionnaire to retrieve self-reported oral health and hygiene information (**Study II**). Questions related to oral health condition included the number of teeth lost after age 20 years, age at first tooth lost after age 20 years, use of fixed dentures, number of teeth filled, daily frequency of brushing teeth, discomfort when eating particular foods, and food avoidance due to tooth or gum problems.

Study III used family structure information from the questionnaire. Information on family included the number, sex, and age of parents, siblings and children, and cancer history in other relatives. For first-degree relatives, we then obtained information on cancer diagnosis regardless of their vital status at the time of interview. When a diagnosis of cancer had been reported in family members, we further obtained information including the type of cancer, age, hospital, diagnostic basis, and whether the diagnosis had been confirmed. If family members were deceased, we obtained information on reasons, years, and their ages at death. For second-degree relatives, we only obtained information on relatives who were diagnosed with cancers.

Because information on first-degree family history of NPC was self-reported, information bias thus cannot be avoided. To check the magnitude of this bias, we used two approaches. First, we compared the NPC incidence rate of control families to that of the general population in the study area in 2011. Second, we validated information from all 41 NPC cases and 5 controls with a self-reported first-degree family history of NPC from Sihui County (where a population-based cancer registry is located), along with 52 randomly selected cases and 48 randomly selected controls (up to 5 subjects within each age stratum) who did not report a first-degree family history of NPC. We linked information on NPC diagnosis of the first-degree relatives (whose names and other identifying information were provided by relatives/local village doctors of cases and controls) from all 146 subjects to the Sihui Cancer Registry and/or medical records provided by village doctors.

4.3 LABORATORY METHOD – AFFINITY ANTIBODY SUSPENSION BEAD ASSAYS (STUDY IV)

Due to the sample availability, for **Study IV**, plasma samples from 174 cases and 175 controls in Taiwan were used for proteomic testing. All samples were prepared according to standard procedures and stored at -80 °C until usage. We used two 384-well assay plates containing six 96-well microtiter plates for testing, and plate layout design is presented in **Figure 9**. The Taiwan samples were randomized across six different plates according to a plate layout that allowed an equal number of samples per plate and a balanced distribution of disease status (i.e. early stages, late stages, and healthy controls) and quality control samples across plates.

Taiwan #1	Taiwan #3
Taiwan #2	

Taiwan #4	Taiwan #6
Taiwan #5	

Figure 9 Plate layout design

4.3.1 Antibodies and Bead Array Generation

A total of 384 antibodies (including 2 positive controls and 2 negative controls) used in the present study were generated within the Human Protein Atlas project [164]. Protein selection was made based on prior knowledge of disease associations according to literature [165-180], and by including inflammatory markers and proteins expressed in possible tumorigenesis pathways [167, 181].

As previously described [182], bead arrays were created by diluting 1.6 µg in 100 µL of 0.05 M MES buffer (pH 4.5) prior to coupling to color-coded magnetic beads (MagPlex-C, Luminex) to create one 384plex bead array in suspension. The coupling of each antibody on the beads was confirmed via R-phycoerythrin-conjugated donkey anti-rabbit IgG antibody (Jackson ImmunoResearch), where all antibodies revealed median fluorescence intensity (MFI) values of 5 000−20 000 AU.

4.3.2 Plasma Profiling

Plasma samples stored at -80 °C were thawed at 4 °C, centrifuged for 10 min at 3000 rpm, and then transferred into 96-well microtiter plates. All samples were then diluted 1:10 in PBS and labeled with biotin, using liquid handling system (Selma, CyBio) as previously described [183]. The biotinylated samples were subsequently diluted 1:50 in an assay buffer composed of 0.5% (w/v) poly (vinyl alcohol) and 0.8% (w/v) polyvinylpyrrolidone (Sigma) in 0.1% casein in PBS (PVXC) supplemented with 0.5 mg/mL nonspecific rabbit IgG (Bethyl), yielding a total sample dilution of 1:500. Diluted samples were heat-treated for 30 min at 56 °C and left to cool at room temperature for 15 min. Forty-five µL of heat-treated samples was then combined and added to 5 µL of the antibody suspension bead array in two 384-well microtiter plates (Greiner BioOne). The locations of the Taiwan samples were equally split across the two 384-well assay plates, resulting in a balanced number of samples and a similar distribution of disease subgroups across the two plates.

The incubation took place overnight on a shaker (Grant) at room temperature and in the dark. Beads were washed with 3 × 60 µL PBS-T (1 × PBS pH 7.4, 0.05% Tween20) using a plate washer (EL406, Biotek) and incubated with 50 µL of 0.4% paraformaldehyde in PBS for 10 min. Beads were then washed with 3 × 60 µL of PBS-T before adding 50 µL of 0.5 µg/mL R-phycoerythrin labeled streptavidin (Invitrogen) in PBS-T and incubated for 20 min. Finally, beads were washed with 3 × 60 µL and measured in 60 µL PBS-T. Samples were measured in a FlexMap3D instrument (Luminex) were a minimum of 50 bead events per antibody were counted. MFI values were obtained for data analysis to allow relative measurement of proteins binding to the beads.

4.4 STATISTICAL METHODS

4.4.1 Conditional Logistic Regression Model (Study I)

In **Study I**, we used conditional logistic regression to estimate the ORs and corresponding 95% CIs for risk of NPC or IM associated with early life exposure, conditioning on year of birth and sex. We included parental region of residence and parental occupation as the confounding factors. We tested for dose-response trends using an ordinal number of total siblings, older siblings, or younger siblings.

In the NPC study, we conducted subgroup analyses by histopathological type. To reduce the selection bias caused by the incompleteness of Multi-Generation Register data for residents who were born before 1932 or who died between 1968 and 1990, we also conducted sensitivity analyses by limiting subjects who born in 1941 or later, or by restricting to cases diagnosed with NPC in 1991 or later and their respective matched controls,

4.4.2 Unconditional Logistic Regression Model (Studies II, III, IV)

The associations between oral health indicators (**Study II**), a family history of NPC (**Study III**), and NPC risk were estimated by ORs and the corresponding 95% CIs derived from multivariate unconditional logistic regression models. Frequency-matching variables of age (in 5-year groups), sex, and residential area (Zhaoqing, Wuhzou or Guiping/Pingnan) were included in minimally-adjusted multivariable model.

We used a two-step approach to determine variables in the fully-adjusted multivariable model. In addition to frequency-matching variables, covariates were firstly considered based on prior knowledge and others were considered if they changed the minimally-adjusted OR of NPC for exposures by more than 10%. We then utilized forward stepwise confounder selection, in which the effect of adding one confounder at a time was evaluated.

In **Study II**, to control for residual confounding effect caused by smoking, we conducted a subgroup analysis which was restrict to non-tobacco users. We tested for linear trends with ordinal variables in the models, using the median value within each category. Likelihood ratio tests for interaction terms were used to compare logistic regression models with and without an interaction term between oral health indicators and each potential effect modifier.

In **Study III**, unconditional logistic regression models were used to estimate the associations between a positive family history of NPC or other cancers, overall and for specific relatives. To evaluate potential influence of reporting bias, using information on sensitivity and specificity of self-reported first-degree family history of NPC among cases and controls obtained from a validation study, we conducted a sensitivity analysis to estimate the OR adjusted for this bias. The population attributable risk for first-degree family history of NPC adjusted for potential confounding factors was also estimated [184].

In **Study IV**, unconditional logistic regression models were used to estimate the ORs (third quartile vs. lowest quartile) and 95% CI to examine the difference of individual biomarker between NPC cases in early stages and healthy controls. All models were adjusted for age and sex and controlled for a false discovery rate (FDR) at 5% to obtain the corrected *P* values.

4.4.3 Cox Proportional Hazards Regression Model (Study III)

In **Study III**, in addition to aforementioned logistic regression model, we used Cox proportional hazards regression models, to estimate hazard ratios (HRs) with 95% CIs in association with a positive family history of NPC among first-degree relatives, adjusting for

age (as the time scale), sex, and geographic area, and stratifying baseline hazards by familial relationship. Different from traditional case-control study design using logistic regression model to estimate the OR as the relative risk, this method can help us to fully account for the family size and structure of each respondent, and the affected relative's age at onset. We firstly attempted to reconstruct a kin cohort, using family structure information reported by the index person. From 42,024 first-degree relatives, we excluded 1243 with an unknown age at death or cancer diagnosis; leaving 40,781 relatives in the cohort analysis. To avoid the influence of familial aggregation, a method described by Lee was used to account for the intracluster dependence [185]. We also estimated the nonproportionality of hazards [186] and no violations were observed.

4.4.4 Cumulative Risk Estimates (Study III)

We used two approaches to calculate the absolute risk (i.e. cumulative risk) for individuals with a positive family history of NPC. First, because of the population-based design and the relative low incidence of NPC, i.e. the OR is a good estimate of relative risk, we were able to quantify cumulative risks of NPC for the first-degree relatives, by multiplying their age-specific OR estimates by the age-specific NPC cumulative risks in the underlying population in 2011. Second, to further account for the family size and structure of each index person, Kaplan-Meier method was used to derive the cumulative risk of NPC. Disease-specific survival curves were plotted. Follow-up duration of all first-degree relatives was considered from birth until his/her age at the date of interview, age at death, age at diagnosis of NPC or cancers other than NPC, whichever occurred first.

4.4.5 Marker Selection and Diagnostic Performance (Study IV)

We used two approaches to conduct the quality control and normalization. First, we detected sample outliers by robust principal component analysis by using the "rrcov" R package [187]. In the present study, no sample was removed based on the graphic diagnostics. Second, we compared median values of intraclass correlation coefficient (ICC) from six schemes using two normalization approaches and found that the data which subjected to probabilistic quotient normalization (PQN, to consider within-plate variation) [188] and unsmoothed multi-normalization (MA, to consider across-plate variation (PQN-MA-Ind) [189].

After data processing, the comparisons were made with a focus on NPC cases in early stages and controls, and the receiver operating characteristic (ROC) curves were constructed using different combinations of sensitivity and specificity. The areas under the curves (AUCs) and with corresponding 95% CIs were calculated for the selected biomarkers. We also calculated sensitivity, specificity, PPV, and negative predictive value (NPV), and their corresponding 95% CIs, by using the optimal values determined from the ROC curves as the cutoffs. Finally, we calculated AUC and 95% CI according to a multivariable logistic regression models including all selected biomarkers.

Data were analyzed using R version 3.1.2 (http://www.R-project.org/), SAS version 9.4 (SAS Institute, Cary, NC), and Stata version 13.1 (Stata Corporation, College Station, TX).

5 RESULTS

5.1 STUDY I

Males were more likely to develop both NPC and IM than females in Sweden. The median age at NPC diagnosis was higher than that at IM diagnosis (49 years vs. 18 years). The associations between early-life family structure and risk of NPC and that of IM had different patterns (**Table 1**). We observed an increased NPC risk with an increasing number of siblings. Compared with those who did not have any siblings, those with three or more siblings had an adjusted OR of 2.03 (95% CI = 1.23, 3.35, P_{trend} = 0.006). Exposure to older siblings was more likely to contribute to an increased risk of NPC, as suggested by a significant dose-response trend with an increasing number of older siblings (P_{trend}= 0.019) but no such trend was observed with an increasing number of younger siblings (P_{trend}= 0.166). For IM, in contrast, we found an inverse association for IM risk associated with an increasing number of siblings (P_{trend} < 0.001), older siblings (P_{trend} < 0.001), and younger siblings (P_{trend} < 0.001).

Table 1. Odds ratios (ORs) and their 95% confidence intervals (CIs) for nasopharyngeal carcinoma and infectious mononucleosis according to familial characteristics in two nested case-control studies in Sweden. *

Variables	Nasopharyngeal carcinoma			Infectious mononucleosis		
	Cases (n=251)	Controls (n=1,255)	OR (95% CI)	Cases (n=11,314)	Controls (n=56,570)	OR (95% CI)
	%	%		%	%	
Sibship size						
0	9.5	16.1	referent	7.6	6.7	referent
1	29.9	32.0	1.59 (0.97, 2.62)	46.4	42.1	0.96 (0.88, 1.04)
2	28.3	25.0	1.94 (1.17, 3.22)	32.0	33.8	0.82 (0.76, 0.89)
≥ 3	32.3	26.9	2.03 (1.23, 3.35)	14.0	17.4	0.72 (0.65, 0.79)
P_{trend}			0.006			< 0.001
No. of older siblings						
0	42.6	48.8	referent	41.1	42.1	referent
1	35.5	30.8	1.54 (1.11, 2.14)	37.7	36.0	0.91 (0.87, 0.96)
2	14.7	13.0	1.65 (1.06, 2.57)	15.9	15.8	0.84 (0.78, 0.89)
≥ 3	7.2	7.3	1.50 (0.83, 2.71)	5.2	6.1	0.73 (0.66, 0.81)
P_{trend}			0.019			< 0.001
No. of younger siblings						
0	36.7	41.9	referent	48.6	43.3	referent
1	31.9	31.9	1.15 (0.81, 1.63)	35.9	37.3	0.83 (0.78, 0.87)
2	15.9	16.3	1.03 (0.67, 1.58)	12.4	14.8	0.73 (0.68, 0.79)
≥ 3	15.5	10.8	1.51 (0.95, 2.40)	3.1	4.6	0.63 (0.56, 0.71)
P_{trend}			0.166			< 0.001

* ORs and 95% CIs were conditioned on matching variables (year of birth and sex), and further adjusted for parents' region of residence, parents' occupation, and other factors listed in the table except for sibship size.

We found null associations between maternal or paternal age and NPC risk, and between paternal age and risk of IM. On the other hand, an increased risk for IM was associated with an increased maternal age. Stratified analyses by histopathological types showed that the ORs for those with one or more siblings increased among squamous cell carcinoma and undifferentiated carcinoma, compared with those having no siblings (**Table 2**), although the associations were slightly stronger for the latter. No such association was observed for other or unclassified NPC. The associations estimated from two sensitivity analyses did not change substantially.

Table 2. Odds ratios (ORs) and their 95% confidence intervals (CIs) for nasopharyngeal carcinoma by histopathological type in a nested case-control study in Sweden. *

Sibship size	Squamous cell carcinoma	Undifferentiated carcinoma	Other or unclassified
	OR (95% CI)	OR (95% CI)	OR (95% CI)
0	referent	referent	referent
1	1.99 (0.96, 4.14)	1.71 (0.51, 5.75)	1.20 (0.51, 2.82)
2	1.93 (0.90, 4.12)	2.74 (0.84, 8.98)	1.79 (0.76, 4.24)
≥ 3	2.15 (1.04, 4.46)	3.24 (0.99, 10.62)	1.33 (0.53, 3.29)
P_{trend}	0.106	0.021	0.374

* ORs and 95% CIs for nasopharyngeal carcinoma were conditioned on matching variables (year of birth and sex), and further adjusted for parents' region of residence, parents' occupation, maternal age, and paternal age.

5.2 STUDY II

Baseline characteristics of the study population are shown in **Table 3**. The distributions of residential areas and sex were similar among cases and controls, whereas cases were slightly younger than controls at enrollment, and were more likely to be less educated, live in cottages, have blue-collar jobs, have a first-degree family history of NPC, have ever smoked, and have consumed salt-preserved fish at least weekly in 2000-2002.

Table 3. Characteristics of nasopharyngeal carcinoma (NPC) cases and controls

Characteristics	Cases (n=2528)		Controls (n=2596)		*P* value *
	No.	%	No.	%	
Residential area					0.28
Zhaoqing	1283	50.8	1321	50.9	
Wuzhou	688	27.2	664	25.6	
Guiping/Pingnan	557	22.0	611	23.5	
Sex					1.00
Male	1858	73.5	1908	73.5	
Female	670	26.5	688	26.5	
Age at diagnosis/referral (years)					
20-29	82	3.3	80	3.1	0.007
30-39	423	16.7	372	14.3	
40-49	910	36.0	890	34.3	
50-59	686	27.1	729	28.1	
60-74	427	16.9	525	20.2	
Education level (years)					0.003
≤ 6	1005	39.8	932	35.9	
7- 9	1013	40.1	1040	40.1	
10-12	403	15.9	483	18.6	
> 12	107	4.2	141	5.4	
Current housing type					<0.001
Building (concrete structure)	1818	71.9	2019	77.8	
Cottage (clay brick structure)	700	27.7	574	22.1	
Boat	10	0.4	2	0.1	
Current occupation					<0.001
Unemployed	77	3.1	95	3.7	
Farmer	855	33.8	984	37.9	
Blue-collar	1020	40.4	900	34.7	
White-collar	350	13.8	416	16.0	
Other/unknown	226	8.9	201	7.7	
First-degree family history of NPC					<0.001
No	2204	87.2	2482	95.6	
Yes	272	10.8	70	2.7	
Unknown	46	1.8	43	1.7	
Cigarette smoking					0.07
Never	1117	44.2	1213	46.7	
Ever	1410	55.8	1381	53.2	
Tea drinking					<0.001
Less than daily	1614	63.8	1512	58.2	
Daily	911	36.1	1081	41.6	
Salt-preserved fish consumption in 2000-2002					0.02
Yearly or less	1853	73.3	1901	73.2	
Monthly	484	19.2	542	20.9	
Weekly or more	188	7.4	149	5.7	

**P* value was determined by a two-sided t-test. Other *P* values were determined by a chi-square test.

The adjusted ORs for the associations of oral health indicators with NPC risk are shown in **Table 4**. Increased number of teeth lost after age 20 was not associated with a higher risk of NPC in either minimally adjusted models (P_{trend} = 0.46) or fully adjusted models (P_{trend} = 0.98). Nevertheless, among those who had experienced tooth loss, we found a marginal association between earlier age at first tooth loss with an increasing risk of NPC (P_{trend} = 0.08). The fully-adjusted ORs (95% CIs) for those with first tooth loss at 40-49, 30-39, and 20-29 vs. ≥ 50 years were 1.18 (0.91, 1.53), 1.23 (0.93, 1.62), and 1.33 (0.98, 1.79), respectively. Increasing risks of NPC with an increasing number of teeth filled were also observed in both minimally adjusted model (P_{trend} = 0.003) and fully adjusted model (P_{trend} = 0.002). For example, compared with those who had not had any teeth filled, those with 1 to 3 or more than 3 teeth filled had significantly increased risks of NPC, with ORs of 1.25 (95% CI = 1.06, 1.49) and 1.55 (95% CI = 1.13, 2.12), respectively. In contrast, more frequent tooth brushing was inversely associated with the risk of NPC. Compared with less frequent brushing, brushing teeth twice per day or more had a fully adjusted OR of 0.62 (95% CI = 0.55, 0.70). Further adjustment for alcohol intake, a self-reported history of chronic rhinitis, herbal medicine intake, and intake of green leafy vegetables (g/day) and fruits (g/day) in 2000-2002 or detailed smoking history (i.e. never, former, or current plus pack-years) did not significantly change the magnitude of the associations (data not shown).

Among never smokers, the strength of the associations with oral health variables was similar to that in the overall study population, except that earlier age at first tooth loss was not associated with a higher risk of NPC (data not shown).

We selected number of teeth lost after age 20 and frequency of tooth brushing as the two main oral health indicators and tried to test the heterogeneity in the associations by sex, age group, education level, first-degree family history of NPC, or salt-preserved fish consumption. No significant heterogeneity was detected (data not shown).

We did not detect any appreciable changes in the magnitude of associations from two sensitivity analyses, one excluding the 32 replacement controls in Wuzhou who participated in a screening program for NPC, and one limiting to cases interviewed within 30 days of diagnosis (2166 cases, 86% of 2528, data not shown).

Table 4. Odds ratios (ORs) and 95% confidence intervals (CIs) of nasopharyngeal carcinoma associated with oral health in southern China (2010-2014).

Variable	Cases (n=2528)		Controls (n=2596)		Minimally adjusted OR	Fully adjusted OR
	No.	%	No.	%	(95% CI) *	(95% CI) †
Number of teeth lost after age 20 years						
None	1282	50.7	1268	48.8	referent	referent
1 - 3	701	27.7	739	28.5	0.98 (0.86, 1.12)	0.97 (0.84, 1.11)
4 - 13	417	16.5	453	17.5	1.01 (0.85, 1.19)	0.99 (0.83, 1.18)
≥ 14	128	5.1	136	5.2	1.10 (0.84, 1.45)	1.00 (0.75, 1.32)
P for trend					0.46	0.98
Age at first tooth loss (years) ‡						
≥ 50	202	16.2	306	23.0	referent	referent
40 - 49	285	22.9	352	26.5	1.12 (0.87, 1.44)	1.18 (0.91, 1.53)
30 - 39	304	24.4	323	24.3	1.26 (0.96, 1.64)	1.23 (0.93, 1.62)
20 - 29	297	23.8	289	21.8	1.34 (1.00, 1.79)	1.33 (0.98, 1.79)
unknown	158	12.7	58	4.4	-	-
P for trend					0.04	0.08
Number of filled teeth						
None	2077	82.2	2223	85.6	referent	referent
1 - 3	352	13.9	297	11.4	1.25 (1.06, 1.47)	1.25 (1.06, 1.49)
≥ 4	99	3.9	76	2.9	1.39 (1.03, 1.89)	1.55 (1.13, 2.12)
P for trend					0.003	0.002
Frequency of brushing teeth (times/day)						
≤ 1	1696	67.1	1456	56.1	referent	referent
≥ 2	824	32.6	1133	43.6	0.58 (0.52, 0.66)	0.62 (0.55, 0.70)
Irregular	8	0.3	7	0.3	-	-

* Adjusted for sex, age (5-year categories), and residential area (Zhaoqing, Wuzhou, or Guiping/Pingnan).

† Adjusted for sex, age, residential area, education level (≤ 6, 7-9, 10-12, or >12 years), current housing type (building, cottage, or boat), current occupation (unemployed, farmer, blue-collar, white-collar, or other/unknown), first-degree family history of nasopharyngeal carcinoma (yes, no, or unknown), cigarette smoking (ever or never), tea drinking (less than daily or daily), and salt-preserved fish consumption in 2000-2002 (yearly or less, monthly, or weekly or more). We excluded 18 subjects due to missing covariate values.

‡ Among 2574 participants who experienced tooth loss after age 20 years.

5.3 STUDY III

Cases and controls had similar numbers of first-degree relatives (median 8 vs 8), number of siblings (median 4 vs 4), and number of offspring (median 2 vs 2) (**Table 5)**. The sibling sex distribution was similar between cases and controls, but siblings of controls were on average older than those of cases. The median number of offspring and their age and sex distribution were similar between cases and controls, except that sons of controls were slightly older than those of cases.

Table 5. Distribution of family size, number and age of siblings, and number and age of offspring in nasopharyngeal carcinoma cases and controls.

	Cases (No=2499)	Controls (No=2576)
Family size, median (range) *	8 (2-18)	8 (2-18)
Siblings		
No. of siblings, median (range)	4 (0-11)	4 (0-13)
No. of brothers (% siblings)	4749 (51%)	4625 (50%)
Mean age of brothers (years) †	47.8	49.0
No. of sisters (% siblings)	4566 (49%)	4703 (50%)
Mean age of sisters (years) †	48.1	49.0
Offspring		
No. of offspring, median (range)	2 (0-11)	2 (0-8)
No. of sons (% offspring)	3487 (54%)	3671 (55%)
Mean age of sons †	22.3	23.5
No. of daughters (% offspring)	3015 (46%)	3058 (45%)
Mean age of daughters †	23.3	23.6

* Family size includes all first-degree relatives except the index person.

† Age at the date of interview of the index person. For relatives who died before the interview, age at death was used.

Relative risk estimates associated with a positive family history of NPC

We found that people with a first-degree family history of NPC were at a higher risk for NPC, compared to those without, but had no excess risk of other malignancies. Overall, 10.8% of cases had reported a positive family history of NPC among first-degree relatives, versus 2.7% of controls (**Table 6**). Associations were similar between models minimally adjusted for sex, age, residential area, and family size and models fully adjusted for these four variables plus education level, current housing type, current occupation, cigarette smoking, tea drinking, and salt-preserved fish consumption in 2000-2002. The fully-adjusted ORs of NPC associated with ever having any NPC diagnosis among first-degree relatives was 4.6 (95% CI = 3.5, 6.1). NPC risk increased for those with a greater number of affected relatives; the fully adjusted ORs (95% CIs) for having one or at least two affected relatives were 4.4 (3.3, 5.9) and 6.7 (2.8, 16.0), respectively ($P_{trend} < 0.001$). The strength of the association was greater for having a mother (OR=5.6) than a father (OR=3.1) with NPC, and slightly but not significantly greater for having an affected sibling (OR=5.1) than a parent (OR=4.0), and a sister (OR=5.6) than a brother (OR=4.8). A history of NPC in offspring was not associated with a significantly higher risk of NPC (OR=1.6, 95% CI = 0.3, 10.0); however, because of the relative young age (22-24 years) of the offspring of cases and controls, the number of affected offspring was very small.

Table 6. Odds ratios (ORs) and their corresponding 95% confidence intervals (CIs) for risk of nasopharyngeal carcinoma (NPC) in association with family history of NPC in first-degree relatives.

	Cases (n=2499)	Controls (n=2576)	Minimally adjusted OR (95% CIs) *	Fully adjusted OR (95% CIs) †
Family history of NPC in first-degree relatives				
No	2208	2483	referent	referent
Yes	270	69	4.5 (3.4, 5.8)	4.6 (3.5, 6.1)
1 affected	236	63	4.3 (3.2, 5.7)	4.4 (3.3, 5.9)
≥2 affected	34	6	6.6 (2.7, 15.7)	6.7 (2.8, 16.0)
P for trend			<0.001	<0.001
Affected relatives ‡				
Father				
No	2421	2543	referent	referent
Yes	72	26	2.9 (1.8, 4.5)	3.1 (2.0, 5.0)
Mother				
No	2429	2561	referent	referent
Yes	58	11	5.4 (2.8, 10.4)	5.6 (2.9, 10.8)
Parent				
No	2357	2529	referent	referent
Yes	130	37	3.7 (2.6, 5.4)	4.0 (2.7, 5.8)
Brother				
No	2332	2478	referent	referent
Yes	113	26	4.7 (3.0, 7.2)	4.8 (3.1, 7.4)
Sister				
No	2395	2499	referent	referent
Yes	53	10	5.7 (2.9, 11.3)	5.6 (2.8, 11.1)
Siblings				
No	2283	2466	referent	referent
Yes	159	35	5.0 (3.5, 7.3)	5.1 (3.5, 7.4)
1 affected	142	34	4.6 (3.2, 6.8)	4.7 (3.2, 6.9)
≥2 affected	17	1	19.1 (2.5, 144.1)	18.9 (2.5, 142.8)
P for trend			<0.001	<0.001
Offspring				
No	2355	2436	referent	referent
Yes	3	2	1.7 (0.3, 10.4)	1.6 (0.3, 9.9)

* Adjusted for age, sex, residential area, and family size (continuous; see below).

†Adjusted for age, sex, residential area, family size (see below), education level, current housing type, current occupation, cigarette smoking, tea drinking, and salt-preserved fish consumption in 2000-2002.

‡ In analyses of history of NPC in parents, fathers, or mothers, ORs were not adjusted for family size. In analyses of history of NPC in brothers or sisters, siblings, or offspring, ORs were adjusted for number of brothers, number of sisters, number of siblings, or number of offspring, respectively.

A self-reported second-degree family history of NPC was also associated with a higher risk of NPC (OR = 5.3, 95% CI = 3.4, 8.3), and the excess risks increased with a greater number of affected second-degree relatives ($P_{trend} < 0.001$).

Based on the fully adjusted OR, up to 8.6% (95% CI = 5.9%, 11.3%) of all incident NPC cases in the population are attributable to a first-degree family history of NPC.

We identified a total of 20,157 first-degree relatives of cases and 20,624 first-degree relatives of controls. The overall adjusted HR for NPC in first-degree relatives of cases compared with controls was 4.2 (95% CI = 3.2, 5.5). As shown in **Figure 10**, the HR was higher for those with an affected sibling than an affected parent, and it was also slightly higher for those with an affected mother versus a father, and for those with an affected sister versus a brother.

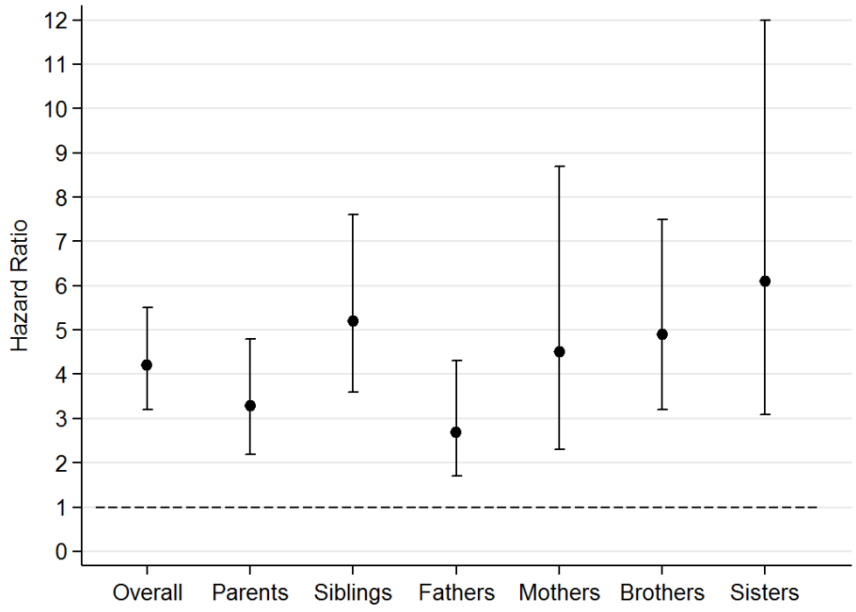

Figure 10 Hazard ratios and 95% confidence intervals for nasopharyngeal carcinoma by relationship of the affected relatives.

Cumulative risk of NPC associated with a positive family history of NPC

The estimated cumulative risks of NPC for people with a positive first-degree family history of NPC are listed in **Table 7**. Because only one sister aged <50 years affected with NPC was reported among male controls and only one mother affected with NPC was reported among female controls, the cumulative risks in these groups were not estimated. We found that between ages 20 and 74, the cumulative risk of NPC for males with any affected first-degree relative was 5.0% (95% CI = 3.6%, 7.0%), ranging from 4.2% with an affected father to 6.3% with an affected sister. For females, between ages 20 and 74, the cumulative risk of NPC with any affected first-degree relative was 1.9% (95% CI = 1.1%, 3.2%), ranging from 0.8% with an affected father to 2.2% with an affected brother.

Table 7. Adjusted odds ratios (ORs) and estimated cumulative risks (%), with corresponding 95% confidence intervals, of nasopharyngeal carcinoma (NPC) by type of relatives affected with NPC and by age (years) and sex of subjects *

	Males			Females		
	- 50 years	- 60 years	- 74 years	- 50 years	- 60 years	- 74 years
General population (2011, %)	0.4	0.7	1.1	0.2	0.3	0.4
Any first-degree relatives						
Adjusted OR	5.7 (3.5, 9.3)	5.3 (3.7, 7.7)	4.7 (3.4, 6.5)	4.3 (2.0, 8.9)	4.5 (2.5, 8.3)	4.6 (2.7, 7.8)
Cumulative risk (%)	2.2 (1.4, 3.6)	3.9 (2.7, 5.7)	5.0 (3.6, 7.0)	0.7 (0.3, 1.5)	1.3 (0.7, 2.4)	1.9 (1.1, 3.2)
Father						
Adjusted OR	3.7 (1.7, 7.8)	4.5 (2.4, 8.6)	3.9 (2.2, 6.8)	2.0 (0.7, 5.6)	1.8 (0.7, 4.4)	1.9 (0.8, 4.3)
Cumulative risk (%)	1.4 (0.7, 3.0)	3.3 (1.8, 6.4)	4.2 (2.4, 7.3)	0.3 (1.2, 9.5)	0.5 (0.2, 1.3)	0.8 (0.3, 1.8)
Mother †						
Adjusted OR	6.1 (2.4, 15.9)	4.5 (2.2, 9.3)	4.4 (2.2, 8.9)	N/A	N/A	N/A
Cumulative risk (%)	2.4 (0.9, 6.2)	3.3 (1.6, 6.9)	4.7 (2.4, 9.5)	N/A	N/A	N/A
Brother						
Adjusted OR	5.2 (2.0, 13.7)	5.2 (2.7, 9.7)	4.6 (2.8, 7.6)	4.0 (1.0, 15.5)	4.2 (1.6, 10.9)	5.4 (2.3, 12.8)
Cumulative risk (%)	2.0 (0.8, 5.3)	3.9 (2.0, 7.2)	4.9 (3.0, 8.1)	0.7 (0.2, 2.6)	1.2 (0.5, 3.2)	2.2 (0.9, 5.3)
Sister ‡						
Adjusted OR	N/A	6.5 (2.5, 16.9)	5.9 (2.6, 13.2)	3.1 (0.6, 15.9)	5.8 (1.3, 26.5)	4.4 (1.2, 15.7)
Cumulative risk (%)	N/A	4.8 (1.9, 12.5)	6.3 (2.8, 14.1)	0.5 (0.1, 2.7)	1.7 (0.4, 7.7)	1.8 (0.5, 6.4)

*Adjusted for age, residential area, family size (see below), education level, current housing type, current occupation, cigarette smoking, tea drinking, and salt-preserved fish consumption in 2000-2002.
In analyses of history NPC in fathers, or mothers, ORs were not adjusted for family size. In analyses of history NPC in brothers or sisters, ORs were adjusted for number of brothers or number of sisters, respectively.

† Only one mother affected with NPC was reported among female controls; therefore, the cumulative risk was not estimated.

‡ Only one sister aged <50 years affected with NPC was reported among male controls; therefore, the cumulative risk in this age group was not estimated

Based on the reconstructed kin cohort, we were able to estimate the lifetime cumulative risks of NPC up to age 74 years among first-degree relatives of cases and controls and results are shown in **Table 8**. The cumulative risk of NPC among relatives of cases was higher than that of controls (3.7% vs. 0.9%). The incidence rate ratio comparing age-specific risk of NPC in relatives of cases vs. controls was as high as 12.4 at ages up to 29 years, and gradually decreased with advancing age to 3.7 at ages up to 74 years.

Table 8. Lifetime cumulative risks (%) and incidence rate ratios, with corresponding 95% confidence intervals (CIs), of nasopharyngeal carcinoma (NPC) among first-degree relatives of cases and controls, by age of relatives

	Age of first-degree relatives (years)				
	- 29	- 39	- 49	- 59	- 74
Controls					
No. of relatives at risk	20,624	15,522	12,542	8,712	5,678
Cumulative No. of NPCs	1	11	42	62	74
Cumulative risk (%)	0.006	0.1	0.4	0.6	0.9
95% CIs	0.001, 0.04	0.08, 0.2	0.3, 0.5	0.5, 0.8	0.7, 1.2
Cases					
No. of relatives at risk	20,157	14,871	11,838	7,994	5,278
Cumulative No. of NPCs	13	68	168	250	291
Cumulative risk (%)	0.08	0.5	1.5	2.7	3.7
95% CIs	0.05, 0.1	0.4, 0.6	1.3, 1.7	2.4, 3.0	3.3, 4.2
Cases versus controls					
Incidence rate ratio	12.4	5.4	3.2	4.6	3.7
95% CIs	1.9, 526.4	2.7, 11.8	2.1, 5.0	2.8, 7.9	1.9, 7.7

Figure 11 shows the relative-specific cumulative risks of NPC up to age 74 years. We found that lifetime cumulative risk up to age 74 years of NPC was greater in siblings of NPC cases (6.3% in brothers, 3.5% in sisters), than in parents (3.4% in fathers, and 2.5% in mothers). In contrast, there were no differences in cumulative risks of NPC between siblings of and parents of population-based controls. For male relatives, brothers had a higher cumulative risk of NPC that fathers of cases, but no such difference was observed among brothers and fathers of controls. No significant differences were found between mothers and sisters of cases or controls (data not shown).

Validation on family history of NPC

We confirmed a positive first-degree family history of NPC according to cancer registry data and/or medical records among 39 of 41 cases and 5 of 5 controls, and for a negative family history for 50 of 52 cases and 45 of 48 controls in Sihui City. The reasons for a lack of confirmation were incorrect contact information (N=3) and inability to recall the names of first-degree relatives (N=4). Among cases, the sensitivity and specificity for self-reported first-degree family history of NPC were 95% (95% CI = 88%, 100%) and 98% (95% CI = 94%, 100%), respectively; among controls, they were 83% (95% CI = 54%, 100%) and 100% (95% CI = 100%, 100%), respectively. Among cases, the positive predictive value and negative predictive value for self-reported first-degree family history of NPC were 97% and 96%, respectively; among controls, they were 100% and 98%, respectively. The number of observed (self-reported) NPC cases between ages 20 and 74 among first-degree relatives of controls was slightly higher than that expected in the general population of the overall study area in 2011 (74 observed vs. 59 expected among controls, standardized incidence ratio=1.2, P=0.07). After correcting for misclassification, the OR associated with having at least one affected first-degree relative was 3.1.

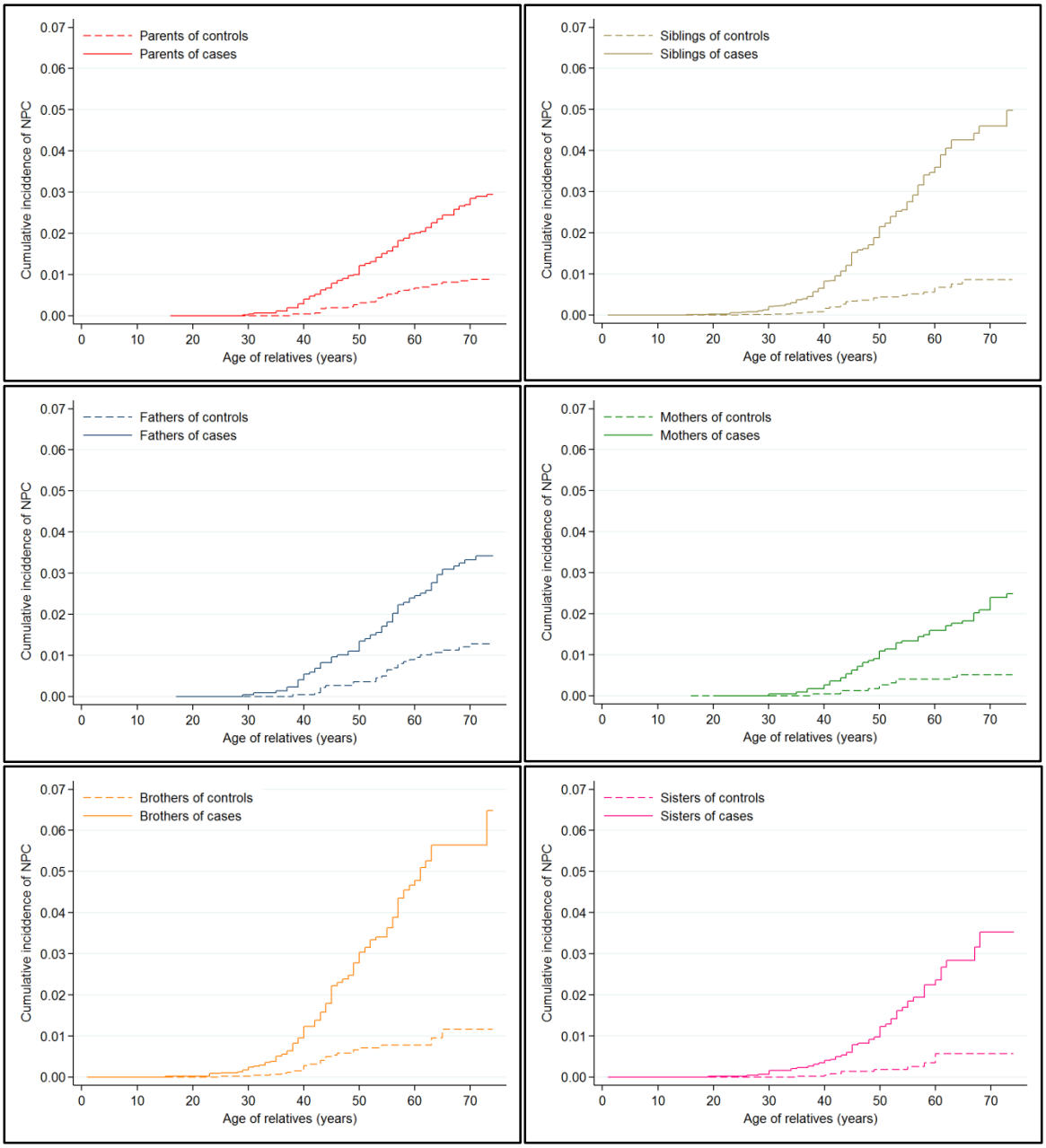

Figure 11 Cumulative incidence of nasopharyngeal carcinoma (NPC) among different relatives of NPC cases and controls.

5.4 STUDY IV

We utilized plasma samples from 174 NPC cases and 175 community-based controls collected in Taiwan. Number of bead counts across all 384 biomarkers (including 2 positive controls and 2 negative controls) are acceptable. We found that PQN-MA-Ind had the highest median ICC across all biomarkers of 0.81, compared with other 5 normalization schemes (**Figure 12**), thus we used PQN-MA-Ind as our normalization method.

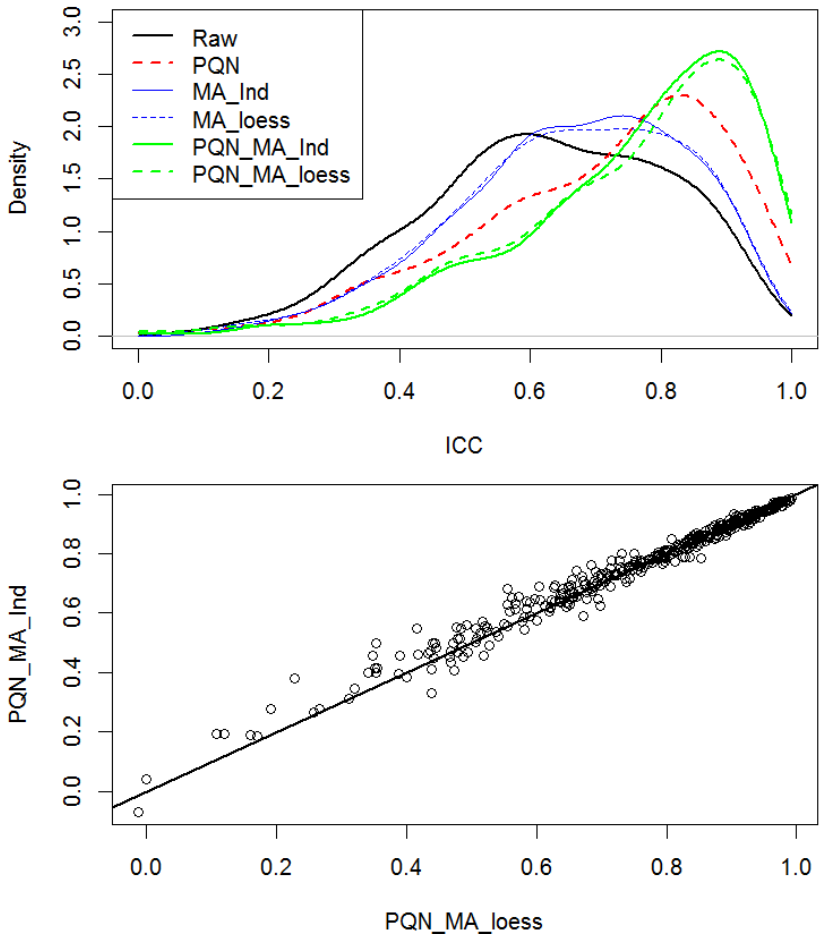

Figure 12 Density plot of ICC distributions and scatter plot and correlation between unsmoothed multi-dimensional normalization after probabilistic quotient normalization (PQN-MA-Ind) and smoothed multi-dimensional normalization after probabilistic quotient normalization (PQN-MA-loess)

Results from unconditional logistic regression models, adjusted for age and sex, showed that nine biomarkers had significant *P* values after controlling false discovery rate at 5%. One biomarker, interleukin 6 receptor (*IL6R*) was excluded because of a low ICC of 0.28, leaving eight biomarkers in the panel. They are cyclin B1 (*CCNB1*), kinase insert domain receptor (*KDR*, also known as vascular endothelial growth factor receptor 2 [*VEGFR2*]), hyaluronan synthase 1(*HAS1*), lymphocyte antigen 6 complex locus K (*LY6K*), interleukin 2 receptor subunit alpha (*IL2RA*), chemokine (C-X-C motif) ligand 10 (*CXCL10*), platelet-derived growth factor beta polypeptide (*PDGFB*), and lectin galactoside-binding soluble 1(*LGALS1*). **Table 9** shows the individual performance of eight selected biomarkers for early detection of NPC. A combination of these eight biomarkers showed a high accuracy to distinguish NPC patients in early stages from controls (AUC = 0.808, 95% CI: 0.745, 0.871) (**Figure 13**).

Table 9. Performance of eight selected biomarkers for early detection of nasopharyngeal carcinoma.

Gene name	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC (95% CI)
CCNB1	42.4 (30.3, 55.2)	80.6 (73.9, 86.1)	45.2 (36.0, 57.9)	78.8 (68.9, 84.8)	0.651 (0.572, 0.729)
KDR (VEGFR2)	69.7 (57.1, 68.9)	61.7 (54.1, 68.9)	40.7 (33.4, 55.1)	84.4 (75.8, 88.1)	0.667 (0.594, 0.740)
PDGFB	69.7 (62.3, 76.4)	65.1 (52.4, 76.5)	84.1 (75.8, 88.2)	44.8 (36.8, 58.1)	0.682 (0.607, 0.757)
LGALS1	60.6 (52.9, 67.8)	66.7 (54.0, 77.8)	82.8 (73.9, 86.9)	38.9 (31.8, 52.8)	0.661 (0.586, 0.736)
HAS1	54.5 (41.8, 66.9)	73.7 (66.5, 80.1)	43.9 (35.7, 56.8)	81.1 (72.0, 86.0)	0.630 (0.545, 0.715)
LY6K	78.8 (67.0, 88.9)	45.7 (38.2, 53.4)	35.4 (28.6, 51.7)	85.1 (75.7, 88.6)	0.639 (0.563, 0.716)
IL2RA	71.2 (58.7, 81.7)	56.0 (48.3, 63.5)	37.9 (31.0, 52.4)	83.8 (74.8, 87.6)	0.646 (0.568, 0.724)
CXCL10	48.5 (36.0, 61.1)	77.1 (70.2, 83.1)	44.4 (35.8, 57.2)	80.0 (70.3, 85.3)	0.640 (0.558, 0.723)

Abbreviations: CI, confidence interval, PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve

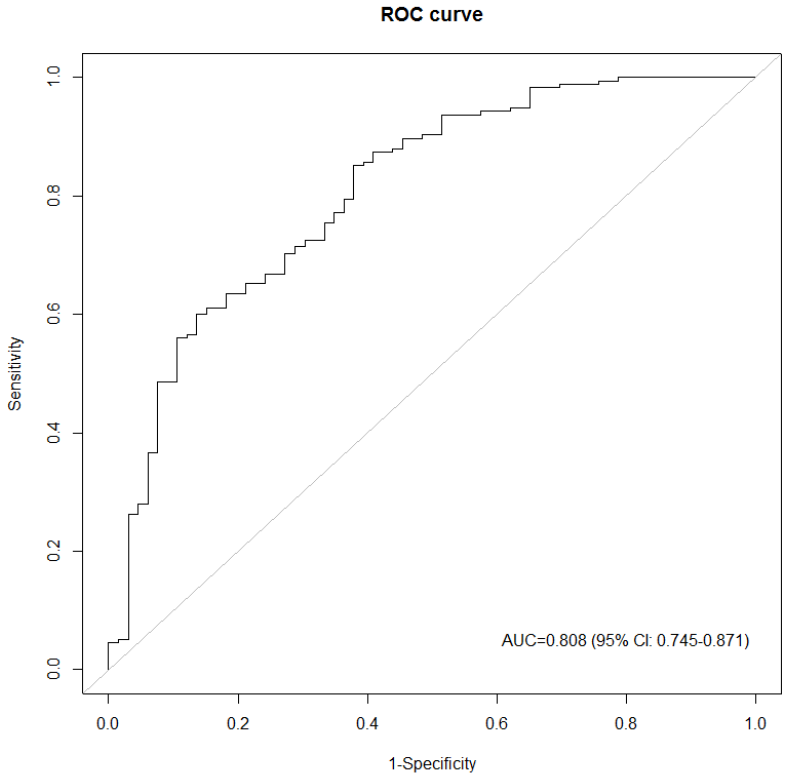

Figure 13 Performance to detect nasopharyngeal carcinoma in early stages. The receiver operating characteristic (ROC) curve from a combination of eight selected biomarkers established from an unconditional logistic regression model.

6 DISCUSSSION

6.1 FINDINGS AND IMPLICATIONS

The results of this thesis indicate that NPC risk is highly related to the early-life social environment, especially for that associated with EBV infection, as further supported by the clearly contrasting findings between NPC and IM in a non-endemic area. In endemic areas, poor oral hygiene indicators were associated with a higher risk of NPC and improvement of oral hygiene might contribute to a lower NPC risk. The excess risk of a positive family history of NPC is higher for a maternal than a paternal history and slightly stronger for a sibling than a parental history, and for a sororal than a fraternal history. Such evidence highlights the need of strategic development for early detection and clinical consultation in NPC high-incidence populations. By applying antibody suspension bead array assays, we have identified eight proteins, which might be a potential panel to facilitate early detection of NPC.

6.1.1 NPC Risk Assessment

Early-Life Risk Factors

So far, most of our knowledge related to risk factors for NPC has been obtained from studies on populations at a high- to intermediate-incidence, studies on low-incidence areas are required to understand the complexity of NPC etiology between endemic and non-endemic areas [17, 190], and between type I and type II/III NPC. In low-incidence areas, however, due to the fact that NPC is an extremely rare disease, it is difficult to achieve enough statistical power to address research questions in terms of risk assessment. The Swedish population registers have provided us valuable resources. We had originally hypothesized that early-life infection with EBV contributes to the pathogenesis of NPC and thus may be associated with the subsequent NPC risk later in life. Although a history of IM, which is caused by delayed infection with EBV, has been proposed to entail a lower risk of NPC, most of the previous studies are based on few cases with such a history [7, 45, 46]. In high-incidence areas, most of individuals are infected with EBV in an early age, precluding us to investigate such association. We therefore utilized early-life family structure as the proxy of timing at infection to test our hypothesis in Sweden, a low-incidence area. A larger number of siblings and a later birth order may serve as indicators of earlier infection with EBV, and can lend insight into whether timing of primary infection with EBV is associated with subsequent risk of NPC [191].

Further evaluation of risk factors for NPC in low-incidence populations will help to reveal differences and similarities in the pathogenesis of types I, II, and III NPC. For example, although infection with EBV is a sufficient cause for type III NPC, its role in type I is uncertain. Well established risk factors for type I NPC are smoking and alcohol drinking, while their roles in type III NPC are less clear. Previous studies in high-incidence areas provided little information on risk factors for type I NPC partly because of the limited number of cases (i.e. only ~1%) in high-incidence areas. Our study showed that the association with number of siblings among cases with type I NPC was weaker than among cases with undifferentiated NPC, although still positive and significant. Age at infection is not the only means through which other siblings could influence NPC risk. Birth order could be an indicator of exposure to other infectious agents, or associated with other exposures such as maternal breast feeding pattern and childhood nutrition [192, 193]. Nevertheless, our results suggested a positive association of larger sibship size, and later birth order with a higher risk of NPC. These NPC-specific findings contrast with data from other EBV-related disorders (i.e., IM and HL) [86, 194], highlighting possible differences in EBV-related risk

factors between the distinct tumor types that should be evaluated in a region such as Southeast Asia where NPC is an endemic disease.

Oral Hygiene

The association between poor oral health and risk of head and neck cancers has been extensive evaluated in studies with cohort design. However, it has not been evaluated for NPC, largely due to the limited number of NPC cases from a cohort study. Case-control design is more suitable to study the association for a rare disease. In this large scale population-based case-control study in NPC high-incidence populations in southern China, although a higher number of teeth lost was not associated with a higher NPC risk, excess risks were observed for other oral health indicators, such as a higher number of filled teeth. The association with earlier age at first adult tooth loss, albeit statistically non-significant, was in the same direction. In contrast, brushing teeth twice per day or more appeared to be inversely associated with NPC risk.

To the best of our knowledge, this is the first population-based case-control study to address the association between poor oral health and NPC risk. Our findings are somewhat in line with those from a hospital-based study in Turkey [118], which showed that both infrequent tooth brushing and having more than 10 decayed teeth were associated with an elevated risk of NPC. The stronger associations in the Turkish study may be due to different study design, confounding factors, bias, chance, as well as the different oral hygiene condition in different populations.

Although we found a null association between the number of adult teeth lost and risk of NPC, biologically plausible interpretations cannot be ruled out. In the present study, due to the relative younger ages of NPC onset in high-incidence populations (40-49 years), ~50% of study subjects had not experienced any adult tooth loss. Thus number of teeth lost may not be a sensitive indicator of oral health in this setting. Alternatively, the positive association with a higher number of filled teeth may suggest that factors related to dental caries are more important in the pathogenesis of NPC.

Family History

The more than 4-fold increase of NPC risk among first-degree relatives of NPC cases is in lines with that reported in other studies [33-36, 121, 122, 129, 130]. Results from high-incidence areas consistently showed that the excess risk was only limited to family history of NPC but not to other non-NPC cancers. Whereas studies from intermediate- [34] and low-incidence populations [36] showed that the elevated risk of NPC among first-degree relatives was not limited to NPC but extended to other cancers, such as cancer of the salivary glands and cervix. The underlying mechanisms are largely unknown. Such inconsistent findings could be explained by different penetrations of NPC susceptibility genes, or variations in attributable environmental or lifestyle risk factors, or both. In high-incidence area, environment risk factors alone are unlikely to contribute such strong familial associations. Studies have demonstrated the role of genetic risk factors in the pathogenies of NPC, such as certain HLA alleles [41, 42, 144].

The present study also showed that the magnitude of association is stronger for people with a sibling history than with a parental history of NPC, and is also higher for people with a maternal history of NPC than a paternal history. These patterns could be explained by shared childhood environment between siblings, or similar dietary habits between mother and child. Our observation of slightly higher relative risks conferred by having an affected female than a male first-degree relative (i.e., a mother or sister as compared with a father or brother,

respectively) may be a chance finding, but it merits further investigation into possible mechanisms.

In contrast to prior reports from China [122, 133], we found no significant modification of the association with family history of NPC by environmental risk factors. Our findings suggest that putative NPC susceptibility genes may act independently of environmental factors in NPC etiology. However, the small number of controls with a positive first-degree family history of NPC and the low power for heterogeneity test make it difficult to draw firm conclusions about the joint effects of family history and environmental risk factors. Pooled studies with larger numbers of study subjects will enable more powerful tests of such interactions.

Different from other studies on family history, we used two complementary approaches to estimate the cumulative risk of NPC. They showed similar results, although each approach has its strength and limitation. Based on a case-control design, we were able to estimate the cumulative risk of NPC between ages 20 and 74 for people with a positive first-degree family history of NPC. We found that approximately 5.0% of men and 1.9% of women with a first-degree family history of NPC will develop NPC, compared with 1.1% of men and 0.4% of women in the general population without such a history. If one was affected with NPC, he/she may be interested to know the cumulative risk for his/her first-degree relatives. We found 6.3% of NPC patients' brothers and 3.5% of sisters up to age 74 years are expected to develop NPC, compared with 3.4% of fathers and 2.5% of mothers. These results could help to determine whether close clinical surveillance for NPC may be beneficial for individuals with a family history of NPC.

6.1.2 Biomarkers for Early Detection and Screening

The identification of disease biomarkers has two implications: it will not only help to identify markers that can distinguish patients from controls, but also help to provide new insights into the biology behind the disease: the selected eight proteins may highlight the important pathways in the carcinogenesis of NPC. For example, cyclin B is in the apoptosis pathway initiated by EBV-encoded LMP1 [195], which is necessary for the progression of the cells into and out of M phase of the cell cycle. Vascular endothelial growth factor (VEGF) is an important signaling protein involved in both vasculogenesis and angiogenesis, which was reported to be widely expressed in NPC specimens [196, 197]. In addition, the links between HAS1, LY6K, and NPC have not been explored before. Our findings may provide new mechanistic insight into the occurrence of NPC.

The development of NPC is thought to involve a complex interplay between environmental risk factors, genetic susceptibility, and EBV infection. Previous studies largely focused on the search for EBV-related biomarkers, but paid less attention to the contributions from the host response. We hypothesized that the host response also plays an important role on the occurrence of NPC. Therefore, we aimed to validate previously reported human proteome and establish a panel of biomarkers that can facilitate early detection and prediction of NPC. The antibody suspension bead array assays, which have been developed for multiplex screening of a large number of proteins in patient cohorts, enable us to obtain the plasma antibody profile more efficiently. We found a panel of eight biomarkers that can achieve a promising result for early detection of NPC. Further studies are needed to develop more sensitive sandwich immunoassays for screening in larger sample sets, and test validity of the selected biomarkers.

6.2 METHODOLOGY ON EPIDEMIOLOGY

6.2.1 Study Design

To identify exposures involved in the development of a rare disease, a case-control design is more efficient than cohort design. For example, in a cohort study among 19,000 individuals conducted in an NPC high-incidence area, only 125 NPC cases were identified after an average of 17 years of follow-up [69]. In Taiwan, an NPC intermediate-incidence area, during a total of 185,587 person-years of follow-up, 33 incident NPC cases were identified from a community cohort comprising 9622 males [35]. In low-incidence area, such as Sweden, we followed about 5.3 million individuals between 1961 and 2009 to obtain a total of 301 NPC cases in **Study I**. For areas without established population registers, cohort design becomes logistically impractical. In addition, for exposure with a small effect size, a large sample size is required to achieve a satisfied statistical power. Sample size is expected to be increased substantially when studying the joint effects of exposures. Conceptually, cohort study design is ideal to address the association between exposure and disease because the exposure information of all individuals in a source population is collected before disease occurrence. Nevertheless, complex diseases such as most types of cancers are thought to involve a complex interplay between environmental factors and genetic susceptibility; the effect size of each individual exposure is usually small. Case-control design becomes the only possible alternative.

Ideally, a case-control study should include all incident cases of the disease under study in a defined population during a specified time period. Under this conceptualization, it requires an existing registry to keep the defined population under surveillance. In the majority of settings around the globe, these fundamental requirements are difficult or impossible to accommodate. In addition, a group of controls without the disease are sampled directly from this source population in such a way (generally randomly) that they inform about the estimate of the prevalence of the exposure and covariates in the person-time that give rise to the cases. As a corollary, many epidemiologic researches are undertaken with suboptimal methods, such as hospital-based and neighborhood-based studies. By using these study designs, information on environmental exposures (particularly with regard to diet, smoking, and medical history) collected is not representative. Even worse, a population and person-time cannot be clearly defined in many case-control studies, leading important human diseases to escape proper investigation of their causes altogether.

For NPC, few strictly population-based case-control studies have as yet been conducted in mainland China [123, 198]. **Study II** and **study III** are based on a population-based case-control study in southern China, where the registries (only available in two counties) do not cover the entire population of the study area. We have provided an example of how population-based epidemiologic research can be successfully conducted in southern China. We took enormous efforts to estimate total number of incident cases in the source population, develop a rapid case ascertainment system, contact randomly selected controls, and collect questionnaire data and biosamples. The overall participating rate of about 85% demonstrated the feasibility of conducting rigorous population-based case-control study in mainland China.

6.2.2 Validity

There are two components of the validity of a study: internal validity and external validity (also known as generalizability). For internal validity, most violations can be classified into three categories: confounding, selection bias, and information bias.

Confounding

Missing information on important confounders is one of the major limitations of studies conducted based on population registers. In **study I**, although we observed an increased risk of NPC with a larger number of total siblings and number of older siblings, the associations could be confounded by other potential confounding factors, such as day care attendance, breast feeding, social economic status, and smoking. In **study II**, information on receipt of professional dental care was not available, which precluded accounting for this potential confounder. There is no formal statistical test to examine whether one potential factor can actually confound an association between exposure of interest and outcome; a lot of approaches, such as directed acyclic graph, are largely based on prior knowledge. However, our knowledge in etiology is usually limited, thus the association we found for observational study cannot be interpreted as the causation.

Selection Bias

Selection bias is one of the major concerns in observational studies, and it causes either an over- or under- estimated association. For example, in **study I**, it is possible that the bias could occur because we "select" NPC cases and IM cases who had had a recorded diagnosis in registers. Because the Swedish Cancer Register is essentially complete, however, it is unlikely that such a selection bias would occur in this NPC study. For IM, we cannot rule out the possibility that the inverse association was limited to those hospitalized IM cases. However, a recent study based on a similar study setting in Denmark [86] has reported similar results irrespective of whether hospitalized or self-reported IM was treated as the outcome. We thus considered this scenario is very rare. In **study II**, although the case-control study has reached a high participating rate among controls (~85%), we cannot rule out the possibility that lower-socioeconomic status (SES) controls with poorer oral health and hygiene were more likely to participate than higher-SES controls who could not be contacted due to employment outside of their hometown. Therefore, in this case, selection bias favoring underestimated ORs could have occurred. In **study III**, if controls who were positive for family history were more likely to participate in an epidemiological research on cancer, that would lead to a biased estimate on the prevalence of a positive family history of study population. In **study IV**, selection bias may also occur because blood samples were available among less than 50% of the parent case-control study. However, because the distributions of age and sex were similar among those with samples and those without, we believe that selection bias occurred due to the unavailability of the samples, if any, should be minimal.

Information Bias (misclassification)

In this thesis, information bias of outcome is less likely to occur because the diagnosis of cases was either based on Swedish Cancer Register or histopathological confirmed. Information bias of exposure is one of the major concerns in the four studies. First, in **study I,** the Multi-Generation Register is incomplete for some individuals, which could result in differential misclassification if the probability of missing data differed between cases and controls [157]. However, our sensitivity analyses restricted by birth date or diagnosis date suggested that such bias, if any, was minimal. In addition, we selected sibship size and birth order as the proxy of timing at infection with EBV, a more direct measure of timing at infection with EBV thus is needed to clarify the misclassification of the exposure. Second, in **study II**, we lacked information on disease symptoms and stage, which may influence some oral health and hygiene measures [199]. If cases were more likely to report a poor oral hygiene due to the presence of the disease, it may lead to an over-estimated association. It may be the same case for **study III** that cases were more likely to report a positive family history of NPC than controls. Therefore, we tried to validate self-reported family history of NPC, although the results were based on only one county, and may not be able to be generalized to other study regions. We also limited the primary analysis to first-degree

relatives to reduce recall bias. The results of our validation study showed that ORs and cumulative risks would be overestimated suggesting that information bias exits. Nevertheless, the striking positive association persisted after adjustment for bias. Third, in **study IV**, biomarkers were measured retrospectively using long term stored samples (collected in 1990s and assayed in 2016), thus we cannot rule out the possibility that results were biased due to the degradation of proteome. Assuming the biomarkers degraded at a similar rate, it may lead to an underestimate (dilution) of the true strength of an association between the biomarker with the risk of NPC. Cross-reaction between proteins resulting from a similar structure or amino acid sequence to our proteins of interest cannot be ruled out, and this may lead to a false positive results. In addition, due to the complex structure of the proteins, we may fail to target the specific domain and the true functionally relevant protein if with a low concentration it could be masked by a non-functional protein with a high concentration, thus resulting in a false negative finding.

7 CONCLUSIONS

- In low-incidence areas, the risk of developing NPC increased among individuals with more siblings, especially older siblings, whereas the risk of developing IM decreased with increasing number of siblings (either younger or older). Childhood social enviroments, especially that associated with an earlier EBV infection, may play a possible pathogenic role in NPC.

- In high-incidence areas, poor oral health may be associated with an elevated risk of NPC, although some associations may be affected by residual confounding of smoking. Prospective cohort design with a more comprehensive measure of oral hygiene conditions could help to clarify the question of whether poor oral health has an important role of NPC development.

- In high-incidence areas, a positive first-degree family history of NPC is associated with a 4-fold increase in risk of NPC. The excess risk was higher for individuals with an affected sibling than for those with an affected parent. The absolute risks quantified, along with results of future studies on other risk stratification factors and cost-benefit analyses, may guide the clinical consultation and development of surveillance policies for this important public health problem in NPC-endemic areas.

- The mulitplex screening of a large number of proteins by antibody suspension bead assays is a powerful method that could identify human proteomic markers for early detection of NPC in additon to EBV-related biomarkers. The validity of the diagnositc performance of selecetd biomarkers is awating further invesitgation.

8 FUTURE PERSPECTIVES

8.1 FOR EPIDEMIOLOGY

Field work training is an essential part for the discipline of epidemiology, whereas it has become less and less available in many training programs. Rigorous epidemiological study design is the key to obtaining unbiased measures for successful scientific research. In NPC research field, however, partly due to the fact that it is endemic in less developed areas, proper facilities (e.g. population register) are lacked to conduct high quality studies. We hope our study can provide an example which will encourage others to undertake similar studies (although not limited to NPC) in populations with similar characteristics. To understand the complexity of etiology, pooling analysis of a few high quality studies is necessary.

8.2 FOR ETIOLOGY AND EARLY DETECTION

A very intriguing question is why almost of all people are infected with EBV, but only some of the infections result in NPC, whereas some others result in HL, BL, and EBV positive gastric cancer. It seems that different populations are susceptible to different types of cancers. For example, areas with a high-incidence of NPC always have a low-incidence of HL, and vice versa. EBV alone is not a sufficient cause of the cancers, and it requires co-factors to act together and contribute the cancer development.

Few studies have tried to address the research question by comparing the etiology across EBV-related cancers. In fact, different risk patterns can be observed for different EBV-related cancers. For example, it seems that IgA response is more important for NPC whereas IgG response is more important for HL [200]. This cumulative evidence highlights that the immune response of host to EBV infection could be critical in explaining the differential risk of developing these EBV-related tumors. However, comprehensive EBV antibody patterns have not been investigated to date. Previous studies of NPC have evaluated both IgG and IgA antibodies against 3 EBV proteins, including EBNA1, VCA, and EA [200]. Although the evidence is strong for a prospective association between NPC risk and EBV-specific antibodies, in particular IgA antibodies, these 3 proteins represent only a small fraction of the approximately 100 open reading frames for EBV. The data evaluating an association between the antibody responses to EBV and HL is even more limited, with only one prospective study specific to EBV-positive HL. Expanding the number of antigens to be evaluated in epithelial and lymphoid tumors in parallel could help to fill this gap and elucidate the role of the immune response to EBV infection in the development of distinct EBV-associated cancers.

Greater EBV genetic diversity other than a few EBV genes such as *LMP1*, *EBNA1*, and *BZLF1* has been observed [201, 202]. With the development of next-generation sequencing technology, we will be able to sequence the whole genome of EBV in a large sample size. Full characterization of viral diversity will help us to understand the patterns of EBV genetic variation and their associations with cancers in different populations.

There are five phases of a biomarker development for early detection of cancer [203]. For NPC, we lacked longitudinally collected samples and randomized controlled trial design to evaluate the true value of biomarkers in the secondary prevention of NPC. More and more large cohort studies with biobanking have been established recently, at least in China. In the future, we may be able to utilize these samples and overcome the limitations of evaluating biomarker only based on cross-sectional samples.

9 ACKNOWLEDGEMENTS

I would like to express my gratitude to all the participants that agree to contribute their information and samples to our studies, and also to all the field work staff, interviewers, doctors, and nurses who helped us to collect questionnaire and samples. Without their contributions, the study would not have started. I would also like to thank all those who have helped me in different ways during my PhD education. Especially, I would like to thank:

Weimin Ye, my main supervisor, for your always support and trust in my ability since the first day we met in Guangzhou, for your guidance in the field work organization, for your vision in the training programs designed for me, and for showing your great enthusiasm and dedication to science, epidemiology, and biostatistics. I really appreciate the tremendous efforts you have put into my projects, especially during my days in the field work in China when the long-distance communications were difficult. I enjoyed every discussion between us. Your honesty, rigorousness, and encouragement I will never forget.

Hans-Olov Adami, my co-supervisor, for your leadership in the case-control study in southern China, for your vision and profound knowledge in epidemiology, your scientific attitude, for your time in your busy schedule to review my manuscripts, to answer my questions, and to improve my work, and for your guidance on my personal development. It's my honor to have the opportunity to work with you.

Ingemar Ernberg, my co-supervisor, for sharing your profound knowledge in microbiology and NPC, for your always enthusiastic encouragement and support.

Su-Mei Cao, my co-supervisor, for taking me as a summer student in your NPC screening project, for introducing me to the world of epidemiology and to Sweden, for providing me support during my days of field work in China.

Fang Fang, my friend, teacher, and role model in epidemiology, for taking care of me at MEB, for your generous helps, priceless advices, and always brilliant ideas on researches. It is privileged to have you to guide me in Swedish register-based studies.

Ellen T. Chang, also my friend, teacher, and role model in epidemiology, for always insightful comments on the manuscript, tolerance of my poor control over grammar, and editing almost all of my manuscripts. I would like to especially express my gratitude to your encouragements during my days in field work. Every time I want to give up, your support and trust brought me back. All of your inputs in our projects are invaluable and inspiring me constantly.

Allan Hildesheim, for being my friend and mentor in NPC epidemiology, for taking me in the National Cancer Institute, for providing me data and samples, and for your belief in my ability. I enjoyed our discussions on every aspect of NPC. I am also deeply impressed by your stories in Taiwan.

My collaborators China, **Qing Liu, Qi-Hong Huang, Shang-Hang Xie, Jian-Yong Shao, Wei-Hua Jia, Yi-Xin Zeng, Yonglin Cai, Guomin Chen, Yuming Zheng, Jian Liao, Yi Zeng, Zhe Zhang, Yufeng Chen, Longde Lin, Guangwu Huang, and Ming-Fang Ji**, for the contributions on the project.

The affinity proteomics group, **Sanna Byström** and **Jochen M. Schwenk**, for helping me to perform such wonderful laboratory experiments, for being always open-minded, and sharing your experiences in biomarker identification.

Alexander Ploner, for being my statistical teacher, and for sharing your statistical and programming skills.

The NCI and Taiwan group, **Anna E. Coghill, Kelly J. Yu, Ruth M. Pfeiffer, Wan-Lun Hsu, Cheng-Ping Wang**, for your trust in my ability, for providing me priceless data sets, for the discussions, statistical helps, and nice work and research atmosphere.

Weimin and Fang Fang group members at MEB, **Ruoqing Chen, Dariush Nasrollahzadeh Nesheli, Jiaqi Huang, Huan Song, Isabella Ekheden, Tracy Peters, Donal Barrett, Nelson Ndegwa Gichora, Tingting Huang, Daniela Mariosa, Donghao Lu, Elisa Longinetti, Qing Shen, Jianwei Zhu, Amanda Regodon Wallin, Ulrika Zagai, Alessandra Grotta, Amelie Plymoth, Xingdong Chen, Hai-Yun Wang, Weiyao Yin, and Lijie Ding**, for being always supportive.

Dear **other MEB members**, for making our department to be such a nice work atmosphere.

My dearest **mother and father**, for all the support.

Qiushi, for everything. You are my life and wings.

10 REFERENCES

1. Ferlay J, et al., *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from:* http://globocan.iarc.fr/, *accessed on June, 2016*
2. Forman D, et al., *Cancer Incidence in Five Continents, Vol. X (electronic version) Lyon, IARC*. 2013.
3. Shanmugaratnam, K. and L.H. Sobin, *The World Health Organization histological classification of tumours of the upper respiratory tract and ear. A commentary on the second edition.* Cancer, 1993. **71**(8): p. 2689-97.
4. Chan JKC, et al., *Nasopharyngeal carcinoma. World Health Organization Classification of Tumours Pathology & Genetics Head and Neck Tumours*, Barnes L, et al., Editors. 2005, Lyon: International Agency for Research on Cancer (IARC). p. p. 85-97.
5. Liu, T., *Issues in the management of nasopharyngeal carcinoma.* Crit Rev Oncol Hematol, 1999. **31**(1): p. 55-69.
6. Malker, H.S., et al., *Occupational risk factors for nasopharyngeal cancer in Sweden.* Br J Ind Med, 1990. **47**(3): p. 213-4.
7. 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.
8. Marks, J.E., J.L. Phillips, and H.R. Menck, *The National Cancer Data Base report on the relationship of race and national origin to the histology of nasopharyngeal carcinoma.* Cancer, 1998. **83**(3): p. 582-8.
9. Niedobitek, G., *Epstein-Barr virus infection in the pathogenesis of nasopharyngeal carcinoma.* Mol Pathol, 2000. **53**(5): p. 248-54.
10. Raab-Traub, N., *Epstein-Barr virus in the pathogenesis of NPC.* Semin Cancer Biol, 2002. **12**(6): p. 431-41.
11. 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.
12. Chiang, C.J., et al., *Incidence and survival of adult cancer patients in Taiwan, 2002-2012.* J Formos Med Assoc, 2016.
13. 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.
14. 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.
15. 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.
16. Tse, L.A., et al., *Incidence rate trends of histological subtypes of nasopharyngeal carcinoma in Hong Kong.* Br J Cancer, 2006. **95**(9): p. 1269-73.
17. Chang, E.T. and H.O. Adami, *The enigmatic epidemiology of nasopharyngeal carcinoma.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(10): p. 1765-77.
18. Bray, F., et al., *Age-incidence curves of nasopharyngeal carcinoma worldwide: bimodality in low-risk populations and aetiologic implications.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(9): p. 2356-65.
19. Lo, K.W., K.F. To, and D.P. Huang, *Focus on nasopharyngeal carcinoma.* Cancer Cell, 2004. **5**(5): p. 423-8.
20. Tang, L.Q., et al., *Establishment and Validation of Prognostic Nomograms for Endemic Nasopharyngeal Carcinoma.* J Natl Cancer Inst, 2016. **108**(1).
21. Chua, D.T., et al., *Treatment outcome after radiotherapy alone for patients with Stage I-II nasopharyngeal carcinoma.* Cancer, 2003. **98**(1): p. 74-80.
22. Al-Sarraf, M., et al., *Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: phase III randomized Intergroup study 0099.* J Clin Oncol, 1998. **16**(4): p. 1310-7.
23. Jackson, C., *PRimary carcinoma of the nasopharynx. a table of cases.* Journal of the American Medical Association, 1901. **XXXVII**(6): p. 371-377.

24. Mousavi, S.M., J. Sundquist, and K. Hemminki, *Nasopharyngeal and hypopharyngeal carcinoma risk among immigrants in Sweden.* Int J Cancer, 2010. **127**(12): p. 2888-92.
25. Buell, P., *The effect of migration on the risk of nasopharyngeal cancer among Chinese.* Cancer Res, 1974. **34**(5): p. 1189-91.
26. Parkin, D.M. and J. Iscovich, *Risk of cancer in migrants and their descendants in Israel: II. Carcinomas and germ-cell tumours.* Int J Cancer, 1997. **70**(6): p. 654-60.
27. Grulich, A.E., M. McCredie, and M. Coates, *Cancer incidence in Asian migrants to New South Wales, Australia.* Br J Cancer, 1995. **71**(2): p. 400-8.
28. McCredie, M., S. Williams, and M. Coates, *Cancer mortality in East and Southeast Asian migrants to New South Wales, Australia, 1975-1995.* Br J Cancer, 1999. **79**(7-8): p. 1277-82.
29. Warnakulasuriya, K.A., et al., *Cancer of mouth, pharynx and nasopharynx in Asian and Chinese immigrants resident in Thames regions.* Oral Oncol, 1999. **35**(5): p. 471-5.
30. Jeannel, D., et al., *Increased risk of nasopharyngeal carcinoma among males of French origin born in Maghreb (north Africa).* Int J Cancer, 1993. **54**(4): p. 536-9.
31. Yu, M.C. and J.M. Yuan, *Epidemiology of nasopharyngeal carcinoma.* Semin Cancer Biol, 2002. **12**(6): p. 421-9.
32. Xie, S.H., et al., *Sex difference in the incidence of nasopharyngeal carcinoma in Hong Kong 1983-2008: Suggestion of a potential protective role of oestrogen.* Eur J Cancer, 2012.
33. Jia, W.H., et al., *Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China.* Cancer, 2004. **101**(2): p. 363-9.
34. Friborg, J., et al., *Cancer susceptibility in nasopharyngeal carcinoma families--a population-based cohort study.* Cancer Res, 2005. **65**(18): p. 8567-72.
35. Hsu, W.L., et al., *Familial tendency and risk of nasopharyngeal carcinoma in taiwan: effects of covariates on risk.* Am J Epidemiol, 2011. **173**(3): p. 292-9.
36. Liu, Z., et al., *Cancer risk in the relatives of patients with nasopharyngeal carcinoma-a register-based cohort study in Sweden.* Br J Cancer, 2015. **112**(11): p. 1827-31.
37. Yu, M.C., *Diet and nasopharyngeal carcinoma.* Prog Clin Biol Res, 1990. **346**: p. 93-105.
38. 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.
39. Liu, Y.T., et al., *Greater intake of fruit and vegetables is associated with lower risk of nasopharyngeal carcinoma in Chinese adults: a case-control study.* Cancer Causes Control, 2012. **23**(4): p. 589-99.
40. 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.
41. Hildesheim, A., et al., *Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan.* J Natl Cancer Inst, 2002. **94**(23): p. 1780-9.
42. 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.
43. Su, W.H., A. Hildesheim, and Y.S. Chang, *Human leukocyte antigens and epstein-barr virus-associated nasopharyngeal carcinoma: old associations offer new clues into the role of immunity in infection-associated cancers.* Front Oncol, 2013. **3**: p. 299.
44. Li, X., et al., *HLA associations with nasopharyngeal carcinoma.* Curr Mol Med, 2009. **9**(6): p. 751-65.
45. Levine, R., et al., *Self-reported infectious mononucleosis and 6 cancers: a population-based, case-control study.* Scand J Infect Dis, 1998. **30**(3): p. 211-4.
46. Melbye, M., et al., *Early primary infection and high Epstein-Barr virus antibody titers in Greenland Eskimos at high risk for nasopharyngeal carcinoma.* Int J Cancer, 1984. **34**(5): p. 619-23.
47. 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.
48. 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.
49. Hung, S.H., et al., *Association of rhinosinusitis with nasopharyngeal carcinoma: a population-based study.* Laryngoscope, 2014. **124**(7): p. 1515-20.

50. Lin, K.T., et al., *Subsequent risk of nasopharyngeal carcinoma among patients with allergic rhinitis: a nationwide population-based cohort study.* Head Neck, 2015. **37**(3): p. 413-7.
51. Hildesheim, A. and C.P. Wang, *Genetic predisposition factors and nasopharyngeal carcinoma risk: a review of epidemiological association studies, 2000-2011: Rosetta Stone for NPC: genetics, viral infection, and other environmental factors.* Semin Cancer Biol, 2012. **22**(2): p. 107-16.
52. Jia, W.H. and H.D. Qin, *Non-viral environmental risk factors for nasopharyngeal carcinoma: a systematic review.* Semin Cancer Biol, 2012. **22**(2): p. 117-26.
53. Fields, B.N., D.M.D.M.-. Knipe, and P.M. Howley, *Fields virology / editors-in-chief, David M. Knipe, Peter M. Howley ; associate editors, Jeffrey I. Cohen ... [et al.].* 6th ed. ed.: Philadelphia : Wolters Kluwer Health/Lippincott Williams & Wilkins, c2013.
54. Young, L.S. and A.B. Rickinson, *Epstein-Barr virus: 40 years on.* Nat Rev Cancer, 2004. **4**(10): p. 757-68.
55. 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.
56. Kangro, H.O., et al., *Seroprevalence of antibodies to human herpesviruses in England and Hong Kong.* J Med Virol, 1994. **43**(1): p. 91-6.
57. Chen, C.Y., et al., *A large-scale seroprevalence of Epstein-Barr virus in Taiwan.* PLoS One, 2015. **10**(1): p. e0115836.
58. 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.
59. Saghafian-Hedengren, S., et al., *Early-life EBV infection protects against persistent IgE sensitization.* J Allergy Clin Immunol, 2010. **125**(2): p. 433-8.
60. Dowd, J.B., et al., *Seroprevalence of Epstein-Barr virus infection in U.S. children ages 6-19, 2003-2010.* PLoS One, 2013. **8**(5): p. e64921.
61. Rickinson, A.B., *Co-infections, inflammation and oncogenesis: future directions for EBV research.* Semin Cancer Biol, 2014. **26**: p. 99-115.
62. Khan, G. and M.J. Hashim, *Global burden of deaths from Epstein-Barr virus attributable malignancies 1990-2010.* Infect Agent Cancer, 2014. **9**(1): p. 38.
63. Old, L.J., et al., *Precipitating antibody in human serum to an antigen present in cultured burkitt's lymphoma cells.* Proc Natl Acad Sci U S A, 1966. **56**(6): p. 1699-704.
64. Nonoyama, M., et al., *DNA of Epstein-Barr virus detected in tissue of Burkitt's lymphoma and nasopharyngeal carcinoma.* Proc Natl Acad Sci U S A, 1973. **70**(11): p. 3265-8.
65. Raab-Traub, N. and K. Flynn, *The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation.* Cell, 1986. **47**(6): p. 883-9.
66. Niedobitek, G., et al., *Epstein-Barr virus and carcinomas: undifferentiated carcinomas but not squamous cell carcinomas of the nasopharynx are regularly associated with the virus.* J Pathol, 1991. **165**(1): p. 17-24.
67. 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.
68. Chien, Y.C., et al., *Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men.* N Engl J Med, 2001. **345**(26): p. 1877-82.
69. Cao, S.M., et al., *Fluctuations of epstein-barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up.* PLoS One, 2011. **6**(4): p. e19100.
70. Coghill, A.E., et al., *High Levels of Antibody that Neutralize B-cell Infection of Epstein-Barr Virus and that Bind EBV gp350 Are Associated with a Lower Risk of Nasopharyngeal Carcinoma.* Clin Cancer Res, 2016.
71. 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.
72. Li, Q.X., et al., *Epstein-Barr virus infection and replication in a human epithelial cell system.* Nature, 1992. **356**(6367): p. 347-50.
73. Chesnokova, L.S. and L.M. Hutt-Fletcher, *Epstein-Barr virus infection mechanisms.* Chin J Cancer, 2014. **33**(11): p. 545-8.
74. Hutt-Fletcher, L.M., *Epstein-Barr virus entry.* J Virol, 2007. **81**(15): p. 7825-32.

75. Tsao, S.W., et al., *The biology of EBV infection in human epithelial cells.* Semin Cancer Biol, 2012. **22**(2): p. 137-43.
76. Gan, Y.J., et al., *Epithelial cell polarization is a determinant in the infectious outcome of immunoglobulin A-mediated entry by Epstein-Barr virus.* J Virol, 1997. **71**(1): p. 519-26.
77. Chang, Y., et al., *Requirement for cell-to-cell contact in Epstein-Barr virus infection of nasopharyngeal carcinoma cells and keratinocytes.* J Virol, 1999. **73**(10): p. 8857-66.
78. Imai, S., J. Nishikawa, and K. Takada, *Cell-to-cell contact as an efficient mode of Epstein-Barr virus infection of diverse human epithelial cells.* J Virol, 1998. **72**(5): p. 4371-8.
79. Sixbey, J.W. and Q.Y. Yao, *Immunoglobulin A-induced shift of Epstein-Barr virus tissue tropism.* Science, 1992. **255**(5051): p. 1578-80.
80. Fleisher, G., et al., *Primary infection with Epstein-Barr virus in infants in the United States: clinical and serologic observations.* J Infect Dis, 1979. **139**(5): p. 553-8.
81. Hjalgrim, H., et al., *Characteristics of Hodgkin's lymphoma after infectious mononucleosis.* N Engl J Med, 2003. **349**(14): p. 1324-32.
82. Callan, M.F., et al., *Large clonal expansions of CD8+ T cells in acute infectious mononucleosis.* Nat Med, 1996. **2**(8): p. 906-11.
83. Callan, M.F., et al., *Direct visualization of antigen-specific CD8+ T cells during the primary immune response to Epstein-Barr virus In vivo.* J Exp Med, 1998. **187**(9): p. 1395-402.
84. 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.
85. Liu, Z., et al., *Sibship size, birth order and risk of nasopharyngeal carcinoma and infectious mononucleosis: a nationwide study in Sweden.* Int J Epidemiol, 2015.
86. Rostgaard, K., et al., *Sibship structure and risk of infectious mononucleosis: a population-based cohort study.* Int J Epidemiol, 2014. **43**(5): p. 1607-14.
87. Chang, E.T., et al., *Number of siblings and risk of Hodgkin's lymphoma.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(7): p. 1236-43.
88. Chatenoud, L., et al., *Number of siblings and risk of hodgkin's and other lymphoid neoplasms.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(2): p. 552.
89. Hsieh, C.C., et al., *Age at first establishment of chronic hepatitis B virus infection and hepatocellular carcinoma risk. A birth order study.* Am J Epidemiol, 1992. **136**(9): p. 1115-21.
90. Kuper, H., et al., *Birth order, as a proxy for age at infection, in the etiology of hepatocellular carcinoma.* Epidemiology, 2000. **11**(6): p. 680-3.
91. Wu, B., et al., *Social stratification and tooth loss among middle-aged and older Americans from 1988 to 2004.* Community Dent Oral Epidemiol, 2014.
92. Beck, J.D., et al., *Prevalence and risk indicators for periodontal attachment loss in a population of older community-dwelling blacks and whites.* J Periodontol, 1990. **61**(8): p. 521-8.
93. Axell, T., *A prevalence study of oral mucosal lesions in an adult Swedish population.* Odontol Revy Suppl, 1976. **36**: p. 1-103.
94. Griffin, S.O., et al., *Burden of oral disease among older adults and implications for public health priorities.* Am J Public Health, 2012. **102**(3): p. 411-8.
95. Abnet, C.C., et al., *Prospective study of tooth loss and incident esophageal and gastric cancers in China.* Cancer Causes Control, 2001. **12**(9): p. 847-54.
96. Stolzenberg-Solomon, R.Z., et al., *Tooth loss, pancreatic cancer, and Helicobacter pylori.* Am J Clin Nutr, 2003. **78**(1): p. 176-81.
97. Abnet, C.C., et al., *Tooth loss is associated with increased risk of gastric non-cardia adenocarcinoma in a cohort of Finnish smokers.* Scand J Gastroenterol, 2005. **40**(6): p. 681-7.
98. Guha, N., et al., *Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies.* Am J Epidemiol, 2007. **166**(10): p. 1159-73.
99. Abnet, C.C., et al., *Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(11): p. 3062-8.
100. Meyer, M.S., et al., *A review of the relationship between tooth loss, periodontal disease, and cancer.* Cancer Causes & Control, 2008. **19**(9): p. 895-907.

101. Fitzpatrick, S.G. and J. Katz, *The association between periodontal disease and cancer: A review of the literature.* Journal of Dentistry, 2010. **38**(2): p. 83-95.
102. Nasrollahzadeh, D., et al., *Gastric atrophy and oesophageal squamous cell carcinoma: possible interaction with dental health and oral hygiene habit.* Br J Cancer, 2012. **107**(5): p. 888-94.
103. 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.
104. Huang, J., et al., *A prospective cohort study on poor oral hygiene and pancreatic cancer risk.* Int J Cancer, 2016. **138**(2): p. 340-7.
105. Scannapieco, F.A., *Periodontal inflammation: from gingivitis to systemic disease?* Compend Contin Educ Dent, 2004. **25**(7 Suppl 1): p. 16-25.
106. Pihlstrom, B.L., B.S. Michalowicz, and N.W. Johnson, *Periodontal diseases.* Lancet, 2005. **366**(9499): p. 1809-20.
107. Karin, M., T. Lawrence, and V. Nizet, *Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer.* Cell, 2006. **124**(4): p. 823-35.
108. Coussens, L.M. and Z. Werb, *Inflammation and cancer.* Nature, 2002. **420**(6917): p. 860-7.
109. Farrell, J.J., et al., *Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer.* Gut, 2012. **61**(4): p. 582-8.
110. Santangelo, R., et al., *Bacterial and viral DNA in periodontal disease: a study using multiplex PCR.* New Microbiol, 2004. **27**(2): p. 133-7.
111. Meurman, J.H. and J. Uittamo, *Oral micro-organisms in the etiology of cancer.* Acta Odontol Scand, 2008. **66**(6): p. 321-6.
112. 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.
113. Slots, J., et al., *Epstein-Barr virus in oral diseases.* J Periodontal Res, 2006. **41**(4): p. 235-44.
114. *Study finds link between oral inflammatory diseases and Epstein-Barr virus.* J Am Dent Assoc, 2009. **140**(2): p. 150-2.
115. Simonian, K., *The role of herpesviruses in periodontal disease.* J West Soc Periodontol Periodontal Abstr, 2003. **51**(1): p. 5-9.
116. Wang-Johanning, F., et al., *Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer.* Int J Cancer, 2013. **134**(3): p. 587-95.
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. Turkoz, F.P., et al., *Risk factors of nasopharyngeal carcinoma in Turkey-an epidemiological survey of the Anatolian Society of Medical Oncology.* Asian Pac J Cancer Prev, 2011. **12**(11): p. 3017-21.
119. 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.
120. Ng, W.T., et al., *Familial nasopharyngeal carcinoma in Hong Kong: epidemiology and implication in screening.* Fam Cancer, 2009. **8**(2): p. 103-8.
121. Yu, K.J., et al., *Cancer patterns in nasopharyngeal carcinoma multiplex families in Taiwan.* Int J Cancer, 2009. **124**(7): p. 1622-5.
122. Ren, Z.F., et al., *Effect of family history of cancers and environmental factors on risk of nasopharyngeal carcinoma in Guangdong, China.* Cancer Epidemiol, 2010. **34**(4): p. 419-24.
123. Yuan, J.M., et al., *Non-dietary risk factors for nasopharyngeal carcinoma in Shanghai, China.* Int J Cancer, 2000. **85**(3): p. 364-9.
124. 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.
125. Liu, Z., et al., *Cancer risk in the relatives of patients with nasopharyngeal carcinoma-a register-based cohort study in Sweden.* Br J Cancer, 2015.
126. Olajos, J., et al., *Familial clustering of nasopharyngeal carcinoma in a non-endemic geographical region. Report of two Hungarian cases and a review of the literature.* Acta Otolaryngol, 2005. **125**(9): p. 1008-13.
127. 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.

128. Liu, Z., et al., *Two Epstein-Barr virus-related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China.* Am J Epidemiol, 2013. **177**(3): p. 242-50.
129. Chen, C.J., et al., *Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency.* Anticancer Res, 1990. **10**(2B): p. 547-53.
130. Ung, A., et al., *Familial and sporadic cases of nasopharyngeal carcinoma in Taiwan.* Anticancer Res, 1999. **19**(1B): p. 661-5.
131. Xie, S.H., et al., *Tobacco smoking, family history, and the risk of nasopharyngeal carcinoma: a case-referent study in Hong Kong Chinese.* Cancer Causes Control, 2015. **26**(6): p. 913-21.
132. Chen, D.L. and T.B. Huang, *A case-control study of risk factors of nasopharyngeal carcinoma.* Cancer Lett, 1997. **117**(1): p. 17-22.
133. 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.
134. Zou, J., et al., *A case-control study of nasopharyngeal carcinoma in the high background radiation areas of Yangjiang, China.* J Radiat Res, 2000. **41 Suppl**: p. 53-62.
135. Zheng, X., et al., *Epstein-Barr virus infection, salted fish and nasopharyngeal carcinoma. A case-control study in southern China.* Acta Oncol, 1994. **33**(8): p. 867-72.
136. Goldgar, D.E., et al., *Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands.* J Natl Cancer Inst, 1994. **86**(21): p. 1600-8.
137. Jia, W.H., et al., *Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China).* Eur J Hum Genet, 2005. **13**(2): p. 248-52.
138. Cao, S.M., et al., *[Clinical analysis of 1,142 hospitalized cantonese patients with nasopharyngeal carcinoma].* Ai Zheng, 2006. **25**(2): p. 204-8.
139. Albeck, H., et al., *Familial clusters of nasopharyngeal carcinoma and salivary gland carcinomas in Greenland natives.* Cancer, 1993. **72**(1): p. 196-200.
140. Zeng, Y.X. and W.H. Jia, *Familial nasopharyngeal carcinoma.* Semin Cancer Biol, 2002. **12**(6): p. 443-50.
141. Cao, S.M., et al., *Familial nasopharyngeal carcinomas possess distinguished clinical characteristics in southern China.* Chin J Cancer Res, 2014. **26**(5): p. 543-9.
142. Ouyang, P.Y., et al., *Prognostic impact of family history in southern Chinese patients with undifferentiated nasopharyngeal carcinoma.* Br J Cancer, 2013. **109**(3): p. 788-94.
143. He, Y.Q., et al., *Household inhalants exposure and nasopharyngeal carcinoma risk: a large-scale case-control study in Guangdong, China.* BMC Cancer, 2015. **15**: p. 1022.
144. 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.
145. Zeng, Y., et al., *Serological mass survey for early detection of nasopharyngeal carcinoma in Wuzhou City, China.* Int J Cancer, 1982. **29**(2): p. 139-41.
146. Zeng, Y., et al., *Follow-up studies on Epstein-Barr virus IgA/VCA antibody-positive persons in Zangwu County, China.* Intervirology, 1983. **20**(4): p. 190-4.
147. Cao, S.M., M.J. Simons, and C.N. Qian, *The prevalence and prevention of nasopharyngeal carcinoma in China.* Chin J Cancer, 2011. **30**(2): p. 114-9.
148. Coghill, A.E., et al., *Epstein-Barr virus serology as a potential screening marker for nasopharyngeal carcinoma among high-risk individuals from multiplex families in Taiwan.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(7): p. 1213-9.
149. Ji, M.F., et al., *Sustained elevation of Epstein-Barr virus antibody levels preceding clinical onset of nasopharyngeal carcinoma.* Br J Cancer, 2007. **96**(4): p. 623-30.
150. Liu, Y., et al., *Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China.* Int J Cancer, 2012. **131**(2): p. 406-16.
151. Coskun, O., et al., *Stress-related Epstein-Barr virus reactivation.* Clin Exp Med, 2010. **10**(1): p. 15-20.
152. Stowe, R.P., D.L. Pierson, and A.D. Barrett, *Elevated stress hormone levels relate to Epstein-Barr virus reactivation in astronauts.* Psychosom Med, 2001. **63**(6): p. 891-5.

153. Xiao, L., et al., *Biomarker discovery of nasopharyngeal carcinoma by proteomics.* Expert Rev Proteomics, 2014.
154. Hanash, S.M., S.J. Pitteri, and V.M. Faca, *Mining the plasma proteome for cancer biomarkers.* Nature, 2008. **452**(7187): p. 571-9.
155. Schwenk, J.M., et al., *Antibody suspension bead arrays within serum proteomics.* J Proteome Res, 2008. **7**(8): p. 3168-79.
156. Schwenk, J.M., et al., *Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays.* Mol Cell Proteomics, 2010. **9**(11): p. 2497-507.
157. Statistics Sweden, *Multi-Generation Register, 2009. A description of contents and quality. Örebro, Sweden: Statistics Sweden*. 2009.
158. Ekbom, A., *The Swedish Multi-generation Register.* Methods Mol Biol, 2011. **675**: p. 215-20.
159. Barlow, L., et al., *The completeness of the Swedish Cancer Register: a sample survey for year 1998.* Acta Oncol, 2009. **48**(1): p. 27-33.
160. Ludvigsson, J.F., et al., *External review and validation of the Swedish national inpatient register.* BMC Public Health, 2011. **11**: p. 450.
161. Verhage, M., *Population and Housing censuses based on a dwelling register and registration on dwellings.* Statistics Sweden, 2010.
162. Fang, F., et al., *Maternal age, exposure to siblings, and risk of amyotrophic lateral sclerosis.* Am J Epidemiol, 2008. **167**(11): p. 1281-6.
163. Richardson, D.B., *An incidence density sampling program for nested case-control analyses.* Occup Environ Med, 2004. **61**(12): p. e59.
164. Uhlen, M., et al., *Towards a knowledge-based Human Protein Atlas.* Nat Biotechnol, 2010. **28**(12): p. 1248-50.
165. Wu, C.C., et al., *Cancer cell-secreted proteomes as a basis for searching potential tumor markers: nasopharyngeal carcinoma as a model.* Proteomics, 2005. **5**(12): p. 3173-82.
166. Yan, G.G., et al., *Identification of novel phosphoproteins in signaling pathways triggered by latent membrane protein 1 using functional proteomics technology.* Proteomics, 2006. **6**(6): p. 1810-1821.
167. Li, F., et al., *A reference map of human nasopharyngeal squamous carcinoma proteome.* Int J Oncol, 2007. **30**(5): p. 1077-88.
168. Cheng, A.L., et al., *Identification of novel nasopharyngeal carcinoma biomarkers by laser capture microdissection and proteomic analysis.* Clin Cancer Res, 2008. **14**(2): p. 435-45.
169. Tong, Y.Q., et al., *BMI-1 autoantibody in serum as a new potential biomarker of nasopharyngeal carcinoma.* Cancer Biol Ther, 2008. **7**(3): p. 340-4.
170. Tong, Y.Q., et al., *Autoantibodies as potential biomarkers for nasopharyngeal carcinoma.* Proteomics, 2008. **8**(15): p. 3185-93.
171. Zhang, L., et al., *Dataset of the plasma membrane proteome of nasopharyngeal carcinoma cell line HNE1 for uncovering protein function.* Acta Biochim Biophys Sin (Shanghai), 2008. **40**(1): p. 55-70.
172. Klibi, J., et al., *Blood diffusion and Th1-suppressive effects of galectin-9-containing exosomes released by Epstein-Barr virus-infected nasopharyngeal carcinoma cells.* Blood, 2009. **113**(9): p. 1957-66.
173. Wu, H.Y., et al., *Proteomics analysis of nasopharyngeal carcinoma cell secretome using a hollow fiber culture system and mass spectrometry.* J Proteome Res, 2009. **8**(1): p. 380-9.
174. Chang, K.P., et al., *Identification of candidate nasopharyngeal carcinoma serum biomarkers by cancer cell secretome and tissue transcriptome analysis: potential usage of cystatin A for predicting nodal stage and poor prognosis.* Proteomics, 2010. **10**(14): p. 2644-60.
175. Ruan, L., et al., *Analysis of EGFR signaling pathway in nasopharyngeal carcinoma cells by quantitative phosphoproteomics.* Proteome Sci, 2011. **9**: p. 35.
176. Barjon, C., et al., *A novel monoclonal antibody for detection of galectin-9 in tissue sections : application to human tissues infected by oncogenic viruses.* Infect Agent Cancer, 2012. **7**(1): p. 16.
177. Liu, Z., et al., *Proteomic features of potential tumor suppressor NESG1 in nasopharyngeal carcinoma.* Proteomics, 2012. **12**(22): p. 3416-25.
178. Parsonage, G., et al., *CXCR6 and CCR5 localize T lymphocyte subsets in nasopharyngeal carcinoma.* Am J Pathol, 2012. **180**(3): p. 1215-22.

179. Tse, K.P., et al., *The relationship between secretory leukocyte protease inhibitor expression and Epstein-Barr virus status among patients with nasopharyngeal carcinoma.* Anticancer Res, 2012. **32**(4): p. 1299-307.
180. Zhang, Y.W., et al., *Role of an MDM4 polymorphism in the early age of onset of nasopharyngeal carcinoma.* Oncol Lett, 2012. **3**(5): p. 1115-1118.
181. Hu, C., et al., *A global view of the oncogenic landscape in nasopharyngeal carcinoma: an integrated analysis at the genetic and expression levels.* PLoS One, 2012. **7**(7): p. e41055.
182. Byström, S., et al., *Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis.*, in *Journal of proteome research*. 2014. p. 4607-4619.
183. Haggmark, A., et al., *Antibody-based profiling of cerebrospinal fluid within multiple sclerosis.* Proteomics, 2013. **13**(15): p. 2256-67.
184. Bruzzi, P., et al., *Estimating the population attributable risk for multiple risk factors using case-control data.* Am J Epidemiol, 1985. **122**(5): p. 904-14.
185. Lin, D.Y., *Cox regression analysis of multivariate failure time data: the marginal approach.* Stat Med, 1994. **13**(21): p. 2233-47.
186. Grambsch, P.M. and T.M. Therneau, *Proportional Hazards Tests and Diagnostics Based on Weighted Residuals.* Biometrika, 1994. **81**(3): p. 515-526.
187. Hubert, M., P.J. Rousseeuw, and K.V. Branden, *ROBPCA: A new approach to robust principal component analysis.* Technometrics, 2005. **47**(1): p. 64−79.
188. Dieterle, F., et al., *Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics.* Anal Chem, 2006. **78**(13): p. 4281-90.
189. Bystrom, S., et al., *Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis.* J Proteome Res, 2014. **13**(11): p. 4607-19.
190. Hildesheim, A. and P.H. Levine, *Etiology of nasopharyngeal carcinoma: a review.* Epidemiol Rev, 1993. **15**(2): p. 466-85.
191. Bach, J.F., *The effect of infections on susceptibility to autoimmune and allergic diseases.* N Engl J Med, 2002. **347**(12): p. 911-20.
192. Oddy, W.H., et al., *Breast feeding and cognitive development in childhood: a prospective birth cohort study.* Paediatr Perinat Epidemiol, 2003. **17**(1): p. 81-90.
193. Pande, R.P., *Selective gender differences in childhood nutrition and immunization in rural India: the role of siblings.* Demography, 2003. **40**(3): p. 395-418.
194. Hjalgrim, H., et al., *Infectious mononucleosis, childhood social environment, and risk of Hodgkin lymphoma.* Cancer Res, 2007. **67**(5): p. 2382-8.
195. Deng, L., et al., *Cells in G2/M phase increased in human nasopharyngeal carcinoma cell line by EBV-LMP1 through activation of NF-kappaB and AP-1.* Cell Res, 2003. **13**(3): p. 187-94.
196. Hui, E.P., et al., *Coexpression of hypoxia-inducible factors 1alpha and 2alpha, carbonic anhydrase IX, and vascular endothelial growth factor in nasopharyngeal carcinoma and relationship to survival.* Clin Cancer Res, 2002. **8**(8): p. 2595-604.
197. Cheng, Y., et al., *Anti-angiogenic pathway associations of the 3p21.3 mapped BLU gene in nasopharyngeal carcinoma.* Oncogene, 2015. **34**(32): p. 4219-28.
198. 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.
199. McMillan, A.S., et al., *Oral health-related quality of life in southern Chinese following radiotherapy for nasopharyngeal carcinoma.* J Oral Rehabil, 2004. **31**(6): p. 600-8.
200. Coghill, A.E. and A. Hildesheim, *Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature.* Am J Epidemiol, 2014. **180**(7): p. 687-95.
201. Kwok, H., et al., *Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy.* PLoS One, 2012. **7**(5): p. e36939.
202. Kwok, H., et al., *Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples.* J Virol, 2014. **88**(18): p. 10662-72.
203. Pepe, M.S., et al., *Phases of biomarker development for early detection of cancer.* J Natl Cancer Inst, 2001. **93**(14): p. 1054-61.