A Genetic Epidemiological Study of Nasopharyngeal Carcinoma

Bingjian Feng

Acknowledgements

The works presented in this PhD thesis were performed in the Genetic Epidemiology Unit of International Agency for Research on Cancer (IARC), Lyon, France, and in the Cancer Center of Sun Yat-Sen University (CCSYSU), Guangzhou, China.

The North African case-control study performed in IARC was supported by the Association for International Cancer Research (grant number 03-252). The genetic studies performed in CCSYSU were supported by the National Outstanding Youth Grant of China, the China Medical Board (CMB 98-678), the Key Projects of National Natural Science Foundation, The National High Technology Research and Development Program of China (863 Program), the National Key Basic Research Project of China (973 Program; G19980510), the Foundation of Guangdong Science and Technology Committee (98001-4, 980950), the Chinese National Natural Science Foundation (30000141), the Chinese National Science and Technique Project (2002BA711A03), and the Shanghai Municipal Foundation for Sciences and Technologies.

Erasmus University Rotterdam and the Department of Epidemiology & Biostatistics provided financial support for the publication of this thesis.

Photos on the cover and title pages of each chapter: "Along the river during Qing-Ming Festival", by Zhang ZeDuan (1085-1145 AD). Handscroll, ink and color on silk, 24.8 X 528.7 cm. Palace Museum, Beijing, China.
Printed by: Optima Grafische Communicatie, Rotterdam, the Netherlands
ISBN: 978-90-8559-329-4

A Genetic Epidemiological Study of Nasopharyngeal Carcinoma

Een Genetische Epidemiologische Studie van Nasopharyngeal Carcinoom

Proefschrift

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 29 November 2007 om 13.30 uur

door

Bingjian feng

geboren te Guangzhou, China

Promotiecommissie

Promotoren:	prof.dr. D.E. Goldgar prof.dr.ir. C.M. van Duijn
Overige leden:	prof.dr. P. Busson prof.dr. J.W.W. Coebergh prof.dr. B.A. Oostra

献给我敬爱的父母

“It is the mark of an educated mind to be able to entertain a thought without accepting it.”

Aristotle (384 BC – 322 BC)

Chapter 2 and 3 are based on the following papers and manuscripts

Chapter 2.1
Feng BJ, Jalbout M, Ben Ayoub W, Khyatti M, Dahmoul S, Ayad M, Maachi F, Bedadra W, Abdoun M, Mesli S, Hamdi-Cherif M, Boualga K, Bouaouina N, Chouchane L, Benider A, Ben Ayed F, Goldgar D, Corbex M*. Dietary Risk factors for Nasopharyngeal Carcinoma in Maghrebian Countries. *Int J Cancer* 2007 Jun 20; 2007; 121(7):1550-5.

Chapter 2.2
Feng BJ*, Khyatti M, Ben Ayoub W, Dahmoul S, Ayad M, Maachi F, Bedadra W, Abdoun M, Mesli S, Bakkali H, Jalbout M, Hamdi-Cherif M, Boualga K, Bouaouina N, Chouchane L, Benider A, Ben Ayed F, Goldgar D, Corbex M. Cannabis Smoking and Domestic Fume Intake are associated with Nasopharyngeal Carcinoma in North Africa. (Submitted to Cancer Epidemiology Biomarkers & Prevention).

Chapter 3.1
Jia WH, Feng BJ, Xu ZL, Zhang XS, Huang P, Huang LX, Yu XJ, Feng QS, Yao MH, Shugart YY, Zeng YX*. Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China. *Cancer* 2004 Jul 15;101(2):363-9.

Chapter 3.2
Feng BJ, Huang W, Shugart YY, Lee MK, Zhang F, Xia JC, Wang HY, Huang TB, Jian SW, Huang P, Feng QS, Huang LX, Yu XJ, Li D, Chen LZ, Jia WH, Fang Y, Huang HM, Zhu JL, Liu XM, Zhao Y, Liu WQ, Deng MQ, Hu WH, Wu SX, Mo HY, Hong MF, King MC, Chen Z, Zeng YX*. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat Genet* 2002 Aug;31(4):395-9.

Chapter 3.3
Chen HK, Feng BJ, Liang H, Zhang RH, Zeng YX*. The susceptibility gene for familial nasopharyngeal carcinoma is mapped on chromosome 4p11-p14 by haplotype analyses. *Chinese Science Bulletin* 2003 Nov;48(21):2327-30.

Chapter 3.4
Feng BJ*, Goldgar DE & Corbex M. TrendTDT - A Transmission Disequilibrium based association Test on Functional Mini/Microsatellites. *BMC genetics* 2007 (in press).

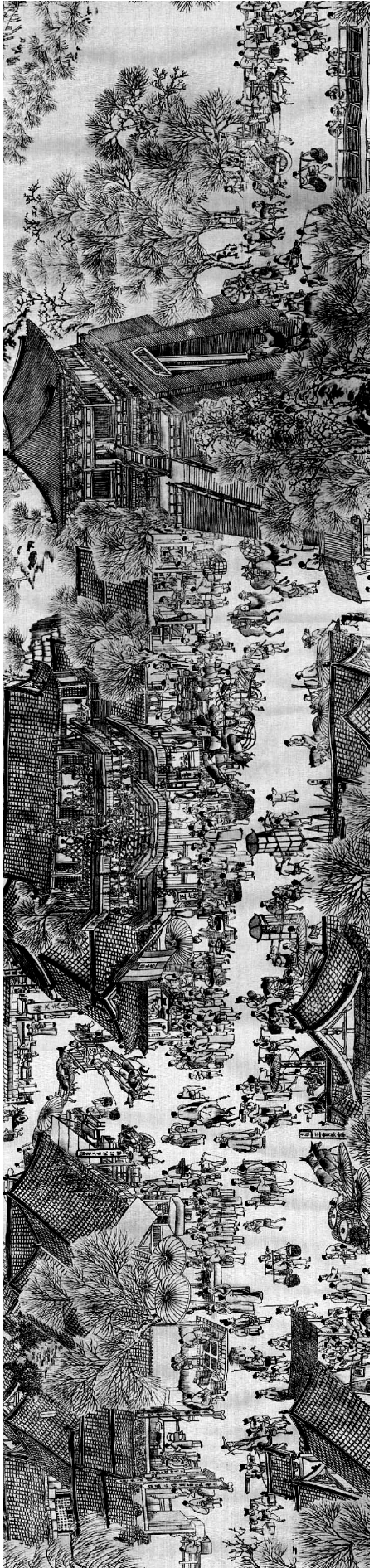
(* Corresponding authors)

Contents

page

Chapter 1

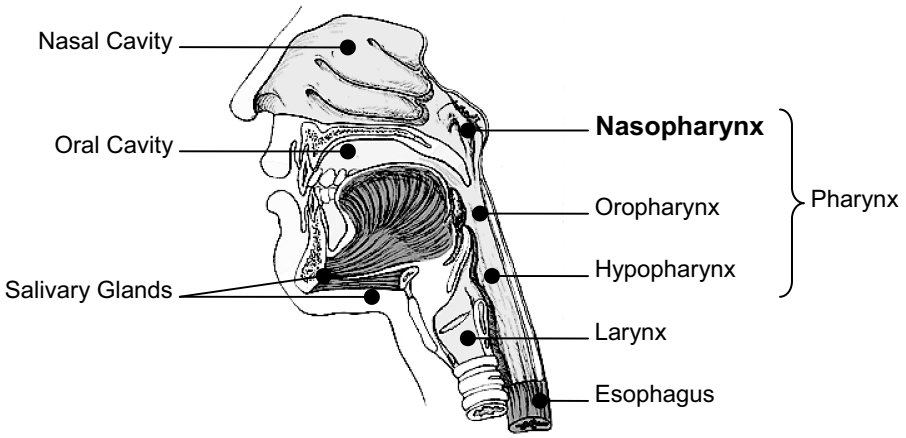
Introduction



Introduction

Nasopharyngeal carcinoma (NPC) is a malignant tumor arising from the epithelial lining of the nasopharynx (**Figure 1**). It is rare in most parts of the world, but much more common in South East Asia, North Africa and Greenland. Endemic ethnic groups include Cantonese Chinese, natives of South East Asia, Arabs living in North Africa, and Eskimos Inuit native Americans[1,2]. The first report of the high incidence rate of NPC ($30/10^5$/year for males and $13/10^5$/year for females) came from Hong Kong, where the majority of the population is Cantonese[3]. In Malaysia, in addition to the ethnic Chinese who demonstrate intermediate risk (16.5 and 7.2 per 10^5 per year)[4], an indigenous population, the Bidayuh, shows the highest incidence rate recorded hitherto ($32/10^5$ for males and $12/10^5$ for females per year)[1]. People from the countries of North Africa also indicate high risk of NPC; these include Algeria, Morocco and Tunisia. In Sétif, Algeria, the incidence rates were $8/10^5$/year for males and $3/10^5$/year for females[5] in 1990-1993, while in Morocco and Tunisia, they may be slightly lower (The age and year adjusted incidence per 100,000 person years in Tunisia is 7.5 for men and 3.3 for women)[6]. In contrast, the incidence for most other countries is less than $0.5/10^5$/year (**Figure 2**). In Asian high-risk populations, the incidence rises in adolescence and peaks at 45-55 years, while in North Africa, the Maghrebian population is characterized by a bimodal age distribution, with one peak occurring in the teens and the other at age 40-50 [7].

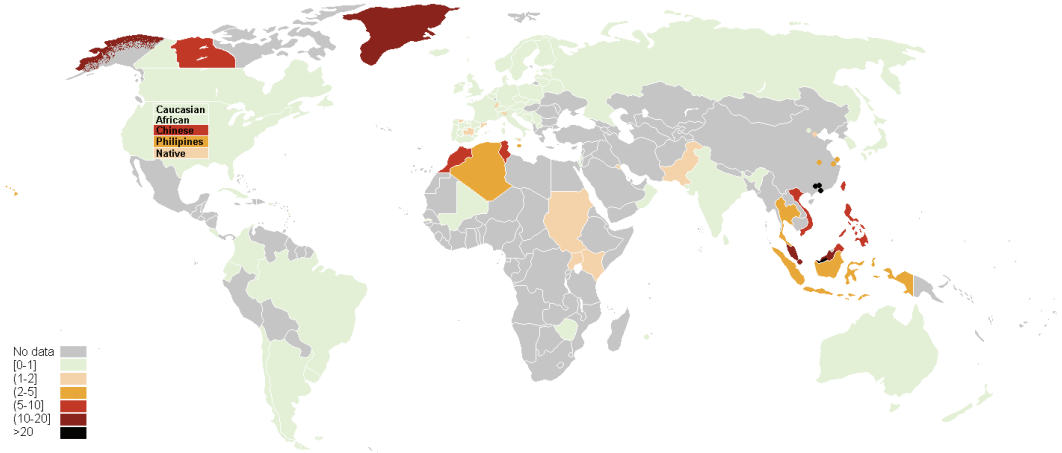
Figure 1: anatomic picture of the nasopharynx.



According to the most recent World Health Organization classification of tumors[8], nasopharyngeal carcinoma can be classified into three histological types: I) keratinizing squamous cell carcinoma; II) non-keratinizing carcinoma; and III) Basaloid squamous cell carcinoma. Type II NPC is further divided into differentiated squamous cell carcinoma and undifferentiated carcinoma. Relative proportions of these histological types are different between endemic and non-endemic regions: in the United States, NPC appears most commonly as keratinizing squamous cell carcinomas[9], while undifferentiated non-keratinizing carcinoma is the major histological type in South East Asia[1,10,11] and North Africa[12]. Basaloid squamous cell carcinoma is relatively uncommon in both endemic and non-endemic areas.

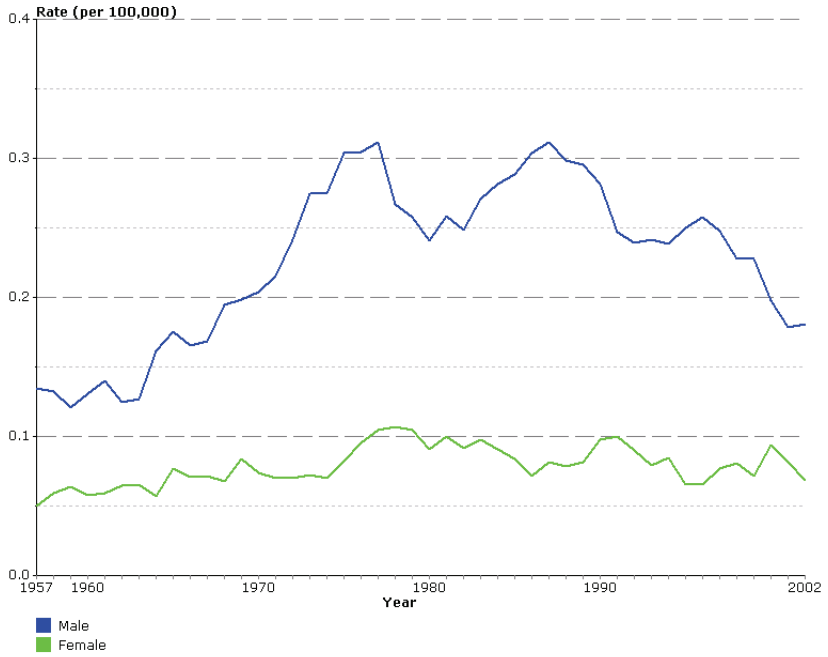
In the Netherlands, the incidence rate of NPC has been low during the last 50 years. During the period from 1993 to 1997, the age standardized incidence rate of NPC in the Netherlands is $0.5/10^5$/year for males and $0.2/10^5$/year for females[13], and the age standardized mortality rate is $0.3/10^5$/year for males and $0.1/10^5$/year for females (**Figure 3**). The age standardized 5-year relative survival of NPC in the Netherlands is 49% for males and 62% for females (http://www.eurocare.it/). The EUROCARE project also has indicated that, in Europe during the period 1978-1989, prognosis of NPC is better for younger patients, or for patients with squamous cell carcinoma than those with undifferentiated carcinoma[14]. It is observed that the population of NPC patients in the Netherlands contained more people of Asia or North-Africa origin[15]. A patient cohort from the Leiden University Medical Centre (LUMC, the Netherlands) between January 1998 and January 2005 indicates that, most of the cases (95.5%) were undifferentiated carcinoma[16]. In this study, half of the patients were European Dutch, 23% were of Indonesian origin and 27% were of North African origin. Chinese patients were not seen in this cohort, which can be expected as Chinese immigrants form only 0.3% of the whole population, while Indonesian account for 2.4%, and Moroccan account for 1.9%. Thus half the NPC cases occurred in only ~4% of the Dutch population, confirming the high rates of disease in South Asia and North African populations. As the migrant population from these areas ages, one might expect to see an increase in the NPC incidence rates in the Netherlands. It will be interesting to see if high rates of NPC persist in second and third generations of the original Indonesian and Moroccan migrants as this can provide ecological evidence of the role of genetic vs. environmental factors in NPC as these generations adopt the diet and lifestyles of their European Dutch counterparts.

Figure 2: Age standardized incidence rate of NPC.



Note: Data are taken from Cancer Incidence in Five Continents vol. VIII[13], except the following countries or areas: Alaska[17], Chile[5], Greenland[18,19], Indonesia[20], Kenya[21], Malaysia[1,4,22], and Tunisia[6].

Figure 3: Age-standardized mortality rate of NPC in the Netherlands.



Note: This figure is created by the WHO mortality database (http://www-dep.iarc.fr/).

Dietary risk factors

Cantonese-style salted fish, especially those consumed in weaning or during childhood, have been widely reported to be associated with a 2~3 fold increased risk of NPC, in populations habited in Tianjin[23], Shanghai[24], Guangxi[25,26], Guangzhou[27], Hong Kong[28], in Malaysia[29-31], as well as other populations including Thai[32] and Inuits[17]. In a study of 250 cases and 250 controls from Hong Kong, the odds ratio for having Cantonese styled salted fish in weaning was 7.5[3.9-14.8], and the OR for weekly consumption at age 10 was 37.7 [14.1-100.4][28].

Besides salted fish, other preserved foods and condiments have been shown to be related to higher NPC risk in many regions, including fermented fish sauce, salted shrimp paste, moldy bean curd, and preserved plum among Cantonese in Guangzhou, China[27], salted duck eggs, salted mustard green, brine fermented radish root, dried fish, fermented bean paste among Chinese in Guangxi, China[25], Salted eggs among Malaysia Chinese[31], protein-containing preserved foods among Shanghai Chinese[24], salted soy bean, canned pickled vegetables, salted tuber, salted mustard greens among Singapore Chinese[33], and salted shrimp paste among Tianjin Chinese[23].

Preserved foods have also been found to be risk factors in populations from North African countries, including quaddid (dried mutton stored in oil), harissa (very spicy condiment prepared with red pepper, olive oil, garlic, caraway), and toklia (basic stewing preparation, contain red pepper, black pepper, garlic, salt, oil, caraway and coriander) in Tunisia[12], pickled fruit and vegetables, dried and salted fat, and rancid butter in Sétif, Algeria[34], and rancid butter and khelii (dried meat, salty, spicy, cooked and preserved in a mixture of oil and melted bovine greases) in Morocco[35].

Despite the divergent distribution of the high-risk populations both geographically and ethnically, they show some common aspects of diet. Among Cantonese, Arab from Maghreb, and Inuits, the common point is that patients tend to consume more preserved foods and dried meats in childhood[36]. In addition, various epidemiological studies consistently found that fresh fruit and leafy vegetables are protective factors against NPC in Chinese[24-27,31,37] as well as some low risk populations[38]. A recent meta-analysis has confirmed the harmful effect of adulthood consumption of preserved vegetables and protective effect of fresh vegetables [39].

Main ingredients in the foods that have a potential role in the carcinogenic process have been one of the study concerns in previous publications. Investigators have analyzed food samples from endemic areas, and found that N-nitrosamines (N-nitrosodimethylamine (NDMA), N-nitrosopyrrolidine (NPYR) and N-nitrosopiperidine (NPIP)), compounds that are classified as probably or

possibly carcinogenic to humans by the International Agency for Research on Cancer (IARC) [40], were detected in Tunisian stewing base ("toklia"), quaddid, Chinese salted fish, brine fermented vegetables and Greenland dried fish[41-44]. Further more, *in vitro* nitrisation study indicated that quantity of N-nitrosamines could be fermented within human stomach from their precursors contained in Chinese salted fish and Tunisian spice[45]. Recently, a case-control study conducted in Taiwan found that, the highest quartile intake of nitrosamines from meat, fish and preserved vegetables during weaning can confer a 4 fold increased risk of NPC (odds ratio 3.9 [1.4-10.4]), which is much higher than any individual food items alone, including salted fish. All these results support the hypothesis that nitrosamines and their precursors are the responsible ingredients in the associated foods.

Besides N-nitrosamines, other substances in these preserved foods could also be the cause of NPC. It has been found that extracts of Cantonese salted fish, harissa and quaddid from Tunisia has inducers of Epstein-Barr virus (EBV) early antigens in Raji cells[46], however, there is no correlation between level of N-nitrosamines and EBV-inducing activities, before and after nitrosation[45], so that the EBV inducers are thought to have chemical structures different from N-nitrosamines. Afterward, lignin-containing high molecule complexes were isolated from harissa and were shown to be strong EBV inducers[47]. Nevertheless, exactly what substances in these foods are reactivating EBV in the human body is not yet known.

Tobacco, alcohol and fume intake

Associations between cigarette smoking and increased risk of NPC has been consistently reported in some low incident populations, such as North American[48-52]. However, results have been controversial in other high incidence areas. In South China, a two-fold increased risk by exposure to 30+ pack-year tobacco intake was reported in Guangzhou[53], one of the most endemic regions of the world; whereas, lack of association has also repeatedly been observed in the same district[54-58]. In other regions of South-East Asia, associations were reported among Chinese population habited in Shanghai[59], Taiwan[60,61], Singapore[11] and Malaysia[62], but not among populations habited in Thailand[63]. There has been a lack of consistency, even when the same population has been examined in multiple studies, e.g., Malaysian Chinese[62,64] and Singapore Chinese[11,65]. In North Africa, the relationship between tobacco intake and NPC risk has not been widely studied.

In regard to alcohol consumption, most, but not all studies reported no association[57-61,63-70], exceptions include studies in Malaysia[71], Morocco[35] and the United States[50,51]. Interplay between cigarette and alcohol intake was studied in the Singapore cohort[11], finding that alcohol consumption did not add further to the risk of NPC among smokers. In the US study, the association with alcohol

consumption appeared to be stronger in elderly subjects[51]. In addition to cigarette smoking and alcohol consumption, it has been postulated that NPC patients in South China were more exposed to domestic fume intake, by poor ventilation in kitchen (no windows, no chimney), kitchen in the main room, or wood fire cooking[26,37,56].

Epstein-Barr Virus

The involvement of Epstein-Barr virus (EBV) in NPC was implicated by the elevated anti-EBV antibody titer in patients than in controls[72]. More importantly, a prospective study from Taiwan showed that the antibodies against EBV capsid antigen and neutralizing antibodies against EBV-specific DNase can present years before the development of nasopharyngeal carcinoma, and was associated with a 32.8 fold increased risk of NPC[73]. In cellular level, EBV have also been detected as monoclonal episomes in almost all cancer cells of non-keratinizing NPCs[74,75], which implies that NPC was clonal expanded from a single cell infected with EBV. All these facts lead to the suggestion of a causal role of EBV in the etiology of NPC. However, it does not mean that EBV infection of epithelial cells initiate the development of nasopharyngeal carcinoma, since the stable EBV infection may require an undifferentiated cellular environment[76], and it is not clear how the undifferentiating was triggered.

Familial clustering

Familial aggregation of nasopharyngeal carcinoma has been widely documented in the Chinese population: more than 5% of the NPC patients have a positive first-degree family history of NPC in high-risk areas such as Hong Kong (7.2%), Yulin (6.0%), and Guangzhou (5.9%)[53,56]. A typical family cluster was found in Guangdong China, which has 15 family members affected with NPC, 2 family members with hepatic cell carcinomas and 1 with breast carcinoma[77]. However, these studies were limited in size, and comparison of the risk of NPC between first degree relatives (FDRs) and general populations were not reported yet. The familial clustering of nasopharyngeal carcinomas could be explained by similar environmental exposures within a family such as dietary or specific EBV strains, and/or genetic predispositions.

Genetic Studies

The involvement of EBV in pathogenesis of NPC inspired the association studies of Human Leukocyte Antigen (HLA) genes, which encode proteins to identify and present foreign antigens, including EBV peptides. HLA-A has consistently been reported to be related to the increased risk of NPC, by both association studies[78-83] and a linkage study[84] in South East Asian population. A recent fine mapping study has located the susceptibility gene to a region of less than 150kb around

HLA-A[85]. However, it is not yet known that whether the NPC susceptibility gene is HLA-A per se or another gene that is in linkage disequilibrium with HLA-A.

Many other genes were implicated to be associated with NPC, such as those related to carcinogenic metabolism (CYP2E1[86-90], GSTM1[91], NOR1[92], NQO1[93]), DNA repair (XRCC1[94,95]), cell cycle regulation (TP53[96,97], NGX6[98], CCND1[99,100], CDKN1A[101]), or EBV receptors (PIGR[102]). However, most of these genes were reported from small-scale case-control studies, and confirmations are limited. Therefore, they need to be carefully interpreted.

Scope of this thesis

In this thesis I studied the environmental and genetic risk factors contributing to the etiology of nasopharyngeal carcinoma. Chapter 2 focuses on the environmental factors in North Africa. It was performed within a large-scale multi-centric case-control study among Maghrebian countries, including Morocco, Tunisia and Algeria. Life styles unique to this region such as specific foods, way of tobacco intake, and cooking appliances were investigated with detailed adjustment for confounders.

Chapter 3 describes the genetic studies of nasopharyngeal carcinoma from South China. First, the genetic contribution to NPC was investigated by familial clustering study of NPC patients. Second, a genome wide search for NPC susceptibility gene(s) by linkage analyses was conducted. Third, a methodology study is presented. Although this method stems from the idea of association trend test on relationship between functional microsatellites and disease, this method could be used in SNP based association studies when SNP haplotypes are appropriately transformed into multi-allelelic markers. One of the potential usages of this method could be the HLA-NPC association test, where the HLA-A alleles are transformed into allele lengths according to their genetic distances from the most common allele in the population.

References

1. Devi, B.C., Pisani, P., Tang, T.S. & Parkin, D.M. High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo Island. *Cancer Epidemiol Biomarkers Prev* **13**, 482-486 (2004).
2. Yu, M.C. Nasopharyngeal carcinoma: epidemiology and dietary factors. *IARC Sci Publ*, 39-47 (1991).
3. Muir, C.S., Waterhouse, J., Mack, T., Powell, J. & Whellan, S. *Cancer Incidence in Five Continents Vol. V*, (International Agency for Research on Cancer, Lyon, France, 1987).
4. Armstrong, R.W., Kannan, K.M., Dharmalingam, S.K. & Ponnudurai, J.R. Incidence of nasopharyngeal carcinoma in Malaysia, 1968--1977. *Br J Cancer* **40**, 557-567 (1979).
5. Parkin, D.M., Whelan, S.L., Ferlay, J. & Storm, H. *Cancer Incidence in Five Continents Vol. VII*, (International Agency for Research on Cancer, Lyon, France, 1997).
6. Ellouz, R., Cammoun, M., Zaouche, A. & Ben Attia, R. [Cancers of the rhinopharynx in Tunisia. Epidemiological and anatomo-clinical aspects]. *Bull Cancer* **62**, 295-306 (1975).
7. Ellouz, R., Cammoun, M., Attia, R.B. & Bahi, J. Nasopharyngeal carcinoma in children and adolescents in Tunisia: clinical aspects and the paraneoplastic syndrome. *IARC Sci Publ*, 115-29 (1978).
8. Barnes, L., Eveson, J.W., Reichart, P.A. & Sidransky, D. *Pathology and Genetics of Head and Neck Tumours (World Health Organization Classification of Tumours)*, 430 (IARC Press, International Agency for Research on Cancer, Lyon France, 2005).
9. Ou, S.H., Zell, J.A., Ziogas, A. & Anton-Culver, H. Epidemiology of nasopharyngeal carcinoma in the United States: improved survival of Chinese patients within the keratinizing squamous cell carcinoma histology. *Ann Oncol* **18**, 29-35 (2007).
10. Tse, L.A., Yu, I.T., Mang, O.W. & Wong, S.L. Incidence rate trends of histological subtypes of nasopharyngeal carcinoma in Hong Kong. *Br J Cancer* **95**, 1269-73 (2006).
11. Friborg, J.T. et al. A prospective study of tobacco and alcohol use as risk factors for pharyngeal carcinomas in Singapore Chinese. *Cancer* **109**, 1183-91 (2007).
12. Jeannel, D. et al. Diet, living conditions and nasopharyngeal carcinoma in Tunisia--a case-control study. *Int J Cancer* **46**, 421-425 (1990).
13. Parkin, D.M., Whelan, S.L., Ferlay, J., Teppo, L. & Thomas, D.B. *Cancer Incidence in Five Continents Vol. VIII*, (International Agency for Research on Cancer, Lyon, France, 2002).
14. Jiong, L., Berrino, F. & Coebergh, J.W. Variation in survival for adults with nasopharyngeal cancer in Europe, 1978-1989. EUROCARE Working Group. *Eur J Cancer* **34**, 2162-6 (1998).

15. Balm, A.J., Plaat, B.E., Hart, A.A., Hilgers, F.J. & Keus, R.B. [Nasopharyngeal carcinoma: epidemiology and treatment outcome]. *Ned Tijdschr Geneeskd* **141**, 2346-50 (1997).
16. Kalpoe, J.S., Dekker, P.B., van Krieken, J.H., de Jong, R.J. & Kroes, A.C. Role of Epstein-Barr virus DNA measurement in plasma in the clinical management of nasopharyngeal carcinoma in a low risk area. *J Clin Pathol* **59**, 537-41 (2006).
17. Lanier, A. et al. Nasopharyngeal carcinoma in Alaskan Eskimos Indians, and Aleuts: a review of cases and study of Epstein-Barr virus, HLA, and environmental risk factors. *Cancer* **46**, 2100-2106 (1980).
18. Albeck, H. et al. Epidemiology of nasopharyngeal and salivary gland carcinoma in Greenland. *Arctic.Med Res* **51**, 189-195 (1992).
19. Nielsen, N.H., Mikkelsen, F. & Hansen, J.P. Nasopharyngeal cancer in Greenland. The incidence in an Arctic Eskimo population. *Acta Pathol Microbiol Scand [A]* **85**, 850-858 (1977).
20. Soeripto, Jensen, O.M. & Muir, C.S. Cancer in Yogyakarta, Indonesia: relative frequencies. *Br J Cancer* **36**, 141-8 (1977).
21. Clifford, P. Carcinogens in the nose and throat: nasopharyngeal carcinoma in Kenya. *Proc R Soc Med* **65**, 682-6 (1972).
22. Muir, C.S., Evans, M.D. & Roche, P.J. Cancer in Sabah (Borneo) a prelimanry survey. *Br J Cancer* **22**, 637-45 (1968).
23. Ning, J.P., Yu, M.C., Wang, Q.S. & Henderson, B.E. Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China. *J Natl Cancer Inst* **82**, 291-296 (1990).
24. Yuan, J.M. et al. Preserved foods in relation to risk of nasopharyngeal carcinoma in Shanghai, China. *Int J Cancer* **85**, 358-363 (2000).
25. Yu, M.C., Mo, C.C., Chong, W.X., Yeh, F.S. & Henderson, B.E. Preserved foods and nasopharyngeal carcinoma: a case-control study in Guangxi, China. *Cancer Res* **48**, 1954-1959 (1988).
26. 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* **69**, 508-514 (1994).
27. Yu, M.C., Huang, T.B. & Henderson, B.E. Diet and nasopharyngeal carcinoma: a case-control study in Guangzhou, China. *Int J Cancer* **43**, 1077-1082 (1989).
28. Yu, M.C., Ho, J.H., Lai, S.H. & Henderson, B.E. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res* **46**, 956-961 (1986).
29. Armstrong, R.W., Armstrong, M.J., Yu, M.C. & Henderson, B.E. Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese. *Cancer Res* **43**, 2967-2970 (1983).
30. Armstrong, R.W. & Eng, A.C. Salted fish and nasopharyngeal carcinoma in Malaysia. *Soc Sci Med* **17**, 1559-1567 (1983).
31. Armstrong, R.W. et al. Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. *Int J Cancer* **77**, 228-235 (1998).

32. Sriamporn, S., Vatanasapt, V., Pisani, P., Yongchaiyudha, S. & Rungpitarangsri, V. Environmental risk factors for nasopharyngeal carcinoma: a case-control study in northeastern Thailand. *Cancer Epidemiol Biomarkers Prev* **1**, 345-348 (1992).
33. Lee, H.P. et al. Preserved foods and nasopharyngeal carcinoma: a case-control study among Singapore Chinese. *Int J Cancer* **59**, 585-590 (1994).
34. Laouamri, S., Hamdi-Cherif, M., Sekfali, N., Mokhtari, L. & Kharchi, R. [Dietary risk factors of nasopharyngeal carcinoma in the Setif area in Algeria]. *Rev Epidemiol Sante Publique* **49**, 145-156 (2001).
35. Ammor, S., Baali, A., Hubert, A. & Cherkaoui, M. Epidemiological approach of the nasopharyngeal carcinoma in the region of Marrakech. *Antropo* **5**, 39-47 (2003).
36. Hubert, A. & de The, G. [Dietary behavior, way of life, and nasopharyngeal cancer]. *Bull Cancer* **69**, 476-482 (1982).
37. Chen, D.L. & Huang, T.B. A case-control study of risk factors of nasopharyngeal carcinoma. *Cancer Lett* **117**, 17-22 (1997).
38. Farrow, D.C. et al. Diet and nasopharyngeal cancer in a low-risk population. *Int J Cancer* **78**, 675-679 (1998).
39. Gallicchio, L. et al. Adulthood consumption of preserved and nonpreserved vegetables and the risk of nasopharyngeal carcinoma: a systematic review. *Int J Cancer* **119**, 1125-35 (2006).
40. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, 354 (International Agency for Research on Cancer, Lyon, 1987).
41. Poirier, S. et al. Occurrence of volatile nitrosamines in food samples collected in three high-risk areas for nasopharyngeal carcinoma. *IARC Sci Publ*, 415-419 (1987).
42. Poirier, S. et al. Volatile nitrosamine levels in common foods from Tunisia, south China and Greenland, high-risk areas for nasopharyngeal carcinoma (NPC). *Int J Cancer* **39**, 293-296 (1987).
43. Zou, X. et al. Volatile N-nitrosamines in salted fish samples from high- and low-risk areas for NPC in China. *Chin Med Sci J* **7**, 201-204 (1992).
44. Zou, X.N., Lu, S.H. & Liu, B. Volatile N-nitrosamines and their precursors in Chinese salted fish--a possible etological factor for NPC in china. *Int J Cancer* **59**, 155-158 (1994).
45. Poirier, S. et al. Volatile nitrosamine levels and genotoxicity of food samples from high-risk areas for nasopharyngeal carcinoma before and after nitrosation. *Int J Cancer* **44**, 1088-1094 (1989).
46. Shao, Y.M. et al. Epstein-Barr virus activation in Raji cells by extracts of preserved food from high risk areas for nasopharyngeal carcinoma. *Carcinogenesis* **9**, 1455-1457 (1988).
47. Bouvier, G. et al. Characterization of macromolecular lignins as Epstein-Barr virus inducer in foodstuff associated with nasopharyngeal carcinoma risk. *Carcinogenesis* **16**, 1879-1885 (1995).

48. Chow, W.H., McLaughlin, J.K., Hrubec, Z., Nam, J.M. & Blot, W.J. Tobacco use and nasopharyngeal carcinoma in a cohort of US veterans. *Int J Cancer* **55**, 538-40 (1993).
49. Mirabelli, M.C. et al. Occupational exposure to chlorophenol and the risk of nasal and nasopharyngeal cancers among U.S. men aged 30 to 60. *Am J Ind Med* **37**, 532-41 (2000).
50. Nam, J.M., McLaughlin, J.K. & Blot, W.J. Cigarette smoking, alcohol, and nasopharyngeal carcinoma: a case-control study among U.S. whites. *J Natl Cancer Inst* **84**, 619-22 (1992).
51. Vaughan, T.L. et al. Nasopharyngeal cancer in a low-risk population: defining risk factors by histological type. *Cancer Epidemiol Biomarkers Prev* **5**, 587-93 (1996).
52. Zhu, K., Levine, R.S., Brann, E.A., Gnepp, D.R. & Baum, M.K. Cigarette smoking and nasopharyngeal cancer: an analysis of the relationship according to age at starting smoking and age at diagnosis. *J Epidemiol* **7**, 107-11 (1997).
53. Yu, M.C., Garabrant, D.H., Huang, T.B. & Henderson, B.E. Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China. *Int J Cancer* **45**, 1033-9 (1990).
54. Li, C.C., Yu, M.C. & Henderson, B.E. Some epidemiologic observations of nasopharyngeal carcinoma in Guangdong, People's Republic of China. *Natl Cancer Inst Monogr* **69**, 49-52 (1985).
55. Ng, T.P. A case-referent study of cancer of the nasal cavity and sinuses in Hong Kong. *Int J Epidemiol* **15**, 171-5 (1986).
56. Yu, M.C., Ho, J.H., Lai, S.H. & Henderson, B.E. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res* **46**, 956-61 (1986).
57. Zheng, X., Yan, L., Nilsson, B., Eklund, G. & Drettner, B. Epstein-Barr virus infection, salted fish and nasopharyngeal carcinoma. A case-control study in southern China. *Acta Oncol* **33**, 867-72 (1994).
58. Zou, J. et al. A case-control study of nasopharyngeal carcinoma in the high background radiation areas of Yangjiang, China. *J Radiat Res (Tokyo)* **41 Suppl**, 53-62 (2000).
59. Yuan, J.M. et al. Non-dietary risk factors for nasopharyngeal carcinoma in Shanghai, China. *Int J Cancer* **85**, 364-9 (2000).
60. Chen, C.J. et al. Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency. *Anticancer Res* **10**, 547-53 (1990).
61. Cheng, Y.J. et al. Cigarette smoking, alcohol consumption and risk of nasopharyngeal carcinoma in Taiwan. *Cancer Causes Control* **10**, 201-7 (1999).
62. Armstrong, R.W. et al. Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles, formaldehyde and heat. *Int J Epidemiol* **29**, 991-8 (2000).
63. Sriamporn, S., Vatanasapt, V., Pisani, P., Yongchaiyudha, S. & Rungpitarangsri, V. Environmental risk factors for nasopharyngeal carcinoma: a case-control

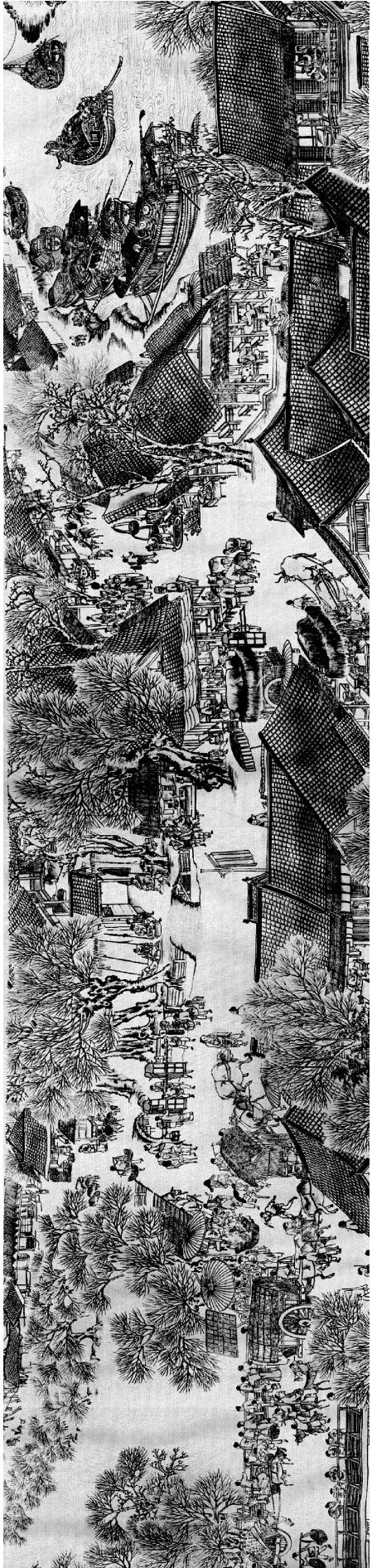
study in northeastern Thailand. *Cancer Epidemiol Biomarkers Prev* **1**, 345-8 (1992).
64. Armstrong, R.W., Armstrong, M.J., Yu, M.C. & Henderson, B.E. Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese. *Cancer Res* **43**, 2967-70 (1983).
65. Shanmugaratnam, K., Tye, C.Y., Goh, E.H. & Chia, K.B. Etiological factors in nasopharyngeal carcinoma: a hospital-based, retrospective, case-control, questionnaire study. *IARC Sci Publ*, 199-212 (1978).
66. Geser, A., Charnay, N., Day, N.E., de-The, G. & Ho, H.C. Environmental factors in the etiology of nasopharyngeal carcinoma: report on a case-control study in Hong Kong. *IARC Sci Publ*, 213-29 (1978).
67. Henderson, B.E., Louie, E., SooHoo Jing, J., Buell, P. & Gardner, M.B. Risk factors associated with nasopharyngeal carcinoma. *N Engl J Med* **295**, 1101-6 (1976).
68. Lin, T.M. et al. Retrospective study on nasopharyngeal carcinoma. *J Natl Cancer Inst* **51**, 1403-8 (1973).
69. Mabuchi, K., Bross, D.S. & Kessler, II. Cigarette smoking and nasopharyngeal carcinoma. *Cancer* **55**, 2874-6 (1985).
70. Ning, J.P., Yu, M.C., Wang, Q.S. & Henderson, B.E. Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China. *J Natl Cancer Inst* **82**, 291-6 (1990).
71. Armstrong, R.W. et al. Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. *Int J Cancer* **77**, 228-35 (1998).
72. Henle, W. et al. Antibodies to Epstein-Barr virus in nasopharyngeal carcinoma, other head and neck neoplasms, and control groups. *J Natl Cancer Inst* **44**, 225-31 (1970).
73. Chien, Y.C. et al. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med* **345**, 1877-82 (2001).
74. Niedobitek, G., Agathanggelou, A. & Nicholls, J.M. Epstein-Barr virus infection and the pathogenesis of nasopharyngeal carcinoma: viral gene expression, tumour cell phenotype, and the role of the lymphoid stroma. *Semin Cancer Biol* **7**, 165-74 (1996).
75. Raab-Traub, N. & Flynn, K. The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. *Cell* **47**, 883-9 (1986).
76. Knox, P.G., Li, Q.X., Rickinson, A.B. & Young, L.S. In vitro production of stable Epstein-Barr virus-positive epithelial cell clones which resemble the virus:cell interaction observed in nasopharyngeal carcinoma. *Virology* **215**, 40-50 (1996).
77. Zhang, F. & Zhang, J. Clinical hereditary characteristics in nasopharyngeal carcinoma through Ye-Liang's family cluster. *Chin Med J (Engl)* **112**, 185-7 (1999).

78. Pimtanothai, N., Kangwanshiratada, O. & Charoenwongse, P. Serological analysis of human leukocyte antigens-A and -B antigens in Thai patients with nasopharyngeal carcinoma. *J Med Assoc Thai* **86 Suppl 2**, S237-41 (2003).
79. Hildesheim, A. et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst* **94**, 1780-9 (2002).
80. Lu, C.C. et al. Genetic susceptibility to nasopharyngeal carcinoma within the HLA-A locus in Taiwanese. *Int J Cancer* **103**, 745-51 (2003).
81. Ooi, E.E., Ren, E.C. & Chan, S.H. Association between microsatellites within the human MHC and nasopharyngeal carcinoma. *Int J Cancer* **74**, 229-32 (1997).
82. Goldsmith, D.B., West, T.M. & Morton, R. HLA associations with nasopharyngeal carcinoma in Southern Chinese: a meta-analysis. *Clin Otolaryngol Allied Sci* **27**, 61-7 (2002).
83. Hu, S.P. et al. Further evidence for an HLA-related recessive mutation in nasopharyngeal carcinoma among the Chinese. *Br J Cancer* **92**, 967-70 (2005).
84. Lu, S.J. et al. Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature* **346**, 470-1 (1990).
85. Lu, C.C. et al. Nasopharyngeal carcinoma-susceptibility locus is localized to a 132 kb segment containing HLA-A using high-resolution microsatellite mapping. *Int J Cancer* **115**, 742-6 (2005).
86. Hildesheim, A. et al. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst* **89**, 1207-12 (1997).
87. Hildesheim, A. et al. Cytochrome P4502E1 genetic polymorphisms and risk of nasopharyngeal carcinoma: results from a case-control study conducted in Taiwan. *Cancer Epidemiol Biomarkers Prev* **4**, 607-10 (1995).
88. Yang, X.R. et al. Evaluation of risk factors for nasopharyngeal carcinoma in high-risk nasopharyngeal carcinoma families in Taiwan. *Cancer Epidemiol Biomarkers Prev* **14**, 900-5 (2005).
89. Kongruttanachok, N. et al. Cytochrome P450 2E1 polymorphism and nasopharyngeal carcinoma development in Thailand: a correlative study. *BMC Cancer* **1**, 4 (2001).
90. He, Z.M., Chen, Z.C., Yuan, J.H. & Wang, S.L. Expression of Cytochrome p450 2E1 in nasopharynx and polymorphism analysis. *Chinese Journal of Pathophysiology* **16**, 1060 (2000).
91. Nazar-Stewart, V. et al. Glutathione S-transferase M1 and susceptibility to nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* **8**, 547-51 (1999).
92. Xiong, W. et al. Studies of Association Between Nasopharyngeal Carcinoma and Single-Nucleotide Polymorphisms in NOR1, a Novel Oxidored2nitro Domain-containing Protein Gene. *Prog. Biochem. Biophys.* **30**, 401-5 (2003).
93. Wu, D.H. [Quinone oxidoreductase gene polymorphism and genetic susceptibility to nasopharyngeal carcinoma]. *Di Yi Jun Yi Da Xue Xue Bao* **22**, 1126-7 (2002).
94. Cho, E.Y. et al. Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. *Cancer Epidemiol Biomarkers Prev* **12**, 1100-4 (2003).

95. Cao, Y. et al. Polymorphisms of XRCC1 genes and risk of nasopharyngeal carcinoma in the Cantonese population. *BMC Cancer* **6**, 167 (2006).
96. Yung, W.C., Ng, M.H., Sham, J.S. & Choy, D.T. p53 codon 72 polymorphism in nasopharyngeal carcinoma. *Cancer Genet Cytogenet* **93**, 181-2 (1997).
97. Tiwawech, D., Srivatanakul, P., Karaluk, A. & Ishida, T. The p53 codon 72 polymorphism in Thai nasopharyngeal carcinoma. *Cancer Lett* **198**, 69-75 (2003).
98. Xiong, W. et al. Single-nucleotide polymorphisms in NGX6 gene and their correlation with nasopharyngeal carcinoma. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **34**, 512-5 (2002).
99. Deng, L. et al. Cyclin D1 polymorphism and the susceptibility to NPC using DHPLC. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **34**, 16-20 (2002).
100. Catarino, R.J. et al. Association of the A870G cyclin D1 gene polymorphism with genetic susceptibility to nasopharyngeal carcinoma. *Head Neck* **28**, 603-8 (2006).
101. Tsai, M.H., Chen, W.C. & Tsai, F.J. Correlation of p21 gene codon 31 polymorphism and TNF-alpha gene polymorphism with nasopharyngeal carcinoma. *J Clin Lab Anal* **16**, 146-50 (2002).
102. Fan, Q. et al. [Correlation of polymeric immunoglobulin receptor gene polymorphisms to susceptibility of nasopharyngeal carcinoma]. *Ai Zheng* **24**, 915-8 (2005).

Chapter 2

Environmental Contributions to Nasopharyngeal Carcinoma



Chapter 2.1

Dietary Fisk Factors
for Nasopharyngeal Carcinoma
in Maghrebian Countries

Dietary risk factors for nasopharyngeal carcinoma in Maghrebian countries

Bing-Jian Feng[1,2], Majida Jalbout[1], Wided Ben Ayoub[3], Meriem Khyatti[4], Sami Dahmoul[5], Messaoud Ayad[6], Fatima Maachi[4], Wided Bedadra[7], Meriem Abdoun[7], Sarah Mesli[6], Mokhtar Hamdi-Cherif[7], Kada Boualga[6], Noureddine Bouaouina[5], Lotfi Chouchane[8], Abdellatif Benider[9], Farhat Ben Ayed[3], David Goldgar[1,2] and Marilys Corbex[1,10]*

[1]*International Agency for Research on Cancer, Lyon, France*
[2]*Department of Dermatology, University of Utah, Salt Lake City, USA*
[3]*Association Tunisienne de Lutte Contre le Cancer, Tunis, Tunisia*
[4]*Institut Pasteur du Maroc, Casablanca, Morocco*
[5]*Service de Radiothérapie, CHU Farhat Hached, Sousse, Tunisia*
[6]*Centre anti-cancer de Blida service de radiothérapie oncologique, Blida, Algeria*
[7]*Service d'épidémiologie, CHU de Sétif, Sétif, Algeria*
[8]*Laboratoire d'immuno-oncologie moléculaire faculté de médecine, Monastir, Tunisia*
[9]*Centre d'oncologie Ibn rochd, Casablanca, Morocco*
[10]*World Health Organization, Eastern Mediterranean Regional Office, Cairo, Egypt*

North Africa is one of the major Nasopharyngeal Carcinoma (NPC) endemic regions. Specific food items unique to this area were implicated to be associated with NPC risk, but results were inconsistent. Here we have performed a large-scale case-control study in the Maghrebian population from Tunisia, Algeria and Morocco. From 2002 to 2005, interviews were conducted on 636 cases and 615 controls. Controls were hospitalized individuals from 15 non-cancer hospital departments, or friends and family members of non-NPC cancer subjects, matched by center, childhood household type (rural or urban), age and sex. Conditional logistic regression is used to evaluate the risk of factors. In results, consumption of rancid butter, rancid sheep fat and preserved meat not spicy (mainly quaddid) were associated with significantly increased risk of NPC, while consumption of cooked vegetables and industrial preserved fish was associated with reduced risk. Other foods such as fresh citrus fruits and spicy preserved meat (mainly osban) in childhood, industrial made olive condiments in adulthood, were marginally associated. In multivariate analyses, only rancid butter, rancid sheep fat and cooked vegetables were significantly associated with NPC. In regard to possible causative substances, our results implicate the involvement of butyric acid, a potential Epstein-Barr virus (EBV) activator.

Key words: nasopharyngeal carcinoma; diet; Tunisia; Algeria; Morocco

Nasopharyngeal carcinoma (NPC) is rare in most areas of the world but is much more common in certain populations. In Hong Kong, the incidence rates were estimated to be $21.4/10^5$/year in males and $8.3/10^5$/year in females,[1] where the majority of the population is Cantonese. In Guangdong province of China, the highest incidence rates in males and females were $30.94/10^5$/year and $13.00/10^5$/year in Sihui county,[2] where Cantonese is also the primary ethnic group. In Malaysia, in addition to the ethnic Chinese who demonstrate intermediate risk (16.5 and 7.2 per 10^5 per year),[3] an indigenous population, the Bidayuh, shows the highest incidence rate recorded hitherto ($32/10^5$ for males and $12/10^5$ for females per year).[4] People from the countries of North Africa also demonstrate high risk of NPC; these include Algeria, Morocco and Tunisia. In Sétif, Algeria, the incidence rates were $8/10^5$/year for males and $3/10^5$/year for females[5] in 1990–1993, while in Morocco and Tunisia, they may be slightly lower (The age and year adjusted incidence per 100,000 person years in Tunisia is 7.5 for men and 3.3 for women).[6] In contrast, the incidence for most other countries is less than $0.5/10^5$/year. In Asian high-risk populations, the incidence rises in adolescence and peaks at 45–55 years, while in North Africa, the Maghrebian population is characterized by a bimodal age distribution, with one peak occurring in the teens and the other at age 40–50 years.[7]

Major risk factors for NPC include Epstein-Barr virus reactivation, diet and genetic susceptibility.[8,9] It has been reported that lower socio-economic level is an important confounding factor for NPC in China, Macau, Greenland and Tunisia.[9] In intermediate to high risk Chinese populations, consumption of salted fish, especially during weaning in childhood, is associated with elevated risk of NPC.[10–16] Similar results have been found in other endemic populations, such as Thai[17] and Inuits.[18] In North African countries, associated foods include quaddid (dried mutton stored in oil), harissa (very spicy condiment prepared with red pepper, olive oil, garlic, caraway) and toklia (basic stewing preparation, contain red pepper, black pepper, garlic, salt, oil, caraway and coriander) in Tunisia,[19] pickled fruit and vegetables, dried and salted fat, and rancid butter in Sétif, Algeria,[20] and rancid butter and khelii (dried meat, salty, spicy, cooked and preserved in a mixture of oil and melted bovine greases) in Morocco.[21] However these results have to be considered with caution as none of the 3 studies included more than 80 cases and, detailed adjustment for possible confounders and multivariate analysis was not performed in the Algerian and the Moroccan studies.

Despite the divergent distribution of the high-risk populations both geographically and ethnically, they show some common aspects of diet. Salted fish is common and confirmed in Chinese patients from many regions, such as Tianjin,[11] Shanghai,[15] Guangxi,[14,22] Guangzhou,[13] Hong Kong,[12] Malaysia,[10,23,24] as well as other populations including Thai[17] and Inuits.[18] Among Cantonese, Arab from the Maghreb populations, and Inuits, the common factor is that patients tend to consume more preserved foods and dried meats in childhood than controls.[25] In addition, various epidemiological studies consistently found that fresh fruit and leafy vegetables are protective factors against NPC in Chinese[13–15,22,24,26] as well as some low risk populations.[27] A recent meta-analysis has confirmed the harmful effect of adulthood consumption of preserved vegetables and protective effect of fresh vegetables in various populations from South-East Asia North Africa and the United States.[28]

Besides the common features of studies in different populations, there are some specific findings, unique to North Africa. Associated foods, such as toklia, harissa, quaddid, dried and salted fat,

Grant sponsor: Association for International Cancer Research; Grant number: 03-252.
***Correspondence to:** WHO/EMRO, Abdul Razzak Al-Sanhouri Street, PO Box 7608 Nasr City, Cairo 11371, Egypt. Fax: +202 27 65 415.
E-mail: corbexm@emro.who.int
Received 8 November 2006; Accepted after revision 22 January 2007
DOI 10.1002/ijc.22813
Published online 20 June 2007 in Wiley InterScience (www.interscience.wiley.com).

and rancid butter,[19] are not common in any other NPC endemic regions. However, none of the above food items have been confirmed by all 3 studies from the same region. Therefore, confirmation of these findings is necessary by a larger study with sufficient power.

To address these problems, we have performed a large-scale multi-centric case-control study in the Maghrebian population from Tunisia, Algeria and Morocco. In this article, we examine the roles of dietary factors, in order to validate previously hypothesized food items, and to uncover additional NPC risk factors which may have been missed in previous studies due to lack of power and detailed adjustment for confounders.

Methods

Study population

All incident cases diagnosed between 2001 and 2004 in 5 hospitals were identified by the clinicians in the oncology and radiotherapy departments, and were invited for interview. These hospitals are public oncology centers (Tunis in Tunisia, Casablanca in Morocco, and Blida in Algeria) or the oncology service of public hospitals (Sétif in Algeria and Sousse-Monastir in Tunisia). In Morocco and Tunisia, it is estimated that roughly 10% of all cases are treated in private clinic rather than public hospitals, and therefore are not able to be recruited in this study; In Algeria, cancer patients are not seen in the private sector. The cases from private clinic have a potentially higher socio-economic status (SES) than the majority of cases seen in the public sector. In addition to the incident cases, prevalent cases were also recruited, in order to increase the size for a genetic association study. We designate those cases recruited within 1 year of diagnosis as "incident cases," otherwise "prevalent cases." For both cases, subjects less than 15 years old were excluded. A total of 636 cases were enrolled in the study (475 incident; 160 prevalent; 1 unknown). The median lag time between interview and diagnosis was 2 months for incident cases and 32 months for prevalent cases. Regarding histological type, almost all (92%) were undifferentiated carcinoma. The 615 controls were hospitalized individuals from 15 non-cancer hospital departments (61%), including gastroenterology, endocrinology, rheumatology, orthopedics, neurology, nephrology, pulmonary, hematology, cardiology, emergency medicine, surgical, ophthalmology, dermatology, gynecology and infection disease departments; or friends and family members of non-NPC cancer subjects (39%) who came to the hospitals as to visit or to take care of the patients. Controls were frequency matched by center, age, sex and household type (urban/rural) in childhood. The possible bias of higher SES and reduced geographical area in those visitors has been seriously evaluated prior to the recruitment and has been anticipated to be minor. Informed consent was obtained from each participant, and the IARC ethical committee has approved the study protocol. Among the individuals invited to participate (both cases and controls), more than 90% were successfully interviewed. The primary reason for non-participation was old age.

Data collection

Interviews were conducted by trained personnel using the identical questionnaires in each center. The questions covered demographic and ethnic information, medical history, consanguinity, familial cancer history, living conditions, exposure to chemicals and smokes, alcohol and tobacco consumption, and dietary consumption in both childhood and adulthood. For food items, subjects were asked to choose from 6 intake frequency categories: "1–2 times per day," "3–4 times per week," "1–2 times per week," "1–3 times per month," "less than 10 times per year" and "never".

Socio-economic variables

In this study, SES was inferred by household type, category of lodging, occupation and educational level (from 1: no education to 5: university). Occupations were classified into 3 levels ((*i*) manual-basic occupation; (*ii*) artisan; (*iii*) professional or managerial job) based on the literal and intellectual requirement of the job, and therefore it is subject to classification bias. However, this classification was blind to the case/control status, so that this bias should be non-differential. "Household type" was designated as rural or urban residency. In contrast with developed countries, the countries of the Maghreb have a large economic disparity between rural and urban areas, where rural areas are less developed and more traditional in lifestyle, so that the likelihood of misclassification is very low. To capture information about household density we constructed a variable, "category of lodging" with 3 levels: "≤2 persons/room"; "2.1–4 persons/room"; and ">4 persons/room or gourbi (makeshift lodging in rural areas)." This variable is thought to be able to capture better the SES than the original types of housing (gourbi, dar or apartment). In the analysis of the dietary exposures, we use these 6 variables for adjustment in the conditional logistic regression analyses, *i.e.* "category of lodging" in childhood and adulthood, "household type" in childhood and adulthood, "occupation" and "education." We also adjusted for exposure to toxic substances (chemical fertilizer, pesticides, fumes, dust, formaldehyde or other chemical substances) as an ever/never variable.

TABLE I – DEMOGRAPHIC CHARACTERISTICS OF THE STUDY POPULATION

	Case (%)	Control (%)
Center		
Tunisia		
Tunis	152 (24)	147 (24)
Sousse	115 (18)	83 (14)
Morocco		
Casablanca	207 (33)	202 (33)
Algeria		
Sétif	51 (8)	67 (11)
Blida	111 (17)	116 (19)
Age[1]		
<25	102 (16)	83 (13)
25–34	78 (12)	92 (15)
35–44	143 (22)	158 (26)
45–54	178 (28)	130 (21)
55–64	96 (15)	93 (15)
≥65	39 (6)	59 (10)
Sex		
Male	450 (71)	419 (68)
Female	186 (29)	196 (32)

Age at diagnosis for cases and age at interview for controls.

Statistical analysis

All statistical analyses were carried out using STATA 9.0 (STATA Corp. College Station, TX). Conditional logistic regression with strata defined by sex and centers was used to evaluate the association of specific dietary factors with NPC, adjusting for age, the 6 SES measures and exposure to toxic substances. Odds ratios and corresponding 95% confidence intervals were calculated. For foods that were rarely consumed, frequency categories were pooled such that every cell in the contingency table had more than 4 observations. If this still could not be reached after pooling, the variable was dropped from the analyses. Reference levels were assigned to the least exposed level. For exposures with more than 2 levels, the Armitage trend test was used to assess significance. Multivariate analysis was performed using stepwise forward selection, with inclusion/exclusion significance level of 0.01. Adjustment variables such as SES, age and exposure to toxic substances were always kept in the model.

Results

Demography

From January 2002 to March 2005, a total of 636 cases and 615 controls were interviewed. The demographic breakdown by sex, age and study center are shown in Table I. The vast majority

TABLE II – ASSOCIATION OF SOCIO-ECONOMIC STATUS (SES) WITH NPC

SES variables	Case (%)	Control (%)	OR	95% CI		*p*-trend
Household type in childhood						0.3125
Urban (ref)	269 (42)	241 (39)	1.0	–	–	
Rural	367 (58)	374 (61)	0.9	0.7	1.1	
Category of lodging in childhood						0.0209
≤2 persons/room (ref)	169 (27)	189 (31)	1.0	–	–	
2.1–4 persons/room	195 (31)	194 (32)	1.2	0.9	1.6	
>4 persons/room or gourbi	261 (42)	228 (37)	1.4	1.1	1.9	
Household type in adulthood						0.6031
Urban (ref)	393 (64)	377 (63)	1.0	–	–	
Rural	220 (36)	223 (37)	0.9	0.7	1.2	
Category of lodging in adulthood						0.0076
≤2 persons/room (ref)	374 (62)	381 (66)	1.0	–	–	
2.1−4 persons/room	155 (26)	152 (26)	1.1	0.8	1.5	
>4 persons/room or gourbi	77 (13)	48 (8)	1.8	1.2	2.7	
Education						0.2543
None (ref)	204 (32)	211 (35)	1.0	–	–	
Primary school	203 (32)	158 (26)	1.2	0.9	1.6	
Secondary school	114 (18)	114 (19)	0.9	0.6	1.3	
High school	70 (11)	86 (14)	0.7	0.5	1.1	
University	45 (7)	41 (7)	1.1	0.6	1.7	
Occupation						0.1260
Manual (ref)	394 (62)	338 (56)	1.0	–	–	
Litterary/artisan	156 (25)	199 (33)	0.7	0.5	1.9	
Professional/Managerial	81 (13)	72 (12)	1.0	0.7	1.4	

Analyses were stratified by sex and center, adjusted for age.

(89%) of subjects were of Arab ethnicity, with others belonging to Berber or other ethnic minority groups. There was no significant difference between cases and controls regarding ethnic groups, countries or recruitment centers. Seven hundred seventy (63.5%) subjects were living in urban areas at the time of recruitment with a smaller proportion reporting urban environment in childhood (41%). No differences in household type were found between cases and controls. The mean age at onset of the cases is 42.6 years old, comparable to the mean age at interview of the controls (43.5 years old; $p = 0.34$). The age at diagnosis of the cases ranged from 11 to 81 years, with an apparent bimodal distribution with peaks at ~20 years and ~40 years, similar to previous reports from cancer registry data in this population.[7] This may suggest heterogeneous etiology of the NPC development in this population. There was no significant difference between the proportion of cases and controls reporting consanguinity in parents (18.3% *vs.* 19.8%, $p = 0.48$).

SES

Among the 6 variables of SES, "household type" (urban *vs.* rural) was not associated with NPC as it was used as a matching variable for the cases and controls. However, "category of lodging" was significantly associated with NPC (Table II), thus confirming the association between SES and NPC risk in previous publications. Although the association between occupation and education level with NPC was not significant (Table II), we included all 6 variables as adjustment variables in the remaining analyses to capture the maximum possible confounding effect.

Dietary

Table III presents the relationship between dietary factors and NPC risk. Consumption of rancid butter and rancid sheep fat are significant risk factors for NPC, while consumption of cooked vegetables was inversely associated with NPC. While all 3 of these associations were observed in both adulthood and childhood, their effect in adulthood (ORs for consumption frequency ≥3 times/week: 3.0, 3.2, 0.6) are much stronger than those in childhood (ORs for consumption frequency ≥3 times/week: 1.8, 2.1, 0.8). Preserved meat including quaddid (dried meat stored in oil) and osban (stomach lining filled with garlic and other spices, which may be either fried or boiled), citrus fruits and industrial olives condiments (olives stored in vinegar and water with salt, aged for at least 8 weeks, stored for months to years) were also associated with increased NPC risk. High consumption of industrial preserved fish (unsalted canned fish) in both childhood and adulthood appears to be inversely associated with NPC. In the multivariate logistic regression analyses (Table IV), dietary factors that remained significant were rancid butter, rancid sheep fat and cooked vegetables in adulthood. Restriction of the analyses to the incident cases only, did not materially change the results (data not shown). Stratified or separate analyses in males or females, age group ≤30 or >30 years old, also did not find significant evidence of heterogeneity in the ORs for the dietary components that were significant in the overall analysis, although the female and ≤30 strata were quite small and consequently had low power.

Discussion

In this study, 636 cases and 615 controls from 3 Maghrebian countries were investigated for dietary association with NPC. When dietary components were analyzed individually, we found that consumption of rancid butter and rancid sheep fat were significantly associated with NPC risk. Cooked vegetables had an inverse association, especially in adulthood. Other foods such as fresh citrus fruits and spicy preserved meat in childhood, industrial made olive condiments in adulthood, preserved meat not spicy and industrial preserved fish in both age periods were associated with NPC. In this questionnaire, both spicy preserved meat and preserved meat not spicy include quaddid and osban. Although the level of spice varies among districts, generally osban is much more spicy than quaddid, and therefore, preserved meat that is not spicy consists primarily of quaddid, while spicy preserved meat mainly consists of osban.

Compared with the other 3 studies from the same region, this study confirms the reported increased risk of NPC associated with rancid butter, rancid sheep fat and quaddid. In addition, we have newly identified potential associations of osban, industrial made olive condiments, and cooked vegetables with NPC. In contrast to

TABLE III – ASSOCIATION OF DIETARY FACTORS WITH NPC

Exposure levels	Childhood						Adulthood					
	Case	Ctrl	OR	95% CI		Trend-*p*	Case	Ctrl	OR	95% CI		Trend-*p*
Rancid butter						0.0023						1E–05
<10 times/year (ref)	312	329	1.0	–	–		350	362	1.0	–	–	
<3 times/week	190	194	1.2	0.9	1.7		173	148	1.6	1.2	2.3	
≥3 times/week	129	87	1.8	1.2	2.6		57	26	3.0	1.7	5.2	
Rancid sheep fat						0.0003						2E–07
<10 times/year (ref)	427	478	1.0	–	–		387	421	1.0	–	–	
<3 times/week	144	104	1.6	1.2	2.2		129	70	2.2	1.5	3.1	
≥3 times/week	54	29	2.1	1.2	3.5		45	18	3.2	1.7	5.9	
Cooked vegetables						0.0588						0.0017
<3 times/week (ref)	200	155	1.0	–	–		148	94	1.0	–	–	
≥3 times/week	430	455	0.8	0.6	1.0		434	464	0.6	0.4	0.8	
Citrus fruits						0.0308						0.1499
<10 times/year (ref)	80	93	1.0	–	–		38	52	1.0	–	–	
<3 times/week	385	380	1.3	0.9	1.8		335	312	1.4	0.8	2.2	
≥3 times/week	166	138	1.6	1.0	2.5		209	189	1.5	0.9	2.6	
Toklia						0.9442						0.6455
<10 times/year (ref)	255	277	1.0	–	–		212	215	1.0	–	–	
<3 times/week	143	149	1.0	0.7	1.4		148	143	1.0	0.7	1.5	
≥3 times/week	141	132	1.0	0.7	1.5		120	119	0.9	0.6	1.3	
Harissa						0.9203						0.7566
<10 times/year (ref)	457	473	1.0	–	–		381	358	1.0	–	–	
<3 times/week	140	108	1.1	0.8	1.6		128	88	1.2	0.8	1.7	
≥3 times/week	24	26	0.8	0.4	1.5		28	31	0.7	0.4	1.3	
Spicy preserved meat (osben)						0.0251						0.3472
<10 times/year (ref)	564	564	1.0	–	–		509	469	1.0	–	–	
≥10 times/year	37	16	2.2	1.1	4.6		17	8	1.5	0.6	3.8	
Preserved meat not spicy (Quaddid)						0.0024						0.005
<10 times/year (ref)	420	496	1.0	–	–		420	464	1.0	–	–	
≥10 times/year	110	70	1.8	1.2	2.7		55	29	2.0	1.2	3.4	
Merguez, khelli						0.4654						0.4009
<10 times/year (ref)	584	570	1.0	–	–		519	493	1.0	–	–	
≥10 times/year	35	29	1.2	0.7	2.2		33	20	1.3	0.7	2.4	
Home made olives condiments						0.7566						0.9283
<10 times/year (ref)	255	272	1.0	–	–		240	228	1.0	–	–	
<3 times/week	204	196	1.1	0.8	1.4		193	175	1.0	0.7	1.3	
≥3 times/week	173	143	1.1	0.8	1.5		135	104	1.0	0.7	1.5	
Industrial olives condiments						0.6892						0.0285
<10 times/year (ref)	371	346	1.0	–	–		246	230	1.0	–	–	
<3 times/week	203	229	0.9	0.6	1.1		259	285	1.0	0.7	1.3	
≥3 times/week	56	33	1.5	0.9	2.5		74	37	2.0	1.3	3.3	
Industrial preserved fish						0.0264						0.0045
<10 times/year (ref)	621	585	1.0	–	–		539	485	1.0	–	–	
≥10 times/year	10	23	0.4	0.2	0.9		20	38	0.4	0.2	0.7	

Analyses were stratified by sex and center, adjusted for age, socio-economic status variables, and exposure to toxic substances.

TABLE IV – MULTIVARIATE ANALYSES OF DIETARY FACTORS

Dietary factors	OR	95% CI	
Rancid butter			
<10 times/year	1.0	–	–
<3 times/week	1.5	1.1	2.2
≥3 times/week	2.5	1.4	4.5
Rancid sheep fat			
<10 times/year	1.0	–	–
<3 times/week	1.9	1.3	2.7
≥3 times/week	2.6	1.4	4.9
Cooked vegetables			
<3 times/week	1.0	–	–
≥3 times/week	0.6	0.4	0.8

Analyses were stratified by sex and center, adjusted for age, socio-economic status variables, and exposure to toxic substances. *p* values were calculated by likelihood ratio test. Model was built by stepwise forward with inclusion criteria $p \leq 0.01$.

previous findings, toklia was not associated with NPC in our study, and we can exclude an effect of greater than a 50% increased risk. We estimated that the power to detect an increased risk similar to that previously reported for toklia with our sample size and observed consumption prevalence (40%) is almost 100%, so that it is unlikely that this is because of lack of power. The same is observed for harissa, which had been reported to be associated with a 5-fold increased NPC risk if eaten more than twice a week.[19] Because the preparation of Harissa could vary a lot between Morocco, Tunisia and Algeria, it was analyzed separately for each country, with no effect found in any individual country. Therefore, it is unlikely that the variation in harissa preparation is the explanation of its lack of association. Other differences in our study compared to previous reports are that preserved fish was found to be inversely associated in this article, rather than positively associated as expected; high consumption of fresh citrus fruits was found to be a risk factor of NPC, which contrast with many other reports.[13–15,22,24,26,27] It should be noted that, the preparation of industrial preserved fish in Maghrebian countries (unsalted canned fish) are quite different from those in South-East Asia, where sea fish are highly salted and sun dried.

Although these associated foods are significant in terms of 95% confident intervals, they must be interpreted carefully because of the multiple testing problem. In regard of this, *p* values for trend tests were also calculated for these variables. The association with rancid butter and rancid sheep fat have very significant *p* values ($p = 0.00001$, $p = 0.0000002$, respectively), and thus are unlikely

to happen by chance even in 50 independent tests. Cooked vegetables, quaddid and industrial preserved fish have intermediate significant levels, whereas fresh citrus fruits, osban and olive condiments have marginal p values. In light of this, the associations of fresh citrus fruits, osban and olive condiments could be spurious findings resulted from multiple testing.

In multivariate analyses, when we included all nominally significant variables, significance level of most factors decreased remarkably, except cooked vegetables, which was even more significant. Variables that remained in the model are rancid butter, rancid sheep fat and cooked vegetables, all consumed in adulthood. The association with other preserved foods found in the adjusted univariate analyses disappeared with the addition of rancid butter, rancid sheep fat and cooked vegetables into the model, as did the association with fresh citrus fruits. It is interesting to see that although rancid butter and rancid sheep fat are positively correlated with each other (individuals who consume rancid butter also tend to consume rancid sheep fat, with similar frequencies, spearman's correlation rho = 0.35, $p < 0.0001$), they both remained in the model with very significant p values.

Dose-response relationship is an important consideration in causality inference. In this study, from the odds ratios of different level of consumptions in childhood and adulthood, a clear dose-response pattern was observed for rancid butter and rancid sheep fat, both in univariate and multivariate analyses. This confirms the validity of the findings of rancid fat. Other factors with more than 2 levels had not shown a clear monotonic dose response pattern, especially harissa and toklia.

Possible confounding, validity and bias

Since SES is highly influential on diet, it could be a strong confounder in dietary association analyses. In this article, the SES confounder effect was controlled in 2 ways. First, the cases and controls were matched by household type in childhood. The reason for choosing this in the matching is that, exposures during childhood are more important for disease development than those in adulthood, and that household type is highly related to the childhood SES status. This matching will reduce the difference of childhood SES between cases and controls, leading to less confounder effect. Second, a series of variables that can infer SES status were used in the adjustment in conditional logistic regression, including the "category of lodging," a strong SES index. Some other SES indices such as occupation and education were not significantly associated with NPC, which could be partly explained by the matching in the study design (data not shown). Nevertheless, we used all of these variables in the adjustment, to minimize any possible confounding; however, it is still likely that not all of the confounding effect was captured in our analyses. In fact, some of the associations found between NPC and consumption of specific food items could partly be explained by residual SES confounding.

Another concern of the validity of the study design is that controls in this study were taken from hospitalized individuals and family members of non-NPC cancer patients. These subjects may have higher chance than the general population to share a common environmental exposure with NPC patients, if any other disease has similar etiology with NPC. However, we believe that given the diversity of clinical services from which the controls were recruited, such effect has been highly diluted. In addition, the sharing of a common environmental exposure between cases and controls could only decrease the power of the study, rather than causing a spurious result. Therefore, this is unlikely to be a major threat to the validity of our conclusions.

Similarly, the fact that some of the cases were recruited some time after their NPC diagnosis could result in survival bias. However, 75% of our cases were interviewed within 1 year of diagnosis, and almost all within 3 years. Further analyses restricted only to incident cases did not materially change the results.

Lastly, as in any retrospective study, our results could be at least partly explained by recall bias, although this problem should be minor, since this is a relatively unsophisticated population in terms of perceived relationship between diet and health, and particularly since the cases in our study would be expected to have no knowledge at all of previously postulated NPC risk factors.

Responsible ingredients

Main ingredients in the foods that have a potential role in the carcinogenic process have been one of the study concerns in previous publications. Investigators have found that *N*-nitrosamines (*N*-nitrosodimethylamine (NDMA), *N*-nitrosopyrrolidine (NPYR) and *N*-nitrosopiperidine (NPIP)), compounds that are classified as probably or possibly carcinogenic to humans (International Agency for Research on Cancer, monograph Vol.17 Suppl. 7, 1987), were detected in toklia, quaddid, Chinese salted fish, brine fermented vegetables and Greenland dried fish,[29–32] or could be fermented within human stomach from their precursors contained in Chinese salted fish and Tunisian spice.[33] Besides N-nitrosamines, lignin-containing high molecule complexes were isolated from harissa and were shown to be strong EBV inducers.[34]

In this article, neither *N*-nitrosamine-containing toklia, nor EBV-inducer-containing harissa were significantly associated with NPC, in comparison with rancid butter and rancid sheep fat. Importantly, higher level of *N*-nitrosamines in rancid fat has not been demonstrated, which suggests some other disease causing chemicals in this population. A possible compound is butyric acid, which is also named *n*-Butanoic Acid. The glyceride form of butyric acid makes up 3 to 4% of butter, and is released into free butyric acid by hydrolysis when it becomes rancid. Butyric acid is known to be able to activate Epstein-Barr virus in the B-lymphoid cells into lyric cycle,[35] and therefore, could be related to NPC.

Vegetables and fruits are well established as protective factors for most epithelial cancers. Here we can now confirm this for NPC also. It is interesting that only cooked vegetables were significantly associated, while raw vegetables were not; if anything, their effects were to increase NPC risk. These 2 dietary components could be different in terms of the specific vegetables consumed, the portion sizes, or their correlation with SES. Further analysis shows that while consumption of raw vegetables is positively associated with higher SES ($p = 0.0001$), consumption of cooked vegetables is not ($p = 0.12$).

Conclusions

In conclusion, this study is the first large-scale epidemiology study that has been performed in the Maghrebian countries of North Africa to infer associations between dietary factors and the risk of Nasopharyngeal Carcinoma. It has provided a clearer view of the dietary risk factors in the NPC etiology, particularly with regard to rancid butter and sheep fat. Although the specific substances contained in these foods that cause NPC are not yet known, results of this study have provided the key information necessary to design a primary preventive strategy aimed at reducing risk factor exposure for nasopharyngeal carcinoma, especially in Maghrebian countries, where NPC has a strong social impact as it is more affected in young, male individuals, typically the main economic support of a family.

Acknowledgements

We thank our colleagues, Ms. Valérie Gaborieau for efficient data management, Ms. Colette Bonnardel for data entry and checking, Ms. Hélène Renard and Mr. Othman Yaqoubi for database design, Dr. Dominique Jeannel and Dr. Annie Sasco for help in designing the questionnaire, and Dr. Mohamed Maamer for its efficent help to the study in Tunis.

References

1. Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB. Cancer incidence in five continents, vol. VIII. Lyon, France: International Agency for Research on Cancer, 2002.
2. Jia WH, Huang QH, Liao J, Ye W, Shugart YY, Liu Q, Chen LZ, Li YH, Lin X, Wen FL, Adami HO, Zeng Y, 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;6:178.
3. Armstrong RW, Kannan KM, Dharmalingam SK, Ponnudurai JR. Incidence of nasopharyngeal carcinoma in Malaysia, 1968–1977. Br J Cancer 1979;40:557–67.
4. Devi BC, Pisani P, Tang TS, Parkin DM. High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo Island. Cancer Epidemiol Biomarkers Prev 2004;13:482–86.
5. Parkin DM, Whelan SL, Ferlay J, Storm H. Cancer incidence in five continents, vol. VII. Lyon, France: International Agency for Research on Cancer, 1997.
6. Ellouz R, Cammoun M, Zaouche A, Ben Attia R. [Cancers of the rhinopharynx in Tunisia. Epidemiological and anatomo-clinical aspects]. Bull Cancer 1975;62:295–306.
7. Ellouz R, Cammoun M, Attia RB, Bahi J. Nasopharyngeal carcinoma in children and adolescents in Tunisia: clinical aspects and the paraneoplastic syndrome. IARC Sci Publ 1978:20;115–29.
8. Hildesheim A, Levine PH. Etiology of nasopharyngeal carcinoma: a review. Epidemiol Rev 1993;15:466–85.
9. Jeannel D, Bouvier G, Hubbert A. Nasopharyngeal carcinoma: an epidemiological approach to carcinogenesis. Cancer Surveys 1999;33: 125–55.
10. Armstrong RW, Armstrong MJ, Yu MC, Henderson BE. Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese. Cancer Res 1983;43:2967–70.
11. Ning JP, Yu MC, Wang QS, Henderson BE. Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China. J Natl Cancer Inst 1990;82:291–96.
12. Yu MC, Ho JH, Lai SH, Henderson BE. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. Cancer Res 1986;46:956–61.
13. Yu MC, Huang TB, Henderson BE. Diet and nasopharyngeal carcinoma: a case-control study in Guangzhou, China. Int J Cancer 1989; 43:1077–82.
14. Yu MC, Mo CC, Chong WX, Yeh FS, Henderson BE. Preserved foods and nasopharyngeal carcinoma: a case-control study in Guangxi, China. Cancer Res 1988;48:1954–9.
15. Yuan JM, Wang XL, Xiang YB, Gao YT, Ross RK, Yu MC. Preserved foods in relation to risk of nasopharyngeal carcinoma in Shanghai, China. Int J Cancer 2000;85:358–63.
16. Zheng X, Yan L, Nilsson B, Eklund G, Drettner B. Epstein-Barr virus infection, salted fish and nasopharyngeal carcinoma. A case-control study in southern China. Acta Oncol 1994;33:867–72.
17. Sriamporn S, Vatanasapt V, Pisani P, Yongchaiyudha S, Rungpitarangsri V. Environmental risk factors for nasopharyngeal carcinoma: a case-control study in northeastern Thailand. Cancer Epidemiol Biomarkers Prev 1992;1:345–48.
18. Lanier A, Bender T, Talbot M, Wilmeth S, Tschopp C, Henle W, Henle G, Ritter D, Terasaki P. Nasopharyngeal carcinoma in Alaskan Eskimos Indians, and Aleuts: a review of cases and study of Epstein-Barr virus, HLA, and environmental risk factors. Cancer 1980;46:2100–06.
19. Jeannel D, Hubert A, de Vathaire F, Ellouz R, Camoun M, Ben Salem M, Sancho-Garnier H, de The G. Diet, living conditions and nasopharyngeal carcinoma in Tunisia—a case-control study. Int J Cancer 1990;46:421–25.
20. Laouamri S, Hamdi-Cherif M, Sekfali N, Mokhtari L, Kharchi R. [Dietary risk factors of nasopharyngeal carcinoma in the Setif area in Algeria]. Rev Epidemiol Sante Publique 2001;49:145–56.
21. Ammor S, Baali A, Hubert A, Cherkaoui M. Epidemiological approach of the nasopharyngeal carcinoma in the region of Marrakech. Antropo 2003;5:39–47.
22. Zheng YM, Tuppin P, Hubert A, Jeannel D, Pan YJ, Zeng Y, de The G. Environmental and dietary risk factors for nasopharyngeal carcinoma: a case-control study in Zangwu County, Guangxi, China. Br J Cancer 1994;69:508–14.
23. Armstrong RW, Eng AC. Salted fish and nasopharyngeal carcinoma in Malaysia. Soc Sci Med 1983;17:1559–67.
24. Armstrong RW, Imrey PB, Lye MS, Armstrong MJ, Yu MC, Sani S. Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. Int J Cancer 1998;77:228–35.
25. Hubert A, de The G. Dietary behavior, way of life, and nasopharyngeal cancer. Bull Cancer 1982;69:476–82.
26. Chen DL, Huang TB. A case-control study of risk factors of nasopharyngeal carcinoma. Cancer Lett 1997;117:17–22.
27. Farrow DC, Vaughan TL, Berwick M, Lynch CF, Swanson GM, Lyon JL. Diet and nasopharyngeal cancer in a low-risk population. Int J Cancer 1998;78:675–9.
28. Gallicchio L, Matanoski G, Tao XG, Chen L, Lam TK, Boyd K, Robinson KA, Balick L, Mickelson S, Caulfield LE, Herman JG, Guallar E, et al. Adulthood consumption of preserved and nonpreserved vegetables and the risk of nasopharyngeal carcinoma: a systematic review. Int J Cancer 2006;119:1125–35.
29. Zou XN, Lu SH, Liu B. Volatile *N*-nitrosamines and their precursors in Chinese salted fish—a possible etological factor for NPC in china. Int J Cancer 1994;59:155–8.
30. Poirier S, Hubert A, de The G, Ohshima H, Bourgade MC, Bartsch H. Occurrence of volatile nitrosamines in food samples collected in three high-risk areas for nasopharyngeal carcinoma. IARC Sci Publ 1987;84:415–19.
31. Poirier S, Ohshima H, de The G, Hubert A, Bourgade MC, Bartsch H. Volatile nitrosamine levels in common foods from Tunisia, south China and Greenland, high-risk areas for nasopharyngeal carcinoma (NPC). Int J Cancer 1987;39:293–96.
32. Zou X, Li J, Lu S, Song X, Wang X, Guo L, Li J, Lin Q, Ye J. Volatile *N*-nitrosamines in salted fish samples from high- and low-risk areas for NPC in China. Chin Med Sci J 1992; 7:201–4.
33. Poirier S, Bouvier G, Malaveille C, Ohshima H, Shao YM, Hubert A, Zeng Y, de The G, Bartsch H. Volatile nitrosamine levels and genotoxicity of food samples from high-risk areas for nasopharyngeal carcinoma before and after nitrosation. Int J Cancer 1989;44:1088–94.
34. Bouvier G, Hergenhahn M, Polack A, Bornkamm GW, de The G, Bartsch H. Characterization of macromolecular lignins as Epstein-Barr virus inducer in foodstuff associated with nasopharyngeal carcinoma risk. Carcinogenesis 1995;16:1879–85.
35. Bauer G. Quantitative analysis of the cooperative effect between inducers of Epstein-Barr virus antigen synthesis. J Gen Virol 1983; 64(Part 6):1337–46.

Chapter 2.2

Cannabis Smoking and Domestic Fume Intake are associated with Nasopharyngeal Carcinoma in North Africa

Abstract

A multi-centric case-control study of Nasopharyngeal Carcinoma (NPC) was performed in North Africa (Tunisia, Algeria, and Morocco). From January 2002 to March 2005, we interviewed 636 cases and 615 controls, frequency matched by center, age, sex, and household type in childhood (urban/rural). Conditional logistic regression with strata defined by sex and centers was used, adjusting for age, socio-economic status (SES), and previously reported dietary risk factors. Association between NPC risk and cigarette smoking was marginally significant. Association with indoor domestic fume intake in childhood was strong, especially when cooking with kanoun (compact oven runs on charcoal). Although results have been insignificant by conditional logistic regression adjusted for age, SES, diet and tobacco, restricted permutation analysis indicates that cannabis smoking was significantly associated with increased NPC risk, and this association was not due to the strong correlation between cannabis and tobacco consumption. In multivariate analyses, with an inclusion criteria of $p<0.01$ and adjusted for age and SES, four variables remained significant in the model: consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood, and cooking with kanoun during childhood. No effects of alcohol or other forms of tobacco consumption, such as neffa (snuff) and chicha (hookah or water-pipe), on NPC risk were observed.

Introduction

Nasopharyngeal carcinoma (NPC) is a rare cancer type in most regions of the world, but is more common in South-East Asia, North Africa and the Inuit populations of Greenland, Canada, and Alaska. The highest age-standardized incidence rates (ASR) hitherto reported, $32/10^5$/year for males and $12/10^5$/year for females, were recorded in an indigenous population in Malaysia, the Bidayuh [1]. People from the countries of western North Africa (Morocco, Algeria and Tunisia) also demonstrate high risk of NPC. In Algeria, the ASR's were $2.7/10^5$/year for males and $1.3/10^5$/year for females in 1993-1997 [2], while in Tunisia, they may be slightly higher (The ASR's in Tunisia were 4.6 for men and 2.2 for women in 1996) [3].

Major risk factors for NPC include Epstein-Barr virus (EBV) infection, diet, and genetic susceptibility [4-7]. Association between low socio-economic level and NPC has been observed throughout all endemic areas [5] and is well established, though without real understanding of the underlying cause. Poor diet, early exposure to EBV as well as higher exposure to fumes and chemicals could account for this association.

Association between cigarette smoking and increased risk of NPC has been consistently reported in some low incidence populations, such as North American [8-12], where keratinizing squamous cell carcinomas are the predominant histological type of NPC [13]. However, results have been controversial in high incidence areas, where the majority of NPC tumors are undifferentiated [1,14,15]. In South China, a two-fold increased risk by exposure to 30+ pack-year tobacco intake was reported in Guangzhou [16], one of the most endemic regions of the world; whereas, lack of association has also repeatedly been observed in the same district [17-21]. In other regions of South-East Asia, such associations were reported among Chinese populations in Shanghai [22], Taiwan [23,24], Singapore [15] and Malaysia [25], but not among populations habited in Thailand [26]. There has been a lack of consistency, even when the same population has been examined in multiple studies, e.g., Malaysian Chinese [25,27] and Singapore Chinese [15,28]. In North Africa, the relationship between tobacco intake and NPC risk has not been widely studied. The only report hitherto was a small-scale case-control study conducted in Morocco, in which the odds ratio for tobacco was estimated to be 2.9, but the association was not statistically significant due to the small size of the study [29].

In endemic areas, local forms of tobacco intake are often implicated by doctors and patients as a primary cause of NPC. These include "kretek" in Indonesia, "neffa" (snuff) in North Africa, and "water pipe" in Guangdong, China. Neffa is a powder mix of tobacco leaf and other additives (calcium phosphate and lime), which is a traditional habit mainly among men of Morocco, Algeria and Tunisia. Its association with NPC has never been studied in North Africa, even though, it is perceived as a risk factor of NPC in the local medical community. We thus included in-depth enquiry about snuff consumption in our study.

It has been suggested from a U.S. study that, the association with cigarette smoking is restricted to differentiated squamous cell carcinoma, but not the undifferentiated and nonkeratinizing carcinomas [11]. On the other hand, a recent cohort study concluded that long term smoking (>40 years) could confer a two-fold increased risk in Singapore Chinese, in which undifferentiated NPC was the dominant histological type, and similar results were obtained from the analyses on undifferentiated tumor strata [15]. Therefore, the relationship between smoking and undifferentiated NPC risk has been inconsistent among studies in different populations.

Marijuana (dried buds or flowers of cannabis) is one of the most prevalent illicit drugs associated with cancer risk [30]. When consumed, cannabis comes in either herbal form (marijuana), resinous form (hashish), or in oil form. In North Africa, marijuana is either smoked alone, or together with tobacco (kiff). Even when marijuana is smoked alone, it produces many of the carcinogens and co-carcinogens that can be found in tobacco smoke [30]. Therefore, cancer studies have been focused on tobacco related cancer sites, including head and neck cancer and lung cancer. However, epidemiology studies have been difficult, partly because of the strong confounder effect of cigarette smoking, and also because of the low report rates due to its illegal status. In fact, most of the studies did not yield significant results after adjustment for tobacco use [30]. Association of marijuana smoking or other forms of cannabis consumption with NPC risk have not been previously reported.

In regard to alcohol consumption, most, but not all, studies reported no association [6], exceptions include studies in Malaysia [31], Morocco [29] and the United States [10,11]. Interplay between cigarette smoking and alcohol intake was studied in the Singapore cohort [15], finding that alcohol consumption did not add further to the risk of NPC among smokers. In the US study, the association with alcohol consumption appeared to be stronger in elderly subjects [11].

It has been postulated that NPC patients in South China were very exposed to domestic fume intake, by poor ventilation in kitchen (no windows, no chimney), kitchen in the main room, or wood fire cooking [19,32,33]. Further, one Chinese study found that poor ventilation (absence of windows) can significantly increase the risk of NPC conferred by domestic wood fire [32]. However, since place of cooking and ventilation of kitchen are highly associated with social economic status (SES), these analyses were subject to the confounding effect of SES. In fact, in some other studies from South China, the association disappeared when the multivariate analyses included dietary risk factors [5], which could also have a link with social economic status or determine the form of cooking.

Non-dietary risk factors have not previously been widely studied in North Africa. One study from Morocco reported the significant association of alcohol consumption, but this study included only 32 cases and 48 controls, and the significant level was marginal [29]. In Tunisia, having the kitchen in the main room during childhood conferred a four fold increased risk of NPC in the crude analyses, although not statistically significant [34].

In our previous dietary investigation on Maghrebian populations, we showed that consumption of rancid butter, rancid sheep fat, and quaddid (preserved meat) were associated with increased NPC risk, while cooked vegetables were inversely associated with NPC risk. In multivariate analyses adjusted for socio-economic status, consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood remained significant ($p<0.01$) [35].

In this paper, we present the results of our investigations on non-dietary risk factors in a large multi-centric case-control study in the Maghrebian population from Tunisia, Algeria, and Morocco. The aim of this study was to examine the role of tobacco, alcohol, cannabis consumption and domestic fume intake in the etiology of NPC in this population, with detailed adjustment of social economic status and associated dietary factors. Life styles specific to this area and related to the above factors, such as chicha (or hookah, an oriental tobacco pipe with a flexible tube attached to a container of water through which smoke is cooled and moisturized), snuff (powdered tobacco leaf, consumed by sniffing or chewing), kanoun (a light and compact sized oven runs on charcoal) and tabouna (a volcano-shaped traditional oven in North Africa and the Middle East) cooking, were also investigated.

Methods

Study population

Details of the study population have been described elsewhere [35]. In brief, all incident cases diagnosed between 2001 and 2004 in 5 hospitals were identified by the clinicians in the oncology and radiotherapy departments, and were invited for interview. These hospitals are public oncology centers (Tunis in Tunisia, Casablanca in Morocco, and Blida in Algeria) or the oncology service of public hospitals (Sétif in Algeria and Sousse-Monastir in Tunisia). In addition to the incident cases, prevalent cases were also recruited. We designate those cases recruited within one year of diagnosis as “incident cases”, otherwise “prevalent cases”. For both cases, subjects less than 15 years old were excluded. The controls were hospitalized individuals from 15 non-cancer hospital departments (61%) or friends and family members of non-NPC cancer subjects (39%), frequency matched by center, age, sex, and household type (urban/rural) in childhood. Among the individuals invited to participate (both cases and controls), more than 90% were successfully interviewed. The primary reason for non-participation was old age. Informed consent was obtained from each participant, and the IARC (International Agency for Research on Cancer) ethical committee approved the study protocol.

Data collection

Interviews were conducted by trained personnel using identical questionnaires in each center. For the assessment of tobacco consumption, the questions included age of starting, age of quitting, total time of cessation in between, and number of cigarettes per day. From these variables, we calculated the estimated total pack-years of cigarette smoking. Use of other forms of tobacco intake such as pipe, chicha and snuff (either by chewing or by sniffing) was assessed independently of cigarette smoking. Alcohol consumption information included age of beginning and number of cups per week for 5 categories of alcoholic beverage (beer or cider, wine, Porto or Martini, Liqueurs, Whisky or digestive). Average percentage of alcohol for these categories were estimated to be 6.5%, 12.5%, 20%, 32% and 45%, and the average volume of cup were estimated to be 25cl, 12cl, 4cl, 4cl and 4cl. The total ethanol intake in one week was then calculated from these numbers combined with number of cups per week. Questions regarding cannabis consumption included the form of cannabis (herbal, resinous, oil or other),

mode of consumption (smoking, smoking with tobacco, ingestion, drinking, sniffing or inhaling), intake frequency (times per month), age of beginning and age of quitting.

Sources of fume intake in North African countries are diverse, and include incense, burned perfume, cooking with kanoun or tabouna, wood fire cooking, wood fire heating, or occupational fume intake. Usage frequency of these items (never, sometimes, often) during both childhood and adulthood were inquired. For the purpose of our analyses, the frequency was combined into an ever-never variable. Questionnaire data also included information on ventilation status that may affect the fume intake of the study individuals. In particular, kitchen ventilation was classified as ventilated (presence of windows or chimney, or cook outside the house) or unventilated (absence of windows and chimney, or cooking in the main living space).

Socio-economic variables

In this study, socio-economic status was inferred by household type (rural or urban residency), category of lodging ("<=2 persons/room"; "2.1-4 persons/room"; or ">4 persons/room or gourbi"), occupation (1=manual-basic occupation; 2=artisan; 3=professional or managerial job) and educational level (from 1= no education to 5=university). The rationale for use of these variables and their association with NPC risk have been described in detail in our previous publication [35].

Statistical Analysis

All statistical analyses were carried out using STATA 9.0 (STATA Corp. College Station, TX). Conditional logistic regression with strata defined by sex and centers was used to evaluate the association of specific factors with NPC, adjusting for age, the socio-economic status (SES) measures, and associated dietary factors (consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood)[35]. Odds ratios and corresponding 95% confidence intervals were calculated. For variables with more than two levels of exposures, the Armitage trend test was used to assess significance. Multivariate analyses were performed by stepwise inclusion, adjusting for age and SES, starting with associated dietary risk factors from our previous study and significant variables regarding tobacco, alcohol, cannabis consumption and domestic fume exposure.

To test whether cannabis is associated with NPC given its strong correlation with tobacco, a restricted permutation analysis was performed. In each replication, 77 individuals were randomly selected without regard to case-control status as cannabis consumers, among which there were 75 smokers and 2 nonsmokers (i.e., identical frequencies to those in the true data set). An association test was then carried out by conditional logistic regression, adjusted for age, SES and diet. From 10,000 repeats, the proportion of times that the odds ratio is equal to or higher than 1.8 (odds ratio of true cannabis intake) is taken as the p-value of the association test on cannabis, given its correlation with cigarette smoking.

Results

Demography

From January 2002 to March 2005, a total of 636 cases and 615 controls were interviewed. Details of the demographic breakdown by sex, age and study center, are presented in our previous paper [35]. In summary, there were no significant differences between cases and controls regarding ethnic groups, countries, recruitment centers, or household type during childhood. Of the 636 cases, 475 were incident, 160 were prevalent and 1 unknown. Regarding histological type, almost all (587, 92%) cases were non-keratinizing undifferentiated carcinoma, the rest were 14 (2.2%) non-keratinizing differentiated carcinoma, 5 (0.8%) keratinizing squamous cell carcinoma, and 30 (4.7%) of unknown histological type.

Tobacco intake

Table 1 presents the association of NPC risk with cigarette smoking. The overall relative risk of ever smoking compared to non-smoking was 1.3, although this falls short of statistical significance. Age at initiation and time of quitting were not significantly associated with risk of NPC. However, when the dose of cigarette intake (smoking history, cigarettes per day, lifetime cigarette intake) was examined, there was a significantly increased risk of NPC in unadjusted analyses, even when only undifferentiated NPC were considered. Interestingly, most of these associations diminished after adjustment for SES and dietary factors, with only cigarettes-per-day remaining marginally significant by trend test (**Table 1**).

Table 1: Cigarette smoking and risk of NPC.

Exposure Levels	Case(%)	Ctrl(%)	Crude #				Adjusted @			
			O.R.	95% C.I.		trendP	O.R.	95% C.I.		trendP
Smoking						0.107				0.156
Non-smoker (ref.)	344(55)	365(60)	1.00	-	-		1.00	-	-	
Ever-smoker	282(45)	240(40)	1.25	0.95	1.65		1.26	0.92	1.73	
Smoking Status						0.174				0.171
Non-smoker (ref.)	344(55)	365(61)	1.00	-	-		1.00	-	-	
Former smoker quit ≥10 yrs	64(10)	59(10)	1.29	0.84	1.98		1.20	0.74	1.93	
Former smoker quit <10 yrs	49(8)	36(6)	1.54	0.95	2.49		1.38	0.82	2.32	
Current smoker	169(27)	141(23)	1.22	0.89	1.66		1.27	0.89	1.82	
Begin age of Smoking						0.089				0.303
Non-smoker (ref.)	344(55)	365(61)	1.00	-	-		1.00	-	-	
>21 years old	61(10)	44(7)	1.52	0.97	2.38		1.49	0.91	2.44	
17-21 years old	112(18)	113(19)	1.03	0.74	1.45		1.17	0.79	1.73	
1-16 years old	109(17)	79(13)	**1.51**	**1.05**	**2.18**		1.26	0.84	1.89	
Aggregate smoking history §						**0.049**				0.238
Non-smoker (ref.)	344(55)	365(61)	1.00	-	-		1.00	-	-	
1-20 years	126(20)	110(18)	1.18	0.84	1.64		1.32	0.89	1.95	
>20 years	156(25)	122(20)	**1.45**	**1.04**	**2.02**		1.24	0.86	1.80	
Cigarettes per day §						**0.005**				**0.035**
Non-smoker (ref.)	344(55)	365(60)	1.00	-	-		1.00	-	-	
1-12 cigarettes	82(13)	84(14)	1.04	0.72	1.50		1.03	0.67	1.58	
13-22 cigarettes	136(22)	117(19)	1.26	0.90	1.75		1.31	0.90	1.91	
>22 cigarettes	64(10)	39(6)	**1.80**	**1.13**	**2.85**		1.59	0.96	2.65	
Cigarettes in lifetime §						**0.008**				0.110
Non-smoker (ref.)	344(55)	365(61)	1.00	-	-		1.00	-	-	
1-40 pack-years	236(38)	191(32)	1.32	0.99	1.75		1.33	0.96	1.84	
>40 pack-years	46(7)	41(7)	1.24	0.76	2.03		1.04	0.60	1.80	

Note: # Analyses were stratified by sex and center, adjusted for age. @ Analyses were stratified by sex and center, adjusted for age, SES measures and associated dietary factors. § Trend-test is calculated by using continuous values. Significant values are in bold fonts.

Other forms of tobacco

In the study population, the average beginning age of snuff intake was 5 years later than that of cigarette smoking (23.7 vs. 18.7 years old), while the prevalence of snuff intake was much lower (11% vs. 42%). Although cigarette smoking was similarly popular in the three countries (38.8%, 44.9% and 42.4% in Algeria, Tunisia and Morocco, separately), snuff intake was more frequent in Algeria than in Tunisia or Morocco (27.3% vs. 5.6% or 5.2%). It is observed that snuff was more consumed in men than in women (140 vs. 3), in cigarette smokers than in non-smokers (17.6% vs. 6.8%, $p=3\times10^{-9}$), in rural areas than in cities (17.6% vs. 8.5%, $p=2\times10^{-6}$), in less educated people ($p<0.0001$), or in populations with lower professional levels ($p=0.0024$). These suggest that consumption of snuff is associated with lower SES levels. Among the snuff consumers, 70% consumed by chewing, 31% by sniffing, with very few who utilized both

forms of intake (1.4%). Before and after adjustment for SES and diet, association of snuff with NPC risk was far from significant (odds ratio=1.0 [0.7-1.6] after adjustment for SES), and so was the dose response (snuff/week: trend p=0.96 ; lifetime snuff: trend p=0.56, after adjustment for SES).

Contrary to snuff, chicha consumption was associated with higher SES, such as higher educational level (p=0.0004), higher professional level (p=0.01), better living condition (p=0.03) and urban household type (2.5% in urban and 0.9% in rural, p=0.05). Chicha showed no association with NPC risk, before and after adjustment for SES and diet, nor in individual age strata.

Alcohol consumption

There was no association of NPC with alcohol consumption assessed as ever/never, either in crude analyses or after adjustment for SES and diet (Odds Ratio=1.2 [0.8-1.6]). There was no evidence of a dose response when specific quantities were examined, nor was there an association by type of alcohol consumed (p=0.60 for beer and p=0.90 for all alcohol beverage). Similar results were obtained when cases were stratified by histological type or age.

Fume intake

Occupational fume intake was not common among the studied population, only 35 cases (6%) and 16 controls (2%) responded that they were exposed to fumes as part of their employment, and the corresponding test does not yield significant results (odds ratio=1.5 [0.8-2.9]). However, domestic fume exposure by usage of a kanoun oven in cooking during childhood was highly associated with increased NPC risk (**Table 2**). It remained significant after adjusting for social economic status and foods. In addition, there was a trend towards more exposure to fumes in cases than in controls through use of a wood fire cooking and poor ventilation in the kitchen during childhood, although the corresponding tests fell short of nominal statistical significance. In addition, incense usage during adulthood was associated with reduced NPC risk (odds ratio=0.6 [0.4-0.9], **Table 2**).

Table 2: Exposure to domestic fumes and risk of NPC.

Exposure	Childhood					Adulthood				
Levels	case(%)	ctrl(%)	O.R.	95% C.I.		case(%)	ctrl(%)	O.R.	95% C.I.	
Incense										
Never (ref.)	188 (30)	189 (31)	1.00	-	-	201 (35)	187 (33)	1.00	-	-
Ever	440 (70)	419 (69)	1.05	0.72	1.54	375 (65)	376 (67)	**0.63**	**0.42**	**0.93**
Burned perfume										
Never (ref.)	203 (42)	221 (47)	1.00	-	-	211 (47)	200 (47)	1.00	-	-
Ever	284 (58)	245 (53)	1.41	0.99	2.00	235 (53)	230 (53)	0.83	0.57	1.20
Kanoun										
Never (ref.)	140 (22)	203 (33)	1.00	-	-	319 (56)	365 (65)	1.00	-	-
Ever	485 (78)	408 (67)	**1.86**	**1.28**	**2.72**	251 (44)	197 (35)	1.41	0.98	2.04
Tabouna										
Never (ref.)	263 (43)	273 (46)	1.00	-	-	370 (65)	362 (66)	1.00	-	-
Ever	355 (57)	318 (54)	1.12	0.81	1.54	199 (35)	188 (34)	1.01	0.72	1.43
Wood fire cooking										
Never (ref.)	217 (37)	257 (43)	1.00	-	-	423 (80)	455 (83)	1.00	-	-
Ever	368 (63)	335 (57)	1.38	0.95	2.00	108 (20)	93 (17)	1.23	0.82	1.85
Wood fire heating										
Never (ref.)	251 (44)	272 (46)	1.00	-	-	412 (78)	450 (82)	1.00	-	-
Ever	324 (56)	321 (54)	1.01	0.71	1.42	113 (22)	98 (18)	1.17	0.79	1.72
Place of cooking †										
Ventilated (ref.)	469 (75)	495 (81)	1.00	-	-	507 (90)	492 (88)	1.00	-	-
Not ventilated	154 (25)	113 (19)	1.29	0.93	1.80	58 (10)	65 (12)	0.78	0.51	1.20

Note: Analyses were stratified by sex and center, adjusted for age, SES measures and associated dietary factors. Significant values are in bold fonts. † "Ventilated" means chimney or windows present in kitchen, or cooked outside the house; "Not ventilated" means chimney or windows absent in kitchen, or cooked in the main room.

Cannabis consumption

Among the 1251 study subjects, 79 (6.3%) reported ever consumption of cannabis. Frequency of consumption was much higher in Morocco (15%) than in Algeria (2.1%) or Tunisia (2.3%). It was consumed almost exclusively by males (78 males, 1 female), and was more consumed in individuals with lower living condition (p=0.0008), lower educational (p=0.008) or professional level (p=0.006), but was equally consumed in rural and urban areas (p=0.65). These suggest that consumption of cannabis is associated with lower SES levels. In addition, almost all cannabis users smoked cigarettes (97%), but among the individuals who consume both cigarettes and cannabis, the dose correlation between cigarettes-per-day and cannabis-per-month was not significant (spearman rank correlation = -0.22, p=0.09). After adjustment for age, SES and diet, cannabis intake was significantly associated with NPC risk (p=0.04), with a marginally insignificant dose-response relationship (trend p=0.06 for cannabis-per-month). Among the two major forms of consumption, smoking cannabis together with tobacco had a higher odds ratio than smoking without tobacco (**Table 3**). When all these analyses were further

adjusted for cigarettes-per-day, findings became insignificant. However, when the confounding with tobacco was better accounted for in the restricted permutation analysis, the effect of cannabis on NPC risk was statistically significant (p=0.027).

Table 3: Cannabis (hashish/kiff) consumption and risk of NPC.

Exposure-Levels	case(%)	ctrl(%)	O.R.	95% C.I.	
Cannabis consumption					
Never (ref.)	572 (92)	580 (96)	1.00	-	-
Ever	52 (8)	27 (4)	**1.80**	**1.03**	**3.15**
Frequency of Cannabis intake					
Never (ref.)	573 (93)	580 (96)	1.00	-	-
<30 times/month	17 (3)	10 (2)	1.61	0.68	3.82
≥30 times/month	27 (4)	14 (2)	1.99	0.93	4.26
Lifetime consumption					
Never (ref.)	589 (94)	589 (97)	1.00	-	-
<2000 times	16 (3)	9 (1)	1.62	0.68	3.88
≥2000 times	21 (3)	7 (1)	**2.75**	**1.07**	**7.06**
Form of Cannabis					
Never (ref.)	572 (92)	580 (96)	1.00	-	-
Herbal	46 (7)	23 (4)	1.76	0.97	3.17
Resinous	4 (1)	3 (1)	2.03	0.33	12.6
Mode of consumption †					
Never (ref.)	572 (92)	580 (96)	1.00	-	-
Smoking	12 (2)	9 (1)	1.06	0.41	2.73
Smoking with Tobacco (kiff)	38 (6)	16 (3)	**2.30**	**1.16**	**4.56**

Note: Analyses were stratified by sex and center, adjusted for age, SES measures and associated dietary factors. Significant values are shown in bold fonts. † Ingestion mode of consumption is not included because of low frequency of observations.

Multivariate analyses

With an inclusion criteria of $p<0.05$ and when adjusted for age and social economic status, consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood (final dietary variables in the multivariate analyses in our previous study [35]), usage of kanoun during childhood, incense during adulthood and cigarettes-per-day remained in the multivariate model (**Table 4**). Among these, the three dietary variables and kanoun were highly significant ($p<0.005$), while incense and cigarette intake were marginal and would have not entered the model if a more stringent criteria ($p<0.01$) had been used, which is more reasonable under the consideration of testing multiple exposures.

Table 4: Multivariate analyses.

Exposures	O.R.	95% C.I.		p-value
Rancid butter (Adulthood)				**0.0006**
<10 times/year (ref.)	1.00	-	-	
<3 times/week	1.43	0.99	2.06	
≥3 times/week	2.89	1.55	5.39	
Rancid sheep fat (Adulthood)				**0.0002**
<10 times/year (ref.)	1.00	-	-	
<3 times/week	1.72	1.18	2.50	
≥3 times/week	2.48	1.29	4.80	
Cooked vegetables (Adulthood)				**0.0012**
<3 times/week (ref.)	1.00	-	-	
≥3 times/week	0.57	0.40	0.81	
Kanoun (Childhood)				**0.0015**
Never (ref.)	1.00	-	-	
Ever	1.84	1.25	2.70	
Incense (Adulthood)				0.02
Never (ref.)	1.00	-	-	
Ever	0.63	0.42	0.94	
Cigarettes per day				0.03
Non-smoker (ref.)	1.00	-	-	
1-12 cigarettes	1.01	0.66	1.57	
13-22 cigarettes	1.24	0.84	1.82	
>22 cigarettes	1.68	0.99	2.84	

Note: Analyses were stratified by sex and center, adjusted for age and SES. Stepwise inclusion method was used, with inclusion criteria of $p<0.05$. Significant p values ($p<0.01$) were in bold fonts.

Discussion

Main findings

In this study, 636 cases and 615 controls from 3 Maghrebian countries were recruited for investigation on non-dietary factors associated with NPC, concentrating on tobacco, alcohol, domestic fumes and cannabis intake. In contrast to the previous publication from Morocco, this study did not confirm the association with alcohol. The effect of cigarette smoking was marginally significant, while effects of the other forms of tobacco intake such as chicha and snuff were not evident. As far as domestic fume intake was concerned, this paper confirmed the association of poor ventilation in kitchen and newly discovered the possible harmful effect of usage of kanoun, especially those during childhood. Although results have been insignificant by conditional logistic regression analysis adjusted for age, SES, diet and tobacco, restricted permutation analysis indicates that cannabis smoking was significantly associated with increased NPC risk. In multivariate analyses, kanoun was the only non-dietary factor that had a p-value less than 0.01.

Possible Confounding, Validity, and Bias

It is known that socio-economic status is associated with higher risk of NPC, so that it could be a confounder in analyses of all factors that have a link with SES. In this paper, for example, tobacco intake, alcohol consumption, cannabis smoking and exposure to indoor fumes could all be affected by SES level. In our study, the SES confounder effects were controlled by matching controls with cases on household type during childhood, and by adjustment for the SES indicators in analyses. Since that household type is highly associated with SES, and that exposures during childhood are more important for disease development than those during adulthood, this matching will reduce the difference of childhood SES between cases and controls, leading to less confounder effect. The inclusion of the six SES indices in conditional logistic regression analyses will further minimize the residual confounding.

Another source of potential confounding is dietary factors. Many studies have reported that smokers tend to have a diet pattern containing more saturated fat and less fresh fruit or vegetables [36]. Dietary differences between smokers and nonsmokers were also observed in our study: smokers consumed more rancid sheep fat (p=0.002 for adulthood and

p=0.0002 for childhood), more rancid butter (p=0.05) and less cooked vegetables (p=0.03) during childhood than non-smokers. These food items are of special concern, because they have also been proven to be associated with NPC risk, so that they could be confounders in a smoking study. In addition, it is possible that dietary patterns were correlated with different modes of cooking. Thus in the analyses of fume intake, it is also necessary to adjust for both social economic status and associated dietary factors. In our recent report, consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood were significantly associated to NPC ($p<0.01$). Therefore, for the non-dietary association study presented here, we adjusted for these three factors to minimize the dietary confounder effect, while adjusting for as few variables as possible.

The inclusion of prevalent cases in this study could possibly lead to survival bias. However, they only occupied a minor proportion of all patients, and their average lag time between diagnosis and interview was not long (32 months). Confining analyses to incident cases only did not substantively change our results, indicating that survival bias is unlikely to be a source of concern in this study.

Cannabis consumption is prohibited by law in the study countries, so that the reported consumption is expected to be underestimated. Nevertheless, since that there is no perceived relationship between cannabis and cancer in the study countries, the under-reporting of cannabis consumption is expected to be the same in cases and controls, and thus can only decrease the significance of the association if there is any. Therefore, the real association of cannabis with NPC may be stronger than the one we observed.

Effect of Cigarette Smoking

The fact that the association of cigarette smoking diminished after adjustment for SES and diet, demonstrates that SES and diet were strong confounders of smoking. Although results have been marginally significant, the involvement of cigarette smoking in NPC etiology still cannot be excluded, because the trend of increased risk of NPC along with the increased doses of cigarette intake per day was still apparent after adjusting for SES and diet, with a marginal trend-p value. However, compared to diet and domestic fume intake, its contribution to NPC risk may not be as strong.

It has been suggested from a low-risk population that association of cigarette smoking is restricted to differentiated squamous cell carcinomas, but not undifferentiated nonkeratinizing carcinomas. In our analyses the contribution of smoking to undifferentiated NPC cannot be definitively excluded. On the other hand, despite the much smaller sample size (19 patients), association of smoking with differentiated squamous cell carcinomas was significant (lifetime cigarette intake trend p=0.026). The highest quartile of both cigarettes-per-day and lifetime cigarette intake were associated with increased NPC risk by 14 fold and 22 fold, respectively. This has raised the possibility that smoking is more important to differentiated squamous cell NPC than to undifferentiated NPC in North Africa. Considering the small sample size of differentiated NPC, and the marginal significance level of the tests, this suggestion needs to be further confirmed.

Effect of fume intake

Our results indicated that usage of kanoun during childhood increases NPC risk two fold. Usage of kanoun during adulthood seems to have less effect and was not significant. Furthermore, poor ventilation of kitchen and wood fire cooking during childhood were associated with 30%-40% increased risk of NPC, although they were not statistically significant. These results suggest that domestic fume intake from cooking is likely to be involved in NPC development, and that those exposures during childhood are more important than those during adulthood.

Effect of cannabis intake

When analyses were adjusted for age, SES and diet, cannabis consumption was significantly associated with NPC risk. Considering that cannabis consumption is highly associated with cigarette smoking, one may expect that this could be a carry-over effect of the tobacco association. Since there were only two individuals in our samples that have consumed cannabis but have not smoked tobacco, analysis in the non-smoker strata was not possible. Thus, to confirm the association of cannabis with NPC risk, we performed a restricted permutation analysis and found that, even given its association with cigarette smoking, cannabis itself was significantly associated with NPC risk (p=0.027). The explanation of this result could be: 1) cannabis is involved in NPC etiology independently of cigarette smoking; and/or 2) cannabis consumption and tobacco smoking has a synergistic effect on NPC risk. Due to the limitation of our sample size, neither of these two hypotheses can be confirmed in

our study. Above all, it is necessary to have a much larger study that includes enough cannabis consumers who has not smoked tobacco.

Other forms of tobacco

In North Africa, it is of interest to study snuff intake since there is an important concern in the community about its association with NPC development. In our study, association between snuff and NPC was far from significant. However, snuff consumption was correlated with important NPC risk factors (SES, sex). These correlations, together with its consumption style (powder is almost directly in contact with nasopharyngeal mucosa) and its geographical distribution (in North Africa, snuff is used mainly in the NPC endemic areas) may explain the false impression of "NPC causing factor" existing in the community.

Conclusions

This study is the first large-scale epidemiology study that has been performed in the Maghrebian countries of North Africa to infer associations between non-dietary factors and the risk of Nasopharyngeal Carcinoma. It has proven the relationship between domestic fume intake and risk of NPC. The specific cooking facility responsible for the major fume exposure was implicated to be kanoun, a light and compact sized oven that use charcoal as fuel. In our study, cigarette smoking only had a marginal effect on NPC risk. Other tobacco intake methods specific to this area such as chicha and snuff, as well as alcohol consumption, did not contribute to the NPC risk in this population. Even given the correlation with cigarette smoking, cannabis consumption is significantly associated with increased NPC risk.

Acknowledgments

Without the funding support of Association for International Cancer Research (grant number 03-252), this project could not have been accomplished. We also thank our colleagues, Valérie Gaborieau for efficient data management, Colette Bonnardel for data entry and checking, Hélène Renard and Othman Yaqoubi for database design, Dominique Jeannel and Annie Sasco for help in designing the questionnaire, and Dr. Mohamed Maamer for his kind and dedicated help in the study in Tunisia.

References

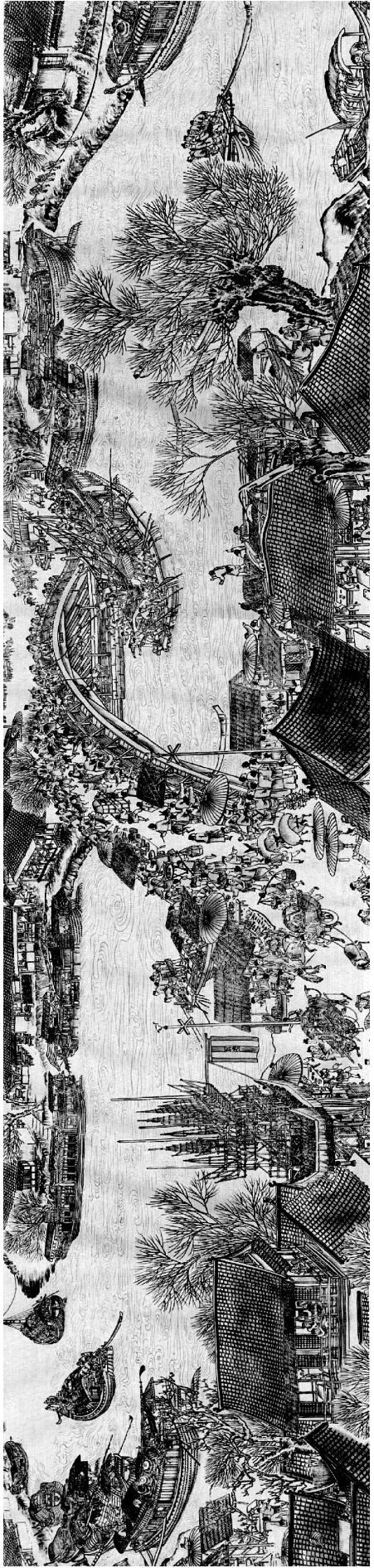
1. Devi, B.C., Pisani, P., Tang, T.S. & Parkin, D.M. High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo Island. *Cancer Epidemiol Biomarkers Prev* **13**, 482-486 (2004).
2. Parkin, D.M., Whelan, S.L., Ferlay, J., Teppo, L. & Thomas, D.B. *Cancer Incidence in Five Continents Vol. VIII*, (International Agency for Research on Cancer, Lyon, France, 2002).
3. Registre des cancers Nord-Tunisie 1995-1998. (Ministère de la santé, Unité de recherche en Epidémiologie des cancers en Tunisie, Institut Salah Azaiz.).
4. Hildesheim, A. & Levine, P.H. Etiology of nasopharyngeal carcinoma: a review. *Epidemiol Rev* **15**, 466-485 (1993).
5. Jeannel, D., Bouvier, G. & Hubbert, A. Nasopharyngeal Carcinoma: An Epidemiological Approach to Carcinogenesis. in *Cancer Surveys* Vol. 33 125-155 (1999).
6. Chang, E.T. & Adami, H.O. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* **15**, 1765-77 (2006).
7. Busson, P., Keryer, C., Ooka, T. & Corbex, M. EBV-associated nasopharyngeal carcinomas: from epidemiology to virus-targeting strategies. *Trends Microbiol* **12**, 356-60 (2004).
8. Chow, W.H., McLaughlin, J.K., Hrubec, Z., Nam, J.M. & Blot, W.J. Tobacco use and nasopharyngeal carcinoma in a cohort of US veterans. *Int J Cancer* **55**, 538-40 (1993).
9. Mirabelli, M.C. et al. Occupational exposure to chlorophenol and the risk of nasal and nasopharyngeal cancers among U.S. men aged 30 to 60. *Am J Ind Med* **37**, 532-41 (2000).
10. Nam, J.M., McLaughlin, J.K. & Blot, W.J. Cigarette smoking, alcohol, and nasopharyngeal carcinoma: a case-control study among U.S. whites. *J Natl Cancer Inst* **84**, 619-22 (1992).
11. Vaughan, T.L. et al. Nasopharyngeal cancer in a low-risk population: defining risk factors by histological type. *Cancer Epidemiol Biomarkers Prev* **5**, 587-93 (1996).
12. Zhu, K., Levine, R.S., Brann, E.A., Gnepp, D.R. & Baum, M.K. Cigarette smoking and nasopharyngeal cancer: an analysis of the relationship according to age at starting smoking and age at diagnosis. *J Epidemiol* **7**, 107-11 (1997).
13. Ou, S.H., Zell, J.A., Ziogas, A. & Anton-Culver, H. Epidemiology of nasopharyngeal carcinoma in the United States: improved survival

of Chinese patients within the keratinizing squamous cell carcinoma histology. *Ann Oncol* **18**, 29-35 (2007).
14. Tse, L.A., Yu, I.T., Mang, O.W. & Wong, S.L. Incidence rate trends of histological subtypes of nasopharyngeal carcinoma in Hong Kong. *Br J Cancer* **95**, 1269-73 (2006).
15. Friborg, J.T. et al. A prospective study of tobacco and alcohol use as risk factors for pharyngeal carcinomas in Singapore Chinese. *Cancer* **109**, 1183-91 (2007).
16. Yu, M.C., Garabrant, D.H., Huang, T.B. & Henderson, B.E. Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China. *Int J Cancer* **45**, 1033-9 (1990).
17. Li, C.C., Yu, M.C. & Henderson, B.E. Some epidemiologic observations of nasopharyngeal carcinoma in Guangdong, People's Republic of China. *Natl Cancer Inst Monogr* **69**, 49-52 (1985).
18. Ng, T.P. A case-referent study of cancer of the nasal cavity and sinuses in Hong Kong. *Int J Epidemiol* **15**, 171-5 (1986).
19. Yu, M.C., Ho, J.H., Lai, S.H. & Henderson, B.E. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res* **46**, 956-61 (1986).
20. Zheng, X., Yan, L., Nilsson, B., Eklund, G. & Drettner, B. Epstein-Barr virus infection, salted fish and nasopharyngeal carcinoma. A case-control study in southern China. *Acta Oncol* **33**, 867-72 (1994).
21. Zou, J. et al. A case-control study of nasopharyngeal carcinoma in the high background radiation areas of Yangjiang, China. *J Radiat Res (Tokyo)* **41 Suppl**, 53-62 (2000).
22. Yuan, J.M. et al. Non-dietary risk factors for nasopharyngeal carcinoma in Shanghai, China. *Int J Cancer* **85**, 364-9 (2000).
23. Chen, C.J. et al. Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency. *Anticancer Res* **10**, 547-53 (1990).
24. Cheng, Y.J. et al. Cigarette smoking, alcohol consumption and risk of nasopharyngeal carcinoma in Taiwan. *Cancer Causes Control* **10**, 201-7 (1999).
25. Armstrong, R.W. et al. Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles, formaldehyde and heat. *Int J Epidemiol* **29**, 991-8 (2000).
26. Sriamporn, S., Vatanasapt, V., Pisani, P., Yongchaiyudha, S. & Rungpitarangsri, V. Environmental risk factors for nasopharyngeal carcinoma: a case-control study in northeastern Thailand. *Cancer Epidemiol Biomarkers Prev* **1**, 345-8 (1992).

27. Armstrong, R.W., Armstrong, M.J., Yu, M.C. & Henderson, B.E. Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese. *Cancer Res* **43**, 2967-70 (1983).
28. Shanmugaratnam, K., Tye, C.Y., Goh, E.H. & Chia, K.B. Etiological factors in nasopharyngeal carcinoma: a hospital-based, retrospective, case-control, questionnaire study. *IARC Sci Publ*, 199-212 (1978).
29. Ammor, S., Baali, A., Hubert, A. & Cherkaoui, M. Epidemiological approach of the nasopharyngeal carcinoma in the region of Marrakech. *Antropo* **5**, 39-47 (2003).
30. Hashibe, M. et al. Epidemiologic review of marijuana use and cancer risk. *Alcohol* **35**, 265-75 (2005).
31. Armstrong, R.W. et al. Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. *Int J Cancer* **77**, 228-35 (1998).
32. 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* **69**, 508-514 (1994).
33. Chen, D.L. & Huang, T.B. A case-control study of risk factors of nasopharyngeal carcinoma. *Cancer Lett* **117**, 17-22 (1997).
34. Jeannel, D. et al. Diet, living conditions and nasopharyngeal carcinoma in Tunisia--a case-control study. *Int J Cancer* **46**, 421-425 (1990).
35. Feng, B.J. et al. Dietary risk factors for nasopharyngeal carcinoma in Maghrebian countries. *Int J Cancer* **121**, 1550-5 (2007).
36. Dallongeville, J., Marecaux, N., Fruchart, J.C. & Amouyel, P. Cigarette smoking is associated with unhealthy patterns of nutrient intake: a meta-analysis. *J Nutr* **128**, 1450-7 (1998).

Chapter 3

Genetic Contributions to Nasopharyngeal Carcinoma



Chapter 3.1

Familial Risk and Clustering of Nasopharyngeal Carcinoma in Guangdong, China

Familial Risk and Clustering of Nasopharyngeal Carcinoma in Guangdong, China

Wei-Hua Jia, M.D., Ph.D.[1]
Bing-Jian Feng, M.S.[1]
Zong-Li Xu, Ph.D.[2]
Xiao-Shi Zhang, M.D., Ph.D.[1]
Ping Huang[1]
Li-Xi Huang[1]
Xing-Juan Yu[1]
Qi-Sheng Feng[1]
Ming-Hong Yao[3]
Yin Yao Shugart, Ph.D.[4]
Yi-Xin Zeng, M.D., Ph.D.[1]

[1] Department of Experimental Research, Cancer Center, Sun Yat-Sen University, Guangzhou, China.

[2] School of Public Health, Sun Yat-Sen University, Guangzhou, China.

[3] Department of Information, Cancer Center, Sun Yat-Sen University, Guangzhou, China.

[4] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland.

Supported by a grant from the Chinese National Natural Science Foundation (30000141) and by Chinese National Key Basic Research Project (Project 973; G19980510) and Chinese National Science and Technique Project (2002BA711A03) grants.

Address for reprints: Yi-Xin Zeng, M.D., Ph.D., Cancer Center, Sun Yat-Sen University, 651 Dong-Feng Road East, Guangzhou 510060, China; Fax: (011) 86 2087343295; E-mail: yxzeng@gzsums.edu.cn

Received October 13, 2003; revision received March 19, 2004; accepted April 20, 2004.

BACKGROUND. Previous studies have suggested that genetic susceptibility may play an important role in the etiology of nasopharyngeal cancer (NPC). However, to date, few large-scale studies have been conducted on familial risk and clustering of NPC in a high-risk area of China.

METHODS. In the current study, 2252 patients with NPC who were treated at the Cancer Center of Sun Yat-Sen University in Guangdong Province, China, were identified as probands. Family histories of NPC and other malignancies were observed in first-degree relatives (FDRs) and second-degree relatives, and other information was obtained through interviews. One thousand nine hundred and three Cantonese families were selected for further investigation. To assess familial aggregation, the authors used standardized incidence ratios (SIRs) to measure the risk of NPC for FDRs and compared the observed number of cases with the number predicted by population-based frequencies in the Cantonese population of Hong Kong.

RESULTS. The current analysis indicated that families with ≥ 3 relatives who had NPC were distributed predominantly among a high-risk subgroup of the Cantonese population in Guangdong Province and that the frequency of these families was 0.68%. An SIR of 2.09 (95% confidence interval [CI], 1.80–2.40) was observed among 13,833 FDRs in the high-risk subgroup, and a significantly elevated risk for NPC was observed in FDRs of probands with early age of onset (age < 40 years; SIR, 9.01 [95% CI, 6.10–13.30]). Furthermore, decreased risks of hepatic, lung, esophageal, gastric, and breast carcinoma, as well as malignancy of all sites, were observed among FDRs of probands with NPC when Hong Kong and Shanghai populations were used as reference groups.

CONCLUSIONS. NPC tends to aggregate in Cantonese families in Guangdong Province, and the malignancies in these families appear to be site specific, with no excess of any other malignancy. ***Cancer* 2004;101:363–9.**

KEYWORDS: nasopharyngeal carcinoma, susceptibility, familial nasopharyngeal carcinoma, standardized incidence ratio.

Nasopharyngeal carcinoma (NPC) has a striking geographic and ethnic distribution. The incidence rates are low throughout most of the world, usually < 1 per 100,000 population.[1] However, NPC is among the leading causes of death among the Cantonese population and several other southern Chinese populations.[2] In fact, the incidence rate per 100,000 males is > 40 in some Cantonese-speaking populations of Guangdong Province. It is noteworthy that the rate is similarly high among Cantonese emigrants in countries such as Singapore, Malaysia, Australia, and the United States.[3–6]

Whereas familial aggregation of NPC is uncommon among low-risk or non-Chinese populations,[7–26] it has been documented widely

in the Chinese population.[27–35] One reported family living in Nanhai County of Guangdong Province has 15 family members affected with NPC, 2 family members affected with hepatic cell carcinomas, and 1 family member with breast carcinoma. In southern Chinese populations, several case–control studies have confirmed that the proportion of NPC cases with positive first-degree family history is $> 5\%$, with similar proportions found in high-risk areas such as Hong Kong (7.2%), Yulin (6.0%), and Guangzhou (5.9%).[28,29] In summary, the geographic and ethnic distribution and the strong evidence for familial aggregation suggest that genetic susceptibility may play an important role in the etiology of NPC.

Previous studies on familial clustering have focused mainly on individual reports of specific families or have been case–control epidemiologic studies that were limited in size. The objective of the current study was to determine the risk of NPC and other malignancies in first-degree relatives (FDRs) of probands with NPC in a Cantonese population. In total, 2252 families that were ascertained through probands who were treated at the Cancer Center of Sun Yat-Sen University (CCSYSU; Guangzhou, China) were enrolled. Standard incidence ratios (SIRs) were calculated to estimate the risk of malignancy.

MATERIALS AND METHODS

Participants

Between January 1999 and June 2001, we visited patients with NPC who received treatment in the Department of NPC at CCSYSU, which is the largest center for cancer prevention and treatment in southern China. CCSYSU is located in Guangzhou, the capital of Guangdong Province, and draws patients with NPC from all areas of Guangdong Province. Because each diagnosed patient has a unique personal identifier assigned by the hospital, we were able to obtain patient information upon their enrollment at CCSYSU. After we received permission from their physicians, we contacted the patients regarding participation in the study.

Approximately 5000 patients were eligible during the study period. Each day, we recruited approximately five patients who initially were treated at the Department of NPC, provided that they were eligible and agreed to participate, regardless of their age, gender, dialect, education, geographic location, family history of malignancy, or communication skills. Once the patients consented to join the study, we interviewed them in the hospital. By June 2001, we had completed 2369 interviews. Only 117 of 2369 probands could not answer the questions qualitatively because of severe side effects associated with radiotherapy. Therefore, 2252 probands (95.1%) were enrolled in the study. For the collection process described, we did not anticipate that any possible bias associated with this process would greatly affect the results of a familial aggregation study.

Each patient's diagnosis of NPC was confirmed pathologically, and each consenting patient with NPC was regarded as a proband. All families were ascertained through a single proband. In instances in which more than one family member was receiving treatment for NPC, the first enrolled patient was selected as the proband, and a single questionnaire was administered to avoid duplication of family information.

Interview

A structured questionnaire was administered to the proband in each family; patients were asked to answer verbally. Information on the probands themselves, along with family history for three successive generations (FDRs and second-degree relatives), was obtained. Whenever probands were unsure of their answers, we interviewed additional family members to ensure the accuracy of the information. Questions included the proband's contact information, date of birth, gender, date of diagnosis, birthplace, physician, dialect, and other general characteristics. Family history included any incidences of malignant disease among the proband's relatives, including cause and date of death for relatives who had died, status of occurrence of malignancy, type of malignancy, date of diagnosis, hospital at which diagnosis was made, and records of diagnosis. A pedigree structure was drawn for each family.

Three terms are used in the current report: *blood relative*, *FDR*, and *second-degree relative*. *Blood relatives* are individuals who are related to a proband by blood; in the current study, they included FDRs and second-degree relatives. *FDRs* included parents, siblings, sons, and daughters. Finally, *second-degree relatives* included grandparents, uncles, aunts, nephews, and nieces.

Statistical Methods

SIRs[36] were used to describe the NPC risk for 13,833 FDRs of patients with NPC compared with the general Cantonese population. Each SIR was calculated as the observed number of patients with NPC compared with the expected number derived using age-specific (5-year age groups) and gender-specific rates. For each family, we counted cases of malignant disease and person-years, which were calculated from the date of birth to the date of death, date of diagnosis of NPC, or date on which we visited. The expected number of cases was calculated by multiplying the stratum-spe-

TABLE 1
Number of Nasopharyngeal Carcinoma Cases in All 2252 Families and in 1903 Cantonese Families

No. of NPC cases per family	No. of families (%)[a]					
	Total		High-risk subgroup		Low-risk subgroup	
	In blood relatives	In FDRs	In blood relatives	In FDRs	In blood relatives	In FDRs
1	1944 (86.3)	2045 (90.8)	1629 (85.6)	1710 (89.9)	315 (90.3)	335 (96.0)
2	268 (11.9)	194 (8.61)	235 (12.3)	180 (9.46)	33 (9.46)	14 (4.01)
3	33 (1.47)	12 (0.53)	32 (1.68)	12 (0.63)	1 (0.29)	0 (0.0)
≥ 4	7 (0.31)	1 (0.04)	7 (0.37)	1 (0.05)	0 (0.0)	0 (0.0)

NPC: nasopharyngeal carcinoma; FDRs: first-degree relatives.

[a] Values in parentheses indicate the percentage of the total number of families.

cific rates of NPC in the reference population by the corresponding numbers of person-years in the study cohort and then summing over all calendar strata. Confidence limits of SIR estimates were calculated using the formula developed by Byar.[36] Because the incidence rates of NPC in Hong Kong are similar to those observed in Guangdong Province, we used Hong Kong Chinese data (based on a report by the International Agency for Research on Cancer [IARC]) as our reference population.

RESULTS

Frequency and Distribution of Familial NPC

In the current study, 2252 probands were ascertained—1597 males and 655 females, for a male:female ratio of 2.4:1.0. There were three major ethnic groups: the Cantonese, the Minnan, and the Hakka. The Cantonese population is native, whereas the other two populations emigrated from northern or eastern China hundreds of years ago. Different levels of NPC risk are observed in these three populations. The Cantonese ethnic group and their geographic distribution are consistent with a high risk of NPC. Based on data from a large-scale survey in China in 1980,[37] we divided patients into high-risk and low-risk subgroups of probands based on both the homeland of the probands' grandfathers and their dialect. Using this method, 1903 families were placed in the high-risk subgroup, and 349 families were placed in the low-risk subgroup. All 1903 families were Cantonese. In the high-risk subgroup, 274 patients (14.4%) and 193 patients (10.1%) had family histories of NPC in blood relatives and FDRs (Table 1), respectively; in the low-risk subgroup, 9.7% and 4.0% had family histories of NPC, respectively.

We also found that families with multiple cases were observed predominantly in the high-risk subgroup. Of the 40 families that had at least 3 NPC cases among blood relatives, 39 (2.05%) were in the high-risk subgroup, and only 1 (0.29%) was in the low-risk subgroup. The chi-square value for this linear trend was 7.30 ($P = 0.007$).

Association of Age and Other Specific Factors with Sporadic NPC and Familial NPC

In the current study, we defined *familial NPC* as NPC in an individual with at least 1 affected member among blood relatives. For the purposes of examining differences between familial NPC and sporadic NPC, we compared the characteristics of patients affected by the former type of NPC with the characteristics of patients affected by the latter type. Familial malignancies often had an earlier age of onset. Our results also suggested that the age of NPC occurrence was earlier for probands in families with multiple NPC cases. The average ages at onset were 40.43 years, 41.22 years, 44.36 years, and 44.48 years, respectively, for Cantonese probands from families with ≥ 4 cases, 3 cases, 2 cases, and 1 case of NPC among blood relatives (Table 2). Thus, there also appeared to be a trend toward earlier onset in families with greater numbers of cases, although this trend was not statistically significant ($P = 0.273$).

SIRs in FDRs

SIRs were used to estimate the familial risk of malignancy for FDRs. In total, 1903 Cantonese families were drawn from the high-risk areas of Guangdong (e.g., Sihui, Zhongshan, etc.), where the incidence rate of NPC is similar to the rate in Hong Kong. The age-specific rates are 27.49, 25.69, and 26.77 per 100,000 males for Sihui, Zhongshan, and Hong Kong, respectively, and 10.51, 10.51, and 10.43 per 100,000 females, respectively.[38,39] FDRs were analyzed to estimate the

TABLE 2
Age of Onset in 1903 Cantonese Probands with Nasopharyngeal Carcinoma versus the Number of Affected Blood Relatives[a]

No. of affected relatives	Mean age of onset (yrs)	No. of families	SD
1	44.48	1629	10.72
2	44.36	235	10.63
3	41.22	32	10.03
4	40.43	7	4.69
Total	44.39	1903	10.69

SD: standard deviation.
[a] $P = 0.273$.

TABLE 3
Numbers of Specific Types of First-Degree Relatives Along with Observed and Expected Numbers of Cases of Nasopharyngeal Carcinoma, with Standardized Incidence Ratios and 95% Confidence Intervals for First-Degree Relatives, in 1903 Cantonese Families with Familial Nasopharyngeal Carcinoma[a]

Relative	No.	O	E	SIR (95% CI)
Brother	3504	63	29.0	2.17 (1.70–2.78)
Sister	3478	33	11.4	2.91 (2.07–4.08)
Father	1903	56	38.5	1.46 (1.12–1.89)
Mother	1903	43	15.6	2.76 (2.05–3.72)
Total	10,788	195	94.5	2.06 (1.78–2.37)
All FDRs[b]	13,833	200	95.8	2.09 (1.81–2.40)
Age of proband (yrs)				
≥ 40	9560	175	93.0	1.88 (1.61–2.17)
< 40	4273	25	2.77	9.01 (6.10–13.30)

O: observed cases; E: expected cases; SIR: standard incidence ratio; 95% CI: 95% confidence interval; FDRs: first-degree relatives.
[a] A Hong Kong Chinese cohort (from the International Agency for Research on Cancer report[39]) was used as the reference population.
[b] Including sons and daughters.

risk of NPC among the high-risk population. The current study population included 13,833 FDRs of 1903 probands with NPC, including 3504 brothers, 3478 sisters, 3806 fathers or mothers, and 3045 sons or daughters (Table 3). There were 200 observed NPC malignancies, compared with 95.8 expected malignancies, yielding an SIR of 2.09 (95% confidence interval [CI], 1.81–2.40) and indicating a significantly increased risk in the high-risk subgroup. The SIRs for brothers, sisters, fathers, and mothers were analyzed separately, revealing significantly increased risks of 2.17, 2.91, 1.46, and 2.76, respectively (Table 3).

To determine whether the risk of NPC among FDRs differed between the families of probands with early-onset NPC and the families of probands with late-onset NPC, we stratified families according to the age of the proband at the time of NPC diagnosis (age ≥ 40 years or age < 40 years). A significantly elevated risk (SIR, 9.01 [95% CI, 6.10–13.30]) was observed for families in which probands had early-onset NPC, and a decreased risk (SIR, 1.88 [95% CI, 1.61–2.17]) was observed for families in which probands had late-onset NPC (Table 3).

Other Malignancies

A total of 243 other malignancies were observed in FDRs in the high-risk Cantonese subgroup, including 69 hepatic, 55 lung, 25 gastric, 9 esophageal, and 8 breast carcinomas and 77 miscellaneous malignancies (Table 4). There have been few previous reports of the incidence of malignant disease in Guangdong Province. Therefore, we used incidence rates for Hong Kong and Shanghai as a reference for estimating malignancy risks based on the data presented above. There were 69 observed cases of hepatic carcinoma (the second most common malignancy in the current study), compared with 112.42 and 92.86 expected cases (based on the Hong Kong and Shanghai data sets, respectively), suggesting a decreased risk for FDRs in the current study (SIRs, 0.61 [95% CI, 0.48–0.78] and 0.74 [95% CI, 0.58–0.94], respectively). The same trend was observed for lung, gastric, esophageal, and breast malignancies and for malignant disease at all sites.

DISCUSSION

The etiology of NPC has been investigated for more than half a century, and it is generally considered multifactorial, with environmental variables, Epstein–Barr virus (EBV) infection, and genetic factors involved in its pathogenesis. Recently, we reported the results of a genomewide search performed in families from Guangdong Province, China, who were at high risk for NPC.[40] Our findings provided evidence of a major susceptibility locus for NPC on chromosome 4 in a subset of these families. However, to our knowledge, few large-scale surveys of familial NPC in high-risk areas have been performed to date, and no previous study has measured NPC risk in family members of probands with NPC.

In the current study, cases were ascertained from CCSYSU. The majority of patients (95%) were from Guangdong Province and lived in cities, counties, and villages throughout the province. For this reason, the probands with NPC in the current study can be considered representative of the Guangdong population. Because there is no established cancer registry in Guangdong Province, we were unable to obtain age-specific incidence rates of NPC for the Guangdong population. We used the incidence rates reported for the Hong Kong Chinese population as an alternative (IARC, 1988–1992),[39] because Hong Kong is adjacent

TABLE 4
Number of Observed Tumors by Site, with Standardized Incidence Ratios and 95% Confidence Intervals, among All First-Degree Relatives in 1903 Cantonese Families with Familial Nasopharyngeal Carcinoma

Site	Father	Mother	Brother	Sister	Offspring	Total cases observed	SIR1 (95% CI)[a]	SIR2 (95% CI)[b]
Liver	40	9	16	3	1	69	0.61 (0.48–0.78)	0.74 (0.58–0.94)
Lung	30	14	8	3	0	55	0.24 (0.18–0.31)	0.35 (0.26–0.45)
Stomach	14	8	1	2	0	25	0.39 (0.25–0.57)	0.16 (0.11–0.24)
Esophagus	7	0	1	1	0	9	0.23 (0.11–0.44)	0.26 (0.12–0.49)
Breast	0	6	0	2	0	8	0.08 (0.04–0.17)	0.10 (0.05–0.21)
All sites	172	104	103	56	8	443	0.35 (0.32–0.38)	0.48 (0.44–0.53)

SIR: standard incidence ratio; 95% CI: 95% confidence interval.

[a] SIR1 was calculated using Hong Kong Chinese data as the reference.

[b] SIR2 was calculated using Shanghai data (from the International Agency for Research on Cancer report[39]) as the reference.

to Guangdong Province, and their historical link and shared Cantonese dialect suggest common population origins. Thus, the IARC report on cancer incidence in the Hong Kong population was a valid comparative data set for analysis of malignancy risk in the Guangdong population.

Most notable was the observation of a significantly elevated risk of NPC in FDRs of probands with NPC. This finding is consistent with the results of previous case–control studies that were limited in size. Yu et al.[28,29] reported that family history of NPC in an FDR was a significant risk factor for NPC in Cantonese individuals from Guangzhou or Hong Kong, with odds ratios of 6.0 and 4.5, respectively. However, these results were based on small sample sizes (case:control ratio, 306:306 and 250:250 for Guangzhou and Hong Kong, respectively), and the 95% CIs were very large. The current study included 1903 cases, suggesting that our results may provide more accurate estimates of the familial risk of NPC compared with the previous studies. Another notable finding was that the risk of NPC was 9 times greater in FDRs of probands who were affected with NPC before age 40 years. This observation provides evidence that genetic susceptibility may contribute to the occurrence of malignant disease in this small subgroup of families.

Typically, hereditary carcinoma syndromes caused by a single mutated susceptibility gene are characterized by the cooccurrence of different malignancies within pedigrees. Examples include breast and ovarian carcinoma clusters caused by mutations in the *BRCA1* and *BRCA2* genes and combinations of colorectal, endometrial, gastric, and ovarian carcinomas resulting from mutations in DNA mismatch repair genes in families affected with hereditary nonpolyposis colon carcinoma. Although little information on NPC is available in the literature, there is some evidence that other malignancies, such as salivary gland carcinomas and Burkitt lymphoma, may be associated with the development of NPC.[35] However, these malignancies all can be traced to the involvement of EBV in NPC pathogenesis. In the current study, we examined the occurrence of other malignancies in FDRs of probands with NPC in a high-risk subgroup and found no significantly increased risks. The SIRs for all other malignancies examined were less than 1.0 relative to the Hong Kong and Shanghai reference populations. We used the Shanghai report in addition to the Hong Kong report because some malignancies are heavily influenced by environmental factors, such as diet, smoking, and socioeconomic level. Socially, Shanghai may be more similar than Hong Kong to Guangdong Province. Although this was a fairly crude analysis, our results suggest that no increased risk of any other malignancy accompanies NPC, indicating that the familial risk appears to be site-specific.

We also found that the distribution of multicase families was quite unbalanced between the high-risk and low-risk populations. This heterogeneous risk has not been reported previously, and it suggests that the strength of genetic effects on the development of NPC varies among different populations in Guangdong Province. In fact, the genetic backgrounds of the three major ethnic populations of Guangdong are fairly heterogeneous, as was discussed previously. The native Cantonese population exhibited the highest incidence of NPC, whereas the Minnan and Hakka populations had lower risks.[37]

In the current study, families with 2 cases of NPC in FDRs were common (representing 9.2% of all families), but greater numbers of affected individuals in 1 family were less common ($< 1\%$), and there was no evidence of pedigrees in which NPC segregated in a simple Mendelian fashion. NPC typically is regarded

as a complex trait and is believed to be influenced by complex interactions among multiple genes and environmental factors. Indeed, although we have provided a crude description of disease frequency in the current work, future investigations will need to consider parameters, such as gene frequency and penetrance, that can be used in linkage analysis.

Early age of onset has been associated with familial breast carcinoma, ovarian carcinoma, and retinoblastoma.[41–43] Previously, we reported this same phenomenon in NPC based on our analysis of 32 Cantonese-speaking families from Guangdong Province. In that study, an average of 4–5 cases of NPC were observed in each family, and the average age of patients at diagnosis was 35.52 years—much younger than the average age of onset for sporadic cases diagnosed in outpatients at CCSYSU (46.6 years).[35] In the current study, the average age of onset decreased from 44.48 years to 40.43 years as the number of cases of NPC per family increased from 1 to 4, although the absolute trend did not achieve statistical significance. In Taiwan, a moderate-risk area with respect to NPC, a study conducted by Ung et al.[30] (involving 25 familial cases and 350 sporadic cases) found no significant difference between familial cases and sporadic cases in terms of age, ethnicity, disease histology, disease stage, or family history of other malignancies. In the current study, which involved 193 familial cases and 1710 sporadic cases, we detected differences in birthplace and age of onset between patients with familial NPC and patients with sporadic NPC, although clinical and pathologic characteristics appear to be similar in both groups of patients.

REFERENCES

1. Parkin DM, Muir CS. Cancer incidence in five continents. Comparability and quality of data. *IARC Sci Publ.* 1992;120: 45–173.
2. Jeannel D, Bouvier G, Huber A. Nasopharyngeal carcinoma: an epidemiological approach to carcinogenesis. *Cancer Surv.* 1999;33:125–155.
3. Lee HP, Duffy SW, Day NE, Shanmugar K. Recent trends in cancer incidence among Singapore Chinese. *Int J Cancer.* 1988;42:159–166.
4. Prasad U, Rampal L. Descriptive epidemiology of nasopharyngeal carcinoma in peninsular Malaysia. *Cancer Causes Control.* 1992;3:179–182.
5. Burt RD, Vaughan TL, McKnight B. Descriptive epidemiology and survival analysis of nasopharyngeal carcinoma in the United States. *Int J Cancer.* 1992;52:549–556.
6. Grulich AE, McCredie M, Coates M. Cancer incidence in Asian migrants to New South Wales, Australia. *Br J Cancer.* 1995;71:400–408.
7. Coffin CM, Rich SS, Dehner LP. Familial aggregation of nasopharyngeal carcinoma and other malignancies. A clinicopathologic description. *Cancer.* 1991;68:1323–1328.
8. Levine PH, Pocinki AG, Madigan P, Bale S. Familial nasopharyngeal carcinoma in patients who are not Chinese. *Cancer.* 1992;70:1024–1029.
9. Poletti T, Capaldi E. Malignant neoplasms of the nasopharynx and simultaneous clinical occurrence in a mother and son. *Cancro.* 1965;18:132–140.
10. Bell RB, Maguda TA. Nasopharyngeal carcinoma in Caucasian siblings: report of two cases. *J Tenn Med Assoc.* 1970; 63:753–754.
11. Brown TM, Heath CW, Lang RM, Lee SK, Whalley BW. Nasopharyngeal cancer in Bermuda. *Cancer.* 1976;37:1464–1468.
12. Williams EH, de The G. Familial aggregation in nasopharyngeal carcinoma [letter]. *Lancet.* 1974;2:295–296.
13. Nevo S, Meyer W, Altman M. Carcinoma of nasopharynx in twins. *Cancer.* 1971;28:807–809.
14. Lanier AP, Bender TR, Tschopp CF, Dohan P. Nasopharyngeal carcinoma in an Alaskan Eskimo family: report of three cases. *J Natl Cancer Inst.* 1979;62:1121–1124.
15. Snow JB Jr. Carcinoma of the nasopharynx in children. *Ann Otol Rhinol Laryngol.* 1975;84:817–826.
16. Fischer RA, Wharam MD, Kashima HK. Nasopharyngeal carcinoma in siblings. *Entechnology.* 1979;58:93–97.
17. Joncas JH, Rioux E, Wastiaux JP, Leyritz M, Robillard L, Menezes J. Nasopharyngeal carcinoma and Burkitt's lymphoma in a Canadian family. I. HLA typing, EBV antibodies and serum immunoglobulins. *Can Med Assoc J.* 1976;115: 858–860.
18. Gajwani BW, Devereaux JM, Beg JA. Familial clustering of nasopharyngeal carcinoma. *Cancer.* 1980;46:2325–2327.
19. Chizh GI, Rozhkova NI, Dorfman SM. Cancer of the nasopharynx in blood brothers. *Vestn Otorinolaringol.* 1977;4:81–82.
20. Ferouz AS, Mohr RM, Paul P. Juvenile nasopharyngeal angiofibroma and familial adenomatous polyposis: an association? *Otolaryngol Head Neck Surg.* 1995;113:435–439.
21. Albeck H, Bentzen J, Ockelmann HH, Nielsen NH, Bretlau P, Hansen HS. Familial clusters of nasopharyngeal carcinoma and salivary gland carcinomas in Greenland natives. *Cancer.* 1993;72:196–200.
22. Jung PF. Familial tendency of nasopharyngeal carcinoma: a report of cases. *Pac Med Surg.* 1965;73:242–243.
23. Fischer A, Fischer GO, Cooper E. Familial nasopharyngeal carcinoma. *Pathology.* 1984;16:23–24.
24. Albeck H, Nielsen NH, Hansen HE, et al. Epidemiology of nasopharyngeal and salivary gland carcinoma in Greenland. *Arctic Med Res.* 1992;51:189–195.
25. Schimke RN, Collins D, Cross D. Nasopharyngeal carcinoma, aplastic anemia, and various malignancies in a family: possible role of Epstein-Barr virus. *Am J Med Genet.* 1987;27:195–202.
26. Li FP. Familial Burkitt lymphoma and nasopharngeal carcinoma. *Lancet.* 1976;1:687–688.
27. Chen DL, Huang TB. A case-control study of risk factors of nasopharyngeal carcinoma. *Cancer Lett.* 1997;117:17–22.
28. Yu MC, Ho JH, Lai SH, Henderson BE. Cantonese style salted fishes as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res.* 1986;46:956–961.
29. Yu MC, Garabrant DH, Huang TB, Henderson BE. Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China. *Int J Cancer.* 1990;45:1033–1039.

30. Ung A, Chen CJ, Levine PH, et al. Familial and sporadic cases of nasopharyngeal carcinoma in Taiwan. *Anticancer Res.* 1999;19:661–665.
31. Yuan JM, Wang XL, Xiang YB, et al. Non-dietary risk factors for nasopharyngeal carcinoma in Shanghai, China. *Int J Cancer.* 2000;85:364–369.
32. Chen CJ, Liang KY, Chang YS, et al. Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency. *Anticancer Res.* 1990;10:547–553.
33. Zhang F, Zhang JM. Clinical hereditary characteristics in nasopharyngeal carcinoma through Ye-Liang's family cluster. *Chin Med J (Engl).* 1999;112:185–187.
34. Ko JY, Sheen TS, Hsu MM, Lui LT. Familial clustering of nasopharyngeal carcinoma. *Otolaryngol Head Neck Surg.* 1998;118:736–737.
35. Zeng YX, Jia WH. Familial nasopharyngeal carcinoma. *Semin Cancer Biol.* 2002;12:443–450.
36. Coleman MP, Hermon C, Douglas A. Person-Years (PYRS). A Fortran program for cohort study analysis. Lyon: IARC Press, 1989.
37. Li ZQ, Pan QC, Chen JJ. Epidemiology of nasopharyngeal carcinoma. In: Li ZQ, Pan QC, Chen JJ, editors. Nasopharyngeal carcinoma clinical and laboratory researches. Guangzhou: Guangdong Science and Technology Press, 1983:42–56.
38. Min HQ, Wang HM, Zhang EP, Hong MH, editors. Nasopharyngeal carcinoma research. Guangzhou: Guangdong Science and Technology Press, 1998.
39. Ferlay J, Black SL, Whelan SL, Parkin DM, editors. CI5VII: electronic database of cancer incidence in five continents. Volume VII. Lyon: IARC Press, 1997.
40. Feng BJ, Huang W, Shugart YY, et al. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat Genet.* 2002;31:395–399.
41. Hopper JL. Genetic epidemiology of female breast cancer. *Semin Cancer Biol.* 2001;11:367–374.
42. Durocher F, Tonin P, Shattuck-Eidens D, Skolnick M, Narod SA, Simard J. Mutation analysis of the BRCA1 gene in 23 families with cases of cancer of the breast, ovary, and multiple other sites. *J Med Genet.* 1996;33:814–819.
43. Scheffer H, te Meerman GJ, Kruize YC, et al. Linkage analysis of families with hereditary retinoblastoma: nonpenetrance of mutation, revealed by combined use of markers within and flanking the RB1 gene. *Am J Hum Genet.* 1989;45:252–260.

Chapter 3.2

Genome-wide Scan for familial
Nasopharyngeal Carcinoma reveals
Evidence of Linkage to Chromosome 4

Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4

Bing-Jian Feng[1*], Wei Huang[2,7*], Yin Yao Shugart[3*], Ming K. Lee[4*], Feng Zhang[5*], Jian-Chuan Xia[1], Hui-Yun Wang[1], Teng-Bo Huang[5], Shao-Wen Jian[1], Ping Huang[1], Qi-Sheng Feng[1], Li-Xi Huang[1], Xing-Juan Yu[1], Duang Li[1], Li-Zheng Chen[1], Wei-Hua Jia[1], Yan Fang[1], Hui-Ming Huang[5], Jing-Liu Zhu[1], Xiao-Ming Liu[2], Yan Zhao[2], Wang-Qing Liu[2], Mang-Quan Deng[5], Wei-Han Hu[6], Shao-Xiong Wu[6], Hao-Yuan Mo[5], Ming-Fang Hong[5], Mary Claire King[4], Zhu Chen[2] and Yi-Xin Zeng[1]

**These authors contributed equally to this work.*

Published online: 15 July 2002, doi:10.1038/ng932

Nasopharyngeal carcinoma (NPC) occurs with high frequency in Asian populations, especially among people of Cantonese ancestry. In areas with high incidence, NPC clusters in families, which suggests that both geography and genetics may influence disease risk[1–6]. Although the *HLA-Bw46* locus is associated with increased risk of NPC[7,8], no predisposing genes have been identified so far. Here we report the results of a genome-wide search carried out in families at high risk of NPC from Guangdong Province, China. Parametric analyses provide evidence of linkage to the *D4S405* marker on chromosome 4 with a logarithm of odds for linkage (lod) score of 3.06 and a heterogeneity-adjusted lod (hlod) score of 3.21. Fine mapping with additional markers flanking *D4S405* resulted in a lod score of 3.54 and hlod score of 3.67 for the region 4p15.1–q12. Multipoint nonparametric linkage analysis gives lod scores of 3.54 at *D4S405* ($P = 5.4 \times 10^{-5}$) and 4.2 at *D4S3002* ($P = 1.1 \times 10^{-5}$), which is positioned 4.5 cM away from *D4S405*. When Epstein–Barr virus antibody titer was included as a covariate, the lod scores reached 4.70 ($P = 2.0 \times 10^{-5}$) and 5.36 ($P = 4.36 \times 10^{-6}$) for *D4S405* and *D4S3002*, respectively. Our findings provide evidence of a major susceptibility locus for NPC on chromosome 4 in a subset of families.

Linkage analysis has been used to identify many important genes that predispose to cancer[9,10]. In this study, we aimed to recruit high-risk families with two or more affected individuals to search for regions of the genome showing significant linkage to inherited predisposition to NPC. We carried out a genome-wide scan for linkage in Cantonese-speaking families from Guangdong Province, China, where the highest incidence of NPC worldwide has been recorded[2,3]. We examined 20 families, with 2–9 affected members, whose characteristics are summarized in Table 1. The study included 65 affected individuals with an average age at diagnosis of 35.5 years. We genotyped 54 of these individuals using 382 polymorphic microsatellite markers covering 22 autosomes with an average marker density of 10 cM. These markers were taken from the ABI Prism Linkage Mapping Set (Version 2) and the Genethon human linkage map[11].

Assuming that there was at least one gene contributing to disease susceptibility, we carried out two types of analysis: the parametric method and the model-independent nonparametric method, using GENEHUNTER[12]. To improve the information obtained from the markers, multipoint linkage analysis was carried out using all markers on a chromosome as a group. The highest scores for the 22 chromosomes calculated from the 20 NPC pedigrees are shown in Figure 1. Marker *D4S405*, on chromosome 4p12–p15, yielded a maximum multipoint lod score of 3.06. After adjusting for heterogeneity, the hlod score of this locus increased to 3.21. A nonparametric linkage lod score (NPL) of 2.75 was obtained for *D4S405* with $P = 0.005$. We also carried out two-point linkage analysis with LINKAGE[13], which produced slightly lower scores than those from multipoint analyses (data not shown). These results suggested that an NPC susceptibility locus is linked to the short arm of chromosome 4.

Another locus, *D12S345* on chromosome 12, yielded multipoint NPL scores of 4.64 (Fig. 1) and 3.50 for two-point nonparametric analysis (data not shown). The results from parametric analyses did not show strong evidence for linkage of NPC to chromosome 12, and the multipoint hlod score was only 1.58 (Fig. 1); however, the nonparametric score represents possible evidence of a susceptibility locus in this region. Therefore, chromosome 12 was analyzed in more detail.

Table 1 • Characteristics of families with NPC genotyped in whole-genome scan

Category	Result
Number of families analyzed	20
Average number of cases per family (range)	3.25 (2–9)
Average number of genotyped cases per family (range)	2.7 (2–7)
Total number of affected individuals genotyped	54
Total number of unaffected individuals genotyped	111
Mean ± s.d. age at diagnosis (range)	35.5 ± 9.1 yr (20–60 yr)

[1]Cancer Institute, [5]Department of Nasopharyngeal Carcinoma and [6]Department of Radiotherapy, Cancer Center, Sun Yat-sen University, GuangZhou, China. [2]Chinese National Human Genome Center at Shanghai and [7]Rui-Jin Hospital and Health Science Center, Shanghai Second Medical University and Shanghai Institutes of Biological Science–Chinese Academy of Science, Shanghai, China. [3]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA. [4]Department of Medicine and Genetics, University of Washington, Seattle, USA. Correspondence should be addressed to Y.-X.Z. (e-mail: yxzeng@gzsums.edu.cn).

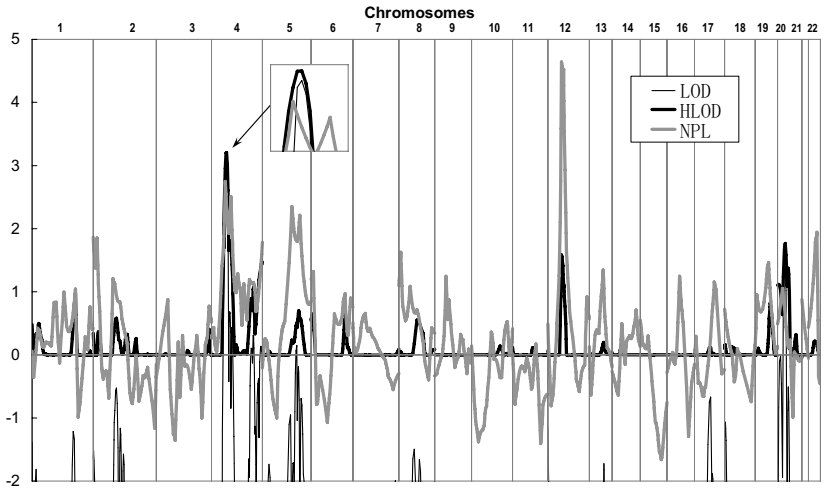


Fig. 1 The multipoint lod, hlod and NPL scores in genome-wide linkage analyses. Multipoint lod, hlod and NPL scores were calculated with GENEHUNTER[12] in a whole-genome scan using 386 polymorphic microsatellite markers covering 22 autosomes. Twenty Cantonese-speaking NPC pedigrees were examined from the Pearl River area, Guangdong Province, China.

To study further the locus on chromosome 4 with suggestive linkage to NPC, we genotyped additional markers spanning 20.06 cM from *D4S3001* to *D4S1592* in the region 4p15–q12 (Fig. 2) in the initial group of 20 NPC pedigrees plus 12 newly recruited Cantonese-speaking families with NPC. This analysis provided additional evidence for linkage in the 4p15.1–q12 region. A maximum two-point lod score of 3.54 and an hlod score of 3.67, calculated with the LINKAGE package, were obtained for *D4S405* (Table 2 and Fig. 3)[13–15]. The multipoint scores for *D4S405* were calculated with FASTLINK[16,17], taking a group of three markers at a time, and the maximum lod and hlod scores were 2.99 and 3.65, respectively (Table 2 and Fig. 3).

We also carried out nonparametric analyses and considered the effect of infection with Epstein–Barr virus (EBV). We used an affected-relative-pair approach with a conditional-logistic model[18] that had been modified further[19]. As described previously[19], the inclusion of a covariate (discrete or continuous) allows for linkage heterogeneity caused by that particular covariate. We used GENIBD and LODPAL from the SAGE package (version β10) in this analysis and calculated the *P* values by previously described methods[20].

The antibody titer of EBV viral capsid antigen (VCA)/IgA has been used for many years for clinical diagnosis and screening[2]. The multipoint lod scores given by the conditional-logistic model[18] with or without the VCA/IgA titer included as a covariate are presented in Table 2 and Fig. 3. Forty-eight affected-relative pairs were included in the analysis without the covariate (analysis a, lod^a), and forty-five pairs were included in the analysis with the covariate (analysis b, lod^b). Both analyses gave strong signals for linkage. In analysis a, the peak lod score (4.20) occurred at *D4S3002* ($P = 1.08 \times 10^{-5}$), which is roughly 4.5 cM away from *D4S405* (where the second-highest lod score occurred, 3.54, $P = 5.3 \times 10^{-5}$). In analysis b, the peak lod score (5.36) occurred at *D4S3002* ($P = 4.36 \times 10^{-6}$) and the second-highest lod score (5.32, $P = 4.76 \times 10^{-6}$) occurred at *D4S3045*, which is 1.5 cM away from *D4S405* (Table 2 and Fig. 3). For *D4S405*, the lod score reached 4.70, with $P = 2.0 \times 10^{-5}$.

To narrow down the most likely location of the predisposing locus, we constructed haplotypes of the affected families for the pedigrees that showed minimum recombination between markers in the region 4p15.1–q12. The recombinants narrowed the candidate region to a 14.21-cM segment on 4p15.1–q12 between markers *D4S2950* and *D4S2916* (Fig. 4*a*,*b*). The minimum shared region of affected members in two larger families, pedigrees 34 and 31, indicated that an NPC susceptibility gene is likely to be located between *D4S2950* and *D4S2916* (Fig. 4*a*,*b*). This is consistent with the results from both parametric and nonparametric analyses.

We used the HOMOG[21] program to estimate the proportion of families showing evidence of linkage. The results suggested that nearly 70% of the families studied had linkage to the 4p15.1–q12 region (data not shown). To increase precision, we constructed the most likely haplotype configurations of the pedigrees on the basis of genotyping analysis (data not shown). In addition to pedigrees 31 and 34, there were 12 families with at least three affected individuals or assumed carriers of whom half shared a disease-associated haplotype. For example, in family 41, five of six individuals with NPC for whom DNA samples were available shared a haplotype. The results of this haplotype study suggest that genetic heterogeneity might have a role in the etiology of NPC in the population studied.

To investigate the positive results obtained for chromosome 12 in the genome-wide scan, we carried out fine mapping with additional markers flanking *D12S345* that spanned 17.8 cM from *D12S1631* to *D12S359* in the region 12p11–q13 for the 20 NPC pedigrees in the original collection and the 12 newly recruited families with NPC. The maximum multipoint NPL score decreased from 4.61 for *D12S345* in the whole-genome scan to 3.17 in fine mapping for *D12S1668*, which is 3.2 cM away from *D12S345* (data not shown). Similarly, the maximum two-point NPL score also decreased from 3.50 to 3.47 for *D12S345*. More importantly, all the parametric analyses resulted in lod scores no higher than the threshold value of 3.0, and the highest two-point hlod score was 2.12 for *D12S1648*, which is 0.1 cM away from *D12S345* (data not shown).

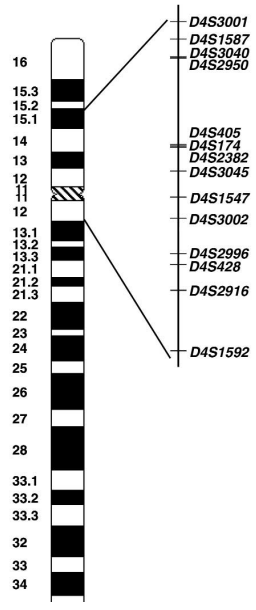


Fig. 2 Scale map of markers on chromosome 4 for fine mapping of the NPC susceptibility locus. Fourteen markers were selected from 4p15.1–q12 and two markers, *D4S405* and *D4S1592*, were used in the genome-wide scan. The distances between markers were plotted according to sex-averaged distances determined with Centre d'Etude du Polymorphisme Humain (CEPH) pedigree data.

Table 2 • Fine mapping on chromosome 4

		Parametric				Nonparametric			
cM	Marker	lod-L	hlod-L	lod-F	hlod-F	lod[a]	*P* value	lod[b]	*P* value
0.00	*D4S3001*	0.404	0.472	–1.013	0.933	0.296	0.243322	1.478	0.033263
1.06	*D4S1587*	1.333	1.387	–1.122	1.048	0.112	0.471978	1.043	0.090474
2.13	*D4S3040*	0.889	0.927	–0.466	1.344	0.114	0.468727	1.069	0.085390
2.13	*D4S2950*	0.882	0.882	–0.578	1.201	0.114	0.468727	1.069	0.085390
7.48	*D4S405*	3.543	3.666	2.989	3.652	3.542	0.000054	4.701	0.000020
7.48	*D4S174*	2.734	2.892	1.697	2.776	3.542	0.000054	4.696	0.000020
7.48	*D4S2382*	2.251	2.363	2.736	3.064	3.542	0.000054	4.696	0.000020
9.08	*D4S3045*	2.647	2.647	2.299	2.849	3.383	0.000079	5.323	0.000005
10.69	*D4S1547*	1.452	1.452	–2.302	1.058	3.396	0.000077	4.304	0.000050
11.95	*D4S3002*	1.002	1.039	–1.617	1.362	4.204	0.000011	5.361	0.000004
14.11	*D4S2996*	0.560	0.563	0.699	1.381	2.993	0.000205	4.043	0.000091
14.77	*D4S428*	1.308	1.308	0.491	1.286	2.397	0.000892	3.414	0.000386
16.34	*D4S2916*	0.488	0.488			1.028	0.029570	2.191	0.006439
20.06	*D4S1592*	0.316	0.316			0.632	0.088126	2.042	0.009070

Here lod-L and hlod-L are the scores from two-point linkage analysis computed by LINKAGE; lod-F and hlod-F are the scores from multipoint linkage analysis computed by FASTLINK; lod[a] is the score for conditional logistic analysis of affected-relative pairs without EBV titer as a covariate (analysis a); lod[b] is the score for conditional logistic analysis of affected-relative pairs including EBV titer as a covariate (analysis b).

By examining the pedigrees contributing to the high NPL score for *D12S345*, we found that pedigree 31 contributed excessively to the NPL scores (data not shown). Using data from *D12S345* and the flanking markers from fine mapping, we inferred haplotypes with SimWalk2 (ref. 22). The haplotype graph, however, did not show a segment that cosegregated with NPC in this pedigree, and is consistent with the low scores derived from parametric analyses (Fig. 4*c*). When we checked the haplotypes carefully, we found that all the affected members in pedigree 31 contained an allele '15' of *D12S345*, an allele '12' of *D12S1663* and an allele '12' of *D12S85*, all of which are relatively common. Therefore, the high NPL score may have been a result of sharing by state (Fig. 4*c*). To increase the information by inferring phases with multiple closely linked markers, we also carried out multipoint linkage analysis with these markers. A maximum hlod score of 0.803 was obtained for this region of chromosome 12 (data not shown). An inspection of the inferred haplotypes with fine-mapping markers indicated that the most likely explanation was that one of the affected siblings inherited a different haplotype from a founding grandparent together with an allele '15' from *D12S345*. Together, these results suggest that the probability of linkage to chromosome 12 is very low.

Previous reports have suggested an association between risk of NPC and certain HLA haplotypes[23]. A linkage study carried out on sibling pairs indicated the existence of a recessive gene for disease susceptibility close to the major histocompatibility complex (MHC) region of chromosome 6 (refs 7,8). Here, however, we did not observe any obvious linkage to chromosome 6 (Fig. 1), which contains the MHC locus. A possible explanation for the differences between those reports and our findings might be that a different group of subjects were studied: those studies were conducted with affected sib-pair families, whereas high-risk NPC pedigrees were examined here.

In summary, both parametric and nonparametric analyses indicate that a susceptibility locus on 4p15.1–q12 may account for a significant subset of hereditary NPC. This provides a solid basis for future studies to identify the NPC susceptibility gene.

Methods

Families. We recruited 32 families with NPC from the Pearl River area in Guangdong Province, China[2]. Twelve of these families were collected after the genome-wide scan was finished and were used only for fine mapping. Because there are three main dialect groups in the Guangdong with different migration origins and the Cantonese-speaking population showed a significantly higher incidence of NPC than the others, for this study we used only pedigrees from the Cantonese-speaking population. Most of these families were recruited from earlier epidemiological studies[2] or from medical records in the Cancer Center, Sun Yat-sen University, Guangzhou, China. Age at diagnosis of NPC was confirmed from medical records or other independent sources, and NPC was confirmed by pathological examination. We obtained aliquots of 5–10 ml of peripheral blood from as many families members as possible, with fully informed consent, and reviewed pathologic slides whenever available. Studies were approved by the ethical review committees of the appropriate institutions.

Genotyping analysis. We prepared genomic DNA from lymphoblastoid cell lines transformed with EBV and used for PCR. Genome-wide linkage analysis was carried out using fluorescently labeled primers (ABI Prism Linkage Mapping Sets, Version 2), which provide an average interval of 10 cM. The primers were labeled with one of three different fluorescent dyes, FAM, HEX and NED (PE Biosystems). The average heterozygosity of the markers selected for the study was 0.75. Each multiplex PCR reaction (5 µl in total) con-

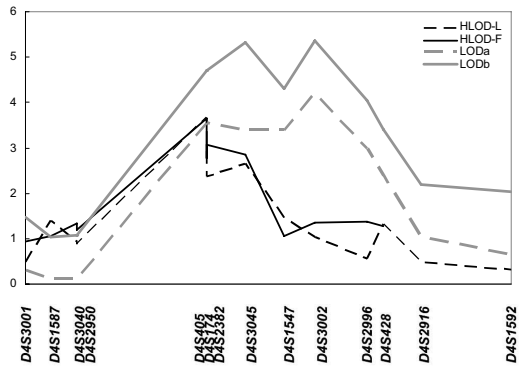
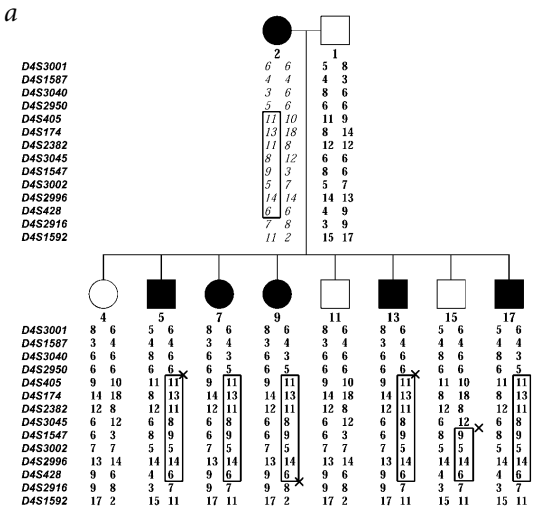
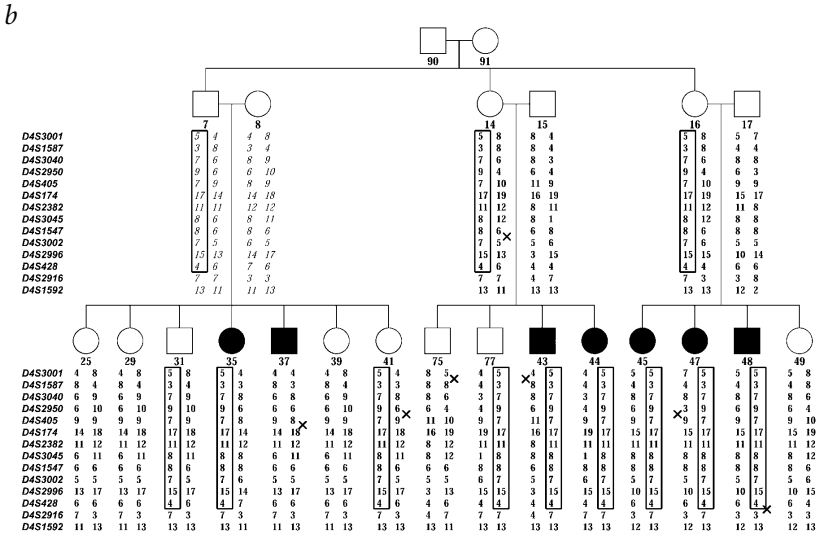
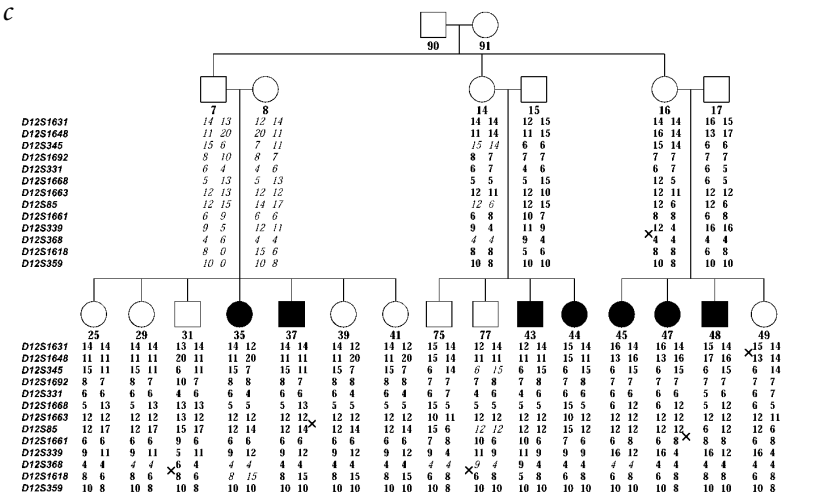


Fig. 3 The lod scores of multipoint and two-point analyses from fine mapping on chromosome 4. hlod-L is the two-point linkage scores calculated with LINKAGE[13–15], and hlod-F is the multipoint linkage scores calculated with FASTLINK[16–17] taking three markers at a time in sequence from the p arm to the q arm, with the distance specified between the markers based on the Marshfield map[24]. lod[a] and lod[b] are affected-relative-pair analysis using a conditional-logistic model[18] with modifications[19], in which the Epstein–Barr virus VCA/IgA antibody titer was included as a covariate for lod[b]. The distances between markers are plotted according to sex-averaged distances determined with CEPH pedigree data.







tained 20 ng of genomic DNA, 1× PCR buffer, 0.25 U of *Taq* DNA polymerase, 200 mM each dNTP, 3 mM $MgCl_2$ and 0.05 μM primers. We carried out PCR in a GeneAmp PCR 9700 instrument as follows: initial denaturing at 94 °C for 12 min; 14 cycles of 30 s at 94 °C, 1 min at 63 °C (decreasing by 0.5 °C per cycle), 90 s at 72 °C; 20 cycles of 30 s at 94 °C, 1 min at 55 °C, 90 s at 72 °C; and a final extension of 10 min at 72 °C. After amplification, we mixed the PCR products with loading buffer and GS-350 ROX standard, denatured the mixture and loaded it onto a 32-cm, denaturing 6% polyacrylamide gel. After 2 h of electrophoresis, the DNA segments were separated by a 377 DNA sequencer (Applied Biosytems). We used Genescan 3.0 (Applied Biosytems) for data collection and Genotyper 2.5 (Applied Biosystems) for microsatellite allele analysis. We checked all genotypes for mendelian segregation in pedigrees.

Linkage analyses. For the genome-wide scan, we used GENEHUNTER[12] to calculate multipoint parametric and nonparametric lod scores, with 386 polymorphic satellite markers in 20 NPC pedigrees of Cantonese speakers. For fine mapping, 12 NPC pedigrees of Cantonese speakers were added to both the parametric and the nonparametric analyses. For multipoint analyses in the genome-wide scan, we used all of the markers in a whole chromosome for computing with GENEHUNTER[12], whereas for multipoint parametric calculation in fine mapping, we used groups of three markers at a time with FASTLINK[16,17]. An affected-relative-pair analysis using a conditional-logistic model that allows the inclusion of covariates[18,19] was also carried out with the GENIBD, LODPAL program from SAGE (version β10). This type of analysis allowed us to examine genetic effects after adjusting for the possible effect of EBV infection. For both parametric and nonparametric analyses, allele frequencies were estimated from unrelated founders of the sampled pedigrees. Autosomal dominant inheritance was assumed for parametric analysis, with a disease-allele frequency of 0.0089. Because antibodies against EBV have been used as indices for early diagnosis of NPC[2], liability classes were assigned into two categories for those individuals in whom EBV antibodies were detected. If there was a high titer of EBV antibodies (VCA/IgA ≥1:80 or EA/IgG ≥0.18), the penetrances were set as 0.1 for sporadic cases and 1.0 for disease haplotype carriers. If the titer of EBV antibodies was low (VCA/IgA <1:80 and EA/IgG <0.18), the penetrances were set as 0.0 for sporadic cases and

Fig. 4 Pedigree structure and haplotypes of family 34 and family 31. Haplotypes were inferred with minimum recombination between markers by GENEHUNTER[12]. Filled symbols represent affected individuals. Inferred genotypes of individuals are shown in italics; other experimentally typed genotypes are in bold. Crosses represent recombinations. Boxes indicate the chromosome region shared by affected members of the pedigree. The markers used for genotyping are listed beside each pedigree. ***a,b,*** Pedigrees 34 (*a*) and 31 (*b*) genotyped with 14 markers from region 4p15.1–q12 of chromosome 4. ***c,*** Pedigree 31 genotyped with 13 markers from region 12p11–q13 of chromosome 12.

0.2 for disease haplotype carriers. If no information about EBV antibodies was available, a penetrance of 73% was assumed on the basis of previous epidemiological genetic data[2] and segregation analysis of a group of high-risk families (W.-H.J. *et al.*, unpublished data). Haplotypes were inferred with minimum recombination between markers. We used HOMOG[21] to determine the proportion of linked families with the assumed genetic model.

URL. SAGE analysis software, http://darwin.cwru.edu/pub/sage.html.

Acknowledgments

We thank colleagues in the Cancer Center, Sun Yat-sen University, for providing information about families with NPC; and K. Jacobs and B. Doan for help with preparing data and test-running the β10 version of SAGE. This work was supported by the China Medical Board, the National Key Basic Research Projects of China, the National Outstanding Youth Grant of China, the Key Projects of National Natural Science Foundation, the National High Technology Project, the Foundation of Guangdong Science and Technology Committee and the Shanghai Municipal Foundation for Sciences and Technologies.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 21 February; accepted 2 June 2002.

1. Vokes, E.E., Liebowitz, D.N. & Weichselbaum, R.R. Nasopharyngeal carcinoma. *Lancet* **350**, 1087–1091 (1997).
2. Huang, T.B. & Min, H.Q. in *Nasopharyngeal Carcinoma Research* (eds Min, H.Q., Wang, H.M., Zhang, E.P. & Hong, M.H) 6–12 (Guangdong Science and Technology Press, Guangzhou, China, 1998).
3. Chen, D.L. & Huang, T.B. A case-control study of risk factors of nasopharyngeal carcinoma. *Cancer Lett*. **117**, 17–22 (1997).
4. Fischer, A., Fisher, G.O. & Cooper, E. Familial nasopharyngeal carcinoma. *Pathology* **16**, 23–24 (1984).
5. Nevo, S., Mexer, W. & Altman, M. Carcinoma of nasopharynx in twins. *Cancer* **28**, 807–809 (1997).
6. Chan, S.H. Aetiology of nasopharyngeal carcinoma. *Ann. Acad. Med. Singapore* **19**, 201–207 (1990).
7. Lu, S.J. *et al.* Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature* **346**, 470–471 (1990).
8. Ooi, E.E., Ren, E.C. & Chan, S.H. Association between microsatellites within the human MHC and nasopharyngeal carcinoma. *Int. J. Cancer* **74**, 229–232 (1997).
9. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
10. Smith, J.R. *et al.* Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* **274**, 137–274 (1996).
11. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* **380**, 152–154 (1996).
12. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
13. Lathrop, G.M. & Lalouel, J.M. Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.* **36**, 460–465 (1984).
14. Lathrop, G.M., Lalouel, J.M., Julier, C. & Ott, J. Strategies for multilocus analysis in humans. *Proc. Natl Acad. Sci. USA* **81**, 3443–3446 (1984).
15. Lathrop, G.M., Lalouel, J.M. & White, R.L. Construction of human linkage maps: likelihood calculations for multilocus linkage analysis. *Genet. Epidemiol.* **3**, 39–52 (1986).
16. Cottingham, R.W. Jr., Idury, R.M. and Schaffer, A.A. Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53**, 252–263 (1993).
17. Schaffer, A.A., Gupta, S.K., Shriram, K. & Cottingham, R.W. Jr. Avoiding recomputation in linkage analysis. *Hum. Hered.* **44**, 225–237 (1994).
18. Olson, J.M. A general conditional-logistic model for affected-relative-pair linkage studies. *Am. J. Hum. Genet.* **65**, 1760–1769 (1999).
19. Goddard, K.A., Witte, J.S., Suarez, B.K., Catalona, W.J. & Olson, J.M. Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. *Am. J. Hum. Genet.* **68**, 1197–1206 (2001).
20. Self, S.G. & Liang, K.Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J. Am. Stat. Assoc.* **82**, 605–610 (1987).
21. Ott, J. Linkage probability and its approximate confidence interval under possible heterogeneity. *Genet. Epidemiol.* **1** (Suppl.), 251–257 (1986).
22. Sobel, E. & Lange, K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am. J. Hum. Genet.* **58**, 1323–1337 (1996).
23. Henderson, B.E. *et al.* Risk factors associated with nasopharyngeal carcinoma. *N. Engl. J. Med.* **295**, 1101–1106 (1976).
24. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. & Weber, J.L. Comprehensive human genetic map: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).

Chapter 3.3

The Susceptibility Gene for familial Nasopharyngeal Carcinoma is mapped on Chromosome 4p11-p14 by Haplotype Analyses

Chinese Science Bulletin 2003 Vol. 48 No. 21 2327—2330

The susceptibility gene for familial nasopharyngeal carcinoma is mapped on chromosome 4p11-p14 by haplotype analyses

CHEN Hankui, FENG Bingjian, LIANG Hui, ZHANG Ruhua & ZENG Yixin

Cancer Center, Sun Yet-sun University, Guangzhou 510060, China
Correspondence should be addressed to Zeng Yixin (e-mail: yxzeng@gzsums.edu.cn)

Abstract **In our previous study, one candidate susceptibility locus for familial nasopharyngeal carcinoma (NPC) has been defined to a 14.21-cM region on 4p15.1-q12, whereas the distal minimum boundary of this region remained to be further determined in respect that the two markers D4S2996 and D4S428 were uninformative. In the present study, we carried out a haplotype analysis to identify the exact boundary by using the combination of a set of microsatellite markers and single nucleotide polymorphism (SNP) markers in two major NPC families. We defined the exact distal boundary between D4S1577 and D4S3347, and consequently shortened the susceptibility locus to an 8.29-cM segment on 4p11-p14.**

Keywords: nasopharyngeal carcinoma, tumor susceptibility gene, chromosome 4p11-p14, haplotype analyses.

DOI: 10.1360/03wc0298

Nasopharyngeal carcinoma (NPC) occurs with high incidence in southern China, and is believed to be etiologically involved with Epstein-Barr virus infection, genetic susceptibility and environmental factors. Several efforts have been made to map the NPC susceptibility gene(s). A linkage study with sib-pairs collected from Guangdong and Guangxi indicates the existence of a recessive disease susceptibility gene close to the MHC region on chromosome 6, although HLA antigens themselves were not the cause of NPC[1]. In our previous study, we have mapped a NPC susceptibility locus on chromosome 4p15.1-q12 between microsatellite markers D4S2950 and D4S2916 by conducting a whole genome scan on 20 Cantonese-speaking NPC pedigrees[2]. However, the susceptibility region, 14.21 cM in length according to the Marshfield map, is too big for the cloning of the NPC gene. Of note, the boundary of this region was decided mainly in two NPC pedigrees, 31 and 34, by identifying the minimal microsatellite haplotype shared in affected individuals. The boundary of this susceptibility region close to telemere was defined between markers D42950 and D4S405 by the chromosome recombination in the cases 34-5 and 34-l3, and the distal boundary was determined between D4S428 and D4S2916 by the chromosome crossover occurring in the affected 34-9 and in the case 31-48 from the third nuclear family of pedigree 31. However, in both pedigrees, the two markers near to the distal borderline, namely D4S2996 and D4S428, were both uninformative in microsatellite haplotype analyses, since the parents 34-2 and 31-16 are both homozygous at the two sites. Therefore, the exact position of chromosome recombination might be anywhere between D4S3002 and D4S2916, indicating that this susceptibility region might be narrowed down by further haplotype analysis when more genetic markers are used, if only a high-resolution map of the markers is available. Moreover, it is worth defining the exact crossover position because the ambiguous region between D4S3002 and D4S2916 is more than 5 cM in length and is a relatively large part of the determined susceptibility region. In this study, we conducted a further haplotype analyses in the two families to determine the exact boundary with additional microsatellite markers and single nucleotide polymorphism (SNP) markers, aiming to narrow down this susceptibility locus for familial NPC.

1 Materials and methods

(ⅰ) Samples. Available genomic DNA samples of members from the NPC family 34 (34-2 except) and the third nuclear family of NPC pedigree 31 (Fig. 1) were prepared from lymphoblastoid cell lines transformed with Epsterin-Barr virus and used for microsatellite and SNP genotyping.

(ⅱ) Microsatellite genotyping. In addition to markers D4S3002, D4S2996, and D4S428, which had been genotyped in the previous study, six markers D4S3255, D4S1577, D4S3347, D4S2971, D4S1630, and D4S3254, obtained from NCBI UniSTS database (http://www.ncbi.nlm.nih.gov/entrez/unists), were used for genotyping. The primers were synthesized and labeled with one of the three different fluorescent dyes FAM, HEX and NED (Bioasia, Shanghai). Multiplex PCR reaction system (a total volume 5 μL) contained 20 ng of genomic DNA, 1×PCR buffer, 0.25 U of Hotstart Taq DNA polymerase (Qiagen), 200 μmol/L of each dNTP, 3 mmol/L of $MgCl_2$, 0.05 μmol/L of each primer. PCR was performed in the GeneAmp PCR9700 with the following parameters: initial denaturing at 95℃ for 15 min, followed by 14 cycles: 30 s at 94℃, 1 min at 63℃ (decreasing by 0.5℃ per cycle), 90 s at 72℃ and 20 cycles: 30 s at 94℃, 1 min at 56℃, 90 s at 72℃, with a final extension of 10 min at 72℃. After amplifications, The PCR products were mixed with loading buffer and GS-350 ROX Standard, denatured and then loaded onto a 32 cm, 6% denaturing polyacrylamide gel. After 2-h electrophoresis, the DNA segments were separated by 377 DNA Se-

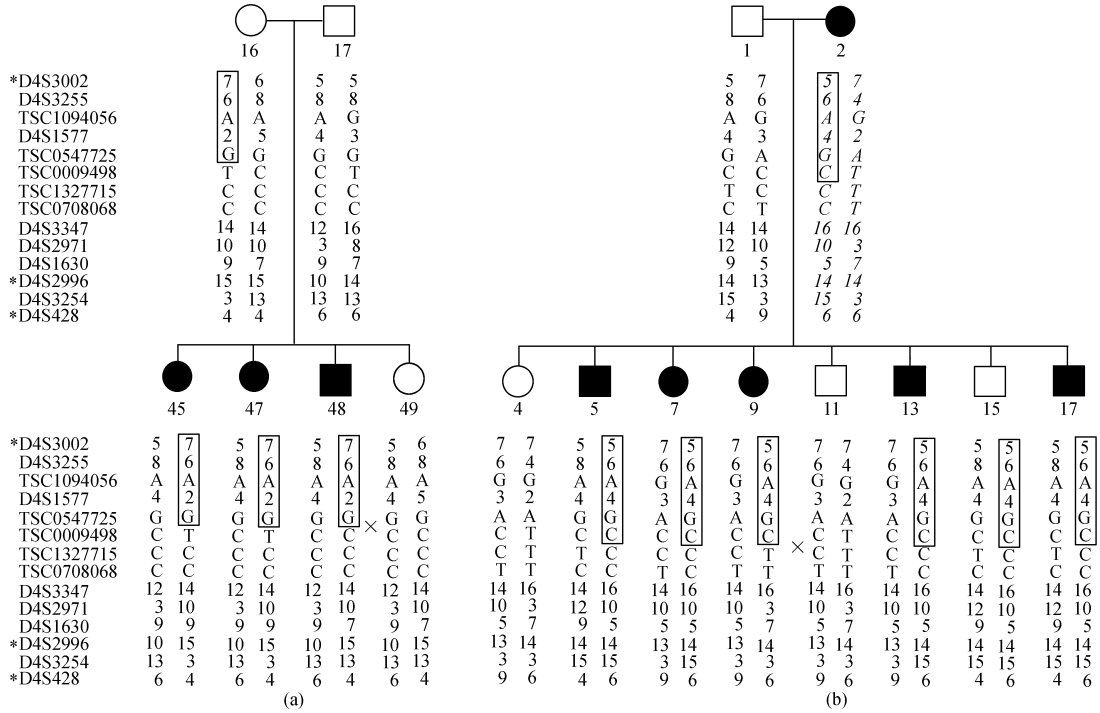


Fig. 1. The combined haplotypes with microsatellite and SNP markers in the nuclear family of NPC pedigree 31 (a) and family 34 (b). Filled symbols represent affected individuals. Experimentally typed genotypes are in bold, and inferred genotypes of individuals are shown in italics. Microsatellite markers used in previous study are labeled with asteroids. Crosses represent recombinations, and boxes indicate the chromosome region shared by affected members of the pedigree.

quencer (Applied Biosytem Inc.). Genescan3.0 software (Applied Biosytem Inc.) was used for collecting data. The microsatellite alleles were analyzed by Genotyper 2.5 software (Applied Biosystem Inc).

(iii) SNP genotyping. In the present study, a total of 12 SNP markers flanking D4S1577 and D4S3347 and covering a 114-kb region were taken from the SNP Consortium (TSC) database (http://www.tsc.schl.org) and were used for SNP genotyping by PCR-direct sequencing. Primers for each SNP marker were designed with Primer3 online software (http:www.cbr.nrc.ca/cgi-bin/primer3_www.cgi) and synthesized by Bioasia Company. The names, alleles, and primers of 12 SNP markers are listed in Table 1. Each 20 μL PCR mixture contained 2.0 μL of 10×PCR Buffer, 0.5 μmol/L each primer, 2.5 mmol/L $MgCl_2$, 250 μmol/L each dNTP, 0.5 U Hotstart Taq (Qiagen), and about 30 ng of DNA template. Amplification was performed with a Touchdown PCR program: started with first denaturing at 95℃ for 15 min, followed by 10 cycles composed of denaturing at 94℃ for 30 s and extension at 72℃ for 1 min, annealing temperature was decreased from 63℃ to 58℃ with decreasing by 0.5℃ per cycle. Then 20 cycles of denaturing at 94℃ for 30 s, annealing at 58℃ for 1 min, extension at 72℃ for 1 min and a final extension at 72℃ for 10 min. The PCR products were purified by QIAquick PCR purification kit (Qiagen, CA). DNA sequencing was performed on the ABI/PRISM 377 DNA Sequencer (ABI, US).

(iv) Haplotype construction. In the two nuclear families, a combination of nine microsatellite markers and informative SNP markers were used in haplotype construction under the rule of the minimum recombination between markers by using GENEHUNTER 2.1. The positional order of both microsatellite and SNP markers were defined according to the Marshfield genetic map and the sequence map on Build 32, NCBI MapViewer.

2 Results

(i) Microsatellite haplotype analyses. We genotyped the six microsatellite markers in each member of the two nuclear families, and constructed the haplotypes in each individual by combining with the genotypes of other

Table 1 12 TSC SNP markers and primers used in PCR amplification

Name	Position (bp)	Allele	Forward primer	Reverse primer
TSC1094056	53004558	a/g	ctgcagagtgctgtgtctgt	ccctgagagtgtctccttca
TSC0283055	53018525	c/t	gtgggagggaagaccattta	agccaaggcacatcaagtta
TSC0547725	53023754	a/g	tttgtcttcctaagtcttggtttt	ggcatggacttgagaggtaa
TSC0009498	53034287	c/t	tcaaatccctaaagagcaaca	tgtggttgaaaacacataagg
TSC0708067	53074843	c/g	ggccaaaatttttacgaggt	taaaatggatgtgggcattc
TSC0873510	53076844	c/t	ttccaaatttctccattgttg	ccagctatagcaacagtgattt
TSC1327715	53084094	c/t	ggctcattgaatgctctacc	tggggctacaaatacacgat
TSC0873514	53088376	a/t	tctgcccatcacttctgatt	tgcaggtgaggtagaatggt
TSC0708068	53096022	c/t	ttgtagatccttttgtgtggtg	ttgttctttgaaaatgtcctttg
TSC0708063	53102712	a/g	aagctctgtgggttctatcg	ctccttcctgtctccacctc
TSC0279721	53107905	a/g	taccaggatgggattttgag	tgcactctgcctctcatgta
TSC0949050	53118488	a/c	ggactgggcattcttacctt	accttaacagaatgaaggacaa

three microsatellite markers that had been typed previously, to determine the exact chromosome recombination in patients 31-48 and 34-9 (Fig. 1). In the nuclear family of pedigree 31, the position of chromosomal crossover in the affected 31-48 could be restricted between D4S1577 and D4S1630, because the two markers D4S3347 and D4S2971 are both homozygous in the parent 31-16 who carried the putitive susceptibility chromosome (Fig. 1(a)). Similarly, the crossover in the affected 34-9 could only be identified between D4S1577 and D4S2971 for the marker D4S3347 could not offer supporting genotypes too (Fig. 1 (b)). In sight of these observations, we could exclude the chromosomal segment between D4S2971 and D4S428 from the susceptibility region shared by affected individuals in both families.

(ii) SNP haplotype analyses. To further identify the exact recombination position, we genotyped a set of 12 SNP markers flanking D4S1577 and D4S3347 and covering a 114-kb region in the two families. We found that five SNP markers were at least heterozygous in one family and therefore much informative in haplotype analyses, whereas the other seven SNP markers were nonpolymorphic in both families for their consistent homozygous genotypes. The recombination position was consequentially identified between TSC0547725 and TSC0009498 in case 31-48, and between TSC0009498 and TSC1327715 in case 34-9. For the three SNP loci were all positioned between D4S1577 and D4S3347, the crossover positions in the two patients were both unambiguously identified between the two microsatellite markers.

(iii) Shortening of the NPC susceptibility locus. With the observations that the boundary close to telemere was defined near the marker D42950 in previous study and the distal boundary near D4S3347 in this study, we could restrict the susceptibility region on 4p11-p14 between D4S2950 and D4S3347, which is within 8.29cM in genetic length and about 15.21 Mb in physical distance, according to MapView Build 32.

3 Discussion

We have previously determined the genetic susceptibility locus for familial NPC on the region of 4p15.1-q12, suggesting that there be at least one gene responsible for the carcinogenesis of NPC on this region. However, a fine mapping of the candidate region is still crucial for the searching of the NPC susceptibility gene, in respect that many genes reside on this 14.21-cM region and it is difficult for further gene cloning if using positional candidate approach. In the present study, we narrowed down the region to 8.29 cM in genetic distance on 4p11-p14 by further haplotype analyses in the two major NPC families. The excluded 4q11-q12 region between D4S3347 and D4S428 contains about 50 known or predicted genes. Among them, some genes, such as KIT[3], KDR[4], PDGFR[5] and NECC1[6], although were reported to be associated with human cancers, would no longer be taken as candidates for the NPC susceptibility gene because of their genomic localization. Of note, the individual 34-15 (48-year old now), who is not affected with NPC so far, shares a part of the region between D4S1547 and D4S428 with other affected members, indicating that the susceptibility gene might reside on the 4p14-p13 segment between D4S2950 and D4S1547, and this region could be considered at priority when using positional candidate approach in gene searching.

With the development of Human Genome Project, the integral sequence map of human genome becomes a powerful tool in genetic disease studies. In the sequence map, the order of the genetic markers could be determined according to their base sequence positions, even if they were in the same genetic distance. In the present study, the physical order of all microsatellite and SNP markers on the same DNA contig NT-022831.11 (GenBank GI 20533721) were defined by their DNA sequence positions.

Microsatellite markers are commonly used in linkage

analysis of single gene diseases for their high heterozygosity, whereas SNP markers, which are distributed in the human genome with higher density and with more genetic stability, have shown a great potential in association studies for gene mapping and candidate gene cloning in complex genetic disorders[7—9]. The TSC database, which was originated from 24 individuals including 6 Asia-Americans, is one of the biggest public SNP databases and contains more than 1.42 million SNPs[10, 11]. In the present study, we used 12 TSC markers flanking D4S1577 and D4S3347 for haplotype analysis on account of no microsatellite markers available between the two markers in UniSTS database. We identified the minimal haplotype shared by the affected individuals with five informative SNP markers, demonstrating that the SNP haplotype analysis is also powerful in gene fine mapping. In short, we narrowed down the NPC susceptibility locus from 14.21cM to 8.29cM by using microsatellite and SNP haplotype analyses in the two NPC families, providing additional useful information for further studies to identify the NPC susceptibility gene.

Acknowledgements This work was supported by the China Medical Board (CMB 98-678), the National "973" Key Basic Research Projects (Grant No. G19980510), and the Foundation of Guangdong Science and Technology Committee (98001-4, 980950).

References

1. Lu, S. J., Day, N. E., Degos, L. et al., Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region, Nature, 1990, 346: 470—471.
2. Feng, B. J., Huang, W., Shugart, Y. Y. et al., Genome-wide scan for familial nasopharyngeal carcinoma and evidence for linkage to chromosome 4, Nat. Genet., 2002, 31: 395—399.
3. Beghini, A., Tibiletti, M., Roversi, G. et al., Germline mutation in the juxtamembrane domain of the KIT gene in a family with gastrointestinal stromal tumors and urticaria pigmentosa, Cancer, 2001, 92: 657—662.
4. Walter, J. W., North, P. E., Waner, M. et al., Somatic mutation of vascular endothelial growth factor receptors in juvenile hemangioma, Genes Chromosomes Cancer, 2002, 33: 295—303.
5. Heinrich, M. C., Corless, C. L., Duensing, A. et al., PDGFRA activating mutations in gastrointestinal stromal tumors, Science, 2003, 299: 708—710.
6. Asanoma, K., Matsuda. T., Kondo, H. et al., NECC1, a candidate choriocarcinoma suppressor gene that encodes a homeodomain consensus motif small star, filled, Genomics, 2003, 81: 15—25.
7. Xi, H.F., Ke, Y.H., Xiao, J.H. et al., Linkage analysis on chromosome 2 in essential hypotension pedigrees, Chin. Sci. Bull., 47(18): 1538—1540.
8. Tan, W., Wu, J.X., Tang, H.J. et al., Expression of cytochrome P450 2E1 in human liver: Relationship between genotype and phenotype in Chinese, Science in China, Ser. C, 2001, 44(4): 356—364.
9. Emahazion, T., Feuk, L., Jobs, M. et al., SNP association studies in Alzheimer's disease highlight problems for complex disease analysis, Trends Genet., 2001, 17(7): 407—13.
10. Collins, F. S., Brooks, L. D., Chakravarti, A., A DNA polymorphism discovery resource for research on human genetic variation, Genome Res., 1998, 8: 1229—1231.
11. Sachidanandam, R., Weissman, D., Schmidt S. C. et al., Map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, Nature, 2001, 409: 928—933.

(Received June 13, 2003; accepted July 29, 2003)

Chapter 3.4

TrendTDT - A Transmission/Disequilibrium based Association Test on functional Mini/Microsatellites

Abstract

Background

Mini- and micro-satellites are associated with human disease, not only as markers of risk but also involved directly in disease pathogenesis. They may play significant roles in replication, repair and mutation of DNA, regulation of gene transcription and protein structure alteration. Phenotypes can thus be affected by mini/microsatellites in a manner proportional to the length of the allele. Here we propose a new method to assess the linear trend toward transmission of shorter or longer alleles from heterozygote parents to affected child.

Results

This test (trend-TDT) performs better than other TDT (Transmission/Disequilibrium Test) type tests, such as TDT_{max} and $TDT_{L/S}$, under most marker-disease association models.

Conclusions

The trend-TDT test is a more powerful association test when there is a biological basis to suspect a relationship between allele length and disease risk.

Background

Variable number tandem repeats (VNTR's) are repetitive DNA sequences widely dispersed in the human genome. They are highly unstable and thus display a remarkable degree of polymorphism. They vary in length from a few to several thousand nucleotides and vary in complexity from simple di-, tri- and tetra-nucleotide repeats (microsatellites) to more complex repetitive elements (minisatellites). VNTR's, mainly microsatellites, have assumed an increasingly important role as markers in the genome and are intensively exploited for gene mapping. But VNTR's could be associated with human disease, not only as markers but also directly involved in disease pathogenesis; indeed, several functions have been suggested for micro- and mini-satellite DNA sequences.

If located within a coding sequence, VNTR's may alter protein structure. For example, expansions of tri-nucleotide microsatellites are responsible for genetic diseases such as X-linked spinal and bulbar muscular atrophy, Huntington disease, type 1 spinocerebellar ataxia, dentatorubral-pallidoluysian atrophy, and Machado-Joseph disease. These diseases are caused by expansion of CAG triplets within protein-coding regions [1].

VNTR's may also regulate gene transcription. Numerous *in vitro* studies have shown that gene transcription may be increased or decreased proportionally to

the number of repeated sequences (i.e. length of alleles) as illustrated in Table 1 (for detailed review, see Kashi et al. [2]). Direct effect of transcriptional modulation on risk of disease has been observed. As an example, the minisatellite *ILPR* (Insulin-Linked Polymorphic Region, $(ACAGGGGTGTGGGG)_n$) located 5' of the Insulin gene is implicated in Insulin-Dependent Diabetes Mellitus [3]. To date, many transcriptional factors have been identified and their binding with minisatellite repeated sequences have been demonstrated. There is increasing evidence that some gene-disease associations are due to functional micro/minisatellites, with the magnitude of susceptibility being related to allele length [4-6].

Table 1 - Micro/minisatellites that regulate gene transcription.

Genes regulated	VNTR localization	Repeat unit	length of alleles	transcription regulation [a]	interacting factor [b]
Microsatellites					
EGF receptor	intron 1	$(CA)_n$	14-21	down	
metalloproteinase 9	promoter	$(CA)_n$	14-23	down	
Pax-6	promoter -1kb	$(AC)_m$ $(AG)_n$	24-36	up	
HLA-DRB	intron 2	$(GT)_m$ $(GA)_n$	(15-22)(4-15)	down	CTCF (ZNF)
NOS2A	promoter -2.5kb	$(CCTTT)_n$	8-18	up	
COL1A2	promoter, intron 1	$(CA)_n(CG)_n(CA)_n$	(14-21)(6-7)(8)	up	
prolactin	promoter	$(TG)_m$ $(CA)_n$	(8-15)(4-13)	down	
phospholipase A2	promoter -595 bp	$(CA)_n$	48	down	
heme oxygenase	promoter -240 bp	$(GT)_n$	16-38	up	
CD30 gene	promoter -400 bp	$(CCAT)_n$	2-12	repression	
Minisatellites					
HRAS	3':1kb after polyA	28 bp repeat	30-84	up	NF-kappa B
Insulin	promoter -596 bp	14 bp repeat	40-157	up	Pur-1
ABO gene	promoter -3.6 kb	43 bp repeat	4-6	up	CBF/NF-Y

Note: [a] The direction of transcription regulation is according to elongation of alleles;
[b] If identified, transcription factors that interact with the micro/minisatellites are mentioned.

The Transmission/Disequilibrium Test (TDT) is a popular method to assess the involvement of a candidate gene or a genome region in the genetic component of a disease, using cases and their parents. The TDT, as originally developed [7], tested the association between a bi-allelic marker and a disease. Many authors have proposed an extension of the TDT to multi-allelic markers, by testing each allele separately [8, 9], by testing symmetry of the transmitted/non-transmitted table [10, 11], by testing marginal homogeneity [12, 13], or by conditional logistic regression [14, 15]. However, all these extensions considered implicitly the multi-allelic marker as a polymorphism without function, that is, the risk of disease was not treated as being correlated with allele repeat length. While this is true for most situations, there are some situations where the multi-allelic marker under study may have a functional effect on the studied disease, and thus this correlation may be present. This may introduce new information that can be taken into account in the test. From a statistical point of view, increased allele length could be understood as an increased dose of exposure to a risk

factor. In contrast to case-control association studies where one can use the classical trend-chi-square (the Cochran-Armitage trend test) to test this hypothesis, available extensions of the TDT to multi-allelic markers do not test such a "dose effect" in family-based association studies. However, case-control studies can be subject to bias produced by hidden population stratification. Therefore, a new statistical method that can test the correlation of allele length with disease susceptibility, and is not sensitive to population stratification is needed. In this paper, we describe a newly developed method to meet this requirement.

Methods

Algorithm

Consider a multi-allelic marker with *k* alleles, which are assumed to be coded as integers proportional to their length. The trend-TDT statistic is based on the length of alleles transmitted from heterozygous parents to their affected children. Let's denote, for each heterozygous parent *i*, t_i the length of the transmitted allele, u_i the length of the untransmitted allele, and x_i the difference between the length of transmitted and untransmitted alleles ($x_i=t_i-u_i$). For family *f*, let n_f be the number of calculated x_i within the family, and define d_f as

$$d_f = \frac{\sum x_i}{\sqrt{n_f}}$$

Under the situation that neither the micro/minisatellite is the cause of the disease, nor is it in linkage disequilibrium with any disease causing genes, then the mean of d_f should be zero, and its variance is

$$V(d_f) = V(\frac{\sum x_i}{\sqrt{n_f}}) = \frac{V(\sum x_i)}{n_f} = \frac{n_f V(x)}{n_f} = V(x)$$

Note that this d_f is actually the mean of x_i weighted by square root of n_f, so that the variance of d_f is equal between families. Hence the test statistic

$$T = \frac{mean(d_f)}{S/\sqrt{N}}$$

asymptotically follows the Student's t distribution with *N-1* degrees of freedom. Here *S* is the estimated standard deviation of the d_f, and *N* is the number of informative families. In case there is a trend toward transmission of shorter alleles, the $mean(d_f)$ will be less than 0, and vice versa. If biological clues indicate that preferential transmission of shorter alleles (or longer alleles) should be observed, the test is one-tailed t test (H_1: $T<0$ or H1: $T>0$); otherwise the test is two-tailed (H_1: $T\neq 0$).

The missing genotype problem is treated according to Curtis [16]. In case both parents are missing, or, one parent is missing and the affected child has the same heterozygote genotype as the other parent, these families are considered

uninformative and are discarded in the analysis. When only one parent is missing but the affected child is homozygote, inclusion of such triads will lead to bias, therefore they are also discarded [16]. In other situations, transmission status of either allele can be inferred, and they are used in the analysis.

Comparison with other methods

Two other methods that can be used in testing association between disease and functional micro/minisatellites are TDT_{max} and $TDT_{L/S}$. TDT_{max} stems from the classical bi-allelic TDT. The statistics corresponding to TDT_{max} is the maximum chi-square value obtained over all alleles:

$$TDT_{\max} = \max_i \frac{(n_{i\bullet} - n_{\bullet i})^2}{(n_{i\bullet} + n_{\bullet i})} \qquad (i = 1...k)$$

Here $n_{i\bullet}$ denote the number of heterozygote parents who transmit an allele *i*, and $n_{\bullet i}$ denote the number of heterozygote parents who has an allele *i* but do not transmit it. Individual TDT is calculated for all alleles, and the maximal value is taken as the TDT_{max}. Although the individual TDT test follows Chi-square distribution with 1 degree of freedom, the TDT_{max} does not. Clearly, this method will not have appropriate type I error due to the selection of the highest Chi-square value. Several methods have been proposed to address the multiple testing problem in TDT_{max}, including empirical p value simulation [9] and modified Bonferroni correction [8]. Since the former method requires enormous number of repetitions to accurately obtain a low p value, in this study, Bonferroni corrected TDT_{max} is used and evaluated.

$TDT_{L/S}$ corresponds to the classical bi-allelic TDT computed on collapsed long alleles vs. collapsed short alleles. In this case, the traditional TDT statistics can be used:

$$TDT_{L/S} = \frac{(b-c)^2}{b+c}$$

where *b* is the number of parents that transmit the long allele but not the short one, and *c* is the number of parents that transmit the short allele but not the long one. It should be noted that some of the heterozygote parents are not counted in the computation if both of their alleles belong to the long allele pool or short allele pool. The specific problem of this approach is the choice of the threshold between "long" and "short" alleles; here we choose the first allele (from shortest to longest) whose cumulative allele frequency is greater than 0.5, so that roughly half of the alleles are long alleles and another half the short ones. We note however that in some cases there be relevant biological data which might suggest a more appropriate threshold.

The cut-of thresholds to reject H0 hypothesis used in these two methods are the same as trend-TDT.

Type I error computations
In order to assess and compare the type I error rates of each of the three tests, we simulated 200 trios (case and both parents) with disease-unrelated microsatellite genotypes. The total number of alleles of this marker is set to 10, with equal allele frequencies. Simulations are performed 1,000,000 times. The proportion of times that calculated p-value is equal to or less than an expected value is plotted against this expected value, in minus logarithm scale. For a correct test statistic, this curve should be exactly the line “y=x”. For a test with higher type I error rate, the curve will be bellow the line “y=x”, and for a conservative test, the curve lies above.

Modeling genotyping errors
The most common genotyping errors in microsatellites were simulated to evaluate their effects on type I error rate of the trend-TDT test. These errors include confusing homozygote and adjacent-allele-heterozygote genotypes in allele banding pattern scoring [17], false homozygotes due to the preferential amplification of shorter alleles over longer alleles (short allele dominance), false homozygotes due to priming site mutations (null allele), offspring gaining one more repeat unit in one of the alleles (microsatellite mutation), and randomly mis-scoring an allele as its adjacent allele due to binning error. In simulation, each of these genotyping error rates was moderately higher than what is usually discovered in real data [18]. The microsatellite was simulated with 10 equally distributed alleles, without association with disease. Type I error rates were then calculated as the proportion of times trend-TDT yielding significant results ($p \leq 0.05$) from 1,000,000 simulations on 200 trios.

Power computations
Power can be estimated by generating samples with a determined pattern of marker-disease association, and by calculating the proportion of these simulations that the null hypothesis is correctly rejected. Here in this paper, we assume a significance level of 0.001. Following this design, we evaluate the power of the trend-TDT and compare it with the power of two other TDT tests: TDT_{max} and $TDT_{L/S}$.

The powers of the three tests were evaluated under different patterns of marker-disease association, parameterized in terms of relative-risk, and under different kinds of multi-allelic markers in terms of the number of alleles and allele frequencies. The different models are presented in Table 2. In these models, the maximum relative risk for any single allele size is always equal to 3, and the prevalence of the disease is fixed at 10%. Calculation of genotype-wide penetrance is based on multiplicative model. All estimates of power were based on 10,000 generated tests on 200 trios, unless otherwise specified.

Table 2 - Alleles frequencies and allelic relative risks in power simulation.

Designation	Allele length						Notes
	1	2	3	4	5	6	
Allele Frequencies							
F6.eq	1/6	1/6	1/6	1/6	1/6	1/6	Equal allele frequencies.
F6.rd	.15	.20	.10	.40	.10	.05	Randomized allele frequencies.
F6.bi	.10	.10	.30	.30	.10	.10	There exist two major alleles.
Relative Risk							
RR(lin)	1	1.4	1.8	2.2	2.6	3	RRs increase linearly along with allele length.
RR(thr4) §	1	1	1	3	3	3	RRs increase above a threshold of allele length.

Note: § The number in the bracket denotes the first allele with higher risk.

Modeling non-functional markers

Situations when VNTR markers are associated with a disease, without linear correlation between allele length and disease risk, are also modeled. In this model, the VNTR marker has 10 alleles, with allele frequencies equally distributed. Relative risks are assigned proportional to allele length, then before each repeat of the simulation, this relative risk vector is permuted. Empirical power is calculated to compare the performance of the statistics before and after permutation, based on 10,000 repeats of simulations on 200 trios.

A computer program for the trend-TDT, TDT_{max}, and $TDT_{L/S}$, can be obtained from http://geocities.com/trntdt/ .

Results

Type I error

As shown in Figure 1, the curve for both trend-TDT and $TDT_{L/S}$ are very close to the diagonal line, showing correct type I error rates in simulation. After Bonferroni correction, the type I error rate of TDT_{max} is nearly correct, although it is still a little conservative. As shown in Table 3, genotyping errors lead to slightly inflated type I error rates for trend-TDT.

Table 3 - Simulated genotyping errors and resultant type I error rates.

Error Models §	Mistypes in total genotypes (%)	Misinheritance in mistyped trios (%)	Type I Err. ($p \leq 0.05$) Rate (95% C.I.)
A	0.50	27	.0504 (.0499-.0508)
B	0.90	59	.0502 (.0498-.0506)
C	2.70	59	.0502 (.0497-.0506)
D	0.90	59	.0500 (.0496-.0505)
E	0.15	41	.0499 (.0495-.0503)
F	0.17	88	.0502 (.0498-.0507)
ALL	5.23	57	.0506 (**.0502-.0511**)

Note: § A: mis-scoring genotype (x,x) as (x,x+2); B: mis-scoring genotype (x,x+2) as (x,x); C: short allele dominance; D: binning error; E: priming site mutation; F: microsatellite mutation; ALL: all of above.

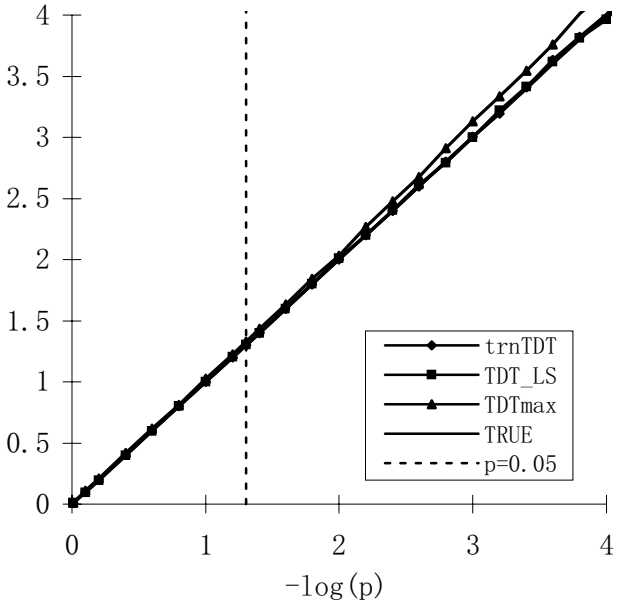


Figure 1 : Type I error rates for trend-TDT, TDT_{max} and $TDT_{L/S}$.
X axes is the expected p value in minus logarithm scale, Y axes is the observed frequencies that the calculated p value is equal to or less than the expected p value, in minus logarithm scale.
The line "TRUE" is the expected curve for a correct test, which should be exactly the line "y=x".

Power

The power of the three tests, trend-TDT, $TDT_{L/S}$ and TDT_{max} on simulated trios are plotted in Figures 2-4. Figure 2 presents the power of the tests under different VNTR/STR models, which vary in terms of the number of alleles at the VNTR (4, 6 or 10 alleles with equal allele frequencies). In each of these models, the relative risk associated with each allele increases linearly with the length of the allele. The trend-TDT is clearly the most powerful test in all situations. An increase in the number of alleles resulted in decreased power for all tests; however, the trend-TDT was the least sensitive to this effect. Figure 3 presents the behavior of the tests under different sets of allele frequencies, assuming a linear relative risk model of the simulated functional VNTR. It can be seen from the figure that the power is higher when the allele frequencies are equally distributed, and is lower when some major alleles exist. This is probably related to the fact that overall heterozygosity (and thus informativeness of the sample) is maximized with equal allele frequencies. Nevertheless, the simulations indicate that the trend-TDT is the least sensitive to the distribution of allele frequencies and is the most powerful for association detection among the three methods.

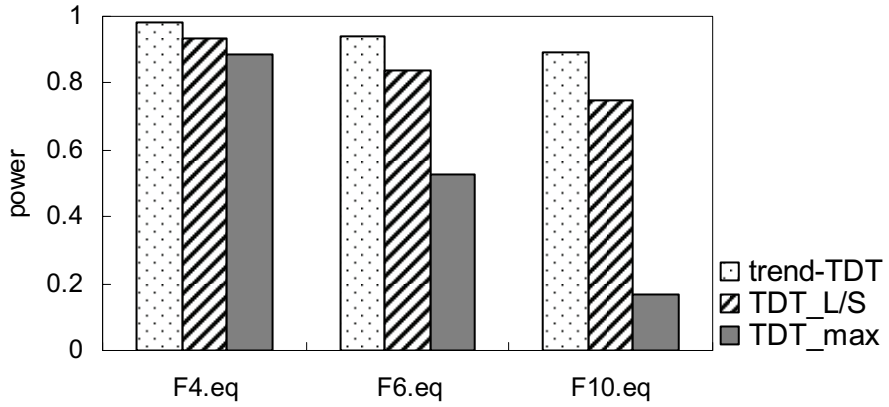


Figure 2 - Power of the three TDT tests under different number of alleles.
Disease risk linearly increases along with the allele length. All allele frequencies are set to equal. Number of alleles are 4, 6 and 10, respectively.

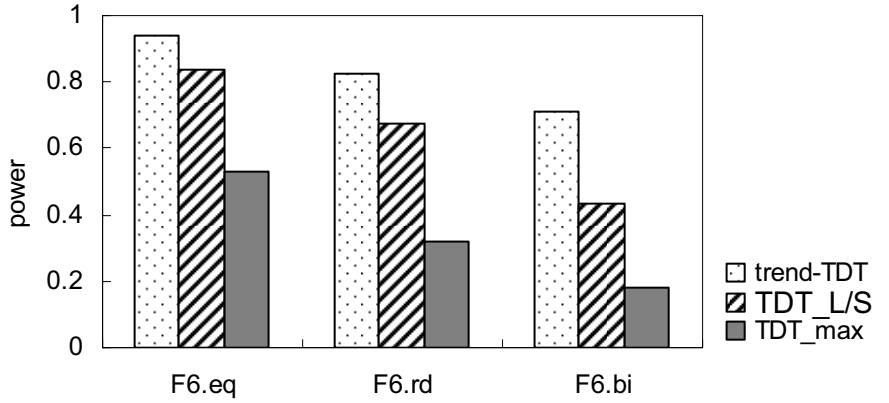


Figure 3 - Power of the three TDT tests under different sets of allele frequencies.
Disease risk linearly increases along with the allele length, i.e. RR(lin) in Table 2. Number of alleles is set to 6. The allele frequencies are equal (F6.eq), random (F6.rd), or uneven, where two major alleles exist (F6.bi).

The behavior of the tests under different marker-disease association models is presented in Figure 4. These models are defined so that relative risks increased linearly ("RR(lin)") or uniformly above a threshold ("RR(thr3)", "RR(thr4)", "RR(thr5)"), according to the increase in VNTR length. The assumed marker is a microsatellite with six equally frequent alleles. In the threshold models, the thresholds for higher relative risk are set to allele 3 ("RR(thr3)"), allele 4 ("RR(thr4)"), or allele 5 ("RR(thr5)"). As shown in Figure 4, the trend-TDT is the most powerful method under the linear model, while under threshold models, the relative performance depends on where the threshold is. When the threshold is close to the shortest or longest allele, the trend-TDT performed much better than $TDT_{L/S}$. When the threshold is exactly in the middle, which is most favorable to $TDT_{L/S}$, the $TDT_{L/S}$ is better. However, in this case both the trend-TDT and $TDT_{L/S}$ have high power and the difference is very small (Figure 4). If the threshold can

be inferred by biologic knowledge of the gene under study, then using the known threshold will lead to much higher power in $TDT_{L/S}$ than the trend-TDT (Figure 4). Under most circumstances, TDT_{max} performed the worst among the tested methods (Figure 2-4), with the only exception that in the RR(thr3) model in Figure 4, TDT_{max} is better than $TDT_{L/S}$.

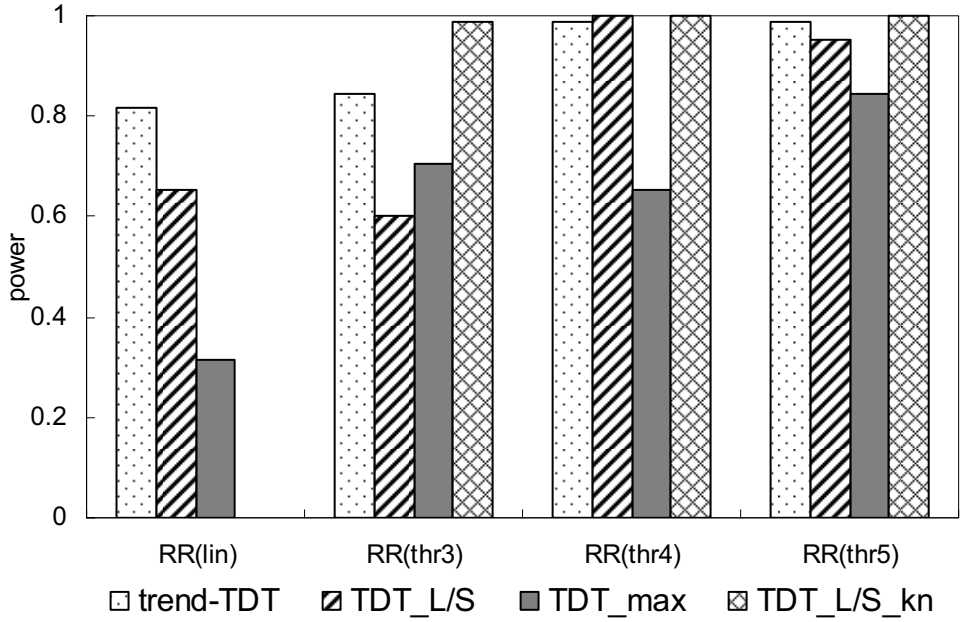


Figure 4 - Power of the three TDT tests under different marker-disease association models.
RR(lin) designates the linear model, RR(thr#) the threshold model, where # denotes the first allele with higher risk (Table 2). Number of alleles is set to 6, with equal allele frequencies, i.e. F6.eq in Table 2. "TDT_L/S" is the TDTL/S method using medium allele length as threshold, "TDT_L/S_kn" is the TDTL/S method when the threshold is known and is used in the test. Power is calculated from 10,000 simulations on 150 trios, using significant criteria 0.001.

When markers are associated with the studied trait, but without a specific trend, the power of TDT_{max} remains unchanged, while the power of both the trend-TDT and $TDT_{L/S}$ decrease markedly (Figure 5). Notably, the trend-TDT and $TDT_{L/S}$ still have some power for association detection. In-depth study of each replicate of the simulation found that the power depends on the trend of the increase/decrease of the relative risk vector: in the most extreme cases where the trend is almost zero, the power of these two tests are equal to type I error rates; however, because in most cases, the trend is not zero, the power of trend-TDT and $TDT_{L/S}$ remain above the type I error level.

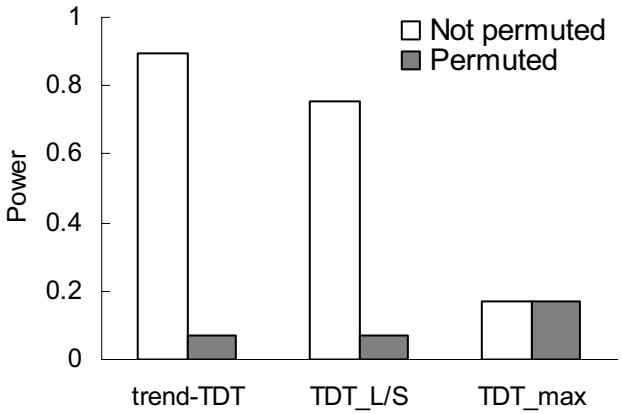


Figure 5 - Power of the TDT tests before and after permuting relative risk vector.
The disease model is linear relative risk of VNTR with 10 alleles.

Discussion

Performance of the tests

As expected, when the relative risks increase proportionally with allele length, the trend-TDT is always more powerful than the other tests, irrespective of the number of alleles or their frequencies. When the RRs increase according to a threshold model, the performances of $TDT_{L/S}$ and trend-TDT depend on the threshold. $TDT_{L/S}$ is more sensitive to the threshold and less powerful when the threshold is close to the longest or shortest allele. When the threshold is close to medium allele length, $TDT_{L/S}$ performs slightly better than the trend-TDT, but both are quite powerful in this situation. The TDT_{max} performs the worst in most situations studied here. This may be because both trend-TDT and $TDT_{L/S}$ use the information on the correlation between allele length and disease risk that is present in the generated disease model.

Choice of the tests

Based on these results, we do not recommend the TDT_{max} for any situation when there could be a relationship between allele length and disease risk. Whether to use trend-TDT or $TDT_{L/S}$ depends on prior knowledge of the functional relationship between allele length and gene function. When the threshold model is biologically true, and this threshold can be inferred by biologic knowledge of the gene under study, then $TDT_{L/S}$ is a better choice. Under all other situations, trend-TDT is recommended. When the threshold model is true but it is not clear where the threshold is, trend-TDT should be used, since by using $TDT_{L/S}$, one either has a multiple testing problem by trying different thresholds, or alternatively has less power for the test by using the median allele length only, which could be wrong biologically. Even when the true threshold is close to the median allele length, the difference between trend-TDT and $TDT_{L/S}$ is so small that it could be

ignored. In other situations when a VNTR is associated with a disease without trend, trend-TDT and $TDT_{L/S}$ are not as powerful, therefore other TDT methods should be used.

Another potential transmission/disequilibrium based test that could take into account the phenotypic response trend toward longer or shorter alleles is conditional logistic regression [19, 20], using a continuous variable for the allele length rather than a categorical one. Preliminary simulations indicate that this test is not as powerful as the trend-TDT test (data not shown); nevertheless, conditional logistic regression could be more beneficial, since it can incorporate various genetic risk models, include other genetic or environmental risk factors, and provide estimates of the risk of the disease conferred by the functional micro/minisatellite. Therefore, both methods might be used depending on the particular study circumstances.

Impact of genotyping errors

Given that genotyping errors may lead to increased type I error rates of TDT tests, several modified TDT statistics were proposed for analysis of single nucleotide polymorphisms [21-25], since it is much easier to model genotyping errors in bi-allelic markers than in multi-allelic markers. It was expected that genotyping errors would also increase the type I error rate of the trend-TDT test. However, simulation has shown that, with reasonable typing error frequencies, the type I error rates were inflated only slightly. The reason might be that genotyping errors in multi-allelic markers can be efficiently detected by Mendelian-inheritance analysis when parental data are available [26]. It should be noted that the extent of type I error is a function of the typing error frequencies, the number of alleles, the allele frequencies, and sample size [22, 27]. Thus, if genotyping errors are observed in a subset of a larger sample of pedigrees (e.g., over 500 affected offspring), statistical methods to address genotyping errors in TDT analysis should be considered to confirm that significant results are not false positives due to undetected genotyping errors. To further eliminate genotyping errors in real data analysis, it is recommended that siblings of the patients are genotyped and/or closely adjacent markers are genotyped, so that more typing errors can be detected as either Mendelian inconsistencies in the former or haplotype double crossovers in the latter.

Conclusions

In summary, we have developed a new statistical test, the trend-TDT test, appropriate for those situations when a) parental data are available; and b) there are multiple alleles at the marker locus hypothesized to be associated with the disease of interest; and, most importantly, c) there is a biological basis to suspect a relationship between allele length and disease risk.

Authors' contributions

BJF carried out the programming, testing and simulation of the methods, and drafted the manuscript. DEG contributed to the design of the study and critical review of the manuscript, MC conceived the study, participated in its design and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr Bruno Fallisard for his scientific input in the design of the test.

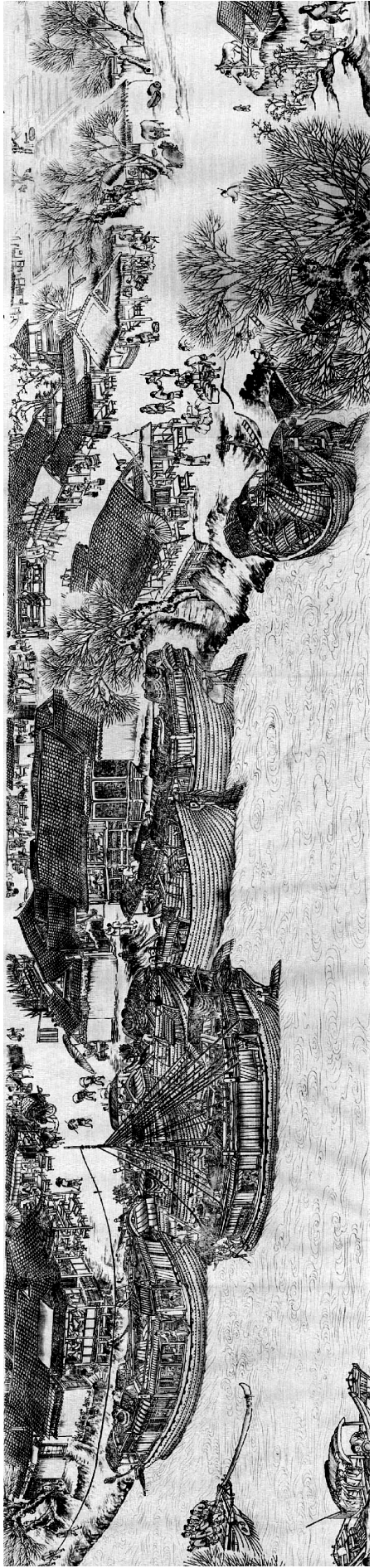
References

1. Ashley CT, Jr., Warren ST: **Trinucleotide repeat expansion and human disease**. *Annu Rev Genet* 1995, **29**:703-728.
2. Kashi Y, King D, Soller M: **Simple sequence repeats as a source of quantitative genetic variation**. *Trends Genet* 1997, **13**(2):74-78.
3. Kennedy GC, German MS, Rutter WJ: **The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription**. *Nat Genet* 1995, **9**(3):293-298.
4. Comings DE: **Polygenic inheritance and micro/minisatellites**. *Mol Psychiatry* 1998, **3**(1):21-31.
5. Gatchel JR, Zoghbi HY: **Diseases of unstable repeat expansion: mechanisms and common principles**. *Nat Rev Genet* 2005, **6**(10):743-755.
6. Greene E, Handa V, Kumari D, Usdin K: **Transcription defects induced by repeat expansion: fragile X syndrome, FRAXE mental retardation, progressive myoclonus epilepsy type 1, and Friedreich ataxia**. *Cytogenet Genome Res* 2003, **100**(1-4):65-76.
7. Spielman RS, McGinnis RE, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)**. *Am J Hum Genet* 1993, **52**(3):506-516.
8. Betensky RA, Rabinowitz D: **Simple approximations for the maximal transmission/disequilibrium test with a multi-allelic marker**. *Ann Hum Genet* 2000, **64**(Pt 6):567-574.
9. Morris AP, Curnow RN, Whittaker JC: **Randomization tests of disease-marker associations**. *Ann Hum Genet* 1997, **61 (Pt 1)**:49-60.
10. Cleves MA, Olson JM, Jacobs KB: **Exact transmission-disequilibrium tests with multiallelic markers**. *Genet Epidemiol* 1997, **14**(4):337-347.
11. Lazzeroni LC, Lange K: **A conditional inference framework for extending the transmission/disequilibrium test**. *Hum Hered* 1998, **48**(2):67-81.
12. Bickeboller H, Clerget-Darpoux F: **Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers**. *Genet Epidemiol* 1995, **12**(6):865-870.
13. Spielman RS, Ewens WJ: **The TDT and other family-based tests for linkage disequilibrium and association**. *Am J Hum Genet* 1996, **59**(5):983-989.
14. Harley JB, Moser KL, Neas BR: **Logistic transmission modeling of simulated data**. *Genet Epidemiol* 1995, **12**(6):607-612.
15. Sham PC, Curtis D: **An extended transmission/disequilibrium test (TDT) for multi-allele marker loci**. *Ann Hum Genet* 1995, **59 (Pt 3)**:323-336.

16. Curtis D, Sham PC: **A note on the application of the transmission disequilibrium test when a parent is missing**. *Am J Hum Genet* 1995, **56**(3):811-812.
17. Hoffman JI, Amos W: **Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion**. *Mol Ecol* 2005, **14**(2):599-612.
18. Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW, Foote SJ: **Identification and analysis of error types in high-throughput genotyping**. *Am J Hum Genet* 2000, **67**(3):727-736.
19. Schaid DJ: **General score tests for associations of genetic markers with disease using cases and their parents**. *Genet Epidemiol* 1996, **13**(5):423-449.
20. Cordell HJ, Clayton DG: **A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes**. *Am J Hum Genet* 2002, **70**(1):124-141.
21. Morris RW, Kaplan NL: **Testing for association with a case-parents design in the presence of genotyping errors**. *Genet Epidemiol* 2004, **26**(2):142-154.
22. Gordon D, Heath SC, Liu X, Ott J: **A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data**. *Am J Hum Genet* 2001, **69**(2):371-380.
23. Gordon D, Haynes C, Johnnidis C, Patel SB, Bowcock AM, Ott J: **A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents**. *Eur J Hum Genet* 2004, **12**(9):752-761.
24. Cheng KF, Chen JH: **A simple and robust TDT-type test against genotyping error with error rates varying across families**. *Hum Hered* 2007, **64**(2):114-122.
25. Bernardinelli L, Berzuini C, Seaman S, Holmans P: **Bayesian trio models for association in the presence of genotyping errors**. *Genet Epidemiol* 2004, **26**(1):70-80.
26. Douglas JA, Skol AD, Boehnke M: **Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data**. *Am J Hum Genet* 2002, **70**(2):487-495.
27. Mitchell AA, Cutler DJ, Chakravarti A: **Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test**. *Am J Hum Genet* 2003, **72**(3):598-610.

Chapter 4

General Discussion



Introduction

Although nasopharyngeal carcinoma (NPC) is a rare cancer in the Netherlands and most other Western countries, the disease has an enormous social impact in a number of endemic areas. During the period from 1998 to 2002, the mortality rate of NPC in Sihui county, South China, was $22/10^5$/year for males and $7.7/10^5$/year for females[1]. In Hong Kong, the situation is better (age standardized mortality rate in 2005 was $6.8/10^5$/year for males and $2.0/10^5$/year for females). Although NPC is the 4th most frequent cancer in Hong Kong males in terms of 0-74 years cumulative incidence rate, the 0-64 years cumulative incidence of NPC ranks the 3rd. Similarly, NPC is the 5th most frequent cancer in men $\leq$ 74 years of age in Singapore Chinese, but in those $\leq$ 64 years NPC ranks the 2nd [2]. The observation that the 0-64 years cumulative rates of NPC rank higher than the 0-74 years cumulative rates, can be explained by the fact that, unlike other major cancer sites (stomach, colon, liver) which incidence increases with age, NPC mostly affects individuals in their 40s, after which the incidence declines with age[1]. In addition, the male to female sex ratio of NPC incidence in most countries ranges from 2 to 3 [2], and males tend to be the major source of income in the affected families[3] in endemic regions. Thus, the occurrence of nasopharyngeal carcinoma is a big burden to the patients and their families. For the purpose of promoting public health, both early diagnosis, which is highly associated with better prognosis[4,5], and disease prevention are very important. Implementation of these measures depends on the elucidation of the etiology of nasopharyngeal carcinoma.

Pathogenesis of nasopharyngeal carcinoma is heterogeneous regarding histological type. Occurrence of squamous cell NPC (keratinizing squamous cell carcinoma and non-keratinizing squamous cell carcinoma) may involve smoking[6] and/or Human papillomavirus (HPV) infection[7,8], whereas non-keratinizing undifferentiated NPC is more likely to be EBV related[4]. Since undifferentiated nasopharyngeal carcinoma (UNPC) is the predominant form of NPC in most endemic areas, it has become the focus of pathogenesis studies. UNPC is developed in a multi-step manner[9,10]. Firstly, low-grade dysplastic lesions occur in normal nasopharyngeal epithelium, which could be the result of chromosomal changes (e.g. loss of chromosome region 3p and 9p) induced by genetic susceptibility and/or carcinogens[10]. The low-grade dysplastic change of the epithelium is thought to precede Epstein-Barr virus infection in the epithelial cells, since EBV can be detected in high-grade dysplasia but not in low-grade dysplasia[10]. After the infection of EBV, a series of cellular signaling cascades are altered, predominantly the STAT3, NFκB and p38-MAPK pathways, transforming the epithelial cells into invasive tumors[11]. During this process, other genetic lesions, such as gain of 12q, loss of 11q, 13q, 14q, 16q, and some other mutations, may also have facilitated the transformation[10]. It is not yet known what risk factors, both environmental and genetic, have taken parts in the process, nor is it clear when and how they are involved.

Environmental contributions

Nasopharyngeal carcinoma is a complex disease where both environmental and genetic components have played a role in its etiology. Suggested risk factors include tobacco, salted fish at weaning, domestic fumes, occupational dust and heat, herbal medicine, Epstein-Barr virus (EBV) activation, and family history of NPC[12]. However, almost all studies were performed in South East Asia or the United States; it is not known whether specific foods are connected with increased NPC risk in North Africa. Efforts have been made to identify associated dietary factors in this region[13-15], but the sample sizes associated with these studies were small and detailed adjustment for social economic status was not applied in some of the studies. In chapter 2, we performed a multi-centric large-scale case-control study in Maghrebian countries, including Tunisia, Algeria and Morocco. In this study, 636 cases and 615 controls, matched by center, childhood household type (rural or urban), age and sex, were recruited from 2002 to 2005. The questionnaires covered demographic and ethnic information, medical history, consanguinity, familial cancer history, living conditions, exposure to chemicals and smokes, alcohol and tobacco consumption, and dietary consumption in both childhood and adulthood. Blood samples were taken for future genetic studies. In chapter 2.1, we focus our analyses on environmental risk factors regarding dietary consumption. After adjustment for social economic measurements, consumption of rancid butter, rancid sheep fat, and quaddid were associated with increased risk of NPC, while consumption of cooked vegetables was inversely associated. In multivariate analyses, only three variables remained in the model, which are consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood ($p<0.01$). Compared with the other three studies from the same region, this study confirmed the increased risk of NPC associated with rancid butter, rancid sheep fat and quaddid, and newly identified the potential inverse association of cooked vegetables. In addition, some of the reported food items were excluded from association with NPC in this study, such as toklia and harissa.

The association of tobacco intake and NPC risk has not been widely studied in North African countries. Moreover, there are concerns about the correlation between local styled tobacco intake method, such as chicha and snuff, and NPC. In chapter 2.2, these questions were addressed with detailed analyses. Our results demonstrated that daily consumption of >22 cigarettes conferred a moderately increased risk of NPC (OR 1.6 [1.0-2.6]). Likewise, multivariate analyses only yielded marginal results for daily cigarette smoking ($p=0.035$). These findings suggest that, the effect of cigarette smoking on the development of nasopharyngeal carcinoma is small in Maghrebian countries, especially when compared to dietary exposures. There was no evidence of association of local tobacco intake (chicha and snuff) and alcohol consumption with NPC in our study.

In addition to cigarette smoking and alcohol consumption, it has been postulated that NPC patients in South China were more likely than controls to be exposed to domestic fume intake, by poor ventilation in the kitchen (no windows, no chimney) or the kitchen being situated in the main room, or wood fire cooking[16-18]. In North Africa, occupational fume intake was not common among the studied population. However, domestic fume exposure by usage of the kanoun charcoal oven in cooking during childhood was strongly associated with increased NPC risk. It remained significant after adjusting for social economic status and food intake. In addition, higher exposure to fumes in cases than in controls implicated by wood fire cooking and poor ventilation of kitchen further implicates the role of domestic fumes in NPC in this population, although the tests were only marginally significant. Beside these findings, incense usage during adulthood was associated with a significant reduction in NPC risk (odds ratio=0.63 [0.42-0.93]). Although results have been insignificant by conditional logistic regression adjusted for age, social economic status, diet and tobacco, restricted permutation test indicated that cannabis smoking was significantly associated with increased NPC risk, and this association was not due to the strong correlation between cannabis and tobacco consumption. The putative determinants identified are in part correlated and therefore may not be independent. In a multivariate analysis using forward/backward selection of determinant at an inclusion criteria of $p<0.05$ and adjusting for age and social economic status, consumption of rancid butter, rancid sheep fat and cooked vegetables during adulthood, usage of kanoun during childhood, incense during adulthood and cigarette intake per day remained in the model.

Genetic contributions

Chapter 3.1 documents the large-scale familial cancer history study conducted in Guangdong China. Between January 1999 and June 2001, 2252 NPC patients seen at the Cancer Center of Sun Yat-Sen University were successfully interviewed, with a structured questionnaire regarding familial cancer history. In this study, 14% of all patients reported a family history of NPC, and 1.8% reported more than one patient in blood-relatives. When compared with the data from cancer registry of Hong Kong, it is estimated that, among the first degree relatives of the NPC probands, the standardized incidence ratio (SIR) is 2.1 [1.8-2.4], and that the risk of NPC are substantially increased among first degree relatives of NPC patients with early age at onset (<40 years, SIR 9.0 [6.1-13.3]). The results of this study confirm the familial clustering of nasopharyngeal carcinoma patients in Guangdong China, and provide evidence that genetic susceptibility may contribute to the occurrence of NPC in Guangdong.

Chapter 3.2 describes the first genome wide linkage study of NPC. Twenty families with multiple cases of NPC were collected from the same region in Guangdong province as our previous study on NPC. In total, 165 individuals from

these families were genotyped for 400 microsatellite markers covering the human genome, which provide an average interval between markers of 10 cM. Parametric analyses assuming a dominant model provided evidence of linkage to the marker D4S405 on chromosome 4 with a logarithm of odds for linkage (LOD) score of 3.06 and a heterogeneity-adjusted LOD (HLOD) score of 3.21. Fine mapping with additional markers flanking D4S405 resulted in a LOD score of 3.54 and HLOD score of 3.67 for the region 4p15.1–q12. Further haplotype analyses (chapter 3.3) have narrowed down the susceptibility gene harboring region to a 15.4Mb segment, between D4S2950 and D4S3347. Currently, there are 168 known genes in this region (http://genome.ucsc.edu/). Mutation screening in some of the candidate genes, such as the toll like receptor genes, did not find any rare mutations in the coding regions. To pinpoint the susceptibility gene on chromosome 4, further narrowing down of the region is necessary through identification of additional linked families.

NPC is a complex disease

Heterogeneity of environmental risk factors

The results of chapter 2.1 reveal different dietary exposures between the two major endemic populations in the world: while the main associated food is Cantonese styled salted fish in South East Asia, it is rancid fat in North Africa. It should be noted that, Cantonese style salted fish are seldom consumed in North Africa, while rancid fat is rarely eaten in South East Asia. In addition, not only are the associated foods diverse, but the main ingredient that is responsible for the NPC susceptibility could also be different. It has been suggested that N-nitrosamines or their precursors were the main NPC causing agent in Cantonese styled salted fish, while the association of rancid fat in North Africa implicates butyric acid. However, whether these two components can increase NPC risk still requires further confirmation. Nevertheless, the differences in dietary risk factors implicate the diverse etiology of NPC in different endemic regions of the world.

Heterogeneity of genetic risk factors

Several linkage studies have been conducted on candidate genes or candidate regions, such as the Human Leukocyte Antigen (HLA) genes on chromosome 6 [19], and the chromosome regions that are frequently lost in tumors (3p21 and 9p) [20]. Linkage of HLA with NPC was reported among multinational resources of families from Guangxi, Hong Kong and Singapore, and linkage of 3p21 was reported from Hunan province of South China. However, these loci were not confirmed in the linkage study described in chapter 3.2, in which families were recruited from Guangdong province of China. Likewise, linkage to chromosome 4 was not evident in the Hunan study[20]. These findings suggest that there is strong genetic heterogeneity of NPC among Southern Chinese populations. Genetically,

the populations inhabited in these regions are complicated and diverse; it is not known whether the NPC families in the above studies belong to the same population.

Interactions

The complexity of the disease is manifested not only by the presence of different risk factors in different populations, but also the interplay of various factors in the same population. It is suggested that exogenous and endogenous nitrosamines are carcinogenic agents of NPC. In vitro studies revealed that fresh fruits and vegetables can reduce endogenous formation of N-nitrosamines[21,22], and that vitamin C can inhibit the DNA damage induced by some nitrosamines that have also been found in Cantonese styled salted fish[23]. These findings raised the possibility of interaction between high consumption of salted fish and low consumption of fresh fruit and vegetables. Interestingly, most dietary investigations from South East Asia have found that NPC patients have consumed less fresh fruits and vegetables than have controls. Although no significant interaction test was reported, these observations underline the interplay between dietary factors.

Beside the studies described above, association studies also have demonstrated potential gene-environment interactions, e.g. Cao et al. reported that a polymorphism of XRCC1 genes could confer increased NPC risk in smokers[24]. XRCC1 is involved in repair of sister chromatid exchange that can be induced by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), a tobacco-specific nitrosamine. Therefore, one could hypothesize that there could be a interconnection between XRCC1 mutations and cigarette smoking. However, the sample size was too small to show a convincing test of the interaction in this study. Similarly, CYP2E1, a demethylase involved in activation of pro-carcinogens N-nitrosodimethylamine (NDMA) and N-nitrosonornicotine (NNN), is repeatedly found to be associated with NPC risk[25-30]. NDMA is one of the carcinogenic nitrosamines contained in salted fish[31], or could be fermented from salted fish[32,33]. NNN is a kind of tobacco-specific N-nitrosamine[34]. Interestingly, the elevated NPC risk associated with CYP2E1 homozygous c2 allele in Taiwan was limited to non-smokers, which puts the interaction of CYP2E1 with tobacco smoke in doubt. Nevertheless, significant evidence for interplay between CYP2E1 and salted fish was not reported, and needs to be further investigated. In addition to gene-environment interaction, gene-gene interactions are also very possible, and should be considered in future investigations.

Limitations

In epidemiologic studies, multiple comparisons can be a serious problem, since many variables are tested in one study. Although this problem is not different from that of multiple groups studying the same variable and it may be argued that multiple testing problem extends that of a single study, the problem cannot be ignored fully, in particular when studying determinant with a low *a priori* of association. Thus using a significance level of 0.05 may be prone to finding false positive results, while Bonferroni correction or false discovery rate (FDR) control could be too conservative. Therefore, normally the cut off value is still set to 0.05, but marginally significant p values should be interpreted carefully. In the second chapter of this thesis, rancid butter, rancid sheep fat and cooked vegetables have very significant p values (p=0.00002, p=0.0000002, p=0.0006, respectively). These are unlikely to happen even in 50 independent tests. However, quaddid, preserved fish, citrus fruits, osban, olive condiments, cigarette smoking and domestic fume intake has intermediate to marginal significant levels. Among these variables, quaddid, cigarette smoking and domestic fume intake are consistent with other independent investigations, but osban, olive condiments, citrus fruits and incense usage are marginally significant and are against the *a priori* or previous publications, so that they might be spurious results.

Beside multiple comparisons, many other factors could lead to false positive findings in epidemiology studies, such as the lack of pre-selection of tested relationships, small sample size, small effect size, problems and bias in exposure measurement, recall bias occurring when patients dedicate their disease to certain food or environmental exposures, many teams in the same field chasing statistical significance and testing putative determinants with a low prior probability, and so on[35]. Unfortunately, the "gold" standard to definitely avoid all these problems is unattainable. However, there are several approaches to improve the reliability of the study results. In this thesis, a large-scale study with international collaborations based on a joint protocol, is more powerful than any smaller study within each country, and could avoid the false findings due to the selection of positive report from multiple independent teams addressing the same question. Moreover, the environmental factors studied in this thesis were restricted to those with high *a priori*, which could increase the pre-study probability of association, and hence the post-study probability in return. The questionnaire was designed with subjective exposure definitions. Nevertheless, exposure levels could be variable during a lifetime, thus the flexibility of exposure measurement could provide spaces for bias. Therefore, interpretation of the results requires careful consideration of the underlying bias.

Another consideration of the validity of the epidemiology study of complex disease is confounders. Examples in the second chapter are socio-economic status in the study of all exposures, and dietary consumptions in the study of tobacco and domestic fumes intake. In this chapter, the confounding effect is

controlled by matching of controls in the study design, and by adjusting for social economic status and dietary variables in the analyses. However, it is still likely that not all of the confounding effect was captured in our analyses.

As in any retrospective study, recall bias caused by differential recall by cases and controls of factors perceived to be risk factors for the disease in question, is possible in our investigation. However, since that there is no perceived relationship between cancer and most of the studied risk factors in the North African population, this is not likely to have been a major problem in our study. Similarly, because of the short lag time between diagnosis and enrollment of patients into the study, survival bias is not a likely explanation of the results.

In chapter 3, the familial clustering of NPC patients in South China could be explained by the sharing of environmental factors within a family, and/or genetic predisposition among the population. The observation that patients with younger age of disease onset have higher level of familial aggregation implies that it is more likely that the genetic component contributes to NPC familial risk. A subsequent complex segregation analyses on the same data confirmed that both environmental and genetic risk factors play a role in etiology of NPC in South China[36].

Linkage analysis is a powerful tool to locate disease predisposition genes in the human genome; however, the lack of recombination events in the studied families may limit its precision. To further pinpoint the location of the gene, it is necessary to recruit more families, and perform family based association test (FBAT), which can capture all meiotic recombination events in the evolution of the disease harboring chromosome, resulting in a small candidate region containing only a few genes. In addition, linkage analysis is not as powerful as association analysis to search for common genetic variants conferring a relatively modest increased risk of disease. Since familial NPC accounts for only a small proportion of all patients in the whole population, most of the NPC cases would have to be explained by low risk mutations in multiple genes. Therefore, to search for all NPC related genes, association analysis on sporadic cases on a genome-wide basis is an important adjunct to the linkage analyses.

Future studies

Environmental exposures are undoubtedly important factors for nasopharyngeal carcinoma development. Here the contribution of rancid fat, fumes and tobacco intake were investigated in North African countries. Nonetheless, there are still some other factors that are of interest that remain to be tested. For example, chronic ear and nose conditions, Chinese herbal medicines and occupational exposure to heat, dust and formaldehyde have been demonstrated to be related to NPC risk[37], although not broadly confirmed. Therefore, it would be necessary to investigate the contribution of medical history, traditional medicines and occupational exposures to NPC in North Africa.

This thesis reveals the association of rancid butter and rancid sheep fat with NPC, implicating the potential involvement of butyric acid as a causal agent. However, whether butyric acid is most likely the underlying factor that is responsible for the association requires further studies. In addition, the involvements of nitrosamines still need to be confirmed. A possible study design would be to assess the nitrosamine and butyric acid content of each food sample, and re-analyze the data in a nitrosamine or butyric acid content oriented way. The identification of the causal agents in food samples is crucial for future study designs, since that it would be a good implication of the underlying biological pathway of the disease development. It is also very important to future campaign to reduce NPC occurrence, locally or globally.

In this thesis, linkage of chromosome 4 in families from Guangdong province was discovered. However, thus far it has been difficult to identify the precise gene responsible for the linkage signal. The reasons for this include: i) the region is relatively large for candidate gene mutation scan; ii) there are too few families for family based association test; and iii) findings from multiple-case families could not be generalized to sporadic cases, thus using sporadic cases in association analysis in this chromosomal region seems unlikely to be successful. In light of this, it is necessary to recruit more families with multiple cases, which may be useful in narrowing down the region. In addition, the elucidation of the mutations depends on the sequencing of the region. The development of high throughput (shot gun) sequencing technology will have this feasible and affordable in the near future.

Considering the fact that only less than 2% of all NPC patients report having three or more relatives with NPC (chapter 3.1), and the strong genetic heterogeneity among the NPC families, it is very likely that the susceptibility gene on chromosome 4 only contributes to a small proportion of all NPC cases in South China. Therefore, to better understand the disease etiology of NPC in this area, it is indispensable for a large-scale association study using sporadic cases. Give thanks to the rapid development of technology and the accomplishment of the HapMap project, it is now possible to conduct an association search with up

to 1,000,000 SNPs covering the whole human genome. However, with this number of multiple testing, one needs a much larger sample size in the study than that used in this thesis. To detect a polymorphism with relative risk of 1.4 from 500,000 SNPs with 80% power, allowing maximum 10 false positives, at least 3000 cases and 3000 controls are needed. A large-scale association study design is also more beneficial in regard to detection of gene-environment and gene-gene interactions, and in turn, addressing interactions may also increase the power of the association test on individual risk factors. Therefore, the use of very detailed questionnaires should be considered in sample recruitment for such studies.

Chapter 3.4 presents a novel transmission/disequilibrium based test on association between micro- / minisatellites and disease risk, the trend-TDT test. This method could be useful when there is a biological basis to suspect a relationship between allele length and gene function, such as the CAG repeat within the Huntingtin gene, the minisatellite ILPR (Insulin-Linked Polymorphic Region, $[ACAGGGGTGTGGGG]_n$) located 5' to the Insulin gene, the CAG repeat within the androgen receptor gene, and etc. Nevertheless, this method could also be applied in other situations that a pseudo-microsatellite can be created, and can test whether this “microsatellite” is proportionally associated with increasing or decreasing disease risk along with its “allele length”. For example, a potential usage of the method could be an association test on HLA-A haplotypes with NPC risk. However, more simulations are necessary to clarify the performance of the test statistics in this situation.

At present, the complex etiology of nasopharyngeal carcinoma is far from completely understood, and a lot of further research will be required in the future. The complexity of the disease requires that researchers from different disciplines collaborate to bring a multidisciplinary focus to the problem, and development of novel methodology is needed. Above all we must ensure that knowledge derived from these studies is translated into public health measures within the large general at-risk population.

References

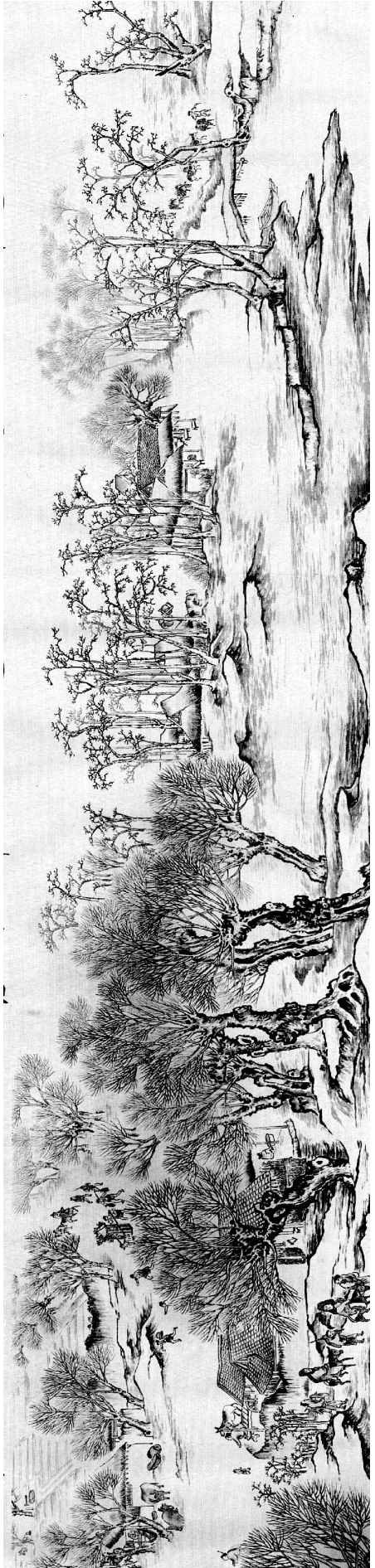
1. Jia, W.H. et al. Trends in incidence and mortality of nasopharyngeal carcinoma over a 20-25 year period (1978/1983-2002) in Sihui and Cangwu counties in southern China. *BMC Cancer* **6**, 178 (2006).
2. Parkin, D.M., Whelan, S.L., Ferlay, J., Teppo, L. & Thomas, D.B. *Cancer Incidence in Five Continents*, (International Agency for Research on Cancer, Lyon, France, 2002).
3. Armstrong, R.W., Armstrong, M.J. & Lye, M.S. Social impact of nasopharyngeal carcinoma on Chinese households in Selangor, Malaysia. *Singapore Med J* **41**, 582-7 (2000).
4. Wei, W.I. & Sham, J.S. Nasopharyngeal carcinoma. *Lancet* **365**, 2041-54 (2005).
5. Liu, M.T. et al. Prognostic factors affecting the outcome of nasopharyngeal carcinoma. *Jpn J Clin Oncol* **33**, 501-8 (2003).
6. Vaughan, T.L. et al. Nasopharyngeal cancer in a low-risk population: defining risk factors by histological type. *Cancer Epidemiol Biomarkers Prev* **5**, 587-93 (1996).
7. Krishna, S.M., James, S., Kattoor, J. & Balaram, P. Human papilloma virus infection in Indian nasopharyngeal carcinomas in relation to the histology of tumour. *Indian J Pathol Microbiol* **47**, 181-5 (2004).
8. Punwaney, R. et al. Human papillomavirus may be common within nasopharyngeal carcinoma of Caucasian Americans: investigation of Epstein-Barr virus and human papillomavirus in eastern and western nasopharyngeal carcinoma using ligation-dependent polymerase chain reaction. *Head Neck* **21**, 21-9 (1999).
9. Chan, A.T., Teo, P.M. & Johnson, P.J. Nasopharyngeal carcinoma. *Ann Oncol* **13**, 1007-15 (2002).
10. Lo, K.W., To, K.F. & Huang, D.P. Focus on nasopharyngeal carcinoma. *Cancer Cell* **5**, 423-8 (2004).
11. Lo, A.K. et al. Epstein-Barr virus infection alters cellular signal cascades in human nasopharyngeal epithelial cells. *Neoplasia* **8**, 173-80 (2006).
12. Jeannel, D., Bouvier, G. & Hubbert, A. Nasopharyngeal Carcinoma: An Epidemiological Approach to Carcinogenesis. in *Cancer Surveys* Vol. 33 125-155 (1999).
13. Jeannel, D. et al. Diet, living conditions and nasopharyngeal carcinoma in Tunisia--a case-control study. *Int J Cancer* **46**, 421-425 (1990).
14. Laouamri, S., Hamdi-Cherif, M., Sekfali, N., Mokhtari, L. & Kharchi, R. [Dietary risk factors of nasopharyngeal carcinoma in the Setif area in Algeria]. *Rev Epidemiol Sante Publique* **49**, 145-156 (2001).
15. Ammor, S., Baali, A., Hubert, A. & Cherkaoui, M. Epidemiological approach of the nasopharyngeal carcinoma in the region of Marrakech. *Antropo* **5**, 39-47 (2003).
16. 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* **69**, 508-514 (1994).

17. Yu, M.C., Ho, J.H., Lai, S.H. & Henderson, B.E. Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res* **46**, 956-61 (1986).
18. Chen, D.L. & Huang, T.B. A case-control study of risk factors of nasopharyngeal carcinoma. *Cancer Lett* **117**, 17-22 (1997).
19. Lu, S.J. et al. Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature* **346**, 470-1 (1990).
20. Xiong, W. et al. A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res* **64**, 1972-4 (2004).
21. Chung, M.J., Lee, S.H. & Sung, N.J. Inhibitory effect of whole strawberries, garlic juice or kale juice on endogenous formation of N-nitrosodimethylamine in humans. *Cancer Lett* **182**, 1-10 (2002).
22. Vermeer, I.T., Moonen, E.J., Dallinga, J.W., Kleinjans, J.C. & van Maanen, J.M. Effect of ascorbic acid and green tea on endogenous formation of N-nitrosodimethylamine and N-nitrosopiperidine in humans. *Mutat Res* **428**, 353-61 (1999).
23. Arranz, N., Haza, A.I., Garcia, A., Rafter, J. & Morales, P. Protective effect of vitamin C towards N-nitrosamine-induced DNA damage in the single-cell gel electrophoresis (SCGE)/HepG2 assay. *Toxicol In Vitro* (2007).
24. Cao, Y. et al. Polymorphisms of XRCC1 genes and risk of nasopharyngeal carcinoma in the Cantonese population. *BMC Cancer* **6**, 167 (2006).
25. Hildesheim, A. et al. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst* **89**, 1207-12 (1997).
26. Hildesheim, A. et al. Cytochrome P4502E1 genetic polymorphisms and risk of nasopharyngeal carcinoma: results from a case-control study conducted in Taiwan. *Cancer Epidemiol Biomarkers Prev* **4**, 607-10 (1995).
27. Yang, X.R. et al. Evaluation of risk factors for nasopharyngeal carcinoma in high-risk nasopharyngeal carcinoma families in Taiwan. *Cancer Epidemiol Biomarkers Prev* **14**, 900-5 (2005).
28. Wang, J., Liu, Z., Lu, X. & Chen, B. [Cytochrome P450 1A1 and cytochrome P450 2E1 gene polymorphisms in Guangzhou Hans]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **17**, 259-61 (2000).
29. He, Z.M., Chen, Z.C., Yuan, J.H. & Wang, S.L. Expression of Cytochrome p450 2E1 in nasopharynx and polymorphism analysis. *Chinese Journal of Pathophysiology* **16**, 1060 (2000).
30. Kongruttanachok, N. et al. Cytochrome P450 2E1 polymorphism and nasopharyngeal carcinoma development in Thailand: a correlative study. *BMC Cancer* **1**, 4 (2001).
31. Poirier, S. et al. Volatile nitrosamine levels in common foods from Tunisia, south China and Greenland, high-risk areas for nasopharyngeal carcinoma (NPC). *Int J Cancer* **39**, 293-296 (1987).
32. Weng, Y.M., Hotchkiss, J.H. & Babish, J.G. N-nitrosamine and mutagenicity formation in Chinese salted fish after digestion. *Food Addit Contam* **9**, 29-37 (1992).

33. Zou, X.N., Lu, S.H. & Liu, B. Volatile N-nitrosamines and their precursors in Chinese salted fish--a possible etological factor for NPC in china. *Int J Cancer* **59**, 155-158 (1994).
34. Hecht, S.S. & Hoffmann, D. Tobacco-specific nitrosamines, an important group of carcinogens in tobacco and tobacco smoke. *Carcinogenesis* **9**, 875-884 (1988).
35. Ioannidis, J.P. Why most published research findings are false. *PLoS Med* **2**, e124 (2005).
36. 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* **13**, 248-52 (2005).
37. Chang, E.T. & Adami, H.O. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* **15**, 1765-77 (2006).

Chapter 5

Summary
Samenvatting
总结



Summary

Nasopharyngeal carcinoma (NPC) is a malignant tumor arising from the epithelial lining of the nasopharynx. It is rare in most parts of the world, but much more common in South East Asia, North Africa and in the Arctic Circle of North America. Among individuals less than 65 years of age, it is the third most prevalent cancer type in Hong Kong, and the second most prevalent in Singapore Chinese. Since the disease mostly affects males in their 40s, the occurrence of nasopharyngeal carcinoma has a strong social impact in endemic areas.

NPC is a complex disease with both environmental and genetic factors playing a role in its development. Suggested risk factors include tobacco, salted fish, domestic fumes, occupational dust and heat, herbal medicine, Epstein-Barr virus (EBV) activation, and a familial history of NPC. In this thesis, environmental and genetic factors in the etiology of nasopharyngeal carcinoma are addressed.

Chapter 1 is a general introduction to nasopharyngeal carcinoma. The basic epidemiology of NPC is presented, and some known risk factors are described, including diet, life-style, Epstein-Barr virus and genetics. Their potential involvement in the carcinogenesis of NPC is briefly discussed. An outline of the research described in this thesis is also given.

Chapter 2 presents the investigation of environmental risk factors for NPC in North Africa, which is part of a multi-center case-control study conducted in Maghrebian countries including Tunisia, Algeria and Morocco. In this study, 636 cases and 615 controls, matched by center, childhood household type (rural or urban), age and sex, were recruited from 2002 to 2005. In this chapter, we identified rancid butter and rancid sheep fat as dietary risk factors, and confirmed the reverse association of cooked vegetables. The results imply that butyric acid is potentially involved in NPC development. In the second part of this chapter, we found that tobacco intake has a marginal effect on NPC risk, which may be more important in squamous cell NPC than in undifferentiated NPC. Intake of domestic fumes produced from cooking with kanoun is another significant risk factor for NPC. In addition, restricted permutation analysis indicated that cannabis smoking was significantly associated with increased NPC risk, and that this association was not due to the strong correlation between cannabis and tobacco consumption.

In chapter 3, genetic studies of NPC are presented. Chapter 3.1 describes the familial clustering of NPC in Guangdong China. Between January 1999 and June 2001, 2252 patients seen at the Cancer Center of Sun Yat-Sen University were successfully interviewed, with a structured questionnaire regarding familial cancer history. In this study, 14% of all patients reported a familial history of NPC, and 1.8% reported more than one patient in blood-relatives. When compared with the data from the cancer registry of Hong Kong, it is estimated that the risk of NPC among first degree relatives of the NPC probands is roughly twice that of the general population, and that the risk is even further increased among first degree relatives of the NPC probands with an early age of disease onset. The results of this study confirm the familial clustering of NPC patients in Guangdong China, and provide evidence that genetic susceptibility may contribute to the occurrence of NPC in Guangdong.

Chapter 3.2 describes a whole genome scan for susceptibility gene(s) for NPC. In this study, 165 individuals from 20 families recruited from Guangdong, China were genotyped for 400 microsatellite markers covering the human genome at an average genetic interval of 10 cM. Parametric linkage analysis using a dominant model provided evidence of linkage to the marker D4S405 on chromosome 4 with a logarithm of odds for linkage (LOD) score of 3.06 and a heterogeneity-adjusted LOD (HLOD) score of 3.21. Fine mapping with additional markers flanking D4S405 resulted in a LOD score of 3.54 and HLOD score of 3.67 for the region 4p15.1–q12. Further haplotype analyses (chapter 3.3) narrowed the region harboring the susceptibility gene down to a 15.4 Mb segment, between D4S2950 and D4S3347.

Chapter 3.4 presents a novel transmission/disequilibrium-based test for association between micro-/minisatellites and disease risk, the trend-TDT test. A potential application of this method could be to test the associations between HLA-A haplotypes and NPC risk.

Finally, the findings of this thesis are discussed in the context of previous publications in chapter 4. Prospective studies are also presented in this chapter.

Samenvatting

Nasofaryngeaal carcinoom (NPC) is een kwaadaardige tumor die voortkomt uit de epitheliale bekleding van de nasofarynx. Het is zeldzaam in grote delen van de wereld maar komt vaker voor in Zuidoost-Azië, Noord-Afrika en het noordpoolgebied van Noord-Amerika. In personen jonger dan 65 jaar is het het op twee na meeste voorkomende kankertype in Hong Kong, Signapore staat op de tweede plaats. Aangezien met name mannen van rond de 40 jaar deze kanker krijgen, heeft NPC een sterke sociale impact op endemische gebieden.

NPC is een complexe ziekte waar zowel omgevings- als genetische factoren een rol spelen. Tabak, gezouten vissen, vochtige ruimtes, stof en hitte op de werkplek, geneeskrachtige kruiden, Epstein-Barr virus (EBV) activering, en een familiegeschiedenis positief voor NPC zijn risico factoren. In dit proefschrift komen zowel omgevingsfactoren als genetische factoren in de ontwikkeling van NPC aan de orde.

Hoofdstuk 1 geeft een algemene inleiding over NPC. De algemene epidemiologie van NPC wordt gepresenteerd, en een aantal bekende risicofactoren worden beschreven, zoals dieet, levensstijl, EBV en genetische factoren. Hun mogelijke rol in de ontwikkeling van NPC wordt kort besproken. Er wordt ook een overzicht gegeven van het onderzoek.

Hoofdstuk 2 beschrijft het onderzoek van omgevingsrisicofactoren voor NPC in Noord-Afrika, dat deel uitmaakt van een multi-center case-controle studie die is uitgevoerd in de Maghrebian landen Tunesië, Algerije en Marokko. Voor deze studie werden 636 patienten en 615 controles geïncludeerd bij wie de diagnose tussen 2002 tot 2005 is vastgesteld. In de analyses werd gecontroleerd op onderzoekcentrum, woonomstandigheden als kind (landelijk of stedelijk), leeftijd en geslacht. In dit hoofdstuk, bevestigen dat wij ranzig boter en ranzig schapenvet als dieetrisicofactor en bevestigden wij de beschermende werking van gekookte groenten. De resultaten geven aan dat boterzuur mogelijk bij de ontwikkeling van NPC l betrokken is. In het tweede deel van dit hoofdstuk, vonden wij dat tabaksopname een marginale risicofactor voor NPC is, die mogelijk belangrijker is in plaveiselcel NPC dan in niet gedifferentieerde NPC. Opname van dampen die tijdens het koken met kanoun vrijkomen is een andere betekenisvolle risicofactor voor NPC. Bovendien wijst de beperkte permutatieanalyse erop dat het roken van cannabis geassocieerd kan worden met een significant verhoogd risico voor NPC,

en dat deze associatie niet toe te schrijven is aan de sterke correlatie tussen cannabis en tabaksconsumptie.

In hoofdstuk 3, worden de genetische studies van NPC gepresenteerd. Hoofdstuk 3.1 beschrijft de familiaire clustering van NPC in Guangdong, China. Tussen januari 1999 en juni 2001 werden 2252 patiënten gezien, die op het Kanker Centrum van de Universiteit van Sun Yat-Sen werden ondervraagd over hun familiegeschiedenis omtrent kanker. In deze studie, meldde 14% van alle patiënten een familiegeschiedenis van NPC, en 1,8% meer dan één patiënt als bloedverwant. In vergelijking tot de gegevens van de kankerregistratie van Hong Kong, blijkt het risico voor NPC onder verwanten van de eerste graad van Chinese NPC probands ongeveer twee keer zo hoog als voor de algemene bevolking. Het risico wordt nog verder verhoogd bij familieleden van probands die op jonge leeftijd NPC hebben gekregen. De resultaten van deze studie bevestigen de clustering van families met NPC in Guangdong, China en suggereren dat genetische aanleg een bijdrage levert aan het voorkomen van NPC in Guangdong.

Hoofdstuk 3.2 beschrijft een genomische zoektocht ("whole genome scan") naar genen die een rol spelen in NPC. In deze studie werden 165 individuen van 20 families uit Guangdong getest met 400 markers die het menselijke genoom met een gemiddeld genetische afstand van 10 cM dekken. Parametrisch koppelingsonderzoek met een dominant overervingmodel leidde tot het identificeren van D4S405 op chromosoom 4 met een LOD score van 3,06 en een LOD score onder heterogeneity (HLOD) van 3.21. Een gedetailleerde analyse met extra markers die D4S405 flankeren, resulteerde in een LOD score van 3,54 en een HLOD score van 3,67 voor het gebied 4p15.1-q12. Verdere haplotype analyses (hoofdstuk 3.3) versmalde het mogelijk gebied van het NPC gevoeligheidsgen tot een segment van 15.4Mb, tussen markers D4S2950 en D4S3347.

Hoofdstuk 3.4 beschrijft een nieuwe op de transmissie/disequilibrium-gebaseerde test, de trendTDT test, om associaties tussen micro-/minisatellites en ziekterisico te vinden. Een mogelijk applicatie van deze methode zou de vergelijking tussen HLA-A haplotypes en het risico op NPC kunnen zijn.

Tot slot worden in hoofdstuk 4 de bevindingen van dit proefschrift besproken in de context van bestaande publicaties. Tevens worden prospectieve studies voorgesteld in dit hoofdstuk.

总结

鼻咽癌(NPC)是一种原发于鼻咽粘膜被覆上皮的恶性肿瘤。它在世界大部分地区是罕见的，但在东南亚、北非和北美洲的北极圈地区更常见。在所有癌症种类中， 鼻咽癌在年龄低于 65 岁的人群中的累积发病率，在香港排行第三，在新加坡华人中排行第二。由于这种疾病主要高发于 40 岁左右的男性，在高发区鼻咽癌的发生有很强的社会影响。

鼻咽癌是一种复杂疾病，环境因素和遗传因素都对它的发生发展产生了影响。可能的风险因素包括烟草、咸鱼、室内烟雾、职业性尘土和热、草药、Epstein-Barr 病毒(EBV)的活化，以及家族史。此论文研究了鼻咽癌的病因学，包括环境因素和遗传因素。

第 1 章是鼻咽癌的简介，它介绍了鼻咽癌基本的描述性流行病学数据，和一些已知的风险因素，包括饮食、生活方式、Epstein-Barr 病毒和遗传。这一章还扼要讨论了这些因素在鼻咽癌发病机理中的作用，并概述了本论文的研究。

第 2 章是对北非的环境风险因素的调查，它是一个在 Maghrebian 国家进行的、多中心合作的、病例－对照研究的一部分，这些国家包括突尼斯、阿尔及利亚和摩洛哥。在这项研究中，我们从 2002 年到 2005 年收集了 636 个病例和 615 个对照，这些病例－对照根据收集医院、童年时期家庭类型(农村或都市)、年龄和性别进行了匹配。在这一章里，我们发现馊黄油和馊绵羊油脂是鼻咽癌的饮食风险因素，并证实了熟蔬菜和鼻咽癌风险的反向关联。这些结果提示，丁酸（馊黄油的一种成分）是潜在的鼻咽癌致病因素。在这个章节的第二部份，我们发现烟草对鼻咽癌的发生只有微弱的影响，并且其对鳞狀細胞癌的作用比对非分化癌的作用也许更重要。用 kanoun 烹调引起的室内烟雾是另一个重要的鼻咽癌风险因素。另外，受限的置换检验显示，吸大麻烟跟鼻咽癌的发病风险显著关联，并且，大麻烟与香烟消费的相关性并不能解释这种关联。

第 3 章是鼻咽癌的遗传学研究。其中第 3.1 节描述了鼻咽癌在中国广东的家族聚集性现象。在 1999 年 1 月和 2001 年 6 月之间，从中山大学的肿瘤防治中心我们成功地采访了 2252 名患者，并完成了有关家族癌症史的调查。在这项研究中，14%的患者有鼻咽癌家族史，1.8%有超过一个血亲患鼻咽癌。在把这些数据与香港癌病资料统计中心的数据进行比较后，估算出鼻咽癌先证者的一级亲属的患病风险是普通人群的大致两倍，并且此风险在早发

的鼻咽癌先证者的一级亲属中进一步提高。这项研究的结果证实了中国广东鼻咽癌患者的家族聚集性现象，并为鼻咽癌的遗传易感性提供了证据。

第 3.2 节是对鼻咽癌易感基因的全基因组连锁分析扫描。在这项研究中，我们用 400 个微卫星标记对从中国广东收集的 20 个高发家系的 165 个个体进行了基因分型，这些标记以 10 厘摩的间距覆盖了整个人类基因组。参数连锁分析显示 4 号染色体的 D4S405 标记跟鼻咽癌连锁，其 LOD 值是 3.06，HLOD 值是 3.21。用更多的标记进行精细定位后，在 4p15.1-q12 区域的最高 LOD 值达到 3.54，最高 HLOD 值达到 3.67。对此染色体区域的进一步单体型分析(第 3.3 节)把易感基因的候选区域缩窄到一个长度为 15.4Mb 的区段，位于微卫星标记 D4S2950 和 D4S3347 之间。

第 3.4 节提出了一个基于传递不平衡的新统计方法（trend-TDT），用于检验微卫星和小卫星标记与疾病风险的关联。这个方法有可能用于对 HLA-A 单体型的鼻咽癌关联分析。

最后，第 4 章对此论文的研究结果进行了讨论，并提出了将来的研究方向。

Acknowledgments

Many people have helped me in my research. Without their guidance and collaborations, this thesis will not be possible.

First of all I sincerely thank prof. David E. Goldgar. With your help, I got the chance to go to Lyon and research on nasopharyngeal carcinoma in the International Agency for Research on Cancer; and the chance to go to Salt Lake City and study in the University of Utah. In research, you have given me a lot of smart advices in various fields of study, and detailed suggestions in scientific writing. These are so important for my career. Thank you for the accommodations in Salt Lake City, driving in the avenues in search of an apartment, looking for a wheelchair when I cannot walk, the thanks-giving dinner, and so many other helps that cannot be listed comprehensively.

Special thanks to prof. Cornelia van Duijn. You provided me the scholarship and the opportunity to study genetic epidemiology, toward both Doctor of Science and Doctor of Philosophy degrees. Your insightful and detailed comments to my papers and thesis have taught me how a scientific paper should be written. The participation in the GRIP project is invaluable to me in learning genetic epidemiology. Thank you for the nice party at your place, and the encouragement to run Marathon.

I am so grateful to dr. Marilys Corbex, who let me play a part in the case-control study of nasopharyngeal carcinoma in North Africa, and the methodology study of trend-TDT. Although we were not working face to face most of the time, our discussion about diet and tobacco has been enlightening to me. I am really happy that I can work on environmental factors, which will be helpful for my future studies. Your hospitality, generosity, enthusiasm, kindness have made you always a pleasure to work with!

I am greatly thankful to the members of my doctoral reading and evaluation committee: prof. Jan Willem W. Coebergh, prof. Pierre Busson, and prof. Ben A. Oostra; the large committee members and experts: prof. Jeanine J. Houwing Duistermaat, prof. Vincenzo Bonifati, and dr. Yurii Aulchenko. Thanks for the time you have taken, and the criticisms and suggestions you have raised.

My sincere appreciation is given to prof. Yi-Xin Zeng. With your support and guidance I started the research on genetic epidemiology of NPC, my favorite subject. This work, as well as the collaboration with other researchers in Guangzhou, is crucial to my PhD. You have been patient with me, confident in me, and considerate of me, which I will never forget. I am also indebted to prof. Wei-hua Jia: you afforded me the opportunity to work on familial clustering of nasopharyngeal carcinoma. Our works have been really happy. More than a colleague, you are also a true friend of mine.

Researchers in the International Agency for Research on Cancer have substantially contributed to my education. They include, but by no means limited to, dr. Paul Brennan, dr. Paolo Boffeta, dr. James McKay, dr. Rayjean Hung, dr. Sean Tavtigian, dr. Mia Hashibe and so on. For the African study, I sincerely appreciate dr. Wided Ben Ayoub, dr. Meriem Khyatti, dr. Sami Dahmoul, dr. Messaoud Ayad, dr. Fatima Maachi, dr. Wided Bedadra, dr. Meriem Abdoun, dr. Sarah Mesli, dr. Mokhtar Hamdi-Cherif, dr. Kada Boualga, dr. Noureddine Bouaouina, dr. Lotfi Chouchane, dr. Abdellatif Benider, dr. Farhat Ben Ayed, dr. Mohamed Maamer and dr. Hanae Bekkali for the collaboration in Tunisia, Algeria and Morroco. Many thanks to dr. Majida Jalbout for the discussion about dietary tradition of Tunisia, Ms. Valérie Gaborieau for efficient data management, Ms. Colette Bonnardel for data entry and checking, Ms. Hélène Renard and Mr. Othman Yaqoubi for database design, and dr. Dominique Jeannel and dr. Annie Sasco for the help in designing the questionnaire.

Many people have friendly contributed to my happiness in France. Charles Hsu is super funny, but very considerate and generous. It is really my pleasure to come to know you and hang out with you, especially to Turkey. Kasia, being your office-mate was one of my joyful experiences in France, and the Jazz à Vienne festival is unforgettable. Your advice to me did change my life toward a positive direction. Paolo and James, thanks for the skiing trip and snowboarding lecture in Italy. James was greatly helpful in my last few days in France in 2006. My gratitude extends to Carolyn Dresler, Fabrice Odefrey, Yvette Granjard, Stéphanie Monnier, Meïli Baragatti, Manuela Marron, Gary Clifford, Anne Cust and her husband, Ghislaine Scelo, Amir Sapkota and his wife, Wei-min Tong, Min Dai, Xiu-ju Sun, Hong Shi, Bing-chen Zhu, Dong-fang Gai, Yun-qi Xue, and many other friends and colleagues.

I thank all the people I have been working with in Rotterdam. Yurii and I have thoughtful discussions on methodology and diabetes. Alejandro pushed me to run from his home to the Erasmus Medical Center. Fan and Luba have been very helpful, and I really enjoyed Fan's sea food and Luba's tortilla. Mojgan and Nahid have very nice Persian cooking and Persian music. Pascual has fine recipes of Spanish sangria. Aaron and Angela gave me suggestions on printing thesis. Fakhredin and Stefano, they have always been kind and amiable. All of you guys are wonderful people.

In the writing of the thesis, Judith Brouwer and Margaret Jacobs have helped me translate the summary into Dutch. Macoy McMurray helped me find a Dutch speaker in Salt Lake City, and Alexis Deforge helped me design the cover page of the thesis. Fan and Luba, you have been of great help to me, I am so grateful to your kindness. Macoy, it's so joyful to meet you and the other wonderful people in Salt Lake City, your hospitality make me feel home in a new place.

In the research conducted in Guangdong China, a lot of people have helped us in the sample recruitment; these include dr. Feng Zhang, dr. Teng-Bo Huang, dr. Hui-Ming Huang, dr. Mang-Quan Deng, dr. Hao-Yuan Mo and dr. Ming-Fang Hong from the department of nasopharyngeal carcinoma, Cancer Center of Sun Yat-sen University; dr. Wei-Han Hu and dr. Shao-Xiong Wu from Department of Radiotherapy, Cancer Center of Sun Yat-sen University; and Ms. Li-Zhen Chen, Shao-Wen Jian, Mr. Ping Huang, Mr. Qi-Sheng Feng, Ms. Li-Xi Huang, Ms. Xing-Juan Yu, Duang Li, prof. Wei-Hua Jia, prof. Yan Fang and dr. Jing-Liu Zhu from the Department of Laboratory, Cancer Center of Sun Yat-sen University. Particularly, dr. Feng Zhang and Ms. Li-Zhen Chen have spent a lot of time and efforts in visiting NPC families far from Guangzhou, many thanks. In addition, Ms. Li-Zhen Chen, Mr. Qi-Sheng Feng, Ms. Li-Xi Huang, and Ms. Xing-Juan Yu have efficiently carried out the DNA extraction, genotyping and other laboratory work. During my study, prof. Mary Claire King, prof. Yin Yao Shugart, prof. Zhu Chen, prof. Wei Huang, dr. Ming K. Lee, dr. Wan-Qing Liu and dr. Xiao-Ming Liu have given me suggestions and have critically reviewed one of my papers.

I thank all the patients and other persons participated in the environmental and genetic study of nasopharyngeal carcinoma, both in China and in North Africa. Although their participation is not directly beneficial to them, the findings of these studies can promote public health in the future.

No language can adequately express my gratitude to my parents. You have been very supportive, considerate and encouraging. I would not be able to graduate, without your love and your sacrifice, a debt I can never repay.

List of Publications

Feng BJ*, Goldgar DE & Corbex M. TrendTDT - A Transmission Disequilibrium based association Test on Functional Mini/Microsatellites. *BMC genetics* 2007 (in press).

Arias-Vasquez A, de Lau L, Pardo L, Liu F, **Feng BJ**, Bertoli-Avella A, Isaacs A, Aulchenko Y, Hofman A, Oostra B, Breteler M, van Duijn C. Relationship of the Ubiquilin 1 gene with Alzheimer's and Parkinson's disease and cognitive function. *Neurosci Lett*. 2007 Aug 31;424(1):1-5.

Hung RJ, Moore L, Boffetta P, **Feng BJ**, Toro JR, Rothman N, Zaridze D, Navratilova M, Bencko V, Janout V, Kollarova H, Szeszenia-Dabrowska N, Mates D, Chow WH, Brennan P. Family History and the Risk of Kidney Cancer: a Multicenter Case-control Study in Central Europe. *Cancer Epidemiol Biomarkers Prev*. 2007 Jun;16(6):1287-90.

Liu F, Arias-Vásquez A, Sleegers K, Aulchenko YS, Kayser M, Sanchez-Juan P, **Feng BJ**, Bertoli-Avella AM, van Swieten J, Axenovich TI, Heutink P, van Broeckhoven C, Oostra BA, van Duijn CM. A genome-wide screen for late onset Alzheimer's disease in a genetically isolated Dutch population. *Am J Human Genet*. 2007 Jul;81(1):17-31.

Feng BJ, Jalbout M, Ben Ayoub W, Khyatti M, Dahmoul S, Ayad M, Maachi F, Bedadra W, Abdoun M, Mesli S, Hamdi-Cherif M, Boualga K, Bouaouina N, Chouchane L, Benider A, Ben Ayed F, Goldgar D.E. & Corbex M. Dietary Risk factors for Nasopharyngeal Carcinoma in Maghrebian Countries. *Int J Cancer*. 2007; 121(7):1550-5

Weng DS, Li JT, Mai SJ, Pan ZZ, **Feng BJ**, Feng QS, Huang LX, Wang QJ, Li YQ, Yu XJ, Chen SP, He J, Xia JC. Identification of a new target region on the long arm of chromosome 7 in gastric carcinoma by loss of heterozygosity. *World J Gastroenterol*. 2006 Apr 21;12(15):2437-40.

Xia JC, Weng DS, Li JT, Qin HD, Mai SJ, **Feng BJ**, Fan Q, Feng QS, Huang LX, Yu XJ, Pan ZZ, Li YQ, Wang QJ, Zhan YQ, Chen SP, He J, Huang WL, Wu PH, Zeng YX. Loss of heterozygosity analysis of a candidate gastric carcinoma tumor suppressor locus at 7q31. *Cancer Genet Cytogenet*. 2006 Apr 15;166(2):166-72.

Jiang RC, Qin HD, Zeng MS, Huang W, **Feng BJ**, Zhang F, Chen HK, Jia WH, Chen LZ, Feng QS, Zhang RH, Yu XJ, Zheng MZ, Zeng YX. A functional variant in the transcriptional regulatory region of gene LOC344967 cosegregates with disease phenotype in familial nasopharyngeal carcinoma. *Cancer Res*. 2006 Jan 15;66(2):693-700.

Jia WH, Collins A, Zeng YX, **Feng BJ**, Yu XJ, Huang LX, Feng QS, Huang P, Yao MH, Shugart YY. Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance. *Eur J Hum Genet.* 2005 Feb;13(2):248-52.

Jia WH, Xu ZL, **Feng BJ**, Huang LX, Yu XJ, Feng QS, Zhang F, Zeng YX. [Epidemiological study of nasopharyngeal carcinoma risk in relatives of high-risk families in Guangdong] *Ai Zheng*. 2004 Jul;23(7):767-70. Chinese.

Jia WH, **Feng BJ**, Xu ZL, Zhang XS, Huang P, Huang LX, Yu XJ, Feng QS, Yao MH, Shugart YY, Zeng YX. Familial risk and clustering of nasopharyngeal carcinoma in Guangdong, China. *Cancer*. 2004 Jul 15;101(2):363-9.

Jiang JH, Jia WH, Chen HK, **Feng BJ**, Qin HD, Pan ZG, Shen GP, Huang LX, Feng QS, Chen LZ, Lin DX, Zeng YX. Genetic polymorphisms of CYP2A13 and its relationship to nasopharyngeal carcinoma in the Cantonese population. *J Transl Med.* 2004 Jun 29;2(1):24.

Chen HK, **Feng BJ**, Liang H, Zhang RH, Zeng YX. The susceptibility gene for familial nasopharyngeal carcinoma is mapped on chromosome 4p11-p14 by haplotype analyses. *Chinese Science Bulletin.* 2003; 48(21):2327-30.

Jia WH, Shao JY, **Feng BJ**, Zeng YX. Genetic Component Involved in Nasopharyngeal Carcinoma Development. *Cancer Reviews: Asia-Pacific*. 2003; 1(1):31-49.

Shugart YY, **Feng BJ**, Collins A. The power and statistical behaviour of allele-sharing statistics when applied to models with two disease loci. *J Genet*. 2002 Dec;81(3):99-103.

Feng BJ, Huang W, Shugart YY, Lee MK, Zhang F, Xia JC, Wang HY, Huang TB, Jian SW, Huang P, Feng QS, Huang LX, Yu XJ, Li D, Chen LZ, Jia WH, Fang Y, Huang HM, Zhu JL, Liu XM, Zhao Y, Liu WQ, Deng MQ, Hu WH, Wu SX, Mo HY, Hong MF, King MC, Chen Z, Zeng YX. Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat Genet*. 2002 Aug;31(4):395-9.

Zhang XS, Song KH, Mai HQ, Jia WH, **Feng BJ**, Xia JC, Zhang RH, Huang LX, Yu XJ, Feng QS, Huang P, Chen JJ, Zeng YX. The 30-bp deletion variant: a polymorphism of latent membrane protein 1 prevalent in endemic and non-endemic areas of nasopharyngeal carcinomas in China. *Cancer Lett*. 2002 Feb 8;176(1):65-73.

Shao JY, Wang HY, Huang XM, Feng QS, Huang P, **Feng BJ**, Huang LX, Yu XJ, Li JT, Hu LF, Ernberg I, Zeng YX. Genome-wide allelotype analysis of sporadic primary nasopharyngeal carcinoma from southern China. *Int J Oncol*. 2000 Dec;17(6):1267-75.

About the Author

Mr. Bingjian Feng was born on Feb 13, 1976 in Guangzhou, China. After graduating from the Guang-ya High School in 1994, he entered the Sun Yat-sen University, Guangzhou, China, majoring in biochemistry. Soon after that, finding that he loves computer science more than bacteria or virus, he then minored in computer software engineering, and received a National Rank Certificate of Software Engineer at Advanced Level. Indulging himself in computer science, however, did not stop him from earning distinction scholarships for almost each year. In 1998, he graduated and transferred to the Cancer Center of Sun Yat-sen University, where he started the genetic epidemiological research on nasopharyngeal carcinoma, under the auspices of prof. Yi-xin Zeng. With his research achievements, he completed a Master of Science degree in oncology, and won several national awards, including the National Natural Science Award from the State Council of China. In 2004, he relocated to Rotterdam, the Netherlands, to pursue a Doctor of Science degree in genetic epidemiology in the Genetic Epidemiology Unit of the Department of Epidemiology & Biostatistics, Erasmus University Rotterdam, under the supervision of prof. Cornelia van Duijn. During the study he has attained a substantial amount of knowledge about genetic research in isolated populations, Alzheimer's disease, and cooking fish. Thereafter, supported by the Special Training Award, jointly supervised by prof. David Goldgar, dr. Marilys Corbex, and prof. Cornelia van Duijn, he continued working on nasopharyngeal carcinoma toward a Doctor of Philosophy, in the Genetic Epidemiology Unit of the International Agency for Research on Cancer, Lyon, France, where he became a big fan of jazz music, particularly Bossa Nova. In 2006, he went to University of Utah, Salt Lake City, USA, and began research on psoriasis, celiac disease, breast cancer, tennis and snowboarding.

List of Abbreviations

CCSYSU	Cancer Center of Sun Yat-Sen University
EBV	Epstein-Barr Virus
FDR	First Degree Relative
HLA	Human Leukocyte Antigen
HLOD	Heterogeneity-adjusted LOD
IARC	International Agency for Research on Cancer
LD	Linkage Disequilibrium
LOD	Log of the odds
LOH	Loss Of Heterozygosity
NPC	NasoPharyngeal Carcinoma
NDMA	N-nitrosodimethylamine
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
NNN	N-nitrosonornicotine
NPYR	N-nitrosopyrrolidine
NPIP	N-nitrosopiperidine
SES	Social Economic Status
SIR	Standardized Incidence Ratio
SNP	Single-Nucleotide Polymorphism
TDT	Transmission/Disequilibrium Test
UCNT	Undifferentiated Carcinoma Nasopharyngeal Type
UNPC	Undifferentiated NasoPharyngeal Carcinoma