GENETIC POLYMORPHISMS ASSOCIATED WITH NASOPHARYNGEAL CARCINOMA SUSCEPTIBILITY IN MALAYSIAN CHINESE

YEW POH YIN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2012

GENETIC POLYMORPHISMS ASSOCIATED WITH NASOPHARYNGEAL CARCINOMA SUSCEPTIBILITY IN MALAYSIAN CHINESE

YEW POH YIN

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2012

ABSTRACT

Nasopharyngeal carcinoma (NPC) is a malignancy arising from the epithelial lining of the nasopharynx. The high prevalence of NPC in Asian countries and certain ethnic groups suggest the involvement of genetic susceptibility in the development of NPC, therefore identification of NPC susceptibility genes is essential for elucidating their pathogenesis. In the present study, we performed a population based case-control association study to identify the NPC susceptibility genes in a Malaysian Chinese population, by using candidate gene and genome-wide screening approaches in 447 NPC patients and 487 healthy controls. For candidate gene approach, 11 candidate genes which are involved in the xenobiotic metabolism (*CYP1A1, CYP2A13, CYP2E1, EPHX1, EPHX2, GSTM1, GSTT1* and *GSTP1*) and immune response (*PIGR, LPLUNC1* and *SPLUNC1*) were examined for their associations with NPC. Of these 11 candidate genes, *CYP2E1*, *EPHX2* and *SPLUNC1* were found to be associated with NPC susceptibility. Five SNPs in *CYP2E1* (rs2070672, rs3813865, rs3813870, rs8192772 and rs915906) were significantly associated with increased risk of NPC (P = 0.00081-0.0032, OR = 2.92 – 3.69), while the T allele for rs4149244 of *EPHX2* was associated with decreased risk of NPC (P = 0.0013, OR = 0.53). In the case of *SPLUNC1*, 37 SNPs which are in strong linkage disequilibrium (LD) ($r^2 \geq 0.85$) were associated with NPC susceptibility (P = 0.00016 - 0.0018, OR = 1.69 – 1.74). Further investigation of these variants by electrophoretic mobility shift and luciferase reporter assays showed that rs1407019 located in intron 3 of *SPLUNC1* caused allelic difference in the binding of Sp1 transcription factor and affected luciferase activity. This SNP may consequently alter the expression of SPLUNC1 in the epithelial cells. On the other hand, a novel NPC susceptibility gene, *ITGA9*, was identified in this genome-wide association study (GWAS). An intronic SNP of *ITGA9* (rs2212020) showed significant association with NPC susceptibility (crude P = 0.000045, OR = 1.80). We subsequently genotyped

additional 19 tag-SNPs within a 40 kb LD block surrounding this landmark SNP. Among them, SNP rs169111 showed the strongest association with *P*-value of 0.00012 (OR = 3.24, 95% CI = 1.78-5.91). In summary, our study supported the association of *CYP2E1* polymorphisms with NPC, and identified a functional variant of *SPLUNC1* (rs1407019) which may increase the risk of NPC by lowering its expression. Furthermore, two novel NPC susceptibility genes, *EPHX2* and *ITGA9* were identified via candidate gene approach and GWAS respectively.

ABSTRAK

Karsinoma nasofarinks (KNF) merupakan sejenis kanser yang timbul daripada lapisan epithelium nasofarinks. Frekuensi KNF yang lebih tinggi di negara-negara Asia dan kumpulan etnik tertentu amat mencadangkan penglibatan faktor genetic dalam perkembangan KNF, oleh itu pengenalpastian gen kerentanan terhadap KNF adalah penting untuk memahami patogensisnya. Dalam kajian ini, kami mengenalpasti gen kerentanan terhadap KNF dengan perdekatan yang berbeza, pemeriksaan calon gen dan GWAS, dan sejumlah 447 pesakit NPC dan 487 kawalan yang sihat telah dikumpulkan. Bagi pendekatan dengan memeriksa calon gen, kerentanan bagi polymorfisme pada 11 calon gen yang terlibat dalam metabolism xenobiotic (*CYP1A1, CYP2A13, CYP2E1, EPHX1, EPHX2, GSTM1, GSTT1* dan *GSTP1*) dan tindakbalas imun (*PIGR, LPLUNC1* dan *SPLUNC1*) terhadap KNF telah dikaji. Bagi 11 calon gen, hubungan antara polymorfisme pada *CYP2E1*, *EPHX2* dan *SPLUNC1* dengan kerentanan KNF telah dikenalpastikan. Lima SNPs bagi *CYP2E1* (rs2070672, rs3813865, rs3813870, rs8192772 dan rs915906) adalah berkaitan rapat dengan peningkatan risiko untuk menghidapi KNF (P = 0.00081-0.0032, OR = 2.92 – 3.69), manakala alel T bagi *EPHX2* rs4149244 adalah berkaitan dengan pengurangan risiko bagi KNF (P = 0.0013, OR = 0.53). Bagi *SPLUNC1*, 37 SNPs yang berada dalam rantaian ketidaksaimbangan yang kuat (r2 $\geqslant$ 0.85) adalah berkaitan dengan kerentanan NPC (P = 0.00016 - 0.0018, OR = 1.69 – 1.74). Penyelidikan yang selanjutnya dengna menjalankan "electrophoretic mobility shift" dan "luciferase reporter assays" menunjukan bahawa rs1407019 yang berada di intron ketiga bagi *SPLUNC1* menyebabkan perbezaan bagi alel dalam pengikatan faktor transkripsi SP1 dan mempengaruhi aktiviti luciferase. SNP ini mungkin mengubah ungkapan bagi SPLUNC1 yang ada di sel-sel epithelium. Sebaliknya, satu gen kerentanan KNF yang baru, *ITGA9* telah dikenalpasti melalui GWAS. Satu SNP yang berada di intron ITGA9 (rs2212020) berkaitan dengan

kerentanan KNF (crude *P* = 0.000045, OR = 1.80). Seterusnya, kami menyiasat 19 tag-SNPs yang berada dalam satu 40 kb blok LD yang mengelilingi rs2212020. SNP rs169111 menunjukkan perkaitan dengan kerentanan KNF yang kuat dengan *P* nilai 0.00012 (OR = 3.24, 95% CI = 1.78-5.91). Secara ringkas, kajian kami telah menyokong hubungan antara polymorfisme pada *CYP2E1* dengan kerentanan KNF, dan mengenalpasti satu SNP bagi *SPLUNC1* yang mungkin meningkatkan risiko tentang menghidapi KNF dengan mengurangkan ungkapan gen tersebut. Selain itu, dua gen kerentanan yang baru terhadap KNF (*ITGA9* dan *EPHX2*) telah dikenalpasti melalui pemeriksaan calon gen dan GWAS.

ACKNOWLEDGEMENTS

First and foremost, my deepest gratitude goes to my supervisor, Dr. Ng Ching Ching, for giving me the opportunity to carry out this research. Her continued patience and valuable guidance contributed significantly to this work. Special thanks to Professor Dr. Nakamura Yusuke for giving me the opportunity to work in his laboratory and his unfailing assistance, guidance and encouragement.

I am deeply grateful to Dr. Taisei Mushiroda and Dr. Kiyotani Kazuma for their valuable guidance and assistance. Thanks also go to all the members in Laboratory for Pharmacogenetics, RIKEN Center for Genomic Medicine for their contributions and supports in completion of this study.

I would like to thank all the voluntary participants in this project; clinicians and staff at the University of Malaya Medical Centre, NCI Cancer Hospital, Tung Shin Hospital and the Malaysian Nasopharyngeal Carcinoma study group who have helped in procurement of the samples. I greatly appreciate the technical assistance and advice from Dr. Cha Pei Chieng, Dr. Amanda Low Siew Kee, Ms. Puah Suat Moi, Mr. Chin Yoon Ming, Ms. Tai Mei Chee, Ms. Goh Siang Ling and Mr. Charles Khor Seik Soon. Finally, I would like to thank my family for their support and encouragement. Thanks for being there for me.

PUBLICATIONS AND PRESENTATIONS

Publications:

Part of this study has been accepted as publication in:

Ng, C.C., Yew, P.Y., Puah, S.M., Krishnan, G., Yap, L.F., Teo, S.H., Lim, P.V., Govindaraju, S., Ratnavelu, K., Sam, C.K., Takahashi, A., Kubo, M., Kamatani, N., Nakamura, Y. and Mushiroda, T. (2009). A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. *J. Hum. Genet.*, **54(7)**, 392-397.

Yew, P.Y., Mushiroda, T., Kiyotani, K., Krishnan, G., Yap, L.F., Teo, S.H., Lim, P.V., Govindaraju, S., Ratnavelu, K., Sam, C.K., Yap, Y.Y., Khoo, A.S.B., Pua, K.C., Nakamura, Y., The Malaysian NPC Study Group and Ng, C.C. (2011). Identification of a functional variant in SPLUNC1 associated with nasopharyngeal carcinoma susceptibility among Malaysian Chinese. *Molecular Carcinogenesis*. DOI: 10.1002/mc.21857

Manuscript in preparation:

Yew, P.Y., Mushiroda, T., Kiyotani, K.,Krishnan, G., Yap, L.F., Teo, S.H., Lim, P.V., Govindaraju, S., Ratnavelu, K., Sam, C.K., Yap, Y.Y., Khoo, A.S.B., Pua, K.C., Nakamura, Y., The Malaysian NPC Study Group and Ng, C.C. Association study between genetic polymorphisms in xenobiotics metabolizing enzymes and nasopharyngeal carcinoma susceptibility among Malaysian Chinese.

Presentations:

Part of this study has been presented in:

Poster presentation at EBV International conference, 2008, Guang Zhou.

Ng, C.C., Yew, P.Y., Mushiroda, T., Krishnan, G., Nakamura, Y. SPLUNC1 polymorphisms associated with nasopharyngeal carcinoma susceptibility in Malaysian Chinese.

Poster presentation at My1Bio conference, 2010, Malaysia

Yew, P.Y., Ng, C.C., Mushiroda, T., Krishnan, G., Yap, L.F., Khoo, A.S.B., Nakamura, Y., The Malaysian NPC Study Group. SNPs in SPLUNC1 are associated with Nasopharyngeal Carcinoma susceptibility in Malaysian Chinese.

Oral presentation at Biological Sciences Graduate Congress (BSGC) conference, 2010, Malaysia

Yew, P.Y., Ng, C.C., Mushiroda, T., Krishnan, G., Yap, L.F, Khoo, A.S.B., Nakamura, Y., The Malaysian NPC Study Group. Association study of single nucleotide polymorphisms in SPLUNC1 with nasopharyngeal carcinoma susceptibility in Malaysian Chinese.

TABLE OF CONTENTS:

LIST OF FIGURES:

LIST OF TABLES:

ABBREVIATIONS

>	More than
<	Less than
≤	Less than and equal to
≥	More than and equal to
%	Percentage
℃	Degree Celcius
$\chi 2$	Chi-square
3' UTR	3 prime untranslated region
ABI	Applied Biosystems
ASR	Age-standardized rate
BAL	Bronchoalveolarlavege
bp	Base pair
BPI	Bactericidal/ permeability-increasing protein
CETP	Cholesteryl ester transfer protein
CI	Confidence interval
CI5	Cancer Incidence in Five Continents
CF	Cystic fibrosis
Chr.	Chromosome
CT	Computed Tomography
CYP	Cytpchrome P450
CYP1A1	Cytochrome P450, family 1, subfamily A, polypeptide 1
CYP2A13	Cytochrome P450, family 2, subfamily A, polypeptide 13
CYP2E1	Cytochrome P450, family 2, subfamily E, polypeptide 1
DIG	digoxigenin-11-ddUTP
dH_2O	Distilled water

DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
EA	Early antigen
EBNA	EBV nuclear entigen complex
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
EET	Arachidonic epoxide
e.g.	For example
EMSA	Electrophoretic mobility shift assay
EPHX	Epoxide hydrolase
et al.	And other people
etc.	And other things
FBS	Fetal bovine serum
FRET	Fluorescent resonance energy transfer
GST	Glutathione–S–transferase
GSTM1	Glutathione S-transferases Mu -1
GSTP1	Glutathione S-transferases class-pi -1
GSTT1	Glutathione S-transferases Theta -1
GWAS	Genome-wide association study
HapMap	Haplotype Map
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
IARC	International Agency for Research on Cancer
IgA	Immunoglobulin A
IgG	Immunoglobulin G
ITGA9	Integrin alpha 9

kb	Kilobase
LBP	Lipopolysaccharide-binding protein
LD	Linkage disequilibrium
LPLUNC1	Long palate, lung, and nasal epithelium clone 1
LPS	Lipopolysaccharide
µ	Mirco
µg	Microgram
µL	Microliter
µM	Micromolar
MAF	Minor allele frequency
$MgCl_2$	magnesium chloride
mM	Milimolar
mRNA	messenger ribonucleic acid
MRI	Magnetic resonance imaging
MRPs	Multidrug resistance-associated proteins
NAT	N-acetiltransferases
NCBI	National Center for Biotechnology Information
NCR	National Cancer Registry
ng	Nanogram
NPC	Nasopharyngeal carcinoma
OTAP2	Organic anion transporting polypeptide 2
OR	Odds ratio
PAH	Polycylic aromatic hydrocarbons
PBS	Phosphate-Buffered Saline
PCA	Principle component analysis
PCR	Polymerase chain reaction

Pgp	P-glycoprotein
PIGR	Polymeric Ig Receptor
pIg	Polymeric immunoglobulin
PLTP	Phospholipids transfer protein
PLUNC	Palate, lung and nasal epithelium associated
r^2	Correlation of coefficient value of pairwise correlation graph
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulfate
sec	Seconds
SNP	Single nucleotide polymorphism
SPLUNC1	Short palate, lung, and nasal epithelium clone 1
SULT	Sulfotransferases
TBE	Tris-borate EDTA
TE	Tris EDTA
TF	Transcription factor
TNM	Tumour, Node, Metastasis
UMMC	University of Malaya Medical Center
UGT	UDP-glucuronosyltransferases
UV	Ultraviolet
V	Volt
v/v	Volume per volume
VCA	Viral capsid antigen
WHO	World Health Organization
w/v	weight per volume
XME	Xenobiotic metabolizing enzyme

CHAPTER 1

INTRODUCTION

1.1 GENERAL INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a head and neck malignancy with striking racial and geographical distribution. It is common in Southern China, Southeast Asia including Malaysia and Singapore (10-30 per 100,000 people) (Curado *et al.*, 2007; Omar and Ibrahim Tamin, 2011). In Malaysia, Malaysian Chinese showed higher incidence compared to Malay and Indian (Omar and Ibrahim Tamin, 2011). Epstein-Barr virus (EBV) infection, environmental factors such as exposure to food carcinogen or chemical carcinogens, and genetic factors are believed to be involved in the tumorigenesis of NPC (Hildesheim and Lavine, 1993; Armstrong *et al.*, 1998; Thompson and Kurzrock, 2004; Zhou *et al.*, 2007). However, the molecular mechanism of NPC pathogenesis is not fully elucidated.

Identification of susceptible genes to NPC is crucial for improving our understanding of biological and etiologic mechanism involved in the development of NPC. The association between NPC risk and polymorphisms in *CYP2E1* (Jia *et al.*, 2009), *GSTM1* (Nazar-Stewart *et al.*, 1999), *HLA* (Goldsmith *et al.*, 2002), *PIGR* (Hirunsatit *et al.*, 2003) and *PLUNC* (He *et al.*, 2005) genes have been reported, suggesting the involvement of xenobiotic metabolizing genes and immune response genes in this disease.

Nevertheless, genetic study of NPC in Malaysian Chinese compared to other populations is still lacking and only one association study had been published so far (Lee *et al.*, 2007). In the present study, a case-control association study was performed to identify NPC susceptibility genes in Malaysian Chinese population. Both candidate gene and genome-wide screening approaches were adopted in this study. For candidate

gene approach, genetic polymorphisms in 11 candidate genes that involved in the xenobiotic metabolism (*CYP1A1*, *CYP2A13*, *CYP2E1*, *EPHX1*, *EPHX2*, *GSTM1*, *GSTT1* and *GSTP1*), immune response (*PIGR*, *LPLUNC1* and *SPLUNC1*) were investigated for their associations and NPC susceptibility. To the best of our knowledge, these candidate genes have not yet been studied in Malaysian populations.

Besides performing candidate gene approach, we also carried out a genome-wide association study (GWAS) that might be able to screen for the unanticipated gene(s). Overall, this thesis is focus on with genetic susceptibility to the risk of NPC in the Malaysian Chinese population.

1.2 NASOPHARYNGEAL CARCINOMA (NPC)

NPC is a malignancy which originates from the nasopharynx, the upper part of the pharynx (Brennan, 2006) (Figure 1.1). It arises from the mucosal epithelium of the nasopharynx, most often within the lateral nasopharyngeal recess or Fossa of Rosenmuller which is the common site of origin for NPC (Loh *et al.*, 1991; Brennan, 2006).

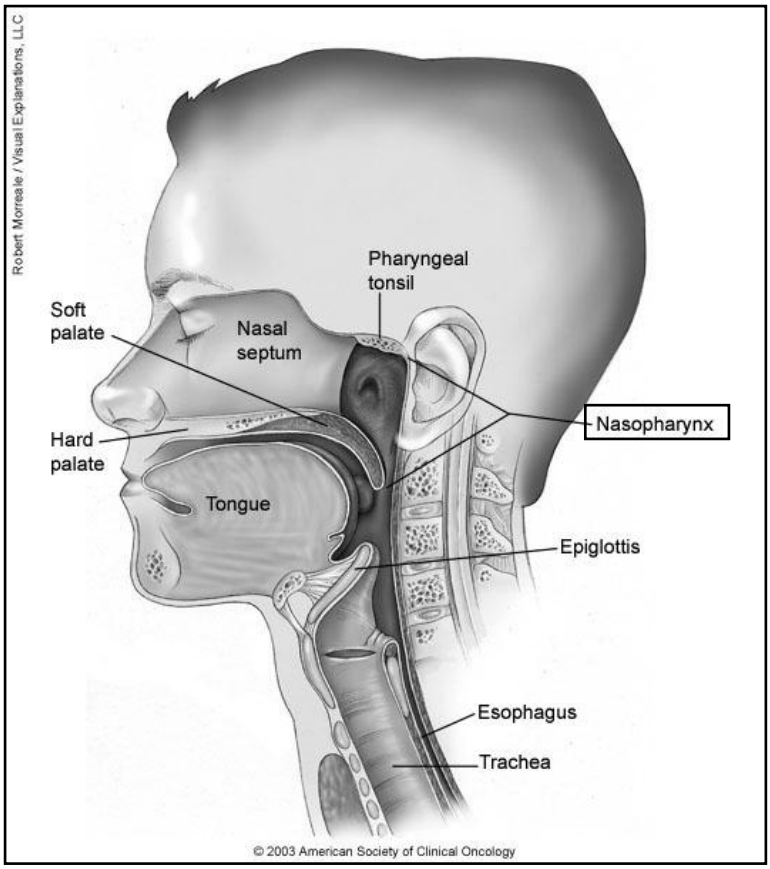


Figure 1.1: Anatomy of the nasopharynx.

Nasopharynx is located behind the nose in the upper part of the throat (Obtained from American Society of Clinical Oncology).

1.2.1 Classification of NPC

Since 1978, World Health Organization (WHO) has proposed a number of different classifications of NPC (Table 1.1). According to the most recent WHO classification of NPC (2005), NPC is classified into three subtypes based on the differentiation status of tumour cells: keratinizing squamous cell carcinoma, non-keratinizing carcinoma which is sub-divided into differentiated and undifferentiated carcinomas, and basaloid squamous cell carcinoma (Barnes *et al.*, 2005; Thompson, 2007). In Malaysia, most of the NPC cases are belong to the non-keratinizing carcinomas (Pua *et al.*, 2008).

Table 1.1: Classification of NPC

World Health Organization (1978)
1. Keratinizing squamous cell carcinoma 2. Non-keratinizing carcinoma 3. Undifferentiated carcinoma
World Health Organization (1991)
1. Keratinizing squamous cell carcinoma 2. Non-keratinizing carcinoma a. Differentiated non-keratinizing carcinoma b. Undifferentiated carcinoma
World Health Organization (2005)
1. Keratinizing squamous cell carcinoma 2. Non-keratinizing carcinoma a. Differentiated non-keratinizing carcinoma b. Undifferentiated carcinoma 3. Basaloid squamous cell carcinoma

(Table adapted from Thompson, 2007)

1.2.2 Staging of NPC

Staging of NPC is based on the clinical and radiologic examination and it is important in planning the treatment options and estimating a patient's outlook for recovery and survival (Deschler and Day, 2008). The Tumour, Node, Metastasis (TNM) staging system of the American Joint Committee on Cancer (Table 1.2) is the most frequent system used to classify the stage of NPC (Table 1.3), combining the information of tumour, lymph nodes and metastasis. Roman numerals 0 to IV are used to describe the progression from earliest to most advanced stage (Chan *et al.*, 2002, Edge *et al.*, 2010).

Table 1.2: The TNM classification of the American Joint Committee on Cancer

Nasopharynx (T)	
TX	Primary tumour can not be assessed.
T0	No evidene of primary tumour.
Tis	Carcinoma *in situ.*
T1	Tumour confined to the nasopharynx, or tumour extends to oropharynx, and/or nasal cavity without parapharyngeal extension.
T2	Soft tissue of oropharynx and/or nasal fossa.
T3	Invades bony structure and/or paranasal sinuses.
T4	Intracranial extension, involvement of cranial nerves, hypopharynx, orbit, or with extension to the infratemporal fossa/masticator space.
Regional lymph node (N)	
NX	Regional lymph nodes can not be assessed.
N0	No regional lymph node metastasis.
N1	Unilateral metastasis in lymph node(s), ≤ 6cm in greatest dimension, above supraclvicular fossa, and/or unilateral or bilateral, retrophayngeal lymph nodes, ≤ 6cm in greatest dimension.
N2	Bilateral metastasis in lymph node(s), ≤ 6cm in greatest dimension, above supraclvicular fossa.
N3	Metastasis in lymph node(s), > 6cm and/or to supraclvicular fossa.
N3a	> 6cm in dimension.
N3b	Extension to supraclvicular fossa.
Distant metastasis (M)	
M0	No distant metastasis.
M1	Distant metastasis.

(Table adapted from Edge *et al.*, 2010)

Table 1.3: Staging of NPC

Stage	T	N	M
0	Tis	N0	M0
I	T1	N0	M0
II	T1	N1	M0
	T2	N0	M0
	T2	N1	M0
III	T1	N2	M0
	T2	N2	M0
	T3	N0	M0
	T3	N1	M0
	T3	N2	M0
IVA	T4	N0	M0
	T4	N1	M0
	T4	N2	M0
IVB	Any T	N3	M0
IVC	Any T	Any N	M0

(Table adapted from Edge *et al*., 2010)

1.2.3 NPC incidence worldwide

According to the Cancer Incidence in Five Continents (CI5) volume IX and GLOBOCAN2008 databases, the incidence rates of NPC are vary markedly worldwide (Curado *et al.*, 2007; Ferlay *et al.*, 2010a). GLOBOCAN2008 estimated a total of 84,400 NPC cases and 51,600 deaths in 2008 (Ferlay *et al.*, 2010b). NPC is rare in most of the countries, where the age-standardized rates (ASR) are generally less than 1 per 100,000 population (Curado *et al.*, 2007; Ferlay *et al.*, 2010a). In California of The United States, low incidence rates are observed in white males and females (0.4 and 0.1 per 100,000 population respectively). In contrast, NPC is a common cancer in Southern China and Southeast Asia, and the highest prevalence of NPC is observed in Southern China (20 – 30 per 100,000 population) (Curado *et al.*, 2007; Ferlay *et al.*, 2010a). NPC occurs more frequently in males compared to females, and the incidence rates are two to three folds higher in males compared to females (Curado *et al.*, 2007; Omar and Ibrahim Tamin, 2011).

1.2.3.1 NPC in Malaysia

Since 2002, four reports on the incidence of NPC have been published by National Cancer Registry (NCR). The previous reports for the 2002 – 2006 period only reported on the cancer occurrence in Peninsular Malaysia. During this period, NPC ranked 2nd or 3rd most frequent cancer among male and 10th to 15th among female in Peninsular Malaysia (Lim *et al*, 2003; Lim and Yahaya, 2004; Omar *et al.*, 2006). On the other hand, the most recent report (2007) has provided the information on cancer occurrence in Peninsular Malaysia, Sabah and Sarawak. According to this report, NPC was the fourth most common cancer in Malaysia. It is the third most common cancer among male and eleventh among female in Malaysia (Omar and Ibrahim Tamin, 2011). A total of 940 NPC patients registered with NCR in 2007, and the ASR was 6.4 and 2.3 per

100,000 population for males and females respectively. NPC incidence is more than two-fold higher in males than in females. Besides that, the Chinese (10.9 and 3.5 per 100,000 population for males and females, respectively) recorded a higher incidence compared to Malay (3.0 and 1.3 per 100,000 population for males and females, respectively) and Indian (1.1 and 0.9 per 100,000 population for males and females, respectively) populations (Omar and Ibrahim Tamin, 2011).

The incidence of NPC in various geographic regions is presented in Figure 1.2. Chinese from Zhongshan, Guangzhou and Hong Kong showed higher incidence of NPC compared to other countries. The incidence rates among Chinese who born in Singapore and United States are lower than those in Southern China, but they still continue to show higher incidence of NPC than other races in the countries in which they live (Curado *et al.*, 2007). For example, in the United States, highest incidence rate is observed among Chinese Americans (1.1 – 9.9 per 100,000 population), followed by Filipino Americans (0.8 – 3.8 per 100,000 population), and finally Caucasians (< 0.5 per 100,000 population) (Figure 1.2). Besides that, there is variation in NPC incidence across the regions in China, where the incidence rate is generally increasing from Northern China (e.g. Harbin city) to Southern China (e.g. Guangzhou). Japan and Korea which are close to Northern China also showed low incidence of NPC (< 1.0 per 100,000 population) (Curado *et al.*, 2007). The disparity in NPC incidence between different populations points to the influence of genetic and environmental factors in the development of NPC. Furthermore, the risk of NPC might vary with degree of racial and social admixture with Southern Chinese. This prompted us to investigate and reveal NPC susceptibility gene(s) among Malaysian Chinese.

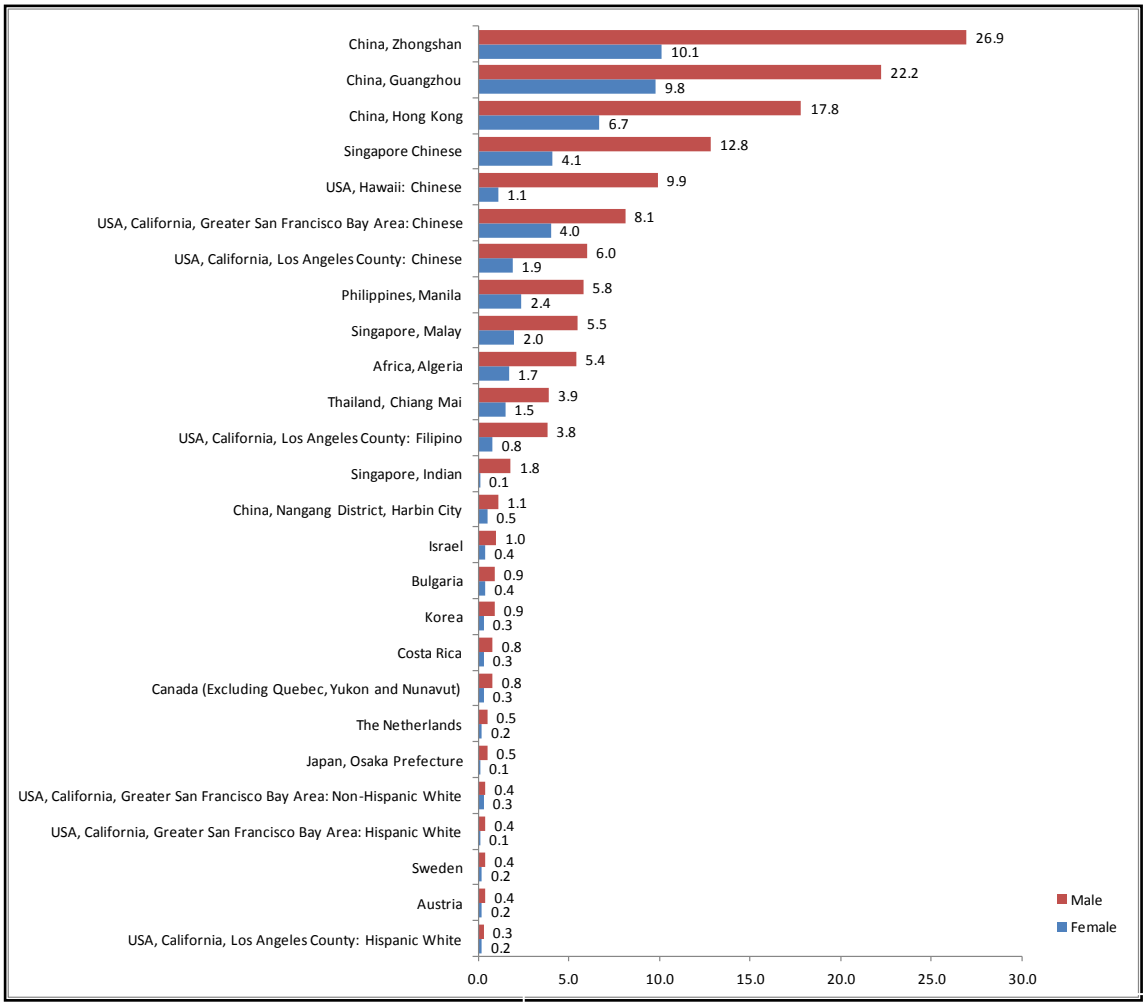


Figure 1.2: Comparison of NPC incidence rates in different populations worldwide. The age-standardized rates (ASRs) for NPC in males and females from several countries are shown this figure. The Chinese show the highest incidence compared to other ethnic groups (Raw data was obtained from Curado *et al.*, 2007).

1.2.3.2 A brief history and origin of Malaysian Chinese

The Malaysian Chinese (马来西亚华人) are Malaysians with Chinese origin. Most of them are the descendants of Chinese who came from the southern parts of China including Fujian, Guangdong and Hainan provinces in the 15th, and the middle of 19th and early 20th centuries (Joshua Project database for Malaysia). Malaysian Chinese who make up 24.6% of the population is the second largest ethnic group in Malaysia.

The Chinese have been settling in Malaysia for many centuries. The first wave of Chinese immigrants came in the early of 15th century. A lot of Chinese merchants and emissaries visited the Malacca Empire, and some of them stayed there as merchants and traders (Tracy, 1993). The diplomatic relations between China and Malacca culminated when Princess Hang Li Po was married to the Sultan Mansur Syah (Iskandar, 1992). A large number of retinue of nobles, companions and servants accompanied the princess to Malacca. Some of these Chinese married the local Malays, and their distinct background, cultures and customs merged together harmoniously. Their descendants are called Baba (male) and Nyonya (female) (Ooi, 2004).

During 19th century and early 20th century, the second and much bigger wave of Chinese immigrants came to Southeast Asia or the Nan yang. The warfare and famine in China caused a lot of Southern Chinese, particularly from Guangdong and Fujian provinces migrate to Singapore and Malaysia (Chander, 1972). Their immigration to Singapore and Malaysia was encouraged by the colonial governments in Southeast Asia, primarily the British who were seeking labourers to develop the economy of its territories. Thousands of Chinese migrated to Peninsular Malaysia and worked as labourers in the tin mines, railway construction, pepper and tapioca farms, rubber and sugar-cane plantations (Campbell, 1971).

The Chinese community in Malaysia is generally made up of various dialect groups such as the Cantonese, Hokkien, Hakka, Teochew, Hainanese etc., who

extensively intermarried with each other. Therefore, Malaysian Chinese can be considered as a heterogeneous group made up of descendants of immigrants from various area of China, majority from the southern parts of China (Lee and Tan, 2000). Nowadays, the vast majority of Chinese stay in Kuala Lumpur, Selangor, Penang, Perak, Johor and Sarawak.

Malaysian Chinese tend to retain the traditional lifestyle of Chinese, including diet (Armstrong *et al.*, 1998). Generally, Malaysian Chinese eat different types of food including Chinese, Indian, Malay and Western cuisines. Some of the Malaysian Chinese foods are similar to the food in Southern China as they are primarily from the cuisine of Fujian, Cantonese and Hakka. Besides that, Malaysian Chinese also like to eat salted fish or preserved foods, which are the traditional food of Southern Chinese, this might explain the higher incidence rate of NPC observed among Malaysian Chinese (Armstrong *et al.*, 1998).

1.2.4 NPC etiology

NPC is a multifactorial and polygenic disease. Recent etiological studies suggested that the development of NPC might be attributed to a complex interaction of EBV infection, environmental factors, dietary exposure and genetic factors (Sriampron *et al.*, 1992; Hildesheim and Lavine, 1993; Armstrong *et al.*, 1998; Thompson and Kurzrock, 2004; Zhou *et al.*, 2007).

a. Epstein-Barr virus (EBV)

The involvement of EBV in NPC was implicated by the increased anti-EBV antibody titer in patients than in controls (Henle *et al.*, 1970; Zheng *et al.*, 1994a; Thompson and Kurzrock, 2004), and later supported by the detection of EBV DNA in the tumour cells (Pathmanathan *et al.*, 1995; Lo, 2001a).

In EBV infection of B cells, it has been well documented that EBV infects B cells by the interaction between gp350/220 (a major viral membrane glycoprotein) and its specific receptor on the surface of B cells, CR2 (CD21) (Niedobitek, 2000; Lee *et al.*, 2007a). The infection of B cells is predominantly latent and it may induce growth transformation of the infected cells. However, the mechanism of EBV entry into epithelial cells is less defined, previous studies suggest that it might be mediated by other cellular receptors such as polymeric immunoglobulin receptor or integrins (Gan *et al.*, 1997; Tugizov *et al.*, 2003). Besides that, EBV may enter the epithelial cells by direct cell-to-cell contact of apical cell membranes with EBV-infected lymphocytes or spread directly across lateral membranes to adjacent epithelial cells after initial infection (Tugizov *et al.*, 2003).

b. Dietary environmental factors

Salted fish and other preserved food including bean pastes and fermented eggs contain high level of nitrosamines, which are known carcinogens (Ward *et al.*, 2000; Yu and Yuan, 2002). Consumption of these food are strongly associated with the risk of NPC (Yu *et al.*, 1989a; Sriampron *et al.*, 1992; Armstrong *et al.*, 1998) and it was further supported by several animal studies (Yu *et al.*, 1989b; Hildesheim and Lavine, 1993). These animals developed nasal cavity tumours after being fed with salted fish which contained carcinogenic nitrosamines / precursors and EBV-activating substances (Shao *et al.*, 1988; Yu *et al.*, 1989b). Furthermore, consumption of these foods at early age seems to be particularly important and increases the risk of NPC (Yu *et al.*, 1986; Yu *et al.*, 1989b).

There are also dietary factors protective against the development of NPC (Armstrong *et al.*, 1998; Yu and Yuan, 2002). These include fresh fruits, tomatoes, orange coloured vegetables including carrots and sweet potatoes and various vegetables including Chinese flowering cabbage, green leafy vegetables (Yu *et al.*, 1989b; Zheng *et al.*, 1994b; Armstrong *et al.*, 1998). The protective effects of tomatoes, oranges / tangerines or other fresh fruits are reported in Guangzhou, China and Hong Kong (Yu *et al.*, 1989a). Moreover, some laboratory-based studies suggest that vitamin C can protect us against NPC as it can inhibit endogenous production of nitrosamines (van Poppel and van den Berg, 1997).

c. Non-dietary environmental factors

Industrial workers exposed to formaldehyde show higher risk of developing NPC (Blair *et al.*, 1986). The carcinogenic potential of formaldehyde on NPC development is further supported by a report from IARC in 1995 (IARC, 1995).

On the other hand, agriculture workers in Thailand exposed to wood dust are at a higher risk of developing NPC (Sriampron *et al.*, 1992). In addition, Armstrong and co-workers (2000) have reported that wood dust but not formaldehyde is associated with elevated risk of NPC in Malaysian Chinese within the vicinity of Selangor and the Federal Territory.

Cigarette smoking is another potential causal factor of NPC. Smokers from the Taiwanese, Shanghai Chinese and the Whites in the United States show increased risk of developing NPC as compared to non-smokers (Nam *et al.*, 1992; Vaughan *et al.*, 1996; Cheng *et al.*, 1999; Yuan *et al.*, 2000) however no association was observed in the Thai population (Sriampron *et al.*, 1992). Alcohol consumption is not linked to NPC development in Thai, Taiwanese and Shanghai Chinese (Sriampron *et al.*, 1992; Cheng *et al.*, 1999; Yuan *et al.*, 2000). However, in the United States, alcohol consumption could increase the risk of developing NPC (Nam *et al.*, 1992).

d. Genetic susceptibility to NPC

High prevalence of NPC in Asian countries and certain ethnic groups such as Chinese suggest that genetic factors may involve in the NPC pathogenesis. Migrant studies, association studies of specific markers and family aggregations of this disease have supported the important role of genetic predisposition in the development of NPC (Hildesheim and Levine, 1993; Zeng *et al.*, 2002). Furthermore, segregation analysis in Taiwanese and a Chinese cohort performed by Chen *et al.*, (1990) and Jia *et al.* (2005) suggests that NPC is in a multifactorial mode of inheritance.

Human leukocyte antigens (*HLA*) have consistently been associated with NPC risk through linkage studies and association studies in a few populations,

including Southern Chinese (Goldsmith *et al.*, 2002), Taiwanese (Hildesheim *et al.*, 2002; Hu *et al.*, 2005), and Tunisians (Li *et al.*, 2007). Furthermore, recent GWAS also suggested *HLA* region as NPC susceptibility loci (Tse *et al.*, 2009; Bei *et al.*, 2010). Researchers hypothesized that *HLA* might involve in the NPC pathogenesis through its cytotoxic T cell recognition and host immune response to EBV infection. However, an earlier linkage study on Chinese sib pairs have suggested a gene which is closely linked to *HLA* region conferred a 21 fold excess risk of NPC (Lu *et al.*, 1990), suggesting that another gene rather than *HLA* alleles is responsible in increasing the risk of NPC.

Several NPC susceptibility loci such as chromosomes 3p21.3 (Xiong *et al.*, 2004) and 4p15.1-q12 (Feng *et al.*, 2002) were identified by linkage studies. Many genes in the region of chromosomes 3p21.3 are involved in the tumourigenesis process. For example, *RASSF1A* gene was suggested as a target tumour suppressor gene in NPC (Chow *et al*., 2004). Besies that, family based association studies have reported that polymorphisms of *COX7B2* and *LOC344967* genes located on chromosome 4p11-p14 are associated with NPC susceptibility (Liang *et al.*, 2004; Jiang *et al.*, 2006). Nevertheless, the identified chromosomal regions are relatively large and many susceptibility genes still remained to be identified. On the other hand, several endeavours have also been made to identify NPC susceptibility genes by performing case-control association studies and investigating the functional related candidate genes. Genetic polymorphisms of *CYP2E1* (Hildesheim *et al.*, 1997; Jia *et al.*, 2009), *GSTM1* (Tiwawech *et al.*, 2005), *PIGR* (Hirunsatit *et al.*, 2003), DNA repair genes including *hOGG1* and *XRCC1* (Cho *et al.*, 2003) were suggested to be associated with increased risk of NPC.

1.2.5 Symptoms, Diagnosis and Treatment

NPC has a high survival rate for stage I or II patients if detected early (Chua *et al.*, 2003; Lee *et al.*, 2005). Thus early detection plays a critical role in reducing the morbidity and metastasis of NPC (Lee *et al.*, 2005). The symptoms of NPC include epistaxis, nasal obstruction, headache, cranial nerve dysfunction, hearing loss or tinnitus (Suzina and Hamzah, 2003; Brennan, 2006; Pua *et al.*, 2008). However, these symptoms are general, not specific to NPC. Therefore, NPC is often misdiagnosed at an early stage and more evident symptoms only manifest at much later critical stages (e.g. stage IV) (Leong *et al.*, 1999; Wei and Kwong, 2010). Most NPC tumours develop on the lateral wall of the nasopharynx, especially in the Fossa of Rosenmuller (Loh *et al.*, 1991; Thompson, 2007) and progress to form a lump in the neck.

NPC is diagnosed using a combination of clinical examination, imaging, serological markers and histology. Doctor will examine the nasopharynx for abnormal growths or bleeding by indirect nasopharyngoscopy or direct nasopharyngscopy (Her, 2001; Brennan, 2006). Biopsy samples will be taken in the event of abnormal growth (Spano *et al.*, 2003). Once NPC is confirmed, imaging test including computed tomography scan (CT scan) and magnetic resonance imaging (MRI) will be performed to determine the NPC stage (Chan *et al.* 2002; Snapo *et al.*, 2003). Serological markers are also used to diagnose NPC. Elevated IgG/IgA antibodies titers against the VCA and EA as well as increased IgG antibody titers against EBNA are detected in NPC patients compared to healthy controls. Plasma EBV DNA load is a reliable biomarker for diagnosis, staging and prognosis of NPC (Lo *et al.*, 2001a; Shao *et al.*, 2004). The plasma EBV DNA level is significantly increased from stage I to stage IV NPC. Besides that, comparison between the levels of post-treatment EBV DNA and pre-treatment EBV DNA had a better prediction for progression free survival. High levels of post-treatment plasma EBV DNA is associated with NPC recurrence (Chan *et al.*, 2002;

Shao *et al.*, 2004). Detection for elevated levels of circulating EBV DNA can be carried out by quantitative PCR in the plasma or serum (Leung *et al.*, 2004).

The primary treatment for NPC is through external radiotherapy as NPC tumours remain relatively radiosensitive (Chan, 2005). The radiotherapy uses the high energy radiation to kill the cancer cells. However, radiotherapy alone is not sufficient to treat a locally advanced NPC. Hence, the chemotherapy is given in combination with radiotherapy (concurrent chemoradiotherapy) to enhance the effectiveness of radiotherapy (Chan, 2005; Baujat *et al.*, 2006). Surgery is sometimes necessary for NPC patients with persistent or relapsed tumour locally in the nasopharynx (Danesi *et al.*, 2007). If cancer has spread to the lymph nodes, doctor may remove the cancerous lymph nodes in the neck. Nevertheless, surgery is an uncommon treatment choice as that area might be inadequate margins of resection (Snapo *et al.*, 2003; Lin *et al.*, 2008; Wei and Kwong, 2010).

1.3 APPROACHES TO GENETIC ANALYSIS OF NPC

In recent years, several approaches have been used to identify NPC susceptibility genes, including linkage analysis (Feng *et al.*, 2002; Xiong *et al.*, 2004) and association studies (Hildesheim *et al.*, 1997; Jia *et al.*, 2009). Linkage analysis is the detection of alleles at a specific chromosomal region that segregate with the disease through a family and involves genome-wide scan of polymorphic markers, mainly microsatellite markers, while association study is performed by comparing the allele or genotype frequencies of genetic variants between cases and controls (Cordell and Clyton, 2005; Witte, 2010).

1.3.1 Genetic association study

In our study, genetic association study was carried out. Compare to linkage analysis, association study has greater power at detecting common variants with small effects in a common multifactorial disease. However, it may require more markers to be examined than linkage analysis (Cordell and Clyton, 2005).

Genetic association study can be hypothesis-driven (candidate gene approach) or a hypothesis-free genome-wide approach (Zhu and Zhao, 2007; Witte, 2010). In candidate gene approach, a hypothesized functional relationship between the genes of interest and known disease etiology is required. On the other hand, genome-wide approach can pinpoint the genes without any prior information or knowledge of their function (Zhu and Zhao, 2007; Witte, 2010).

In recent years, a large number of association studies have been performed to investigate the association between genetic polymorphisms and NPC susceptibility (Zhou *et al.*, 2007). However, most of these NPC susceptibility genes were reported from small scale studies, and in some of these studies, disparities between ethnic groups in addition to geographical locations were observed for the same gene loci (Table 1.4).

Therefore, genetic susceptibility to NPC that related with ethnicity deserves further investigation.

1.3.2 Population based case-control association study

Genetic association studies can be performed in multiple families (family based) or in unrelated individuals (population based case-control). Generally, population based case-control association study is the most widely used approach. It is a powerful tool to detect relatively small genotypic effect, even in modest samples of cases and controls. Furthermore, unrelated subjects of population based association study are easier to be collected than family based study. However, this approach is prone to population stratification which may yield spurious association. Moreover, sometimes inconsistent results are obtained, which probably due to problem of underpowered, racial differences and population heterogeneity (Balding, 2006; Cross *et al.*, 2010). On the other hand, family based association study can overcome the potential confounding effects of population stratification as there is a common genetic background among the family members (Laird and Lange, 2006).

In this study, a population based case-control association study was performed. A significant difference of distribution of allele or genotype of genetic markers among cases and controls is referred as genetic association between this allele or genotype and the disease. A simple way to test for genetic association is performing a fisher exact's test. Besides that, covariates including age or gender can be included by logistic regression analysis (Balding, 2006).

1.4 HUMAN GENETIC POLYMORPHISMS

In recent years, international projects such as the Human Genome Project, the SNP consortium and the International HapMap project have contributed to many databases on genetic polymorphisms, further facilitating the design of case-control association studies (Sherry *et al.*, 2001; International HapMap Consortium, 2010).

Various types of genetic polymorphisms exist in the human genome, including minisatellites (variable number of tandem repeats, VNTRs), microsatellites (short tandem repeats, STRs), deletions, insertions and single base substitution (termed single nucleotide polymorphism, SNP) (Dib *et al.*, 1996; Lewin, 2003; Nakamura, 2009). Genetic polymorphisms are important basis for the differences seen among individuals and they may link to the development of disease. For example, deletion is a polymorphism involving the removal of one or more base pairs (bp) in the DNA sequence and causes the loss of genetic materials (Lewin, 2003; Snustad and Simmons, 2003). Moreover, deletions of certain region of the genes may alter the function of resulting protein and cause disease. For example, deletion of *GSTM1* is suggested to be associated with the increased risk of NPC due to impaired ability to detoxify certain carcinogens (Tiwawech *et al.*, 2005).

1.4.1 Single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are the most common polymorphisms in the human genome. It occurs when a single nucleotide (A, T, G or C) in DNA sequence is altered. SNP is one of the important genetic markers for genetic studies. The abundance and the stability of SNPs suggest that it can be an excellent biological marker (Brookes, 1999). Besides that, SNPs are evolutionarily stable, they do not change much from generation to generation and can be used to follow the inheritance patterns of chromosomal regions (Johnson and Todd, 2000).

In principle, SNPs could be bi-, tri- or tetra-allelic polymorphisms, but tri- and tetra-allelic SNPs are rare, almost all the SNPs are bi-allelic. The binary characteristic of SNPs renders the allelic discrimination of SNPs simple and straightforward (Brooker, 1999). Furthermore, some SNPs are likely to be functional importance including coding SNP which changes the protein structure or function by altering the amino acid sequence (nonsynonymous variants), or SNPs which are involved in the regulatory regions (Cargill *et al.*, 1999; Tabor *et al.*, 2002; Wang *et al.*, 2005). Therefore, SNPs are useful for investigating the differences between individuals or as polymorphic markers to examine the susceptibility of genes to disease through case-control association study.

1.4.2 The International HapMap Project

The International HapMap Project was launched in 2002, this project is established to identify genetic variations in human genome. Furthermore, it aims to develop a haplotype map, which reveals the common patterns of human genetic variation to accelerate genetic research (International HapMap Consortium, 2005; Thorisson *et al.*, 2005; International HapMap Consortium, 2010).

The HapMap project collects the genotypes of SNPs from numerous populations including individuals from Centre d'Etude du Polymorphisme Humain collected in Utah, USA, with ancestry from Northern and Western Europe (CEU), Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT), Yoruba in Ibadan, Nigeria (YRI), African ancestry in the Southwestern USA (ASW), Chinese in Metropolitan Denver, Colorado, USA (CHD), Gujarati Indians in Houston, Texas, USA (GIH), Luhya in Webuye, Kenya (LWK), Maasai in Kinyawa, Kenya (MKK), Mexican ancestry in Los Angeles, California, USA (MXL) and Tuscans in Italy (Toscani in Italia, TSI) (International HapMap Consortium, 2010). Genotyping of large numbers of SNPs in different populations by International HapMap Project has revealed that most of the

human genome is contained in linkage disequilibrium (LD) or haplotype blocks of substantial size. One of the applications of HapMap project to association studies is the selection of tag-SNPs (International HapMap Consortium, 2005).

1.4.3 Tag-SNPs

Ideally, all the putative causative variants for association with disease should be tested in a genetic association study. However, it may require an unfeasibly large number of SNPs to be genotyped (Kruglyak and Nickerson, 2001; The International HapMap Consortium, 2007).

We used the HapMap data for SNP selection in the candidate gene association studies. The prior knowledge of the LD structure in a given population and genomic region of HapMap data is useful in selecting the tag-SNPs that capture majority of the common genetic variation. In such way, genetic information of the full SNP set can be retained without genotype all the SNPs in the LD blocks. This may reduce the amount of genotyping SNPs and costs, and with minimum loss of information (International HapMap Consortium, 2005; Thorisson *et al.*, 2005).

HapMap marker sets capture the overwhelming majority of all common variants at high r^2 which is a measure of LD. Although the haplotype composition, structure and quantity of LD are heterogenous across global populations (Mueller *et al.*, 2005; Sawyer *et al.*, 2005), recent studies have showed that selection of tag-SNPs using the HapMap is generally applicable to similar populations or other populations from the same continent (Ribas *et al.*, 2006; Gu and Chu. 2007). It can capture common variation efficiently in many other independent samples (de Bakker *et al.*, 2006).

1.5 SEARCHING FOR NPC SUSCEPTIBILITY GENES BY CANDIDATE GENE APPROACH

In this study, we examined the association of SNPs in *CYP1A1, CYP2A13, CYP2E1, EPHX1, EPHX2, GSTP1, PIGR, LPLUNC1* and *SPLUNC1* as well as deletion of *GSTM1* and *GSTT1* towards NPC susceptibility in Malaysian Chinese, by performing a case-control association study.

The selected candidate genes can be classified into two categories, namely genes encoding metabolizing enzymes and genes involving in immune response. Genes encoding metabolizing enzymes are further divided into phase I and phase II enzymes. Phase I enzymes comprise *CYP1A1, CYP2A13, CYP2E1, EPHX1* and *EPHX2* that involved in metabolic activation or hydrolase the carcinogens whereas phase II enzymes comprise *GSTM1*, *GSTT1* and *GSTP1* that involved in detoxifying the carcinogens. On the other hand, genes involved in immune response are comprised of *PIGR*, *LPLUNC1* and *SPLUNC1*. These genes and their association with NPC in the studied populations reported by other researchers are summarized in Table 1.4.

a) Genes encoding xenobiotic metabolizing enzymes

i. *CYP1A1, CYP2A13* and *CYP2E1*

ii. *EPHX1* and *EPHX2*

iii. *GTSM1, GSTT1* and *GSTP1*

b) Genes involving in immune response

i. *PIGR*

ii. *LPLUNC1* and *SPLUNC1*

Table 1.4: Summary for previous association studies of the selected candidate genes with NPC susceptibility in different populations

Function	Gene	Chromosomal location	Association	Studied population	References
Xenobiotic metabolizing	*CYP1A1*	Chr. 15q24.1	Negative	Taiwanese	Cheng *et al.*, 2003
			Positive	Southern Chinese	Xu *et al.*, 2009
	CYP2A13	Chr. 19q13.2	Negative	Cantonese	Jiang *et al.*, 2004
				Cantonese	Cao *et al.*, 2011
	CYP2E1	Chr. 10q24.3-qter	Negative	Taiwanese	Hildesheim *et al.*, 1995
			Positive	Taiwanese	Hildesheim *et al.*, 1997
				Cantonese	Jia *et al.*, 2009
	GSTM1	Chr. 1p13.3	Negative	Taiwanese	Cheng *et al.*, 2003
				Chinese	Guo *et al.*, 2008
			Positive	Caucasians	Nazar-Stewart *et al.* 1999
				Chinese	Deng *et al.*, 2005
				Thai	Tiwawech *et al.*, 2005
	GSTM1/GSTT1	Chr. 1p13.3/22q11.23	Negative	Chinese	Guo *et al.*, 2008
			Positive	Chinese	Deng *et al.*, 2005
	GSTT1	Chr. 22q11.23	Negative	Taiwanese	Cheng *et al.*, 2003
				Chinese	Deng *et al.*, 2005
				Chinese	Guo *et al.*, 2008
	GSTP1	Chr. 11q13	Negative	Taiwanese	Cheng *et al.*, 2003
				Chinese	Guo *et al.*, 2010

Table 1.4, continued

Function	Gene	Chromosomal location	Association	Studied population	References
Immune response	*PIGR*	Chr. 1q31-q41	Positive	Thai	Hirunsatit *et al.*, 2003
				Chinese in Thai	Hirunsatit *et al.*, 2003
				Cantonese	Fan *et al.*, 2005
	SPLUNC1	Chr. 20q11.2	Positive	Cantonese	He *et al.*, 2005

* Association between *EPHX1*, *EPHX2* and *LPLUNC1* polymorphisms and NPC susceptibility have not been studied in any population.

1.5.1 Genes encoding xenobiotic metabolizing enzymes (XMEs)

Previous epidemiological studies suggested that intake of certain chemical carcinogens such as nitrosamines which can be found in salted fish or tobacco smoking may contribute to the development of NPC (Sriampron *et al.*, 1992; Hildesheim and Lavine, 1993; Armstrong *et al.*, 1998; Cheng *et al.*, 1999). These chemical carcinogens require metabolic activation to exert their carcinogenic effects. Genes that encode XMEs are involved in the metabolic activation and detoxification of these carcinogens (Ortiz de Montellano, 2005; Jancova *et al.*, 2010).

Generally, enzymes that involved in biotransformation of substances foreign to the body (xenobiotics) are classified into phase I (oxidative) or phase II (conjugative) metabolizing enzymes or phase III transporters involved in the efflux mechanisms (Bozina *et al.*, 2009). Phase I metabolism includes oxidation, reduction, hydrolysis and hydration (Ortiz de Montellano, 2005). These enzymes transform a parent compound to more polar metabolites by unmasking or de novo formation of functional groups such as -OH, -NH_2, -SH. Reactive oxygen is inserted into a lipophilic compound to make it more polar and water-soluble (Ortiz de Montellano, 2005; Jancova *et al.*, 2010). The main enzyme of phase I reactions are cytochromes P450 (CYPs) (Ortiz de Montellano, 2005).

Phase II enzymes including UDP-glucuronosyltransferases (UGTs), Sulfotransferases (SULTs), Glutathione S-transferases (GSTs), and N-acetiltransferases (NATs) conjugate water-soluble moieties such as glucuronic acid, sulfate and glutathione to a lipophilic compound. A new molecule with a lipophilic part and water-soluble part might be generated and this water-soluble compound can be excreted easily. Generally, phase III transporters including P-glycoprotein (Pgp), multidrug resistance-associated proteins (MRPs) and organic anion transporting polypeptide 2 (OTAP2) are

expressed in many tissues including liver, intestine, kidney and brain (Bozina *et al.*, 2009). These enzymes play a role in xenobiotic absorption, distribution and excretion.

Phase I and II enzymes are suggested to be involved in the metabolic activation and detoxification of various environmental carcinogens. Generally, procarcinogens are activated and converted into electrophilic derivatives in the first step. CYPs may activate the procarcinogens to oxidases, epoxygenases reductases, hydrolases, peroxidases and other enzymes. Next, the metabolic products are neutralized by conjugation catalysed by UGTs, SULTs, GSTs or NATs (Ortiz de Montellano, 2005; Bozina *et al.*, 2009). Recent studies have revealed that polymorphisms of phase I and phase II enzymes may be related with carcinogenesis induced by genotoxic xenobiotics (Bozina *et al.*, 2009). Certain genetic polymorphisms of XMEs may affect human response to genotoxic. Genotype of an individual may significantly influence the disposition of certain chemical thereby susceptibility to genotoxic exposure might vary among individuals.

1.5.1.1 Phase I enzymes - Cytochrome P450 (CYP)

CYP comprises a superfamily of enzymes that act on phase I of xenobiotic metabolic transformation, catalysing the oxidative reactions of numerous compounds (Ortiz de Montellano, 2005). CYP is a membrane bound protein and can be found in different tissues. Generally, the highest level of CYP is expressed in the liver, primarily found in the membrane of the endoplasmatic reticulum in the liver. However, extrahepatic tissues including respiratory and gastrointestinal tracts also expressed some CYP enzymes (Ortiz de Montellano, 2005; Bozina *et al.*, 2009).

In human, there are 57 genes and more than 59 pseudogenes that are divided into 18 families of *CYP* genes and 43 subfamilies (Nelson, 2009). Genes that encode CYP enzymes are designated with “CYP” followed by an Arabic number which indicates the

gene family (e.g. CYP2), followed by a capital letter indicating the subfamily (e.g. CYP2E) and an Arabic number denoting the specific gene/protein (e.g. CYP2E1). In addition, the nomenclature guidelines suggest that genes in the same family should share over 40% amino acid identity, whereas members of the subfamilies share more than 55% amino acid identity (Ortiz de Montellano, 2005; Nelson, 2009).

Several enzymes in CYP family including CYP1A1 and CYP2E1 are important in the biotransformation of chemicals and activation of pre-carcinogens. Most of the enzymes in CYP families 1 to 3 have showed inter-individual variability in the catalytic activity (Bozina *et al.*, 2009). Variability in the CYP activity is due to endogenous factors (e.g. age, gender, and morbidity), exogenous factors (e.g. food, smoking) and genetic polymorphisms in *CYP* genes (Van der Weide and Hinrichs, 2006). Several studies revealed that polymorphisms in the CYP genes may cause decreased or increased of enzyme expression. Carrying a variant allele might alter an individual's capability in metabolizing the toxicants (Hong and Yang, 1997; Ortiz de Montellano, 2005; Van der Weide and Hinrichs, 2006). Inter-individual differences in CYPs activity may affect individual susceptibility to cancer risk due to the important role of CYPs in activation and inactivation of procarcinogens.

i. Cytochrome P450, family 1, subfamily A, polypeptide 1 (*CYP1A1*)

CYP1A1 is located on chromosome 15q24.1, and it is primarily expressed in extrahepatic tissues. The expression of CYP1A1 is found in lung, urinary, peripheral bloodcells (lymphocytes, monocytes), bladder and placenta (Hukkanen *et al.*, 1997). Besides that, *CYP1A1* messenger RNA (mRNA) expression has been reported in pancreas, thymus, prostate, small intestine, colon, uterus and mammary gland (Ding and Kaminsky, 2003; Ortiz de Montellano, 2005). CYP1A1 is involved in the first step in the metabolism of polycylic aromatic hydrocarbons (PAH), catalyzing the oxidation of

PAH to epoxides. In addition, it can catalyse the oxidation of several xenobiotic chemicals including caffeine, chlorzoxazone, extrone and theophyline (Bozina *et al.*, 2009).

A number of *CYP1A1* alleles such as *CYP1A1**2B, *2C, *3, *4, *5, *6, *7, *8, *9, *11 which exhibit amino acid changes have been identified. However, the effect of amino acid change to the enzyme activity is still remained to be elucidated. A variant which causes an A to G transition in exon 7 (m2 allele; *CYP1A1*2B*; rs1048943) substituting isoleucine to valine near the heme-binding site of CYP1A1. This alteration of amino acid may lead to an increase in catalytic activity of the protein responsible for metabolism (Crofts *et al.*, 1994). Moreover, smokers with m2 allele are found to have more PAH-DNA adducts compared to smokers without this variant (Mooney *et al.*, 1997). Another variant of *CYP1A1*, m1 allele (*CYP1A1*2A*; rs4646903) shows a T to C mutation in the 3' non coding region. This variant is postulated to enhance the mRNA stability (Crofts *et al.*, 1994), and Georgiadis *et al.* (2005) have showed that it may increase inducibility of the enzyme in response to environmental tobacco smoke.

Recently, several investigations on the association between *CYP1A1* polymorphisms and cancer risk have been carried out. SNPs rs4646903 and rs1048943 of *CYP1A1* were associated with the risk of lung cancer in Japanese (Nakachi *et al.*, 1993) and Chinese populations (Song *et al.*, 2001). Besides that, rs1048943 was suggested to be associated with increased risk of lung cancer in the Brazilian populations (Sugimura *et al.*, 1995). Thus far, association between *CYP1A1* polymorphisms and NPC susceptibility has been examined in the Taiwanese population and Cantonese nuclear families in China (Cheng *et al.*, 2003; Xu *et al.* 2009). Nuclear family is a term used to define a family group which is consisted of father, mother and their children. In the Cheng's study (2003), SNP rs4646903 (m1 allele) was not associated with NPC susceptibility in Taiwanese after examining 172 NPC cases and

218 healthy controls, this finding is in accordance with the Xu's study (2009). On the other hand, Xu's study (2009) suggested that rs1048943 of *CYP1A1* might associate with decreased risk of NPC among the Cantonese nuclear families. Therefore, we would like to examine the association between this gene and NPC susceptibility among Malaysian Chinese.

ii. Cytochrome P450, family 2, subfamily A, polypeptide 13 (*CYP2A13*)

CYP2A13 is one of the members of the human *CYP2A* gene family. It maps on chromosome 19q13.2 and contains 9 exons (Smith *et al.*, 2007). CYP2A13 is an active protein, and it is predominantly expressed at the highest level in the nasal mucosa, followed by lung and trachea (Su *et al.*, 2000; Smith *et al.*, 2007). CYP2A13 involves in the activation of hexamethylphosphoramide, N, N-dimethylaniline, 2'-methoxyacetophenone, N-nitrosomethylphenylamine, and coumarin 7-hydroxylation. It is also highly active in the metabolic activation of 4-methylnitrosamine-1-3-pyridyl-1butanone (NNK) which is a major tobacco specific carcinogen (Su *et al.*, 2000).

Several *CYP2A13* polymorphisms have been identified and one of *CYP2A13* polymorphisms rs8192789 (Arg257Cys) was found in exon 5, leading to a missense mutation and cause amino acid change. This variant significantly decreased the activity to NNK and other substrates, and it was found to be associated with reduced risk for smoking related lung adenocarcinoma in a Chinese population (Wang *et al.*, 2003). Association between *CYP2A13* polymorphisms and NPC was examined by Jiang *et al.* (2004) and Cao *et al.* (2011) in the Cantonese population of Southern China. Both studies suggested that no association was observed between this gene and NPC susceptibility.

Nevertheless, we would like to investigate whether the polymorphisms of *CYP2A13* involve in NPC susceptibility in Malaysian Chinese population, due to its involvement in the metabolism of tobacco smoke procarcinogens.

iii. Cytochrome P450, family 2, subfamily E, polypeptide 1 (*CYP2E1*)

CYP2E1 is located on the chromosome 10q24.3-qter, consists of 9 exons. It is expressed in the extrahepatic tissues including nasopharynx as well as liver. Its expression levels are lower in kidneys, pancreas, brain, lung and intestinal mucosa (de Waziers *et al.*, 1990; Hou *et al.*, 2007). CYP2E1 is an enzyme that responsible for activating most environmental carcinogens including nitrosamines to exert their carcinogenic effects (Nuoso *et al.*, 1992; Hou *et al.*, 2007). Besides that, CYP2E1 was suggested to be involved in the oxidation of other compounds including ethanol, leading to produce reactive free radicals which may affect carcinogenesis. CYP2E1 can reduce the molecular oxygen to highly reactive compounds such as hydrogen peroxide, single oxygen, superoxide anion radical and hydroxyl radical. The enhanced synthesis of active oxygen forms may cause intensified lipid and protein peroxidation, DNA damage and carcinogenesis (Dankoand Chaschin, 2005). CYP2E1 play an important role in human susceptibility to toxicity and carcinogenicity of the environmental or industrial chemicals.

Recently, numerous association studies reported an association between *CYP2E1* polymorphisms and NPC susceptibility (Hildehseim *et al.*, 1997; Kongruttanachok *et al.*, 2001; Jia *et al.*, 2009). Hildesheim *et al.* (1997) reported that a promoter SNP of *CYP2E1* (rs2031920) was associated with NPC susceptibility in Taiwanese population. The variant T allele of rs2031920 was reported to be expressed at higher rate compared to C allele, thereby increasing the risk of NPC. In addition, SNP rs2031920 was reported to be associated with increased risk of other diseases including

lung cancer and lymphoma (Oyama *et al.*, 1997; Sarmanova *et al.*, 2001). On the other hand, Jia *et al.* (2009) suggested that other *CYP2E1* polymorphisms and two haplotypes are significantly associated with NPC in Cantonese-speaking Chinese in China by performing a case-control study and family based association study. These reports prompt us to investigate the association between polymorphisms of *CYP2E1* and NPC susceptibility in Malaysian Chinese.

1.5.1.2 Non CYP phase I enzymes - Epoxide hydrolase (EPHX)

EPHX is classified as non CYP phase I enzymes. EPHX will detoxify the potentially harmful epoxides to diols which are less reactive and easier to be excreted (Sandberg and Meijer, 1996). Epoxides can be found in the environment or generated by other phase I enzymes such as CYP. Some epoxides which are highly reactive electrophilic compounds may react with DNA and proteins, thereby leading to DNA adducts or mutation and ultimately cause cancer. Therefore, both highly and intermediate reactive epoxides need to be detoxified (Sandberg and Meijer, 1996).

EPHX that associated with drug metabolism are categorized into two xenobiotic metabolizing forms: microsomal epoxide hydrolase (mEH, EPHX1) and soluble epoxides hydrolases (sEH, EPHX2) (Beetham *et al.*, 1993). The sequence homology of both EPHX1 and EPHX2 are low, but based on the structural similarities, they are belong to the family α/β hydrolase fold enzymes. EPHX1 will hydrolyze the majority of xenobiotic epoxides whereas EPHX2 is mainly metabolizing the fatty acid epoxides (Ota *et al.*, 1980).

i. Microsomal epoxide hydrolase - Epoxide hydrolase 1 (EPHX1)

EPHX1 with 9 exons covers 35.48 kb genomic sequences in chromosome 1q42.1. The sequences of EPHX1 protein and nucleic acid are found to be conserved evolutionary in humans, rats and rabbits. EPHX1 catalyzes the phase I hydrolysis of both aromatic and aliphatic epoxides to less reactive trans-dihydrodiols which can be conjugated and excreted from the body (Laasanen *et al.*, 2002; Decker *et al.*, 2009). Besides that, it is involved in the activation and detoxification processes and in the metabolism of endogenous and exogenous compounds including androstene oxide, estroxide and styrene oxide (Fandrich *et al*., 1995; Seidegard and Ekstrom, 1997; Oesch *et al*., 2000).

The activity of EPHX1 can be detected in all tissues including microsome, endoplasmic reticulum and integral to membrane. It is highly expressed in lung, liver, kidney, gonads and epithelial cells. In addition, researchers also detect EPHX1 in human adrenal gland, lymphocytes, human blood resting mononuclear leukocytes and adult human hepatocytes (Seidegard and Ekstrom, 1997). Several studies have been performed to investigate the potential association of *EPHX1* polymorphisms with susceptibility to various diseases. *EPHX1* polymorphisms were suggested to be associated with hepatocellular carcinoma (McGlynn *et al.,* 1995), ovarian cancer (Lancaster *et al.*, 1996), colon cancer (Harrison *et al.*, 1999) and smoking-induced lung cancer (Benhamou *et al.*, 1998).

ii. Soluble epoxide hydrolase - Epoxide hydrolase 2 (EPHX2)

EPHX2 gene is located on chromosome 8p21. It covers 45kb and contains 19 exons. It is a multifunctional enzyme with an N-terminal phosphatises activity and a C-terminal with hydrolase activity. EPHX2 is more substrate selective compared to EPHX1. In contrast to EPHX1, it is incapable to metabolize some steroids and PAHs, but it can

catalyze the hydrolysis of trans-substituted epoxides such as trans-stilbene oxide (Decker *et al.*, 2009).

EPHX2 plays a role in the metabolism of preferentially endogenous epoxides such as arachidonic epoxides (e.g. epoxyeicosatrienoicacids; EETs) and linoleic acid epoxides (e.g. leukotoxin). EETs are derived from arachidonic acid through metabolism by CYPs. Inhibition of EPHX2 levels causes decrease in EET degradation which may lower the blood pressure and reduce inflammation and pain. Therefore, inhibition of EPHX2 might be useful in the treatment of hypertension in experimental animals (Imig *et al*., 2002). Besides EETs, leukotoxins can be catalysed by EPHX2 to the actual toxic metabolite leukotoxindiol which may cause failure of multiple organ and adult respiratory distress syndrome after severe body burns (Decker *et al.*, 2009). Leukotoxins may elicit inflammatory symptoms, so inhibition of EPHX2 may have anti-inflammatory effects.

The expression of EPHX2 is detected in many organs including liver, lung, kidney, brain and ovary. The highest activity and expression of EPHX2 are detected in the liver and kidney of human. EPHX2 expression is decreased or lost in several human malignant neoplasms such as liver, renal, colon and prostate cell carcinomas (Enayetallah *et al.*, 2006). Besides that, lower activity and expression of EPHX2 is detected in some human carcinoma cell lines such as HepG2, ACHN, SN12C, Caco-2, Du-145 and Hela cells compared to normal cells (239T cells) (Goldstein *et al.*, 1960; Stone *et al.*, 1978, Aden *et al.*, 1979; Borden *et al.*, 1982). The loss of EPHX2 might cause the loss of hydrolase or phosphatase activity, thereby increases the risk of development of cancers (Enayetallah *et al.*, 2006).

A number of SNPs have been identified in *EPHX2* gene. Some studies have been performed to investigate the potential association of *EPHX2* polymorphisms with the disease susceptibility. Variant Arg287Gln (rs751141) of *EPHX2* has been reported

to be associate with several diseases such as type 2 diabetic. This variant may reduce the stability of EPHX2 protein and epoxide hydrolase activity (Przybyla-Zawislak *et al.*, 2003). *EPHX2* polymorphisms has also been suggested to be associated with coronary heart disease (Lee *et al.*, 2006)

Both *EPHX1* and *EPHX2* genes are important in metabolizing xenobiotic molecules and associated with various cancers, suggesting that they might be involved in the NPC carcinogenesis. To date, both genes have not been studied on their association with NPC susceptibility. Hence, we aimed to investigate if any association exits between *EPHX1*, *EPHX2* polymorphisms and NPC susceptibility in Malaysian Chinese.

1.5.1.3 Phase II enzymes - Glutathione S-transferases (GSTs)

GSTs are a superfamily of phase II metabolic enzymes which involved in the detoxification of reactive electrophilic compounds (Jancova *et al.* 2010). GSTs perform conjugation reactions between reduced glutathione (GSH) and a variety of electrophilic compounds such as products of oxidative stress, environmental pollutants and carcinogens (Singh *et al.*, 2008). GSTs neutralize the electrophilic sites and rendering the products more water-soluble which can be excreted easily. The conjugation may facilitate excretion, so it constitutes a detoxification step (Hayes and Pulford, 1995; Parl, 2005; Jancova *et al.* 2010).

GSTs play important role in cellular protection from environmental and oxidative stress (Mcllwain *et al.*, 2006; Jancova *et al.* 2010). A few studies suggested that genetic polymorphisms of GST may involve in enhancing the defence mechanism against oxidative stress. A number of genetic polymorphisms among GSTs have been identified and may contribute to inter-individual difference in enzyme activity. For instance, individuals with homozygous deletions of either *GSTM1* or *GSTT1* have no

enzymatic functional activity, leading to decreased ability of detoxification. Absent or deficient of GST enzyme activity might increase the risk of cancer due to poor elimination of the electrophilic carcinogens (Rebbeck, 1997; Parl, 2005; Singh *et al.*, 2008). Besides that, GSTs modulate the induction of other enzymes and proteins which are important for cellular functions such as DNA repair. Therefore, GSTs might involve in maintaining cellular genomic integrity and consequently play a role in cancer susceptibility (Geisler and Olshan, 2001).

There are three main subfamilies for GSTs: cytosolic, mitochondrial and membrane bound microsomal. Human cytosolic GSTs are further divided into seven classes based on their biochemical, immunologic and structural properties: Alpha, Mu, Omega, Pi, Sigma, Theta and Zeta (Mcllwain *et al.*, 2006).

i. Glutathione S-transferases Mu -1 (*GSTM1*)

GSTM1 is belonged to Mu class and located on the short arm of chromosome 1 (1p13.3). In general, GSTM1 plays a role in the metabolic detoxification of various reactive chemicals. Three alleles including one deletion allele, *GSTM1**A and *GSTM1**B are identified at the *GSTM1* locus. *GSTM1**A and *GSTM1**B only differ by substitution C to G at base position 534 in exon 7, however no functional difference is found between two alleles (Rebbeck, 1997; Parl, 2005).

GSTM1 activity is present in individuals with at least one copy of either *GSTM1**A or *GSTM1**B allele. Thus, *GSTM1**A and *GSTM1**B are referred as positive conjugator phenotype. Conversely, loss of function of GSTM1 enzyme is a result of homozygous deletion of this gene. The deletion of *GSTM1* might be caused by unequal crossing over of the M1 and M2 loci (Mcllwain *et al.*, 2006). Individual with homozygous deletion genotype are referred as negative conjugator.

The *GSTM1* null genotype is suggested to increase individual sensitivity to various reactive compounds. *GSTM1* null individuals have shown reduced capacity to detoxify genotoxins in tobacco smoke. The *GSTM1* null smokers may exhibit an increased level of DNA adducts compare to *GSTM1* positive smokers (Geisler and Olshan, 2001). Therefore, individuals with the null variants may associate with increased risk of cancer due to impaired ability to detoxify certain carcinogens. Recent meta-analysis of epidemiological studies revealed that deficiency of GSTM1 which caused by homozygous deletions of the gene is associated with the increased risk of lung cancer (Hosgood *et al.*, 2007; Carlsten *et al.*, 2008). To date, several epidemiological studies were carried out to evaluate the association of *GSTM1* null genotype with NPC (Nazar-Sterwart *et al.*, 1999; Cheng *et al.*, 2003; Tiwawech *et al.*, 2005; Deng *et al.*, 2005; Guo *et al.*, 2008). However, inconsistent findings were obtained from these studies (Table 1.4).

ii. Glutathione S-transferases Theta -1 (*GSTT1*)

The theta class of GSTs consists of *GSTT1* and *GSTT2*. *GSTT1* is expressed in the normal esophageal epithelium. It is located on chromosome 22q11.23 (Pemble *et al.*, 1994). Some studies revealed that *GSTT1* polymorphisms may result in production of an enzyme which will activate ethylene oxide, epoxybutanes, halomethanes, and methyle bromide, or no production of enzyme. In addition, it is involved in phase II detoxification of PAHs which are found in tobacco smoke (Pemble *et al.*, 1994; Jancova *et al.* 2010).

Individuals with homozygous deletion genotype of *GSTT1* are considered as negative conjugator phenotype. Individuals with deleted *GSTT1* exhibit decreased catalytic activity and associated with an increased risk of cancer (Pemble *et al.*, 1994; Rebbeck *et al.*, 1997). *GSTT1* null genotype is suggested to be associated with increased

"baseline" level of sister chromatid exchanges (SCEs) in lymphocytes, and show no association with smoking or other known exposures. Deletion of *GSTT1* will lead to the absence of enzyme activity. Therefore, individuals with *GSTT1* null genotype are at increased risk for genotoxic damage from environmental (Rebbeck *et al.*, 1997). An epidemiological study reveals that *GSTT1* null genotype is associated with an increased risk of susceptibility to NPC in Chinese from Guangxi, China (Deng *et al.*, 2005).

iii. Glutathione S-transferases class-pi-1 (*GSTP1*)

GSTP1 (also known as *FAEES3* or *GST3*) is located on long arm of chromosome 11 (11q13) and it is composed of 210 amino acids. Morrow *et al.* (1989) reported that *GSTP1* is approximately 2.8kb, encompassing 7 exons and 6 introns. GSTP1 plays a role in eliminating thymidine and uracil propenal which are the products of DNA oxidation. It is also involved in the inactivation of tobacco-related procarcinogens. GSTP1 is widely expressed in normal human epithelial tissue.

A substitution of A to G in exon 5 causes a replacement of isoleucine to valine at codon 105 (Ile105Val; rs1695). The 105 Val form might change the affinity and enzymatic activity of GSTP1 enzyme. Individuals with Val allele are suggested to possess lower catalytic activity and diminish the capacity for detoxification of carcinogens (Moyer *et al.*, 2008). Individuals with Ile/Val and Val/Val genotypes show higher risk for head and neck cancer compared to those with the Ile/Ile genotype (Olshan *et al.*, 2000). On the other hand, overexpresion of GSTP1 was found in some tumours and drug-resistant cell lines (Hayes and Pulford, 1995), indicating that GSTP1 may play a role in resistance to certain anticancer drugs.

In this study, we aimed at determining whether *GSTM1* null, *GSTT1* null and *GSTP1* polymorphisms are associated with individual susceptibility to NPC in Malaysian Chinese population by performing case-control association study.

1.5.2 Genes involving in immune response

As mentioned in section 1.2.3(a), EBV infection is believed to be involved in the development of NPC (Armstrong *et al.*, 1998; Thompson and Kurzrock, 2004). Normally, EBV infection is not life-threatening and can be controlled by the immune system in a healthy individual. The presence of EBV in epithelial cells and B cells may provoke immune response consisting of antibodies to various virally encoded antigens (Lin *et al.*, 1977; Hadar *et al.*, 1986; Zheng *et al.*, 1994a). An effective immune response against EBV infection is considered vital in the early infection process and it might limit the viral propagation (Khanna *et al.*, 1995). Therefore, genes involved in immune response might play a key role in the development of NPC.

1.5.2.1 Polymeric Ig Receptor Gene (*PIGR*)

The plasma membrane of polarized epithelial cells composed of apical and basolateral. Some basolateral proteins have basolateral sorting signals which can direct the newly made protein from the trans-Golgi network to the basolateral surface. Casanova *et al.* (1991) suggest that first basolateral signal was identified in the polymeric immunoglobulin receptor (PIGR). PIGR is a transmembrane glycoprotein, consisting of an approximately 560 residue extracellular region, a 23-residue transmembrane region and a 103-residue cytoplasmic tail. There are five domains (D1-D5) for the extracellular region of PIGR which show similar sequence with the immunoglobulin variable regions (Mostov *et al.*, 1984; Braathen *et al.*, 2005). PIGR is selectively expressed by mucosal and glandular epithelial cells, including respiratory and intestinal epithelial cells (Brock *et al.*, 2002).

In general, PIGR is involved in the transcellular transport of polymeric immunoglobulin (pIg) into secretion. It plays a role in transcytosis of polymeric IgA or IgM across different types of epithelial cells (Zhang *et al.*, 2000; Braathen *et al.*, 2005).

The pIg contains a short polypeptide (J chain) which is required for binding to the receptor on the basolateral surface of mucosal epithelial cells (Lin *et al.*, 1997; Zhang *et al.*, 2000). The PIGR-pIg complex is endocytosed and trancytosed to the apical cell surface by vesicular transport (Braathen *et al.*, 2005). The protein complex is then cleaved by an unknown protease, and the extracellular domain of PIGR which bound to a pIg molecule is released into mucosa. This released complex is known as secretory immunoglobulin (Zhang *et al.*, 2000). The secretory immunoglobulin might be involved in the adaptive immune defence against inhaled pathogens or microbial infections, thus PIGR is suggested to play a role in the immune system.

However, some pathogens may exploit PIGR in order to enter into the epithelial cells. Zhang and colleagues (2000) suggest that PIGR works as a pneumococcus receptor which may trannslocate these bacteria across the epithelial barrier. Besides that, studies have suggested that EBV may exploit PIGR to invade the epithelial cells through associating with IgA (Gan *et al.*, 1997; Lin *et al.*, 1997). It is also reported that EBV could only infect the PIGR-expressing NPC cells in the presence of IgA (Lin *et al.*, 1997). If the epithelium loses its polarity or *PIGR* is mutated, the viral translocation process might fail, and consequently cause EBV infection. Therefore, PIGR was proposed to be involved in the mechanism of EBV entry into the nasopharyngeal epithelium (Lin *et al.*, 1997; Norderhaug *et al.*, 1999).

In the study by Hirunsatit *et al.* (2003), a missense mutation (*PIGR*1739 C/T; SNP rs291102) was suggested to be associated with the increased risk of NPC among individuals of Thai origin and Chinese origin in Thailand. This variant was proposed to alter the efficiency of *PIGR* in releasing the complex of IgA-EBV and consequently increased the risk of NPC (Hirunsatit *et al.*, 2003). Our study aimed to determine whether there is any association between *PIGR* polymorphisms and NPC susceptibility among Malaysian Chinese.

1.5.2.2 Palate lung and nasal epithelial clone (PLUNC)

The gene cluster of PLUNC family is located on chromosome 20q11.2 (Bingle and Craven, 2002). Members of BPI fold containing family are subdivided into two groups based on their size, i.e., 'short' protein and 'long' protein. 'Short' proteins, comprises SPLUNC1 (256 amino acids), SPLUNC2 (249 amino acids), SPLUNC3 (253 amino acids), while 'Long' proteins, comprises LPLUNC1 (484 amino acids), LPLUNC2 (458 amino acids), LPLUNC3 (463 amino acids) and LPLUNC4 (>469 amino aicds) (Bingle and Craven, 2000; Bingle and Craven, 2002).

PLUNC family proteins are clearly evolutionarily related to each other and to the bactericidal / permeability-increasing protein (BPI), lipopolysaccharide-binding protein (LBP), phospholipids transfer protein (PLTP) and cholesteryl ester transfer protein (CETP) genes which are involved in the host defence mechanism, they exhibit a high degree of overall similarity of exon structure (Bingle and Craven, 2002). Besides that, PLUNC family proteins are predicted to show significant 3D similarity to LBP and BPI. Therefore, the members of the PLUNC family are speculated to play significant roles in host defence in the mouth, nose and upper airways where innate defence is a major requirement for these locations.

All of these proteins except for LPLUNC4, contain putative signal peptides at the N terminus (Bingle and Craven, 2002). "Short" proteins of PLUNC family are suggested to have homology only to the N-terminal domain of BPI, while "long" proteins showed homology to the N- and C- terminal domains of BPI and LBP (Bingle and Craven, 2003). The N-terminal domain of BPI normally confer the LPS binding and bactericidal activity of BPI, whereas the C-terminal domain is involved in the opsonic role of the protein (Elsbach and Weiss, 1998). Two members of the BPI fold containing family, LPLUNC1 and SPLUNC1, are reported to be down-regulated in the NPC

biopsies (Zhang *et al.*, 2003), suggesting that these proteins might be involved in the development of NPC.

i. Long palate, lung, and nasal epithelium clone 1 (*LPLUNC1*)

LPLUNC1 gene which is also known as *BPIFB1*, *MGC14597* and *C20orf114* contains 16 exons and covers 26.7 kb. LPLUNC1 has homology to both N- and C- terminal domains of BPI and LBP. It is highly expressed in normal adult and fetal nasopharyngeal epithelial tissue and in airway epithelial cells which is isolated from cystic fibrosis (CF) patients. Expression of LPLUNC1 is also detected in nasal secretions, bronchoalveolarlavege (BAL) and sputum, and as a major secreted product of cultured tracheobronchial epithelial cells (Zhang *et al.*, 2003; Casado *et al.*, 2005; Candiano *et al.*, 2007). LPLUNC1 levels in BAL are found to be increased in asthmatic patients after segmental allergen challenge (Scheetz *et al.*, 2004), suggesting that LPLUNC1 may play a role in the innate immune response.

A previous study shows that LPLUNC1 is down-regulated in 71% NPC biopsies (Zhang *et al.*, 2003). Furthermore, Yang and colleagues (2007) reported that LPLUNC1 could suppress the growth and proliferation of human nasopharyngeal carcinoma cell line (HNE1) and delayed G1-S phase cell cycle progression of HNE1 cells. These findings suggest that LPLUNC1 might play a role in the pathogenesis of NPC. However, no association study of *LPLUNC1* polymorphism and NPC susceptibility has yet been conducted.

ii. Short palate, lung, and nasal epithelium clone 1 (*SPLUNC1*)

SPLUNC1 (also designated as *BPIFA1*, *PLUNC*, *LUNX*, *YH1*, *SPURT*, or *NASG*) contains 9 exons and covers 7.3 kb (Bingle and Craven, 2000). SPLUNC1 is a secreted protein. It is most prominently expressed in the nose, salivary glands and upper

respiratory tract. Besides that, SPLUNC1 is identified in nasal lavage, nasal mucous, BAL and sputum (Bingle and Craven, 2002). SPLUNC1 is proposed to show putative bactericidal/bacteriostatic functions and able to neutralize endotoxin as it is one of the members of BPI/LBP family (Bingle and Craven, 2003). Ghafouri *et al.* (2004) have showed that it binds to LPS in vitro, supporting its role in innate immune response which may mediate signalling and inflammatory response.

An analysis of nasal cells from patients with and without CF revealed that SPLUNC1 level is elevated in CF (Roxo-Rosa *et al.*, 2006). Lung irritants also cause elevation of SPLUNC1 which can be seen in nasal samples from smokers and epoxy workers (Lindahl *et al.*, 2001). Other studies have suggested that SPLUNC1 was up-regulated in chronic obstructive pulmonary disease and emphysema (Di *et al.*, 2003). Moreover, the expression of SPLUNC1 was reported to be significantly induced in the respiratory epithelium of the nasal cavity following olfactory bulbectomy in rats (Sung *et al.*, 2002). These studies supported the role of SPLUNC1 in inflammatory processes in the upper airways. Besides that, SPLUNC1 (or known as LUNX) is upregulated and has been suggested as a potential marker for micormetastasis for non-small cell lung cancer (NSCLC) (Iwao *et al.*, 2001).

SPLUNC1 was reported to be down-regulated in 70% of NPC biopsies, and transfection of *SPLUNC1* gene into CNE-2Z cells (human NPC cell line) decreased cell proliferation (He *et al.* 2000, Zhang *et al.* 2003), suggesting that it might be involved in the carcinogenesis of NPC. On the other hand, He and co-workers (2005) have performed an association study on *SPLUNC1* (or known as *PLUNC*) polymorphisms and NPC susceptibility among Cantonese-speaking Chinese in China. Their results revealed that that two promoter SNPs of *SPLUNC1* gene (C-2128T and C-1888T; rs2752903 and rs750064, respectively) were associated with susceptibility to NPC (He

et al., 2005). However, the functional importance of these two promoter SNPs are still unknown.

Since the association of *SPLUNC1* polymorphisms and NPC susceptibility has not been studied in Malaysian populations and both *LPLUNC1* and *SPLUNC1* genes are likely to play important roles in NPC development, both candidate genes were examined for their association with NPC among Malaysian Chinese.

1.6 SEARCHING FOR NPC SUSCEPTIBILITY GENES BY GENOME-WIDE SCREENING APPROACH

Genome-wide screening is a common approach in identifying potential SNPs associated with disease susceptibility. The advent of high throughput and parallel SNP genotyping platforms such as Affymetrix SNP Array 5.0 and SNP Array 6.0, Illumina HumanHap300, HumanHap550, Human660W, HumanOmni Express etc. have enabled a large number of GWAS studies being reported for the last 10 years (Matsuzaki *et al.*, 2004; Gurdenson *et al.*, 2005). Furthermore, a large number of variations especially SNPs have been found with the completion of Human Genome project in 2003 and the International HapMap project in 2005. Haplotype block structures across the genome can be defined by using these data (International HapMap Consortium, 2005). The emergence of more data regarding the patterns of coinheritance of markers (LD) through the International HapMap project further facilitate the selection of SNPs for GWAS and stimulated the GWAS of complex disease (International HapMap Consortium, 2005).

Recently, a multistage approach for GWAS is proposed by Hirschhorn and Daly (2005) to minimize the sample size, without sacrificing power. A subset of samples is first genotyped in a discovery stage, examining a full marker set covering the whole genome. In the second stage, an independent set of samples is genotyped with the

strongly associated SNPs, using a less expensive genotyping platform. The data from the first and second stages are combined and analysed. Once genotyping is complete, a number of quality control checks is performed for the genotyped SNPs, including testing for Hardy-Weinberg equilibrium (HWE) and the proportion of samples successfully genotyped (Balding, 2006; Witte, 2010). The association of SNPs with disease is then evaluated using different statistical tests after excluding those SNPs which failed to pass the quality controls. Besides that, a multiple testing correction using Bonferroni correction is performed in GWAS, however, this approach might be conservative as some assayed SNPs are correlated (Pe'er *et al.*, 2008).

In contrast to candidate gene approach, thousands to millions of SNPs across the genome are genotyped simultaneously in the GWAS, this enabled us to discover novel and unanticipated gene associations. Prior to our study, no GWAS was performed to uncover the association of genetic factors to NPC susceptibility. Hence, we conducted a GWAS in Malaysian Chinese population to expand the screening of NPC susceptibility genes. However, two GWAS of NPC were reported later in the Taiwanese and Southern Chinese populations. Both GWAS have identified several NPC susceptibility loci including *HLA-A, HLA-F, GABBR1, TNFRSF19, MDS1-EVI1* and *CDKN2A-CDKN2B* (Tse *et al.*, 2009; Bei *et al.*, 2010).

In the present study, the Illumina HumanHap550 BeadChip which contains 550,000 SNPs on a single microarray was used for genotyping. These 550,000 SNP markers are tag-SNPs identified from both Phase I and Phase II of the International HapMap Project (Steemers and Gunderson, 2005) that captures close to 90% of all existing variations in the human genome with $r^2 > 0.7$ in the European and Asian populations. Illumina's HumanHap 550 uses Infinium assay to efficiently analyse all SNPs on the array chip. This assay utilizes direct hybridization of whole genome amplified genomic DNA to 50mer locus specific primers (Steemers and Gunderson,

2007). After hybridization, each SNP locus is scored by an enzymatic based extension assay using labelled nucleotides. The labels can be detected by staining with a sandwich-based immunohistochemistry assay.

1.7 OBJECTIVES

This project was aimed to identify the gene/genes underlying susceptibility to NPC, leading to a better understanding of the mechanisms involved in the pathogenesis of NPC.

The specific objectives of this study are listed below:

- To investigate polymorphisms in genes involved in xenobiotic metabolisms (*CYP1A1, CYP2A13, CYP2E1, EPHX1, EPHX2, GSTM1, GSTT1* and *GSTP1*) and immune response (*PIGR, LPLUNC1* and *SPLUNC1*) and its association with NPC in Malaysian Chinese.
- To examine the functional significance of the selected SNPs.
- To discover novel NPC-associated genes by genome-wide screening approach.

CHAPTER 2

MATERIALS AND METHODS

2.1 STUDY SUBJECTS

Our study utilized peripheral blood samples from both Chinese NPC patients as well as Chinese healthy controls. The study protocols have been reviewed and approved by the Medical Ethics Committee of University of Malaya Medical Center (UMMC) (reference number: 546.3) and Ethical Committees of the Yokohama Institute, The Institutes of Physical and Chemical Research (RIKEN), Yokohama, Japan. Written informed consent was obtained from all the study subjects.

2.1.1 Blood samples

Approximately 1 - 6 ml of peripheral blood was collected from each study subjects. The collected blood samples were kept in EDTA-coated tubes to prevent clotting, and stored in -20℃ freezer prior to DNA extraction. All participated subjects in this study were unrelated Malaysian Chinese. The patients were pathologically diagnosed as NPC, while the healthy controls consisted of blood donors without family history of cancer. All the healthy controls were recruited with the help of UMMC Blood Bank, collecting from various blood donation campaigns held in Selangor and Kuala Lumpur.

An initial set of 81 NPC patients from UMMC and 147 healthy controls were recruited from September 2006 to January 2007. From March 2007 until 2009, an additional of 366 NPC patients were recruited from UMMC, NCI Cancer Hospital, Tung Shin Hospital, and two NPC referring hospitals of the Malaysian NPC study group, namely Hospital Pulau Pinang and Hospital Kuala Lumpur / Universiti Putra Malaysia (UPM). An additional of 340 controls were recruited with the help of UMMC Blood Bank. The Malaysian NPC study group consists of clinicians and research

scientists, and recruit NPC patients from five major NPC referring hospitals across Malaysia: Hospital Pulau Pinang, Hospital Kuala Lumpur / Universiti Putra Malaysia (UPM), Sarawak General Hospital / Universiti Malaysia Sarawak, Queen Elizabeth Hospital in Sabah, UMMC and Hospital Universiti Sains Malaysia (USM) in Kubang Kerian, Kelantan.

2.2 GENOMIC DNA

2.2.1 DNA extraction

Genomic DNA was extracted from peripheral blood using two methods, namely QIAmp DNA blood mini kit (Qiagen, Valencia, CA) or conventional phenol chloroform extraction method. The choice of method depended on the volume of the blood samples. The QIAmp DNA blood mini kit was used to extract DNA from blood samples with a volume of 200 µl or less, according to the manufacturer's protocol (Qiagen, Valencia, CA).

2.2.2 Conventional phenol chloroform method

Frozen whole blood samples (1 - 6 ml) were thawed, transferred to a 15 ml tube and resuspended in 8 - 12 ml of 1X red blood cell lysis buffer (Appendix B). The solution was mixed by inversion of the tube and then centrifuged at 3,850 x *g* for 10 min at 10 ℃. The supernatant was discarded and the pellet was resuspended in 4 ml of red blood cell lysis buffer. The solution was centrifuged at 3,850 x *g* for 10 min at 10 ℃ and the supernatant was discarded. The pellet was re-washed with red blood cell lysis buffer if the red cells still persists. After that, 80 µl of 10X Proteinase K buffer, 20 µl of 20 mg/ml Proteinase K, 40 µl of 20% SDS solution (Appendix B) and 400 µl of distilled water were added to the pellet. The solution was mixed gently and incubated overnight at 37 ℃.

On the following day, the solution was cooled to room temperature and 200 µl of Sodium Chloride (6M) was added. The mixture was subjected to pulse vortex for 30 seconds and then transferred equally into two 1.5 ml microcentrifuge tubes. Following that, 800 µl of phenol chloroform (1:1, v/v) was added and the solution centrifuged at 16,060 x g for 5 min at 4℃. The aqueous phase containing the genomic DNA was transferred to a new 1.5 ml microcentrifuge tube. DNA was precipitated by adding chilled absolute ethanol followed by incubation at -20℃ for 1 hour. After that, the solution was spun at 16,060 x g, 4℃ for 5 min, and the supernatant discarded. The DNA pellet was washed by 1 ml of 70% Ethanol, centrifuged at 16,060 x g for 5 min at 4℃ and the alcohol discarded once again. The DNA pellet was air-dried and resuspended in Tris EDTA (TE) buffer. Contaminating RNA was removed by adding 0.2 mg/ml RNase A solution, followed by incubation at 37℃ for 30 min. The extracted genomic DNA was analysed by electrophoresis on 0.8% w/v agarose gels and its concentration was measured by Nanodrop-1000 (ND-1000, Thermo Scientific, USA). These extracted DNA were stored at -20℃. Genomic DNA was diluted to 5 ng/µl or 50 ng/µl prior to PCR amplification and genotyping.

2.3 PRIMER DESIGN

Genomic sequence information was obtained from NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). Primers were designed using DNA calculator (http://www.sigma-genosys.com/calc/DNACalc.asp, Sigma-Aldrich, St Louis, MO, USA). In general, primers are between 18 and 25 base pairs in length, and have < 50% GC content, with a melting temperature of 58℃ - 62℃. The primer sequences were evaluated for secondary structure and primer dimer formation using the DNA calculator (Sigma, USA). Primer annealing sites were chosen to be devoid of SNPs and repetitive sequences. The repetitive sequences were identified using a Repeat Masker program

(http://www.repeatmasker.org/cgi-bin/WEBRepeat Masker). The presence of SNP and specificity of the primer sequence were checked using NCBI BLAST tool (http://www.ncbi.nlm.nih.gov/BLAST/) and University College Santa Cruz (USCS) In-silico PCR (http://genome.ucsc.edu/cgi). All the primers were synthesized by Invitrogen Life Technologies or Sigma-Aldrich chemical company.

2.4 CANDIDATE GENE APPROACH

2.4.1 Study subjects for candidate gene approach

A two-stage case-control study was performed to investigate the association between candidate genes and NPC susceptibility, except for *GSTs*. For an initial screening, a total of 81 NPC patients and 147 healthy controls were first examined. These samples were recruited from September 2006 to January 2007. The second batch of samples were obtained much later (from March 2007 until 2009), hence a two-stage screening approach. SNP(s) which showed *P*-value < 0.05 were further genotyped in the second set of samples, consisting of 366 NPC patients and 340 healthy controls (summarized in Table 2.1).

Table 2.1: Sample sets for candidate gene approach

Sample set	Cases	Controls
1st	81	147
2nd	366	340
total	447	487

2.4.2 Selection of polymorphisms and genotyping

SNPs of *CYP1A1, CYP2A13, CYP2E1, EPHX1, EPHX2, GSTP1, PIGR, LPLUNC1* and *SPLUNC1*, and the deletion genotype of *GSTM1*, *GSTT1* were examined for their association with NPC susceptibility in Malaysian Chinese. The selected SNPs were haplotype-tagging SNPs (tag-SNPs) that encompass LD blocks of each candidate genes ($r^2 \geq 0.8$, minor allele frequency (MAF) ≥ 0.05). These tag-SNPs were selected using genotype data of Beijing Han Chinese Population, HapMap Project Phase II dataset (http://hapmap.ncbi.nlm.nih.gov/). In addition, several candidate SNPs were selected from published literature. The list of candidate polymorphisms and the genotyping assay used are summarized in Table 2.2. Random samples were genotyped twice to verify the genotyping consistency, and ≥ 96% concordant results were obtained.

Table 2.2: Candidate polymorphisms postulated to be associated with NPC susceptibility

No	Gene	Polymorphisms	Source	Genotyping assay
1	*CYP1A1*	rs1048943	Tag-SNP, nonsynonymous SNP	Invader
		rs4646422	Tag-SNP, synonymous SNP	
2	*CYP2A13*	rs1645690	Tag-SNP	Invader
		rs7251532	Tag-SNP	
		rs7254343	Tag-SNP	
		rs7255149	Tag-SNP	
		rs12981316	Tag-SNP	
3	*CYP2E1*	rs2031920	Tag-SNP	Invader
		rs2070672	Tag-SNP	
		rs915908	Tag-SNP	
		rs2249695	Tag-SNP	
		rs2515644	Tag-SNP	
		rs8192780	Tag-SNP	
4	*EPHX1*	rs1051740	Tag-SNP	Invader
		rs1877724	Tag-SNP	
		rs2292566	Tag-SNP	
		rs10915884	Tag-SNP	
		rs3753658	Tag-SNP	
		rs3753661	Tag-SNP	
		rs2671272	Tag-SNP	
5	*EPHX2*	rs2741334	Tag-SNP	Invader
		rs721619	Tag-SNP	
		rs751141	Tag-SNP	
		rs4149244	Tag-SNP	
		rs4149252	Tag-SNP	
		rs13262930	Tag-SNP	
		rs7341557	Tag-SNP	
		rs7816586	Tag-SNP	
		rs1126452	Tag-SNP	
		rs7357432	Tag-SNP	
6	*GSTM1*	Deletion	Chen *et al.*, 1997	Multiplex PCR
7	*GSTT1*	Deletion	Chen *et al.*, 1997	
8	*GSTP1*	rs762803	Tag-SNP	Invader
		rs4147581	Tag-SNP	
		rs1695	Functional SNP	

Table 2.2, continued

No	Gene	Polymorphisms	Source	Genotyping assay
9	*PIGR*	rs291085	Tag-SNP	Invader
		rs172361	Tag-SNP	
		rs291102	Tag-SNP	
		rs2275531	Tag-SNP	
		rs291096	Tag-SNP	
		rs3813956	Tag-SNP	
		rs2000057	Tag-SNP	
10	*LPLUNC1*	rs1884882	Tag-SNP	Invader
		rs4911314	Tag-SNP	
		rs1999663	Tag-SNP	
		rs6087476	Tag-SNP	
11	*SPLUNC1*	rs927159	Tag-SNP	TaqMan
		rs1570034	Tag-SNP	
		rs2752903	Tag-SNP, He *et al.*, 2005	
		rs750064	Tag-SNP, He *et al.*, 2005	

2.4.3 Statistical analyses for candidate gene approach

The genotype frequencies of all SNPs were tested for Hardy Weinberg equilibrium (HWE) in the controls by exact test using PLINK (v1.07) (Purcell, 2007). SNP which showed deviation from HWE ($P < 0.01$) was excluded from the subsequent analysis.

The association between *GSTM1, GSTT1* and NPC risk was examined by two-tailed Fisher's exact test, while the association analysis of SNPs of candidate genes was evaluated by a logistic regression analysis adjusted for age and sex using PLINK (v1.07) (Purcell, 2007). The genotype and allele frequencies were compared between NPC patients and healthy controls in allelic, dominant- and recessive-inheritance models. The significance level for the combined analysis was then adjusted using Bonferroni correction for multiple testing. Besides *P*-value, odds ratio (OR) and 95% confidence intervals (CI) were also calculated in the association analysis.

The LD among SNPs in *CYP2E1* and *SPLUNC1* was calculated using Haploview v4.2 (Daly laboratory at the Broad Institute, Cambridge, MA). The pairwise LD was expressed as a squared correlation coefficient (r^2) between two SNPs. The statistical power of each association study was estimated using the module case–control for discrete traits of the Genetic Power Calculator (http://pngu.mgh. harvard.edu/~purcell/gpc/cc2.html; Purcell *et al.*, 2003).

2.5 GENOME-WIDE ASSOCIATION STUDY (GWAS)

2.5.1 Genotyping and quality control

A genome-wide analysis for 111 NPC patients and 260 healthy controls (panel A) was conducted using Illumina HumanHap 550v3 Genotyping BeadChip, according to the manufacturer's protocol (San Diego, CA, USA). The Illumina Infinium genotyping was carried out by Laboratory for Genotyping Development, Center for Genomic Medicine

of RIKEN, Japan, while the statistical analyses of GWAS were performed by Laboratory for Statistical Analysis, Center for Genomic Medicine of RIKEN, Japan.

The overall genotyping call rates of all individuals were ≥ 98%. Among 554,496 SNPs genotyped, about 21,448 SNPs were eliminated from the study as no genotyping data obtained. Furthermore, SNPs with call rates of < 98%, SNPs that showed deviation from HWE in controls ($P < 0.000001$) and monomorphic SNPs were excluded from the subsequent analysis. As a result, 492,131 autosomal SNPs which passed the quality control filters were further analysed.

To verify the genotyping results of the Illumina assay, the top 200 SNPs showing the smallest *P*-values were re-genotyped with all individuals in panel A (111 NPC patients and 260 controls) by high throughput multiplex PCR–invader assay (Third Wave Technologies, Madison, WI, USA) or by direct sequencing using 3730xl DNA analyzer (Applied Biosystems, Foster City, CA). The data obtained by the two assays (Illumina vs Invader assay or direct sequencing) were compared. Only SNPs that passed through the validation process with concordance rates of > 98% were further evaluated using 336 NPC patients and 227 controls (panel B) (Table 2.3).

Table 2.3: Sample sets for genome-wide screening approach

Sample set	Cases	Controls
panel A	111	260
panel B	336	227
total	447	487

2.5.2 Statistical analysis of GWAS

For both GWAS and the replication study, the allele and genotype distributions in cases and controls were compared and evaluated in allelic, dominant- and recessive-inheritance models by two-tailed Fisher's exact test. Statistical analyses were carried out for the data obtained from both panel A and panel B samples. SNPs were rank-ordered

according to the lowest *P*-value in a combined set of samples among these three models. Population stratification for the GWAS data was examined by principle component analysis (PCA) performed using EIGENSTRAT software (Price *et al*., 2006). Four HapMap populations (European (CEU), Africans (YRI) and East-Asians (Japanese and Han Chinese, denoted as JPT+CHB) were used as reference groups in the first PCA analysis. The PCA plots were obtained using the first two components (Eigenvectors 1 and 2), and it indicated that no or low amount of population substructure in our population (Appendix E and F). The quantile-quantile plot (Q-Q plot) of the genome-wide *P*-values was generated by the R program, and the Manhattan plot (Appendix G) was generated using Haploview v4.2 (Daly laboratory at the Broad Institute, Cambridge, MA). For fine mapping association analysis of *ITGA9*, the association between polymorphisms and NPC risk was evaluated by a logistic regression analysis adjusted for age and sex using PLINK (v1.07) (Purcell, 2007).

2.5.3 Fine mapping association analysis of top GWAS loci, *ITGA9*

SNP rs2212020 of *ITGA9* showed the most significant association with NPC susceptibility after replication. The call rates for this landmark SNP (rs2212020) were 0.99 and 1.00 in cases and controls, respectively. The *P*-values for the HWE test in cases and controls (panel A) were 0.761 and 0.223, respectively. To further analyse SNPs within the 40-kb LD region flanking the rs2212020, 19 tag-SNPs with $r^2 \geq 0.80$ and MAF ≥ 0.05 were selected using genotype data of Beijing Han Chinese Population, HapMap Project Phase II dataset (http://hapmap.ncbi.nlm.nih.gov/) and examined by multiplex PCR–invader assay. These SNPs were first genotyped on 111 NPC patients and 260 controls (panel A). SNPs that showed significant association were then genotyped with the panel B samples, consisting of 336 NPC patients and 227 healthy controls.

2.6 GENOTYPING OF *GSTM1* AND *GSTT1*

The genotypes of *GSTM1* and *GSTT1* were determined by performing multiplex PCR which included beta-globin gene as a positive control. Multiplex PCR simultaneously amplified *GSTM1, GSTT1* and beta-globin. The primers for *GSTM1, GSTT1* and beta-globin were obtained from Chen *et al.*, 1997. The following reaction mixture was prepared:

	1X
dH_2O	11.3 µl
10x PCR buffer	3.2 µl
25 mM dNTPs	1.2 µl
50 pmol/µl Forward primers	0.15 µl (0.05 µl from each primer)
50 pmol/µl Reverse primers	0.15 µl (0.05 µl from each primer)
2.5 U/µL ExTaq Hot Start DNA polymerase	2 µl
5 ng/ µl DNA	2 µl
Total	**20 µl**

Multiplex PCR was then performed in an Applied Biosystems thermal cycler (model 9700), according to the following protocol:

94 ℃	5 min	
94 ℃	10 s	40 cycles
60 ℃	20 s	
72 ℃	45 s	
72 ℃	5 min	

PCR products were then electrophoresed on 2% (w/v) agarose gel for 30 min, at 100V. A fragment of 215 bp indicated the presence of *GSTM1*, a fragment of 480 bp indicated

the presence of *GSTT1* and the presence of positive control (beta-globin gene) was indicated by a fragment of 268 bp.

2.7 DNA SEQUENCING OF *CYP2E1* AND *SPLUNC1*

DNA sequencing was performed for *CYP2E1* and *SPLUNC1* genes. Monoplex PCR was first carried out to amplify the target sequence. The following reaction was prepared:

	1X
dH_2O	13.0 µl
10x PCR buffer	3.2 µl
25 mM dNTPs	1.2 µl
50 pmol/µl Forward primers	0.2 µl
50 pmol/µl Reverse primers	0.2 µl
5 U/µL ExTaq DNA polymerase	0.2 µl
5 ng/ µl DNA	2.0 µl
Total	**20 µl**

PCR amplification was performed in an Applied Biosystems thermal cycler (model 9700), according to the following protocol:

94 ℃ 5 min

94 ℃ 30 s

60 ℃ 30 s } 37 cycles

72 ℃ 1 min

72 ℃ 5 min

After amplification, all the PCR products were subjected to 1.5% (w/v) agarose gel electrophoresis to confirm successful amplification of the target gene. PCR products

were purified using PCR Cleanup Filter Plates (Millipore, USA). The purified PCR products were then sequenced using the BigDye® terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA), according to manufacturer's protocol with modifications. The sequencing reaction is as follows:

	1X
dH_2O	7.55 µl
Big dye terminator	0.5 µl
Sequencing buffer	2.8 µl
50 pmol/µl forward or reverse primer	0.1 µl
Purified PCR product	3 µl
Total	14 µl

Cycle sequencing was performed in an Applied Biosystems thermal cycler (model 9700), according to the following protocol:

96 ℃ 2 min

96 ℃ 20 s

50 ℃ 30 s } 25 cycles

60 ℃ 4 min

The sequencing extension products were then subjected to ethanol precipitation. About 100 µl of 80% ethanol was added to each product. The solution centrifuged at 1,109 x g for 40 min, the supernatant discarded, washed with 100 µl of 70% ethanol, centrifuged at 1,109 x g for 15 min, the ethanol was discarded and the samples were air-dried at room temperature for 1 hour to remove excess ethanol. Prior to sequencing in the Applied Biosystems Prism 3700 DNA analyser, the samples were resuspended in 20 µl of Mili-Q water. The DNA sequence obtained will be analysed using Sequencher 4.6 (Gene Codes Corporation, USA) or PolyPhred program (Nickerson *et al.*, 1997).

2.8 MULTIPLEX PCR-INVADER SNP GENOTYPING ASSAY

2.8.1 Multiplex PCR

Multiplex PCR was first performed before Invader assay in order to reduce the amount of required DNA samples for the genotyping of each SNP. A PCR master mixture which contained the following PCR reagents was prepared:

	1X
10x PCR buffer	3.2 µl
25 mM dNTPs	1.2 µl
50 pmol/µl Forward primers	0.05 µl X number of SNPs
50 pmol/µl Reverse primers	0.05 µl X number of SNPs
2.5 U/µL ExTaq Hot Start DNA polymerase	2 µl
5 ng/ µl DNA	2 µl
Add dH_2O to	**20 µl**

Multiplex PCR amplification was performed in an Applied Biosystems thermal cycler (model 9700), according to the following protocol:

95 ℃	5 min	
95 ℃	15 s	40 cycles
60 ℃	45 s	
72 ℃	3 min	

2.8.2 Invader SNP genotyping assay

The probes for Invader assay were designed as described previously (Hsu *et al.*, 2001; Ohnishi *et al.*, 2001). After multiplex PCR was preformed, the PCR products were visualized using 1.5% (w/v) agarose electrophoresis to ascertain successful amplification of the target gene. PCR products were subsequently diluted 10-fold and dispensed into 384-well plates. Each sample was mixed with the reaction mixture containing two allele-specific probes and invader oligonucleotide, invader buffer, FRET probes, ROX and Cleavase enzyme in the following proportions:

	1X
dH_2O	2.125 µl
Invader buffer	0.125 µl
FRET probes	0.125 µl
Cleavase VIII	0.125 µl
0.125mM ROX	0.25 µl
Allele-specific probes/invader oligonucleotide mixture	0.25 µl
Total	**3 µl**

Samples were then incubated in an Applied Biosystems thermal cycler (model 9700) with an initiation denaturation at 95 ℃ for 5 min, followed by incubation at 63 ℃ for 5 min. At the end point of each Invader assay, the fluorescent signal intensities and the ratios of FAM/ROX and VIC/ROX were generated. The endpoint plate read was performed using Applied Biosystems 7900HT Fast Real time PCR system.

2.9 TAQMAN SNP GENOTYPING ASSAY (APPLIED BIOSYSTEMS, FOSTER CITY, CA)

Four tag-SNPs of *SPLUNC1* (rs927159, rs1570034, rs2752903 and rs750064) were genotyped by TaqMan SNP genotyping assay. These TaqMan probes and the reaction reagents were purchased from Applied Biosystems, Foster City, CA. The reaction mixture was prepared as below:

	1X
dH_2O	1.25 µl
TaqMan universal master mixture	2.5 µl
20X primer probe mixture	0.25 µl
5 ng/µl genomic DNA	1 µl
Total	**5 µl**

Non-template controls were included for the PCR reaction. PCR amplification was performed in Applied Biosystems thermal cycler, (model 9700). The PCR conditions were as below:

95 ℃	10 min	
95 ℃	15 s	40 cycles
58 ℃	1 min	
72 ℃	15 s	
72 ℃	5 min	

Same as Invader assay, the fluorescent signal intensities and the ratios of FAM/ROX and VIC/ROX were generated at the end of TaqMan assay. The endpoint plate read was performed using Applied Biosystems 7900HT Fast Real time PCR system.

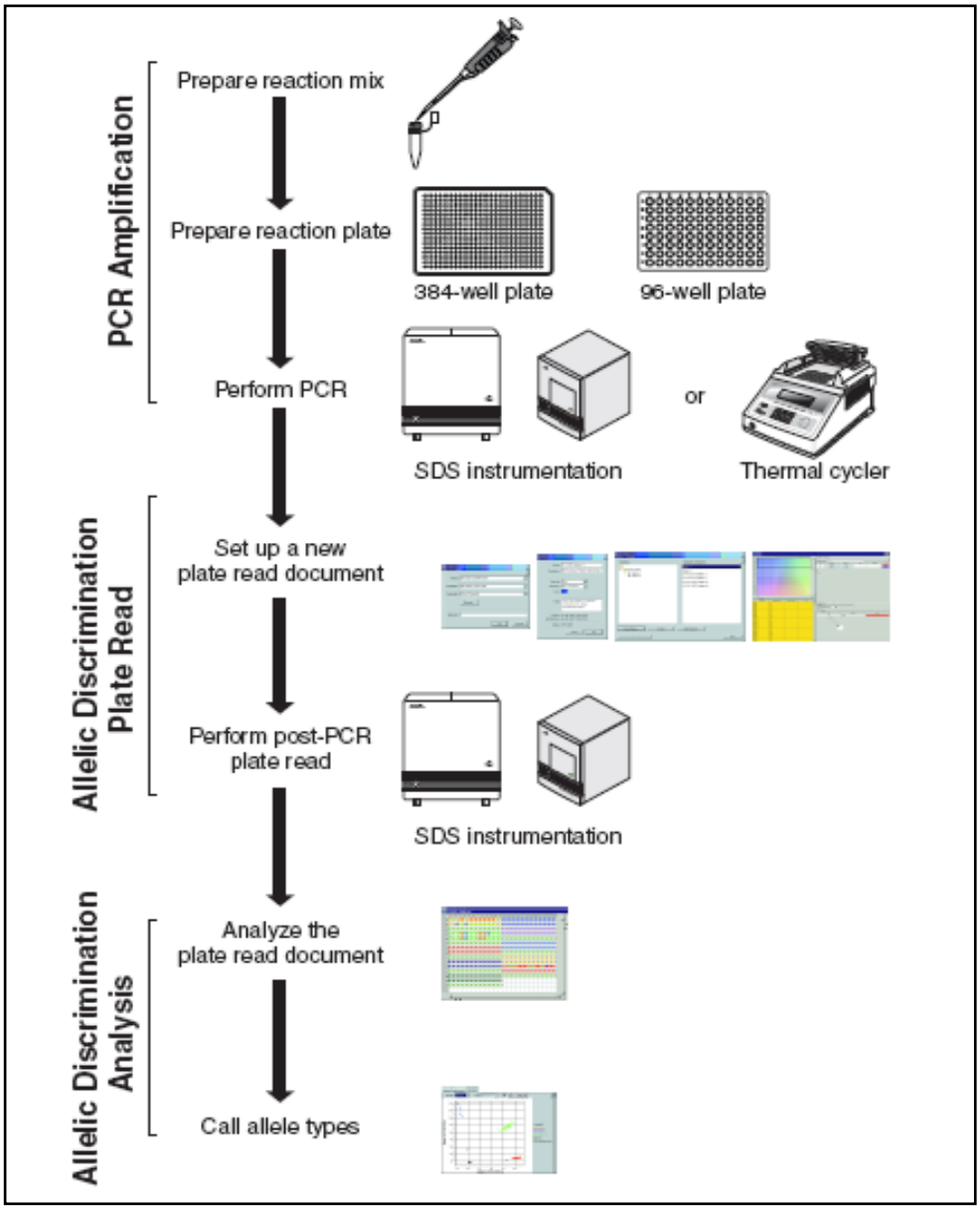


Figure 2.1: The work flow of TaqMan genotyping assay.

A simplified overview of the procedure using TaqMan SNP genotyping assays. TaqMan assay combined PCR amplification and genotyping into one single reaction. The post-PCR plate read were performed using Applied Biosystems 7900HT Fast Real time PCR system (Obtained from TaqMan SNP genotyping assays protocol).

2.10 FUNCTIONAL ANALYSIS OF *SPLUNC1*

2.10.1 Cell culture and reverse transcription (RT)-PCR

HeLa, BT549, MCF7, ZR-75-1, NCI-H520, NCI-H1781, NCI-H2170, A549, HepG2 and HuH-7 cell lines were purchased from the American Type Culture Collection (ATCC, Rockville, MD) while the NPC cell lines (CNE1, HONE1, SUNE1, HONE Akata) were kindly gifted by various laboratories. The cells were maintained in RPMI-1640 (GIBCO®, USA) or Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich Chemical Company, USA) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich Chemical Company, USA) and grown at 37℃ and, 5% Carbon dioxide (CO_2).

Total RNA was isolated from each cell line using RNeasy plus Mini Kit (Qiagen, Valencia, CA) while cDNA was synthesized from total RNA (1 µg) using SuperScriptIII First Strand Synthesis System (Invitrogen, Carlsbad, CA). *SPLUNC1* mRNA expression was determined by semi-quantitative PCR using forward primer: 5'-GCCTGAACAACATCATTGAC-3' and reverse primer: 5'-TGTCCAGAAGACCTTG AATG-3'. GAPDH was used as an internal control. The following reaction mixture was prepared:

	1X
dH_2O	13.0 µl
10x PCR buffer	3.2 µl
25 mM dNTPs	1.2 µl
50 pmol/µl Forward primers	0.2 µl
50 pmol/µl Reverse primers	0.2 µl
5 U/µL ExTaq DNA polymerase	0.2 µl
5 ng/µl cDNA	2.0 µl
Total	**20 µl**

PCR amplification was performed in an Applied Biosystems thermal cycler (model 9700), according to the following protocols:

94 ℃ 5 min

94 ℃ 30 s

60 ℃ 30 s } 37 cycles

72 ℃ 1 min

72 ℃ 5 min

SPLUNC1 expression was low or undetected in most of the cell lines except NCI-H1781. Therefore, NCI-H1781 was used for the following functional studies.

2.10.2 Electrophoretic mobility shift assay (EMSA)

2.10.2.1 Preparation of nuclear extracts

Nuclear extracts were prepared from NCI-H1781 cells grown in the 10 cm^2 culture dish. Culture medium was removed and the cells were washed twice with 10 ml of Phosphate-Buffered Saline (PBS). Lysis buffer was added to the cells and incubated on ice for 5 min. The cells were removed by a sterile scrapper and gently aspirated several times to disrupt cell clumps. The lysate was transferred to a 1.5 ml microcentrifuge tube and viability of the cells was determined using the Trypan blue exclusion test (Strober, 2001).

Cells were collected by centrifuging at 4 ℃, 5 min, 1,000 x g and the supernatant was discarded. Next, 100 µl of nuclear extract buffer was added to the pellet, the mixture was mixed well and incubated on ice for 30 min. After half an hour, the mixture was centrifuged at 4 ℃ for 15 min, 13,000 x g. After centrifugation, the supernatant was transferred to a new 1.5 ml microcentrifuge tube and kept. The BCA protein assay was used for the quantitation of total protein. The BCA protein assay was carried out according to the manufacturer's protocol (Pierce, USA). After quantifying the protein,

nuclear extract buffer was used to dilute the nuclear extracts to the working concentration (1 µg/µl). The nuclear extracts were kept at -80 ℃.

2.10.2.2 Probe preparation

EMSA was performed using DIG Gel Shift Kit, 2nd Generation (Roche Diagnostics, Mannheim, Germany), according to the manufacturer's instructions. Complementary 45 mer oligonucleotides were designed with the SNP site in the middle, flanked by 22 mer upstream and downstream sequences. The annealing reaction consists of 20 µl of each 100 µM oligonucleotide (sense and antisense), 20 µl of TEN-buffer and top up to 200 µl with distilled water was prepared. The mixture was then incubated at 95 ℃ for 10 min. After incubation, it was cooled down slowly to room temperature. Concentration of the double stranded oligonucleotides was measured by Nanodrop 1000 (ND-1000, Thermo Scientific) and diluted to 100 ng/µl with distilled water. The double stranded oligonucleotides were then labelled with digoxigenin-11-ddUTP (DIG), according to the manufacturer's instructions. Labelled oligonucleotides with a final concentration of 4 ng/µl were kept at -20 ℃.

2.10.2.3 Determination of labelling efficiency

Labelling efficiency was determined through a series of dilutions (10^{-1}, $10^{-2}$and 10^{-3}) from the 4 ng/µl of labelled oligonucleotides. Each dilution of the labelled oligonucleotide and the labelled control-oligonucleotide (Roche) was applied to the nylon membrane. Then, the nucleic acid was fixed to the membrane by cross linking with UV light. The membrane was then immersed in 20 ml of washing buffer (Roche) and incubated with shaking for 2 min. After that, the membrane was incubated for 30 min in 10 ml of blocking solution (Roche), subsequently in 10 ml of antibody solution (Roche) for 30 min. Later, the membrane was washed twice with 10 ml of washing

buffer for 15 min, and then equilibrated in 10 ml of detection buffer for 5 min. Next, the membrane was placed in the hybridization bag and 0.1 ml of chemiluminescent alkaline phosphatase substrate (CSPD; Roche) was applied on it. The bag was sealed and the substrate was spread evenly. The membrane was incubated with CSPD working solution for 5 min. After incubation, the excess CSPD working solution was squeezed out from the bag and the edge of the bag was sealed. The membrane was then incubated for 10 min in a 37℃ waterbath to enhance the CSPD chemiluminescent reaction. Finally, the membrane was exposed to X-ray film (Eastman Kodak Co., Rochester, NY) for 10 min and the film was processed by Fuji Medical Film Processing system Cepros SV (Fuji Photo Film Co., Ltd., Japan). The quality of DIG-labelled probe was determined by comparing the dilution series with that of the DIG-labelled control oligonucleotide.

2.10.2.4 Gel shift assay

The reaction mixture was prepared according to the manufacturer's protocol (Roche). The nuclear extracts were incubated with the DIG-labelled probes for 20 min. at 4℃. After incubation, the protein-DNA complexes were separated by electrophoresis on a 7.5% polyacrylamide gel with 0.5 × TBE buffer (Appendix B). The gel was run at 80V for 4 hours. After electrophoresis, electro-blotting and cross-linking were performed. The nucleic acid was fixed to the nylon membrane by cross-linking with UV light. The membrane was then washed with 20 ml of washing buffer for 5 min. The membrane was subsequently incubated for 30 min in 100 ml of blocking solution and 30 min in 20 ml of antibody solution. Next, the membrane was washed with 100 ml of washing buffer for 15 min twice. The membrane was then equilibrated for 5 min in 20 ml of detection buffer. After that, the membrane was placed in the hybridization bag where 1 ml of CSPD working solution was applied on it. The bag was sealed and the substrate

was spread evenly and incubated for 5 min. After incubation, excess CSPD working solution was squeezed out and the edge of the bag was then sealed. The membrane was incubated for 10 min in 37℃ waterbath to enhance the CSPD chemiluminescent reaction. After that, the membrane was exposed to X-ray film (Kodak) for 30 min to 1 hour. The film was then processed by Fuji Medical Film Processing system Cepros SV (Fuji Photo Film Co., Ltd., Japan).

For the competition assay, the nuclear extracts were pre-incubated with a 125-fold excess of unlabelled oligonucleotides before adding the DIG-labelled probes. In the supershift assay, the nuclear extracts were pre-incubated with monoclonal antibody specific to Sp1 (E-3, Santa Cruz Biotechnology, Santa Cruz, CA) for 30 min at 4℃ and then incubated with DIG-labelled probe for another 20 min at 4℃. Each experiment was independently repeated three times.

2.10.3 Luciferase reporter assay of rs1407019

Luciferase reporter assay was performed to investigate the effect of rs1407019 on *SPLUNC1* gene transcription.

2.10.3.1 Vector construction and transient transfections

The pGL4-SV40 was constructed previously in the Laboratory of Pharmacogenetics, RIKEN, Japan. The pGL4-SV40 was constructed by inserting the SV40 promoter into the pGL4 luciferase vector (pGL4.10 [luc2]) (Promega, Madison, WI) at the *Bgl* II/*Hind* III site. A single copy of the 41-bp double stranded oligonucleotides which contains A or G allele of rs1407019 was then cloned into *Xho* I/*Bgl* II site of this pGL4-SV40. After ligation, heat shock transformation of competent *E. coli* (JM109) (Takara Bio Inc, Japan) with the construct was performed, and the transformed cells were plated on LB plates containing carbenicillin (Invitrogen, Carlsbad, CA) and incubated overnight at

37°C. The next day, colony PCR was carried out using the primers RV3- 5'ACTAGCAAAATAGGCTGTCCC3' and gL4R- 5'CTTAATGTTTTTGGCATCTT CCA3'. The following reaction mixture was prepared for colony PCR:

	1X
dH_2O	15.0 µl
10x PCR buffer	3.2 µl
25 mM dNTPs	1.2 µl
50 pmol/µl Forward primers	0.2 µl
50 pmol/µl Reverse primers	0.2 µl
5 U/µL ExTaq DNA polymerase	0.2 µl
Total	**20 µl**

PCR amplification was performed in an Applied Biosystems thermal cycler (model 9700), according to the following protocol:

94 ℃ 2 min

94 ℃ 30 s

57 ℃ 30 s } 35 cycles

72 ℃ 2 min

72 ℃ 5 min

The sequence of the plasmid DNA was then verified by sequencing. Finally, the plasmid with the correct insert sequence was cultured in large scale overnight. The plasmid DNA was extracted using QIAfileter midi and maxi kit (Qiagen), according to the manufacturer's protocol.

For transfection, the NCI-H1781 cells (3×10^6) were nucleofected with 4 µg of luciferase reporter constructs and 1 µg of pGL4-SV40-hRluc (an internal control for transfection efficacy) by nucleofector system, program A23 (Amaxa, Köln, Germany)

and cultured in 24-wells plates. Cells were maintained in RPMI 1640 supplemented with 10% FBS, kept at 37 ℃ in a CO_2 incubator and cultured for 48 hours.

2.10.3.2 Luciferase assay

At 48 hours after transfection, the media was removed from the cells and the cells were rinsed with PBS before 50 µl of 1X passive lysis buffer was added into each well. Cells were then placed on a room temperature shaker for at least 15 min in order to lyse the cells completely. After that, 10 µl of cell lysate was added to each well of a 384-wells plate. The firefly and Renilla luciferase activities were measured in a luminometer (Perkin Elmer, USA) using the Dual-Luciferase Reporter Assay System (Promega), following the manufacturer's instructions. At first, 30 µl of Luciferase Assay Reagent II (LAR II) was added, and firefly luciferase activity was measured. Next, 30 µl of Stop & Glo reagent was added and the Renilla luciferase activity was measured by the luminometer. The luciferase activities were normalized to the renilla activity. The difference of luciferase reporter activities between two alleles was analysed by unpaired Student's t-test (Microsoft). Each experiment was independently repeated three times and each sample was studied in quadruplicate.

2.11 FLOWCHART OF EXPERIMENTS

Figure 2.2 and Figure 2.3 show a schematic diagram of a flowchart to illustrate the related steps and experiments performed in the study.

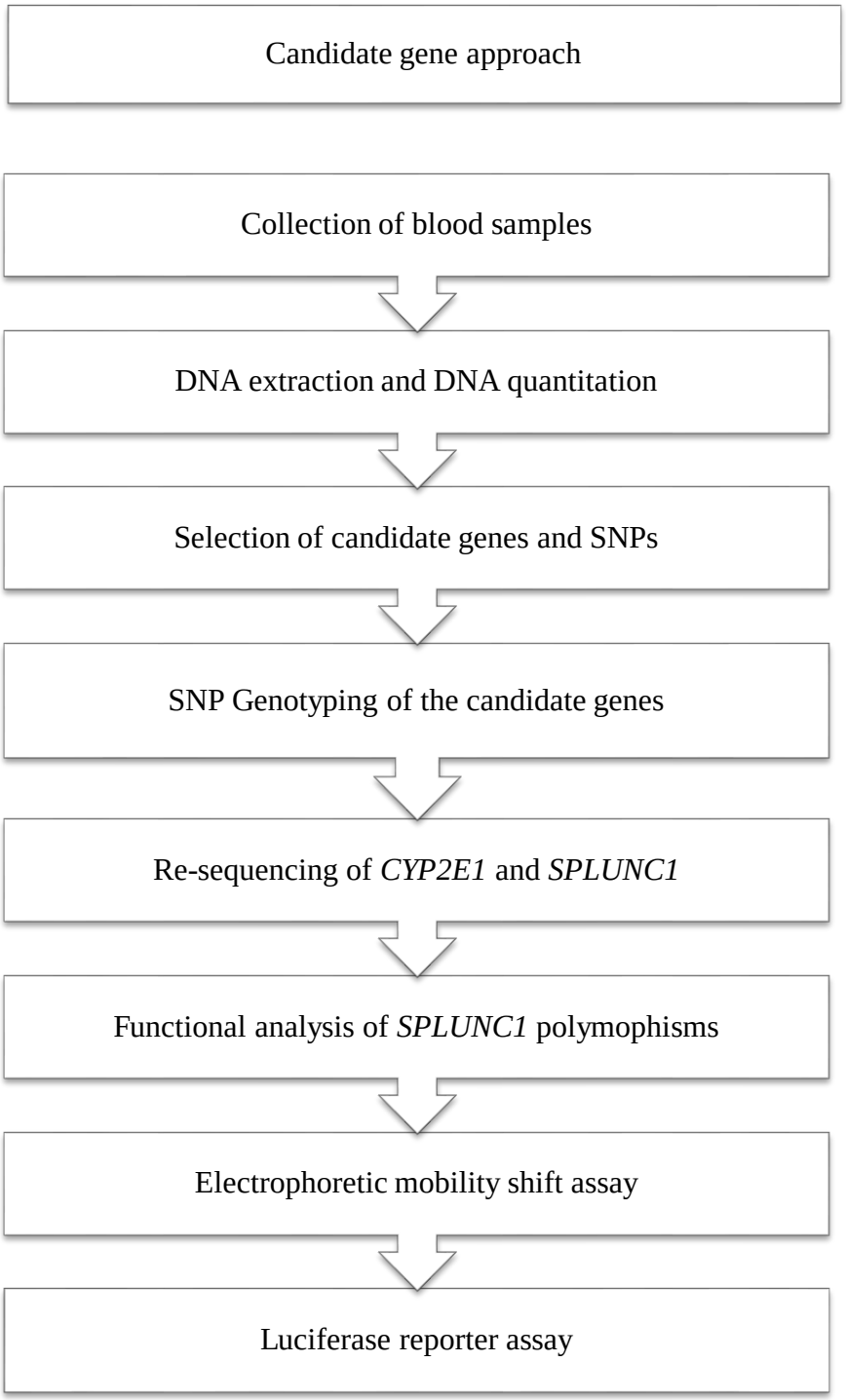


Figure 2.2: A flowchart to illustrate the related steps and experiments performed for the candidate gene approach.

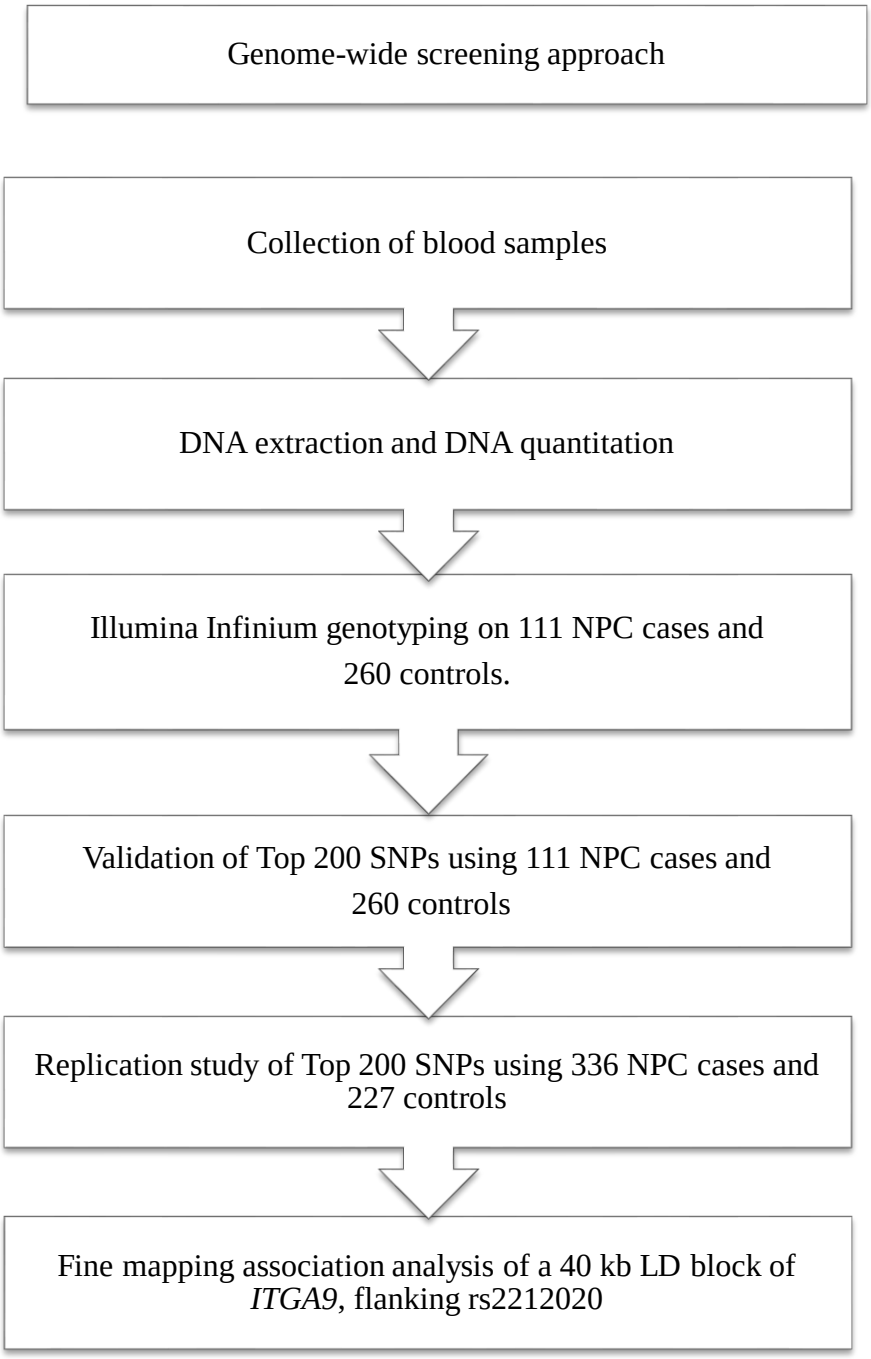


Figure 2.3: A flowchart to illustrate the related steps and experiments performed for the GWAS.

CHAPTER 3

RESULTS

3.1 CHARACTERISTICS OF THE STUDY SUBJECTS

A total of 447 NPC patients and 487 healthy controls were recruited in the present study. Of the 447 cases, 75% (333 cases) were males and their age ranged from 14 to 86 (median = 53). Of the 487 controls, 71% (344 controls) were males, and their ages ranged from 19 to 60 (median = 37). The relevant characteristics of the study subjects are shown in Table 3.1.

The differences in the distributions of age and sex between the NPC patients and controls were evaluated by *t*-test. The mean age in cases was significantly higher than that in controls ($P < 0.0001$) whereas the gender distribution was not significantly different between the two groups ($P = 0.17$).

Table 3.1: Characteristics of study participants in the case-control study

Characteristics	Cases (N)	Controls (N)	*P* value
Gender			
Female	113	143	
Male	333	344	0.17
Unknown	1	0	
Age (years)			
Mean (range)	53.0 (14-86)	37.3 (19-60)	
Median	53	37	< 0.0001
Total	**447**	**487**	

3.2 GENOMIC DNA EXTRACTION

Genomic DNA samples were successfully extracted from the 447 NPC patients and the 487 healthy controls, and approximately 10 µg to 50 µg DNA was obtained from 2 ml - 3 ml of blood. These DNA samples were then diluted to a concentration of 5 ng/µl or 50 ng/µl for subsequent experiments. All the extracted DNA samples were relatively pure, the A_{260}/A_{280} ratios of these genomic DNA were around 1.70 to 1.90. The integrity of the DNA samples were checked by visualizing under 0.8% (w/v) agarose gel electrophoresis and GeneRuler^{TM} Lambda DNA/HindIII DNA ladder (Fermentas, Canada) was used as a size reference. Figure 3.1 shows a representative agarose gel of genomic DNA which were extracted from five individuals. A single DNA band of 23 kb observed in the agarose gel electrophoresis indicating that high molecular weight and intact genomic DNA was obtained from the extraction.

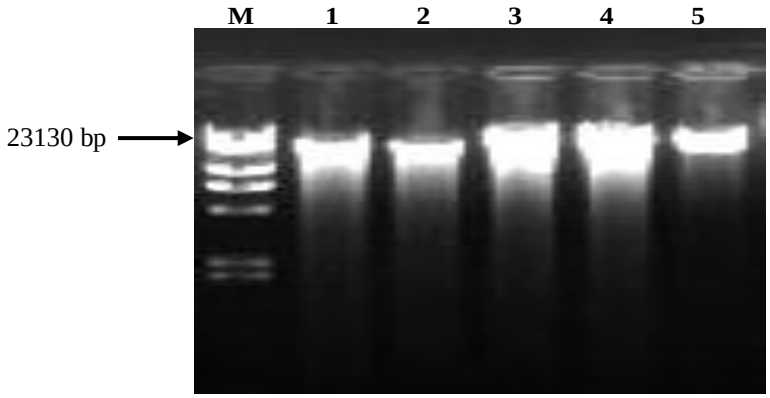


Figure 3.1: A representative agarose gel electrophoresis of genomic DNA extracted from peripheral blood.

M : GeneRuler^{TM} Lambda DNA/HindIII DNA ladder (Fermentas, Canada)

Lanes 1-5 : Genomic DNA

3.3 EXAMINATION OF PCR PRODUCTS BY 1.5% (W/V) AGAROSE GEL ELECTROPHORESIS

The amplified PCR products (monoplex and multiplex PCR) were examined by electrophoresis on 1.5% (w/v) agarose gels. PCR products for genotyping of *GSTM1* and *GSTT1* genotypes were examined by electrophoresis on 2% (w/v) agarose gels (this will be further discussed in section 3.8). Figure 3.2 shows a representative agarose gel of the monoplex PCR products containing SNP rs795520. The 1kb DNA ladder (Invitrogen, Carlsbad, CA) was used as a size reference. No positive amplification was observed in the DNA-free TE buffer (lanes 94 and 96; non-template negative control), absolving the PCR reaction of contamination.

Figure 3.3 shows a representative agarose gel of the multiplex PCR products *SPLUNC1*. A total of 20 SNP regions were simultaneously amplified using 20 pairs of primers in a single reaction. The 1kb DNA ladder (Invitrogen, Carlsbad, CA) was used as a size reference. No amplification was observed in the non-template negative control (lanes 94 and 96).

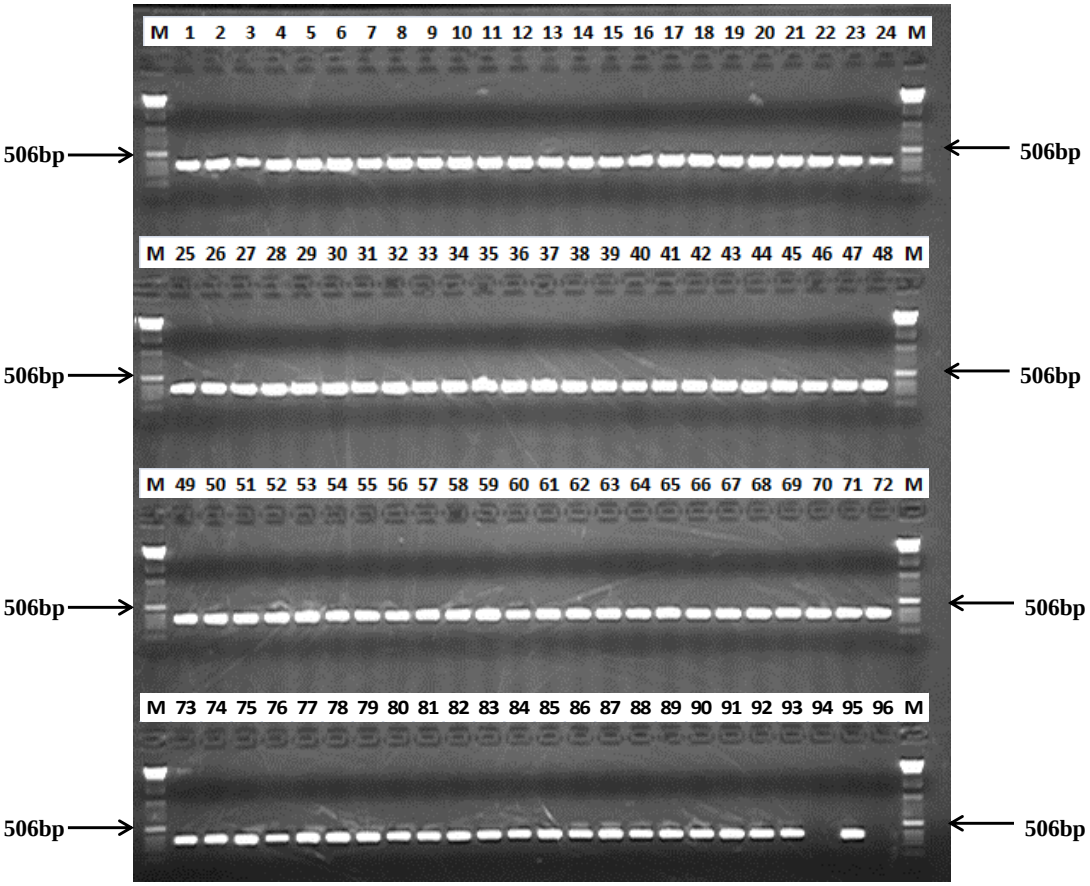


Figure 3.2: Ethidium-bromide-stained 1.5% (w/v) agarose gel of the monoplex PCR products containing SNP rs795520 (size = 341bp).

M : 1kb DNA ladder (Invitrogen, Carlsbad, CA)

Lanes 1-93 & 95 : Amplified PCR products

Lanes 94 & 96 : Non-template negative control

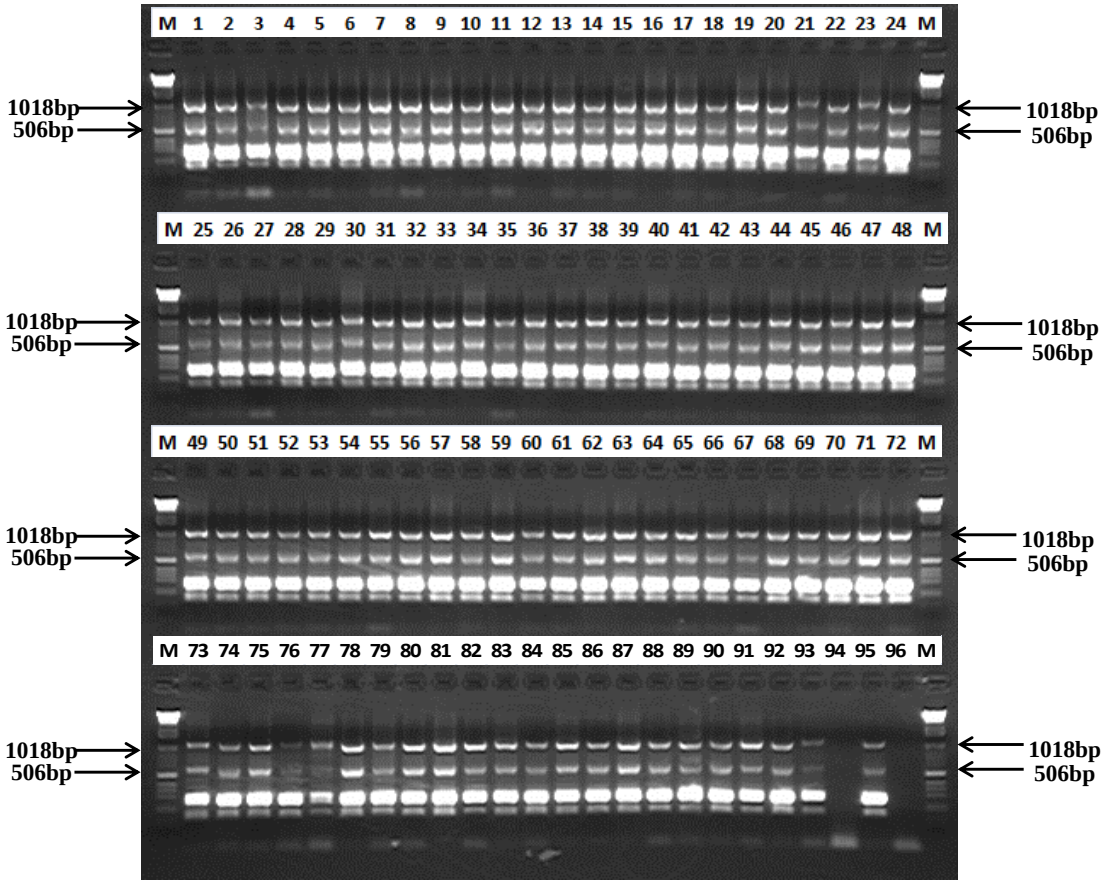


Figure 3.3: Ethidium-bromide-stained 1.5% (w/v) agarose gel of the 20-plex PCR products for *SPLUNC1* (sizes range from 231bp to 1000bp).

M : 1kb DNA ladder (Invitrogen, Carlsbad, CA)

Lanes 1-93 & 95 : Amplified PCR products

Lanes 94 & 96 : Non-template negative control

3.4 INTERPRETATION OF THE TAQMAN AND INVADER ASSAY GENOTYPING RESULTS

Figure 3.4 shows a representative graphical output of a SNP with allele 1 and allele 2 as detected by the Sequence Detection System software of ABI PRISM 7900HT. The genotyping results were obtained by detecting the fluorescent signal intensities of Invader or TaqMan assay and displayed as a scatter plot of Allele X Rn versus Allele Y Rn (Figure 3.4).

The allelic discrimination plot contains four separate and distinct clusters which represented three different genotypes and NTC (Figure 3.4c). The three clustering of different genotypes vary along the horizontal axis (Allele X), vertical axis (Allele Y) or diagonal (Allele X/Allele Y). The cluster of red spot signified a strong VIC/ROX signal and represented those individuals homozygous for allele 1. Heterozygous individuals exhibited increased fluorescence for both reporter dyes (VIC and FAM), they were represented by the cluster of green spot. On the other hand, the cluster of blue spot indicated a strong FAM/ROX signals and symbolized those individuals homozygous for allele 2

Besides that, the normalized reporter signal (Rn) values generated by alleles were checked to ensure the reproducibility of the result and to confirm the allelic discrimination. Allele X Rn represented the Rn for allele 1 (VIC/ROX) whereas allele Y Rn represented the Rn for allele 2 (FAM/ROX) (Figure 3.4b).

.

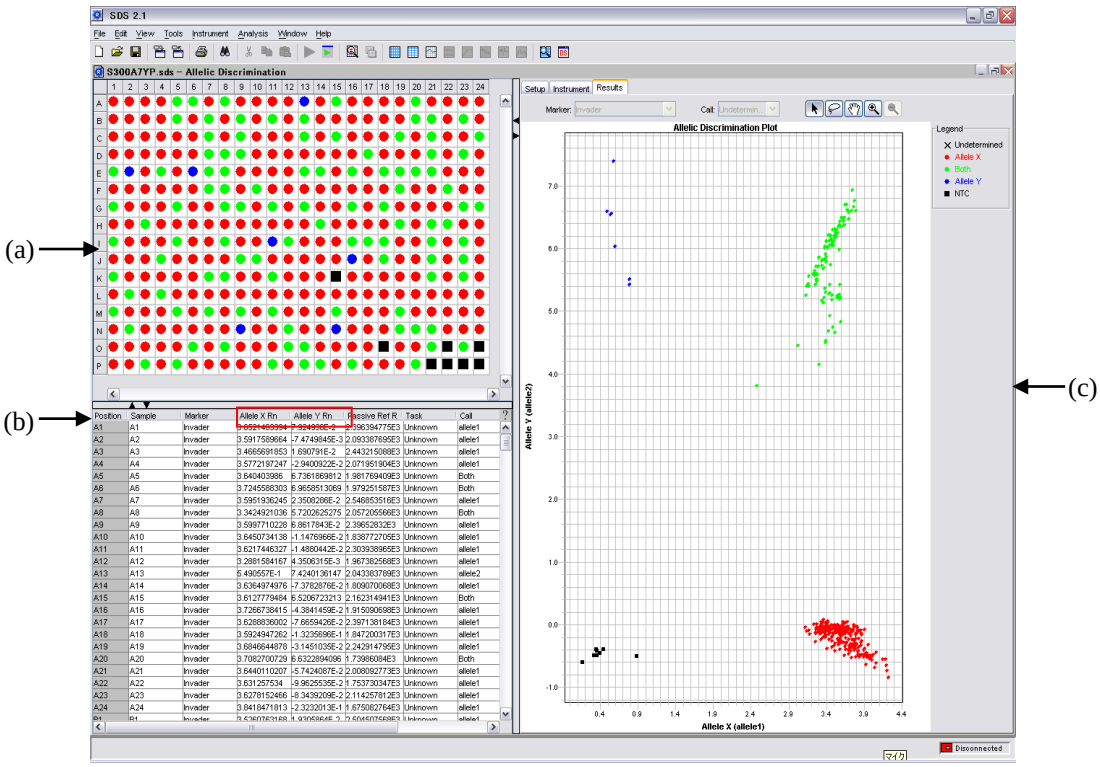


Figure 3.4: A representative graphical output of a SNP detected by the Sequence Detection System software of the ABI PRISM 7900.

(a) Results grid. The genotype calls were displayed as a 384-plate view. (b) Results table. Allelic call, well position, passive reference Rn, Rn value for allele X and Y were displayed in a table format. (c) Allelic discrimination plot. The genotype calls were assigned for each sample by Sequence Detection System software.

- Homozygous allele 1
- Homozygous allele 2
- Heterozygous
- NTC

3.5 INTERPRETATION OF THE DIRECT SEQUENCING RESULTS

The DNA fragments of interest were successfully sequenced by ABI3730*xl* sequencer. Sequence of each sample was analysed by Sequencher 4.6 (Gene Codes Corporation, USA) or PolyPhred program (Nickerson *et al.*, 1997) and compared to detect the presence of SNPs. If there is a SNP, the genotype of a sample can be determined by referring to the chromatogram. Figure 3.5 shows a representative graphical output of the chromatogram and nucleotide sequence data obtained after DNA sequencing.

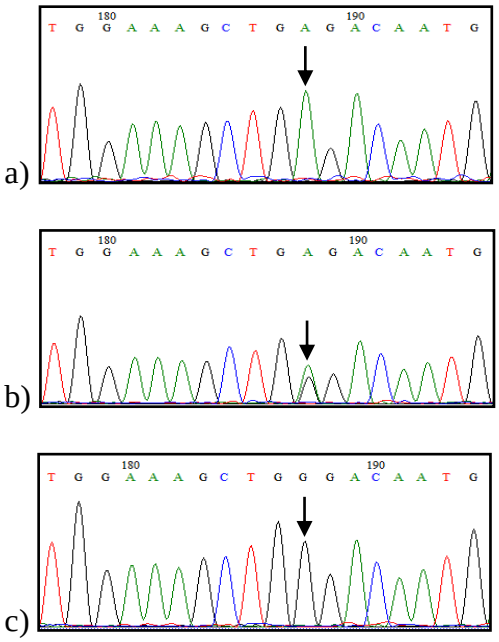


Figure 3.5: Chromatograms showing the sequencing results of SNP rs6059187.

The chromatograms show three different genotypes of rs6059187. The arrows pointing to the peaks indicated the SNP site. (a) Individuals with homozygous A allele (AA). (b) Individual with heterozygous alleles (AG). Heterozygous alleles in the sample are represented by a double peak. (c) Individual with homozygous G allele (GG).

3.6 ASSOCIATION STUDY BETWEEN *CYP1A1, CYP2A13, CYP2E1* POLYMORPHISMS AND NPC SUSCEPTIBILITY

3.6.1 *CYP1A1*

Potential susceptibility of *CYP1A1* was elucidated by examining two tag-SNPs (rs4646422 and m2 allele; rs1048943) through Invader genotyping assay in 81 NPC patients and 147 healthy controls. SNP rs4646422 is a nonsynonymous SNP which causes substitution of Glycine to Aspartic acid.

No SNPs showed deviation from HWE ($P \geq 0.01$) and both SNPs had a call rate of 100%. Similar MAF distribution was observed in the controls and cases (Table 3.2). Logistic regression analysis was used to evaluate the association between both SNPs and NPC susceptibility, adjusting for age and gender. However, case-control analysis of the data suggested that both SNPs might not involve in the pathogenesis of NPC in Malaysian Chinese ($P > 0.0.5$; Table 3.2).

3.6.2 *CYP2A13*

Five tag-SNPs of *CYP2A13* ($r^2 \geq 0.8$) were selected to be examined by Invader genotyping assay in 81 NPC patients and 147 healthy controls. These SNPs were successfully genotyped in all the samples, having a call rate of 100%. Deviation from HWE was observed for SNP rs7254343 of *CYP2A13* in the controls ($P < 0.01$). The distributions of MAF for these tag-SNPs in the case group do not differ significantly from the control group (Table 3.2). The association between *CYP2A13* polymorphisms and NPC risk was estimated with logistic regression analysis, adjusting for potential confounders such as age and gender. However, no association was found between the polymorphisms in *CYP2A13* and the risk of NPC ($P > 0.05$; Table 3.2).

Table 3.2: Association analysis of tag-SNPs of *CYP1A1* and *CYP2A13* to NPC susceptibility in 81 NPC patients and 147 controls

Gene	SNP ID	Location	Alleles[a] 1/2	Cases MAF	Controls MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
CYP1A1	rs1048943 (m2 allele)	Exon	T / C	0.25	0.22	0.93	0.61	0.41	1.03 [0.57-1.86]	1.21 [0.59-2.48]	0.50 [0.098-2.55]
(Chr. 15q24.1)	rs4646422	Exon	C / T	0.17	0.13	0.059	0.068	0.40	2.00 [0.98-4.09]	2.09 [0.95-4.63]	2.98 [0.23-38.44]
CYP2A13	rs1645690	Intron	A / G	0.06	0.08	0.99	0.99	1.00	1.00 [0.35-2.87]	1.00 [0.35-2.87]	NA
(Chr. 19q13.2)	rs7251532	Intergenic	A / G	0.30	0.37	0.28	0.25	0.55	0.76 [0.46-1.25]	0.66 [0.32-1.35]	0.74 [0.27-2.02]
	rs7255149	Intergenic	C / A	0.22	0.26	0.19	0.29	0.23	0.69 [0.40-1.20]	0.67 [0.32-1.41]	0.45 [0.12-1.65]
	rs12981316	Intergenic	C / G	0.19	0.20	0.43	0.51	0.49	0.79 [0.44-1.42]	0.77 [0.36-1.66]	0.60 [0.14-2.55]

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P-value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 81 NPC cases; 147 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele

3.6.3 *CYP2E1*

In the first screening, six tag-SNPs ($r^2 \geq 0.8$) of *CYP2E1* were examined by multiplex-PCR based Invader assay in 81 NPC patients and 147 healthy controls. All SNPs showed no deviation from HWE ($P \geq 0.01$) in controls and had a call rate of 100%. In the first screening, only SNPs rs2070672 and rs2249695 showed *P*-value < 0.05 after the adjustment of age and gender, indicating that these SNPs may be associated with risk of NPC (Table 3.3).

Both SNPs were further genotyped in the second set of samples, consisting of 366 NPC patients and 340 healthy controls, and the association between rs2070672 and NPC susceptibility was successfully replicated ($P = 0.016$; Table 3.4). The combined *P*-values of the two data sets revealed that rs2070672 remained significantly associated with the increased risk of NPC after the adjustment for age and gender ($P = 0.0011$; Table 3.4), and overcame the Bonferroni-corrected significance level of 6 tested SNPs ($P < 0.0083$).

SNP rs2070672 is located at -352 bp upstream of *CYP2E1*. Homozygous for the risk G allele (GG) was present in 7.9% of 447 NPC patients compared to 3.1% of the 487 healthy controls. The GG genotype was associated with an increased risk to NPC compared to GA and AA genotypes (OR = 3.57, 95% CI = 1.66-7.66 for recessive-inheritance model).

Table 3.3: Association analysis of tag-SNPs in *CYP2E1* to NPC susceptibility in 81 NPC patients and 147 controls

Gene	SNP ID	Location	Alleles[a] 1/2	Cases MAF	Controls MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2		Odds ratio [95%CI] vs 11		Odds ratio [95%CI] vs 22	
CYP2E1 (Chr. 10q24.3 -qter)	rs2031920	5' near gene	C / T	0.22	0.20	0.68	0.70	0.80	0.88	[0.49-1.59]	0.86	[0.42-1.8]	0.82	[0.18-3.66]
	rs2070672	5' near gene	A / G	0.28	0.19	0.0038[b]	0.0063[b]	0.11	2.77	[1.39-5.52]	2.88	[1.35-6.14]	6.10	[0.66-56.75]
	rs915908	Intron	G / A	0.18	0.14	0.38	0.30	0.62	1.38	[0.67-2.83]	1.51	[0.69-3.27]	0.35	[0.006-21.29]
	rs2249695	Intron	C / T	0.49	0.37	0.034[b]	0.14	0.045[b]	1.79	[1.05-3.05]	1.84	[0.83-4.11]	2.57	[1.02-6.44]
	rs2515644	3' near gene	C / A	0.49	0.39	0.067	0.13	0.16	1.66	[0.96-2.86]	1.89	[0.84-4.29]	1.96	[0.77-4.96]
	rs8192780	Intergenic	G / T	0.48	0.39	0.112	0.24	0.16	1.55	[0.90-2.65]	1.62	[0.72-3.61]	1.96	[0.77-4.96]

MAF, Minor allele frequency; CI, Confidence interval
P-value, odds ratio for minor allele and 95% CI were adjusted for age and gender
Sample set = 81 NPC cases; 147 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] $P < 0.05$

Table 3.4: Association analysis of rs2070672 and rs2239695 of *CYP2E1* to NPC susceptibility in 447 NPC patients and 487 controls

SNP ID	Location	Alleles[a] 1/2	Sample set	Cases[b] MAF	Controls[b] MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
rs2070672	5' near gene	A / G	1st	0.28	0.19	0.0038[c]	0.0063[c]	0.11	2.77 [1.39-5.52]	2.88 [1.35-6.14]	6.10 [0.66-56.75]
			2nd	0.24	0.21	0.14	0.52	0.016	1.27 [0.93-1.75]	1.14 [0.76-1.70]	2.75 [1.21-6.26]
			total	0.25	0.20	0.0062[c]	0.07	0.0011[c]	1.48 [1.12-1.96]	1.37 [0.97-1.93]	3.57 [1.66-7.66]
rs2249695	Intron	C / T	1st	0.49	0.37	0.034	0.14	0.045	1.79 [1.05-3.05]	1.84 [0.83-4.11]	2.57 [1.02-6.44]
			2nd	0.43	0.40	0.31	0.29	0.58	1.16 [0.88-1.53]	1.25 [0.83-1.89]	1.16 [0.70-1.92]
			total	0.44	0.39	0.068	0.17	0.10	1.25 [0.98-1.59]	1.28 [0.90-1.83]	1.44 [0.93-2.23]

MAF, Minor allele frequency; CI, Confidence interval
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set= 447 NPC cases; 487 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] 1st: cases, N = 81; controls, N = 147
2nd: cases, N = 366; controls, N = 340
1st+2nd: cases, N =447; controls, N = 487
[c] Statistically significant at the Bonferroni-corrected significance level, based on 6 independent tests ($P < 0.0083$)

3.6.3.1 Re-sequencing of *CYP2E1*

Fine mapping association was performed in the *CYP2E1* gene to search whether additional SNPs conferring susceptibility to NPC. We sequenced the *CYP2E1* gene, covering the 15.8 kb genomic region from 5' flanking to 3' flanking which covers the exons and introns. A total of 18 sets of PCR primer were designed for re-sequencing of the *CYP2E1* based on its gene sequence (GenBank Accession No. NC_000010.10) (Appendix C). These 18 fragments were obtained from sequencing of 48 healthy controls and SNP candidates were identified from the assembled sequences using the Polyphred program (Nickerson *et al.*, 1997). The SNP positions were further confirmed by referring to the Human Genome Reference Build 36.3.

In addition to the previously genotyped SNPs, we identified 49 SNPs with MAF $\geq$ 0.05. One of the intronic SNPs (SNP C1) was a novel, previously unreported SNP after cross-checking with the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). The positions of identified SNPs are summarized in Table 3.5 and the pairwise LD pattern between SNPs in *CYP2E1* is shown in Figure 3.6. We then examined the association of these 49 SNPs using 447 NPC patients and 487 healthy controls. All the SNPs had a call rate of more than 97%, and no deviation from HWE was observed except for rs11516897 and rs10857739 ($P < 0.01$). After adjustment for age and gender, four SNPs (rs3813865, rs3813870, rs8192772 and rs915906) which are in strong LD with rs2070672 ($r^2 \geq 0.85$) are significantly associated with NPC susceptibility (0.00081 – 0.0032; Table 3.6). Among these five associated SNPs, rs3813870 showed the strongest association with NPC susceptibility ($P = 0.00081$, OR = 3.69, 95% CI = 1.72-7.94 for recessive-inheritance model; Table 3.6).

The statistical power of this association study was then estimated using the module case–control for discrete traits of the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html; Purcell *et al.*, 2003). A total of 447

NPC patients and 487 healthy controls were investigated to detect the association between *CYP2E1* rs2070672 and NPC susceptibility. It is evident that for effect sizes of OR = 3.57 and MAF = 0.20, the power of this study was more than 80% to detect the association at a significance threshold of 0.0083.

Table 3.5: SNPs detected in *CYP2E1*. Chromosome positions were based on the Human Genome Reference Build 36.3.

SNP ID	Location	Chromosome position	Alleles[a] 1/2
rs3813865	5' near gene	Chr. 10: 135189234	G / C
rs3813866	5' near gene	Chr. 10: 135189324	T / A
rs8192766	5' near gene	Chr. 10: 135189375	T / G
rs3813867	5' near gene	Chr. 10: 135189595	G / C
rs2031921	5' near gene	Chr. 10: 135189863	T / C
rs3813870	5' near gene	Chr. 10: 135189960	A / G
rs2031922	5' near gene	Chr. 10: 135190083	T / C
rs2070673	5' near gene	Chr. 10: 135190557	T / A
rs943975	Intron	Chr. 10: 135192250	T / C
rs1536828	Intron	Chr. 10: 135192305	C / G
rs41299418	Intron	Chr. 10: 135193622	C / A
rs915906	Intron	Chr. 10: 135193728	T / C
rs8192772	Intron	Chr. 10: 135194701	T / C
rs2070674	Intron	Chr. 10: 135195330	C / T
rs8192773	Intron	Chr. 10: 135195964	T / G
rs2070675	Intron	Chr. 10: 135196686	C / T
rs915907	Intron	Chr. 10: 135196917	C / A
rs4646976	Intron	Chr. 10: 135197716	A / G
rs8192775	Intron	Chr. 10: 135198016	G / A
SNP C1	Intron	Chr. 10: 135198168	T / G
rs8192776	Intron	Chr. 10: 135198177	C / T
rs7092584	Intron	Chr. 10: 135198247	C / T
rs6413432	Intron	Chr. 10: 135198534	T / A
rs2864987	Intron	Chr. 10: 135198569	A / G
rs2864986	Intron	Chr. 10: 135198663	C / A
rs2864985	Intron	Chr. 10: 135198747	G / C
rs2864984	Intron	Chr. 10: 135198947	C / A
rs743534	Intron	Chr. 10: 135199216	A / C
rs743535	Intron	Chr. 10: 135199357	G / A
rs1410897	Intron	Chr. 10: 135199527	A / G
rs1329149	Intron	Chr. 10: 135199791	C / T
rs12761234	Intron	Chr. 10: 135199904	T / C
rs8192777	Intron	Chr. 10: 135200861	T / G
rs2070676	Intron	Chr. 10: 135201127	C / G
rs2070677	Intron	Chr. 10: 135201164	A / T
rs2515641	Exon	Chr. 10: 135201352	C / T
rs2515642	Intron	Chr. 10: 135202003	T / C
rs2480259	Intron	Chr. 10: 135202066	A / G

Table 3.5, continued

SNP ID	Location	Chromosome position	Alleles[a] 1/2
rs2480258	Intron	Chr. 10: 135202090	C / T
rs2249694	Intron	Chr. 10: 135202143	G / A
rs2480257	3' UTR	Chr. 10: 135202499	A / T
rs2480256	3' UTR	Chr. 10: 135202504	A / G
rs11516897	3' near gene	Chr. 10: 135202681	T / A
rs10857739	3' near gene	Chr. 10: 135202687	T / G
rs2480255	3' near gene	Chr. 10: 135202845	T / C
rs2480254	3' near gene	Chr. 10: 135202859	T / C
rs1952467	Intergenic	Chr. 10: 135203489	T / G
rs8192779	Intergenic	Chr. 10: 135204111	C / T
rs8192781	Intergenic	Chr. 10: 135204241	T / G

[a] Allele 1 = Major allele, Allele 2 = Minor allele

Table 3.6: Association analysis of fine mapped SNPs of *CYP2E1* to NPC susceptibility in 447 NPC patients and 487 controls

SNP ID	Cases	Controls	*P* value			Odds ratio [95%CI]			$r^{2\,b}$
	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2	vs 11	vs 22	
rs3813865	0.25	0.20	0.0037[a]	0.039	0.0017[a]	1.52 [1.15-2.01]	1.44 [1.02-2.03]	3.24 [1.55-6.74]	0.987
rs3813866	0.21	0.22	0.29	0.33	0.50	0.86 [0.64-1.14]	0.84 [0.59-1.19]	0.77 [0.36-1.65]	0.071
rs8192766	0.46	0.42	0.14	0.64	0.037	1.20 [0.94-1.53]	1.09 [0.76-1.57]	1.58 [1.03-2.43]	0.351
rs3813867	0.21	0.22	0.30	0.35	0.49	0.86 [0.64-1.15]	0.85 [0.60-1.20]	0.76 [0.34-1.67]	0.070
rs2031921	0.21	0.22	0.24	0.27	0.51	0.84 [0.63-1.13]	0.82 [0.58-1.16]	0.77 [0.35-1.69]	0.071
rs3813870	0.25	0.20	0.0056[a]	0.073	0.00081[a]	1.49 [1.12-1.98]	1.37 [0.97-1.94]	3.69 [1.72-7.94]	1.000
rs2031922	0.21	0.22	0.25	0.30	0.45	0.85 [0.63-1.13]	0.83 [0.59-1.18]	0.74 [0.34-1.61]	0.071
rs2070673	0.45	0.42	0.18	0.63	0.070	1.18 [0.93-1.5]	1.09 [0.76-1.57]	1.49 [0.97-2.29]	0.352
rs943975	0.21	0.22	0.19	0.23	0.39	0.83 [0.62-1.1]	0.81 [0.57-1.15]	0.71 [0.33-1.55]	0.065
rs1536828	0.44	0.41	0.16	0.50	0.083	1.19 [0.93-1.52]	1.13 [0.79-1.62]	1.48 [0.95-2.29]	0.298
rs41299418	0.09	0.08	0.33	0.28	0.97	1.23 [0.81-1.87]	1.29 [0.81-2.06]	0.97 [0.19-5.03]	0.306
rs915906	0.25	0.22	0.030	0.22	0.0032[a]	1.35 [1.03-1.78]	1.24 [0.88-1.74]	2.92 [1.43-5.94]	0.881
rs8192772	0.25	0.21	0.020	0.22	0.00082[a]	1.39 [1.05-1.84]	1.24 [0.88-1.74]	3.51 [1.68-7.31]	0.878
rs2070674	0.21	0.21	0.48	0.52	0.64	0.90 [0.67-1.21]	0.89 [0.63-1.26]	0.82 [0.37-1.85]	0.067
rs8192773	0.05	0.05	0.35	0.34	1.00	1.30 [0.75-2.27]	1.32 [0.75-2.31]	NA	0.009
rs2070675	0.46	0.42	0.21	0.65	0.092	1.17 [0.92-1.49]	1.09 [0.76-1.57]	1.45 [0.94-2.23]	0.328
rs915907	0.46	0.42	0.16	0.61	0.059	1.19 [0.93-1.52]	1.10 [0.76-1.58]	1.52 [0.99-2.33]	0.328
rs4646976	0.20	0.17	0.19	0.32	0.19	1.22 [0.91-1.64]	1.20 [0.84-1.71]	1.75 [0.76-4.03]	0.645
rs8192775	0.23	0.22	0.74	0.62	0.75	1.05 [0.78-1.42]	1.09 [0.78-1.54]	0.87 [0.35-2.13]	0.000
SNP C1	0.09	0.09	0.98	0.72	0.17	0.99 [0.66-1.50]	1.09 [0.69-1.71]	0.32 [0.062-1.61]	0.025

Table 3.6, continued

SNP ID	Cases	Controls	*P* value			Odds ratio [95%CI]			r^{2b}
	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2	vs 11	vs 22	
rs8192776	0.43	0.36	0.023	0.15	0.015	1.33 [1.04-1.70]	1.29 [0.91-1.84]	1.78 [1.12-2.84]	0.437
rs7092584	0.45	0.41	0.25	0.64	0.13	1.15 [0.91-1.47]	1.09 [0.76-1.57]	1.40 [0.91-2.16]	0.317
rs6413432	0.27	0.25	0.89	0.71	0.69	1.02 [0.78-1.34]	1.07 [0.76-1.50]	0.87 [0.44-1.73]	0.007
rs2864987	0.27	0.25	0.80	0.60	0.67	1.04 [0.79-1.36]	1.10 [0.78-1.54]	0.86 [0.43-1.72]	0.008
rs2864986	0.27	0.25	0.91	0.71	0.64	1.02 [0.77-1.34]	1.07 [0.76-1.50]	0.85 [0.43-1.70]	0.007
rs2864985	0.20	0.17	0.14	0.28	0.10	1.25 [0.93-1.69]	1.22 [0.85-1.73]	2.05 [0.87-4.82]	0.645
rs2864984	0.26	0.25	0.90	0.70	0.64	1.02 [0.77-1.34]	1.07 [0.76-1.50]	0.85 [0.43-1.70]	0.007
rs743534	0.20	0.18	0.14	0.28	0.11	1.25 [0.93-1.68]	1.22 [0.85-1.73]	1.97 [0.85-4.58]	0.634
rs743535	0.27	0.25	0.81	0.60	0.64	1.03 [0.79-1.36]	1.09 [0.78-1.54]	0.85 [0.43-1.70]	0.007
rs1410897	0.20	0.18	0.15	0.31	0.10	1.24 [0.92-1.67]	1.20 [0.84-1.71]	2.03 [0.86-4.79]	0.639
rs1329149	0.20	0.18	0.15	0.31	0.10	1.24 [0.92-1.67]	1.20 [0.84-1.71]	2.03 [0.86-4.79]	0.639
rs12761234	0.20	0.18	0.15	0.31	0.10	1.24 [0.92-1.67]	1.20 [0.84-1.71]	2.03 [0.86-4.79]	0.639
rs8192777	0.20	0.18	0.19	0.38	0.10	1.22 [0.91-1.64]	1.17 [0.82-1.67]	2.03 [0.86-4.79]	0.633
rs2070676	0.20	0.18	0.17	0.29	0.16	1.23 [0.92-1.66]	1.21 [0.85-1.73]	1.82 [0.79-4.18]	0.634
rs2070677	0.20	0.18	0.14	0.29	0.10	1.25 [0.93-1.69]	1.21 [0.85-1.73]	2.05 [0.87-4.82]	0.639
rs2515641	0.20	0.18	0.14	0.29	0.10	1.25 [0.93-1.69]	1.12 [0.78-1.61]	1.43 [0.94-2.19]	0.639
rs2515642	0.46	0.43	0.18	0.55	0.098	1.18 [0.93-1.50]	1.07 [0.74-1.54]	1.45 [0.95-2.22]	0.333
rs2480259	0.46	0.43	0.23	0.71	0.088	1.16 [0.91-1.48]	1.09 [0.75-1.57]	1.47 [0.96-2.25]	0.333
rs2480258	0.46	0.42	0.20	0.66	0.079	1.17 [0.92-1.50]	1.24 [0.87-1.77]	1.43 [0.92-2.21]	0.337
rs2249694	0.44	0.40	0.10	0.24	0.11	1.23 [0.96-1.56]	1.28 [0.90-1.83]	1.44 [0.93-2.23]	0.364

Table 3.6, continued

SNP ID	Cases	Controls	*P* value			Odds ratio [95%CI]			r^{2b}
	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2	vs 11	vs 22	
rs2480257	0.46	0.42	0.17	0.43	0.13	1.19 [0.93-1.51]	1.16 [0.80-1.67]	1.39 [0.90-2.15]	0.333
rs2480256	0.46	0.42	0.17	0.42	0.14	1.19 [0.93-1.51]	1.16 [0.80-1.67]	1.39 [0.90-2.13]	0.333
rs2480255	0.46	0.42	0.15	0.51	0.077	1.20 [0.94-1.53]	1.13 [0.78-1.63]	1.48 [0.959-2.27]	0.330
rs2480254	0.46	0.42	0.18	0.55	0.092	1.18 [0.93-1.51]	1.12 [0.78-1.61]	1.45 [0.94-2.24]	0.330
rs1952467	0.46	0.42	0.14	0.43	0.092	1.20 [0.94-1.53]	1.16 [0.81-1.67]	1.45 [0.94-2.24]	0.310
rs8192779	0.27	0.25	0.87	0.71	0.73	1.02 [0.78-1.35]	1.07 [0.76-1.50]	0.88 [0.44-1.79]	0.007
rs8192781	0.46	0.42	0.15	0.49	0.083	1.20 [0.94-1.53]	1.14 [0.79-1.64]	1.47 [0.95-2.27]	0.332

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set= 447 NPC cases; 487 controls
[a] Statistically significant at the Bonferroni-corrected significance level, based on 6 independent tests ($P < 0.0083$)
[b] Pairwise LD coefficients (r^2) between SNPs and rs2070672

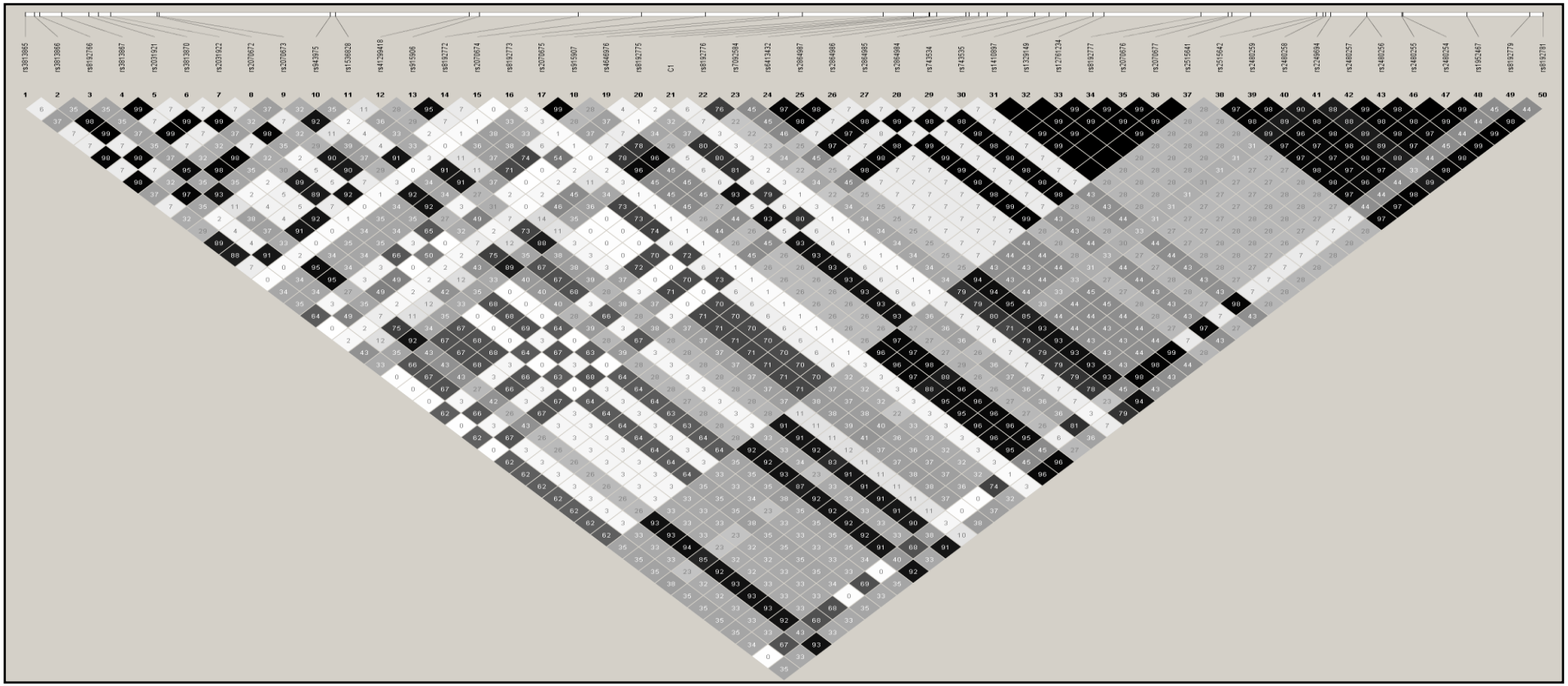


Figure 3.6: Pairwise LD pattern between SNPs in *CYP2E1*.

Value in each diamond represents the pairwise LD coefficients (r^2) calculated using Haploview. The darker the colour, the higher is the r^2 value.

3.7 ASSOCIATION STUDY BETWEEN *EPHX1* AND *EPHX2* POLYMORPHISMS AND NPC SUSCEPTIBILITY

In the first screening, seven tag-SNPs of *EPHX1* ($r^2 \geq 0.8$) and ten tag-SNPs of *EPHX2* ($r^2 \geq 0.8$) were examined by multiplex PCR based Invader assay in 81 NPC patients and 147 healthy controls. All SNPs had a call rate of 100%, and deviation from HWE was observed for SNP rs3753661 of *EPHX1* ($P < 0.01$). After adjusting for age and gender in the logistic regression analysis, only rs4149244 of *EPHX2* showed *P*-value < 0.05 (Table 3.7).

SNP rs4149244 was further genotyped in the second set of samples consisting of 366 NPC patients and 340 healthy controls. The association between rs4149244 and the decreased risk of NPC was successfully replicated ($P = 0.017$; Table 3.8). The combined *P*-values of the two data sets showed that rs4149244 of *EPHX2* remained significantly associated with decreased risk of NPC after the adjustment for age and gender ($P = 0.0013$; Table 3.8), and overcame the Bonferroni-corrected significance level of 17 tested SNPs ($P < 0.0029$). The prevalence of T allele was found to be lower in NPC patients (10%) compared to controls (15%) (OR = 0.53, 95% CI = 0.36-0.78 for allelic-inheritance model) suggesting a protective effect for this SNP.

The statistical power of this association study was then estimated using the module case–control for discrete traits of the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html; Purcell *et al.*, 2003). A total of 447 NPC patients and 487 healthy controls were investigated to detect the association between *EPHX2* rs4149244 and NPC susceptibility. The power of this study was 51% to detect the association at a significance threshold of 0.0029 for effect sizes of OR = 0.53 and MAF = 0.15.

Table 3.7: Association analysis tag-SNPs of *EPHX1* and *EPHX2* to NPC in 81 NPC patients and 147 controls

Gene	SNP ID	Location	Alleles[a]	Cases	Controls	*P* value			Odds ratio [95%CI]		
			1/2	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2	vs 11	vs 22
EPHX1 (Chr. 1q42.1)	rs1051740	Exon	T / C	0.57	0.47	0.12	0.15	0.25	0.66 [0.40-1.11]	0.56 [0.25-1.25]	0.61 [0.26-1.42]
	rs1877724	Intron	C / T	0.31	0.39	0.25	0.15	0.79	0.73 [0.44-1.24]	0.59 [0.29-1.21]	0.87 [0.30-2.49]
	rs2292566	Exon	G / A	0.24	0.32	0.086	0.056	0.68	0.59 [0.32-1.08]	0.49 [0.24-1.02]	0.73 [0.17-3.19]
	rs10915884	Intron	C / T	0.40	0.31	0.086	0.080	0.35	1.58 [0.94-2.65]	1.92 [0.93-4.00]	1.63 [0.58-4.58]
	rs3753658	Intron	G / T	0.15	0.18	0.67	0.44	0.45	0.86 [0.43-1.71]	0.73 [0.33-1.62]	2.10 [0.3-14.65]
	rs2671272	Intron	G / A	0.13	0.09	0.65	0.459	1.00	1.20 [0.55-2.60]	1.39 [0.58-3.31]	NA
EPHX2 (Chr. 8p21)	rs2741334	Intron	G / A	0.21	0.18	0.85	0.90	0.80	1.06 [0.57-2.00]	1.05 [0.50-2.21]	1.26 [0.21-7.57]
	rs721619	Intron	C / G	0.21	0.22	0.81	0.72	0.87	0.93 [0.50-1.72]	0.88 [0.42-1.81]	1.16 [0.2-6.56]
	rs751141	Exon	C / T	0.27	0.23	0.84	0.74	0.87	1.06 [0.60-1.88]	1.13 [0.55-2.32]	0.88 [0.21-3.72]
	rs4149244	Intron	C / T	0.08	0.16	0.035[b]	0.038[b]	1.00	0.39 [0.17-0.94]	0.39 [0.16-0.95]	NA
	rs4149252	Intron	G / T	0.14	0.19	0.17	0.25	1.00	0.60 [0.29-1.25]	0.63 [0.29-1.38]	NA
	rs13262930	Intron	C / G	0.34	0.38	0.34	0.15	0.86	0.78 [0.46-1.31]	0.58 [0.28-1.21]	1.09 [0.39-3.05]
	rs7816586	Intron	G / A	0.44	0.40	0.40	0.31	0.74	1.23 [0.76-1.99]	1.47 [0.70-3.10]	1.16 [0.48-2.76]
	rs1126452	Exon	A / C	0.37	0.40	0.66	0.44	0.87	0.89 [0.53-1.49]	0.75 [0.36-1.56]	1.09 [0.41-2.91]
	rs7357432	Intron	A / C	0.18	0.12	0.21	0.29	0.30	1.58 [0.77-3.23]	1.55 [0.69-3.47]	3.70 [0.32-43.28]
	rs7341557	3’ near gene	G / A	0.21	0.20	0.49	0.41	0.96	1.23 [0.68-2.23]	1.36 [0.65-2.85]	1.04 [0.21-5.07]

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 81 NPC cases; 147 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] $P < 0.05$

Table 3.8: Association analysis of rs4149244 of *EPHX2* to NPC in 447 NPC patients and 487 controls

SNP ID	Location	Alleles[a] 1 / 2	Sample set	Cases[b] MAF	Controls[b] MAF	*P*-value 1 vs 2	*P*-value vs 11	*P*-value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
rs4149244	Intron	C / T	1st	0.08	0.16	0.035	0.038	1.00	0.39 [0.17-0.94]	0.39 [0.16-0.95]	NA
			2nd	0.11	0.14	0.017	0.028	0.13	0.58 [0.37-0.91]	0.58 [0.36-0.94]	0.25 [0.042-1.53]
			total	0.10	0.15	0.0013[c]	0.0020[c]	0.12	0.53 [0.36-0.78]	0.52 [0.35-0.79]	0.27 [0.052-1.41]

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set= 447 NPC cases; 487 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] 1st: cases, N = 81; controls, N = 147
2nd: cases, N = 366; controls, N = 340
1st+2nd: cases, N =447; controls, N = 487
[c] Statistically significant at the Bonferroni-corrected significance level, based on 17 independent effective tests ($P < 0.0029$)

3.8 ASSOCIATION STUDY BETWEEN *GSTM1* NULL, *GSTT1* NULL GENOTYPES AND *GSTP1* POLYMORPHISMS AND NPC SUSCEPTIBILITY

The genotypes of *GSTM1* and *GSTT1* were examined in 274 NPC patients and 442 healthy controls by performing multiplex PCR. The association between *GSTM1* and *GSTT1* and NPC risk was evaluated by Fisher's exact test. The presence PCR products of 215 bp and 480 bp corresponded to the presence of *GSTM1* and *GSTT1* genes respectively, whereas no PCR amplification indicated the corresponding null genotype. Beta-globin amplification is included (expected size of 268 bp) as a positive internal control. Figure 3.7 shows a representative agarose gel of the presence of *GSTM1* and *GSTT1*.

The subjects with homozygous deletion of these genes were classified as null genotypes. No significant difference in the frequencies of *GSTM1* null and *GSTT1* null genotypes was observed between patients and controls. The *GSTM1* gene deletions were detected in 60.5% of NPC patients and 59.3% of healthy controls while about 48.5% of the NPC patients and 45.0% of healthy controls showed *GSTT1* null genotype (Table 3.9). To examine the possible joint effects of *GSTM1* and *GSTT1* null genotypes, the presence of *GSTM1* and *GSTT1* genotypes were classified as the reference group, and *P*-value, ORs and 95% CIs were calculated relative to the *GSTM1* carrier - *GSTT1* carrier. However, no significant risk was found to be associated with any genotype combinations of *GSTM1* and *GSTT1* genes. Our study revealed that 29.9% of NPC patients and 28.2% of healthy controls carry both null genotypes (Table 3.9).

To evaluate whether *GSTP1* is involved in NPC susceptibility, three SNPs were genotyped in 274 NPC patients and 442 healthy controls by Invader assay. No SNPs showed deviation from HWE ($P \geq 0.01$) and all the SNPs had a call rate of 99%. The association between *GSTP1* polymorphisms and NPC risk was evaluated by logistic

regression analysis using PLINK (v1.07) (Purcell, 2007) However, we found no significant association between *GSTP1* polymorphisms and the risk of NPC (Table 3.10).

We then evaluated the combined effect of *GSTP1* (rs1695; Ile105Val) and *GSTM1*, *GSTT1* which are involved in the same biological pathway towards NPC risk. The AG and GG genotypes of rs1695 (Ile/Val and Val/Val) are considered in the same category with both *GSTM1* and *GSTT1* null genotypes as the G allele is suggested to have lower catalytic activity for detoxification of carcinogens (Moyer *et al.*, 2008). The combined effect of these genes was evaluated by Fisher's exact test. The presence of *GSTM1*, *GSTT1* genotypes and AA genotype of rs1695 (Ile/Ile) were classified as the reference group, and *P*-value, ORs and 95% CIs were calculated relative to *GSTM1* carrier -*GSTT1* carrier - rs1695 A/A. However, we failed to detect any association to NPC susceptibility in combinationatorial studies of two (Table 3.11) or three genes (Table 3.12).

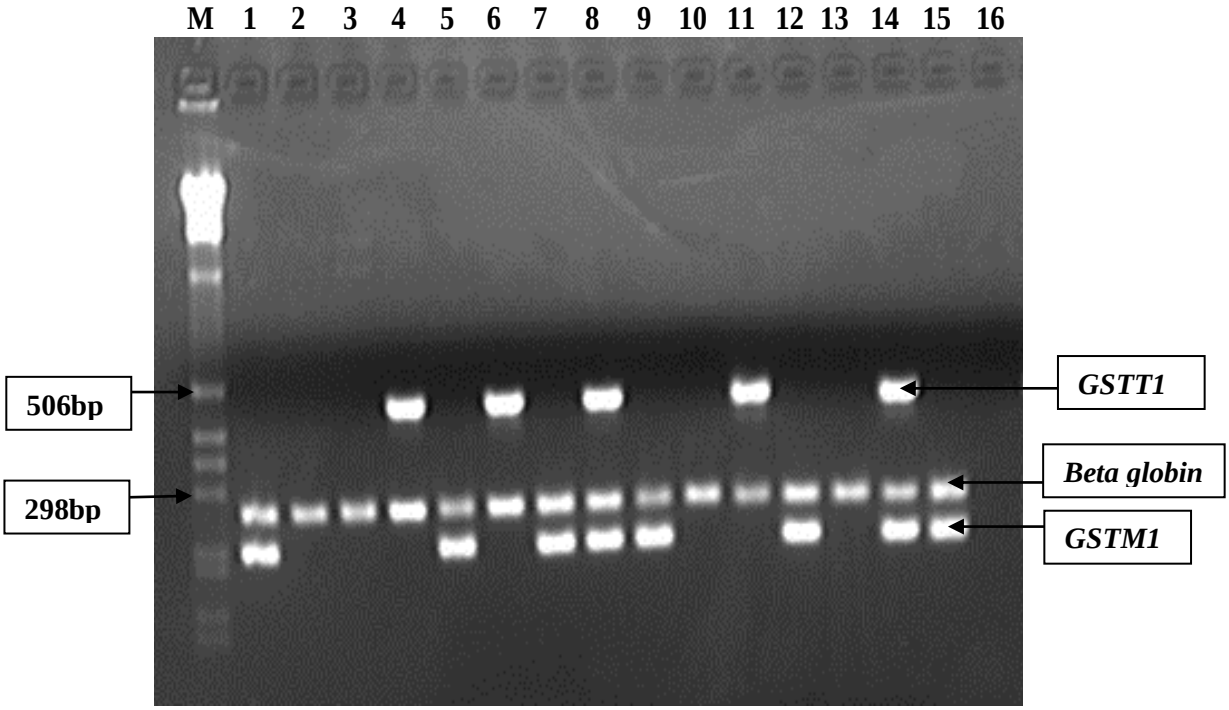


Figure 3.7: Genotyping of *GSTM1* and *GSTT1*.

Genotypes of *GSTM1* and *GSTT1* were examined by performing multiplex PCR protocol. Agarose gel electrophoresis illustrated the amplification of *GSTM1*, *GSTT1* and beta-globin.

M : 1kb DNA ladder (Invitrogen, Carlsbad, CA)

Lane 2, 3, 10 &13 : Individual with deletions of both genes (*GSTM1* and *GSTT1*)

Lane 1, 5, 7, 9, 12 &15 : Individual with only *GSTM1* gene

Lanes 4, 6 &11 : Individual with only *GSTT1* gene

Lanes 8 & 14 : Individual with both genes (*GSTM1* and *GSTT1*)

Lane 16 : Non-template negative control

Table 3.9: Association analysis of *GSTM1* and *GSTT1* genotypes to NPC susceptibility in 274 NPC patients and 442 healthy controls

Gene	Genotype	Cases, n (%)	Controls, n (%)	*P* value	Odds ratio [95% CI]
GSTM1	Present	108 (39.4%)	180 (40.7%)		Reference genotype
(Chr. 1p13.3)	null	166 (60.5%)	262 (59.3%)	0.75	1.06 [0.78-1.44]
GSTT1	Present	141 (51.5%)	243 (55.0%)		Reference genotype
(Chr. 22q11.23)	null	133 (48.5%)	199 (45.0%)	0.40	1.15 [0.85-1.56]
GSTM1 and *GSTT1*	*GSTM1* and *GSTT1* present	57 (20.8%)	106 (23.9%)		Reference genotype
(Chr. 1p13.3/22q11.23)	*GSTM1* deleted, *GSTT1* present	84 (30.7%)	137 (31.0%)	0.59	1.14 [0.75-1.74]
	GSTT1 deleted, *GSTM1* present	51 (18.6%)	74 (16.74%)	0.33	1.28 [0.79-2.07]
	Both *GSTM1* and *GSTT1* deleted	82 (29.9%)	125 (28.2%)	0.39	1.22 [0.80-1.87]

Sample set = 274 NPC cases; 442 Controls

Table 3.10: Association analysis of tag-SNPs of *GSTP1* to NPC susceptibility in 274 NPC patients and 442 controls

Gene	SNP ID	Location	Alleles[a] 1/2	Cases MAF	Controls MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
GSTP1 (Chr. 11q13)	rs1695	Exon	A / G	0.17	0.18	0.77	0.95	0.44	1.06 [0.74-1.52]	0.91 [0.55-1.51]	1.65 [0.47-5.75]
	rs762803	Intron	C / A	0.14	0.14	0.50	0.60	0.46	1.14 [0.76-1.68]	1.09 [0.64-1.87]	1.61 [0.38-6.85]
	rs4147581	Intron	G / C	0.29	0.29	0.24	0.29	0.42	1.20 [0.88-1.62]	1.20 [0.75-1.92]	1.38 [0.63-3.01]

MAF, Minor allele frequency; CI, Confidence interval
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 274 NPC cases; 442 Controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele

Table 3.11: Association of *GSTM1* or *GSTT1* in combination with *GSTP1* genotypes to NPC susceptibility

GSTs genotype		Cases, n	Controls, n	*P* value	Odds ratio [95%CI]
GSTT1	***GSTP1* rs1695**				
Present	AA	105	161		Reference genotype
Present	AG+GG	35	82	0.084	0.65 [0.41-1.04]
Null	AA	84	132	0.93	0.98 [0.68-1.41]
Null	AG+GG	47	66	0.73	1.09 [0.70-1.71]
GSTM1	***GSTP1* rs1695**				
Present	AA	72	115		Reference genotype
Present	AG+GG	34	64	0.61	0.85 [0.51-1.41]
Null	AA	116	178	0.85	1.04 [0.71-1.52]
Null	AG+GG	48	84	0.73	0.91 [0.58-1.45]

Sample set = 274 NPC cases; 442 Controls

Table 3.12: Association of *GSTM1*, *GSTT1* and *GSTP1* triple combination genotypes to NPC susceptibility

GSTs genotype			Cases, n	Controls, n	*P* value	Odds ratio [95%CI]	
GTSM1	***GSTT1***	***GSTP1* rs1695**					
Present	present	AA	42	72		Reference genotype	
Present	present	AG+GG	15	34	0.48	1.32	[0.65-2.71]
Present	null	AA	63	89	0.53	1.21	[0.74-2.00]
Present	null	AG+GG	20	48	0.34	1.40	[0.73-2.67]
Null	present	AA	30	43	0.64	1.20	[0.66-2.18]
Null	present	AG+GG	19	30	0.86	1.09	[0.55-2.16]
Null	null	AA	53	89	1.00	0.98	[0.59-1.63]
Null	null	AG+GG	28	36	0.42	1.33	[0.71-2.49]

Sample set = 274 NPC cases; 442 Controls

3.9 ASSOCIATION STUDY BETWEEN *PIGR* POLYMORPHISMS AND NPC SUSCEPTIBILITY

The genotypes of seven *PIGR* tag-SNPs (rs291085, rs172361, rs291102, rs2275531, rs291096, rs3813956 and rs2000057) were investigated in 81 NPC patients and 147 healthy controls. No deviation from HWE ($P \geq 0.01$) was observed in the control subjects and all of the SNPs had a call rate of 100%.

However, distributions of MAF did not differ significantly among NPC patients and controls. None of the *PIGR* SNPs were significantly associated with NPC susceptibility with all SNPs showing $P > 0.05$ (Table 3.13).

Table 3.13: Association analysis of tag-SNPs of *PIGR* to NPC susceptibility in 81 NPC patients and 147 controls

Gene	SNP ID	Position	Alleles[a]	Cases	Controls	*P* value			Odds ratio [95%CI]		
			1 / 2	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2	vs 11	vs 22
PIGR (Chr. 1q31-q41)	rs291085	Intergenic	A / T	0.40	0.39	0.74	0.87	0.67	1.09 [0.66-1.79]	1.06 [0.51-2.21]	1.22 [0.49-3.05]
	rs172361	Exon	T / C	0.13	0.08	0.61	0.62	0.88	1.25 [0.52-3.00]	1.27 [0.50-3.18]	1.42 [0.02-121.1]
	rs291102	Exon	C / T	0.13	0.08	0.61	0.62	0.88	1.25 [0.52-3.00]	1.27 [0.50-3.18]	1.42 [0.02-121.1]
	rs2275531	Exon	G / A	0.31	0.35	0.77	0.87	0.73	1.08 [0.63-1.87]	1.07 [0.52-2.20]	1.22 [0.39-3.85]
	rs291096	Exon	A / G	0.44	0.42	0.52	0.26	0.84	1.18 [0.71-1.95]	1.55 [0.72-3.35]	0.91 [0.36-2.29]
	rs3813956	Exon	G / A	0.06	0.07	0.81	0.81	1.00	1.14 [0.40-3.23]	1.14 [0.40-3.23]	NA
	rs2000057	Intron	G / A	0.31	0.35	0.81	0.91	0.73	1.07 [0.62-1.84]	1.04 [0.50-2.15]	1.22 [0.39-3.85]

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 81 NPC cases; 147 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele

3.10 ASSOCIATION STUDY BETWEEN *LPLUNC1 AND SPLUNC1* POLYMORPHISMS AND NPC SUSCEPTIBILITY

In the first screening, total eight haplotype-tagging (tag SNPs) ($r^2 \geq 0.8$), four each of *LPLUNC1* and *SPLUNC1* were genotyped in 81 NPC patients and 147 healthy controls. These SNPs were analysed in the logistic regression analysis which included age and gender as covariates using PLINK (Purcell, 2007). Only one promoter SNP of *SPLUNC1* (rs2752903) showed significant *P*-value after the adjustment for gender and age, indicating that there is a possible association of this SNP with the susceptibility to NPC ($P < 0.05$; Table 3.14).

Next, this SNP was further genotyped using the second set of samples, consisting of 366 NPC patients and 340 healthy controls, and the association was successfully replicated ($P = 0.020$; Table 3.15).The combined *P*-values of both data sets revealed that this SNP (rs2752903) remained significantly associated with NPC susceptibility, after the adjustment for age and gender ($P = 0.00097$; Table 3.15), and overcame the Bonferroni-corrected significance level of 8 tested SNPs ($P < 0.00625$). The frequency of individuals with at least one risk C allele (CC, TC) in NPC patients (47.6%) was higher compared to controls (35.9%) (OR = 1.79, 95% CI = 1.27-2.52, dominant effect).

Table 3.14: Association analysis of tag-SNPs in *LPLUNC1 and SPLUNC1* to NPC susceptibility in 81 NPC patients and 147 controls

Gene	SNP ID	Location	Alleles[a]	Cases	Controls	*P* value			Odds ratio [95%CI]					
			½	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2		vs 11		vs 22	
LPLUNC1 (Chr. 20q11.21)	rs1884882	Intergenic	T / C	0.19	0.26	0.13	0.11	0.61	0.63	[0.35-1.15]	0.54	[0.26-1.14]	0.68	[0.16-2.93]
	rs4911314	Intron	C / G	0.20	0.24	0.22	0.13	0.86	0.68	[0.36-1.26]	0.57	[0.28-1.18]	1.16	[0.22-6.27]
	rs1999663	Nonsynonymous coding	C / G	0.25	0.26	0.60	0.82	0.39	1.17	[0.65-2.11]	1.09	[0.53-2.23]	1.85	[0.46-7.47]
	rs6087476	Intergenic	C / T	0.40	0.38	0.65	0.90	0.31	1.12	[0.68-1.85]	0.96	[0.46-1.99]	1.62	[0.64-4.06]
SPLUNC1 (Chr. 20q11.2)	rs927159	Intergenic	C / T	0.15	0.24	0.17	0.18	0.49	0.65	[0.35-1.21]	0.60	[0.28-1.26]	0.56	[0.10-2.97]
	rs2752903	5' near gene	T / C	0.31	0.16	0.022[b]	0.037[b]	0.13	2.06	[1.11-3.83]	2.18	[1.05-4.56]	3.85	[0.67-22.17]
	rs750064	5' near gene	T / C	0.47	0.40	0.51	0.68	0.49	1.19	[0.71-2.00]	1.18	[0.55-2.54]	1.38	[0.55-3.45]
	rs1570034	Intergenic	C / T	0.47	0.41	0.66	0.66	0.79	1.12	[0.67-1.89]	1.19	[0.55-2.58]	1.13	[0.45-2.83]

MAF, Minor allele frequency; CI, Confidence interval
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 81 NPC cases; 147 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] $P < 0.05$

Table 3.15: Association analysis of tag SNP rs2752903 of *SPLUNC1* to NPC susceptibility in 447 NPC patients and 487 controls

SNP ID	Location	Alleles[a] 1 / 2	Sample set	Cases[b] MAF	Controls[b] MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
rs2752903	5' near gene	T / C	1st	0.31	0.16	0.022	0.037	0.13	2.06 [1.11-3.83]	2.18 [1.05-4.56]	3.85 [0.67-22.17]
			2nd	0.27	0.23	0.082	0.020	0.78	1.33 [0.96-1.82]	1.61 [1.08-2.41]	0.90 [0.41-1.96]
			Total	0.28	0.21	0.0036[c]	0.00097[c]	0.52	1.51 [1.14-1.99]	1.79 [1.27-2.52]	1.26 [0.62-2.54]

MAF, Minor allele frequency; CI, Confidence interval
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set= 447 NPC cases; 487 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] 1st: cases, N = 81; controls, N = 147
2nd: cases, N = 366; controls, N = 340
1st+2nd: cases, N =447; controls, N = 487
[c] Statistically significant at the Bonferroni-corrected significance level, based on 8 independent effective tests ($P < 0.0063$)

3.10.1 Re-sequencing of *SPLUNC1*

Fine mapping association through DNA sequencing was performed on *SPLUNC1* to search for additional SNPs potentially associated to NPC susceptibility. A 17-kb LD block including the entire genomic region of *SPLUNC1* was re-sequenced. A total of 19 sets of PCR primer were designed for SNP screening according to the *SPLUNC1* gene sequence (GenBank Accession No. NC_000020.10) (Appendix D). These 19 fragments were amplified from 48 healthy controls and SNP candidates were identified from the assembled sequences using the Polyphred program. The SNP positions were further confirmed by referring to the Human Genome Reference Build 36.3. The identified SNPs were then genotyped in all 447 cases and 487 controls.

In addition to the tag-SNPs genotyped previously, we identified a total of 40 SNPs including one novel SNP (SNP01); however, no SNP was found within the exon region of *SPLUNC1*. The position of identified SNPs is summarized in Table 3.16. A LD block which summarized the pattern of LD for *SPLUNC1* polymorphisms in Malaysian Chinese is shown in Figure 3.8. Most of the SNPs are in high LD with rs2752903 ($r^2 \geq 0.85$). Genotyping of these 40 SNPs in all cases and controls revealed that 36 SNPs, in high LD with rs2752903, are significantly associated with NPC (P = 0.00016-0.0018; Table 3.17). All the SNPs had a call rate of more than 96%, and no deviation from HWE ($P \geq 0.01$) was observed.

The statistical power of this association study was then estimated using the module case–control for discrete traits of the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html; Purcell *et al.*, 2003). A total of 447 NPC patients and 487 healthy controls were investigated to detect the association between *SPLUNC1* ea2752903 and NPC susceptibility. It is evident that for effect sizes of OR = 1.79 and MAF = 0.21, the power of this study was more than 80% to detect the association at a significance threshold of 0.0063.

Table 3.16: SNPs detected in *SPLUNC1*. Chromosome positions were based on the Human Genome Reference Build 36.3

SNP ID	Location	Chromosome position	Alleles [a] (1 /2)
rs1321417	Intergenic	Chr. 20: 31279634	C / T
rs6059169	Intergenic	Chr. 20: 31281265	A / C
rs6059171	Intergenic	Chr. 20: 31282365	G / A
rs6057775	Intergenic	Chr. 20: 31283289	T / G
rs6057776	Intergenic	Chr. 20: 31283311	A / G
SNP S1	Intergenic	Chr. 20: 31283777	C / T
rs6057777	Intergenic	Chr. 20: 31283814	A / C
rs6059173	Intergenic	Chr. 20: 31285204	G / A
rs6057778	Intergenic	Chr. 20: 31285431	G / C
rs1998149	5' near gene	Chr. 20: 31285831	C / A
rs6059177	5' near gene	Chr. 20: 31286081	A / C
rs6059178	5' near gene	Chr. 20: 31286208	G / C
rs750065	5' near gene	Chr. 20: 31286548	T / A
rs2752901	5' near gene	Chr. 20: 31286660	G / A
rs2752902	5' near gene	Chr. 20: 31286685	T / C
rs6141897	Intronic	Chr. 20: 31288326	T / G
rs3787144	Intronic	Chr. 20: 31288430	A / G
rs3787145	Intronic	Chr. 20: 31288513	C / T
rs6059181	Intronic	Chr. 20: 31289034	A / G
rs1407019	Intronic	Chr. 20: 31289458	G / A
rs6059183	Intronic	Chr. 20: 31289688	T / C
rs6141898	Intronic	Chr. 20: 31289791	G / T
rs6059184	Intronic	Chr. 20: 31290126	C / T
rs6059185	Intronic	Chr. 20: 31290393	T / C
rs3746392	Intronic	Chr. 20: 31290991	G / A
rs6141899	Intronic	Chr. 20: 31291100	A / G
rs6059187	Intronic	Chr. 20: 31291926	A / G
rs2273529	Intronic	Chr. 20: 31292744	G / A
rs6059186	Intronic	Chr. 20: 31292915	T / C
rs12480799	Intronic	Chr. 20: 31294373	T / C
rs10875488	Intronic	Chr. 20: 31294434	A / G
rs1047595	3' UTR	Chr. 20: 31294616	T / C
rs2295575	3' UTR	Chr. 20: 31294769	C / A
rs6120186	3' near gene	Chr. 20: 31295406	A / G

Table 3.16, continued

SNP ID	Location	Chromosome position	Alleles [a] (1 / 2)
rs6141900	Intergenic	Chr. 20: 31301024	C / A
rs6059197	Intergenic	Chr. 20: 31302181	A / G
rs911137	Intergenic	Chr. 20: 31302481	G / A
rs6141901	Intergenic	Chr. 20: 31302719	G / T
rs6057786	Intergenic	Chr. 20: 31304113	A / G
rs6059214	Intergenic	Chr. 20: 31312132	A / C

[a] Allele 1 = Major allele, Allele 2 = Minor allele

Table 3.17: Association analysis of fine mapped SNPs of *SPLUNC1*

SNP ID	Cases	Controls	*P* value			Odds ratio [95%CI]						r^{2b}
	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2		vs 11		vs 22		
rs1321417	0.26	0.20	0.0015[a]	0.00062[a]	0.35	1.60	[1.20-2.14]	1.85	[1.30-2.64]	1.45	[0.68-3.15]	0.933
rs6059169	0.28	0.21	0.0021[a]	0.00065[a]	0.39	1.55	[1.17-2.04]	1.84	[1.30-2.61]	1.36	[0.68-2.73]	0.994
rs6059171	0.29	0.21	0.0023[a]	0.00072[a]	0.38	1.54	[1.17-2.04]	1.83	[1.29-2.59]	1.37	[0.68-2.74]	0.981
rs6057775	0.29	0.21	0.0019[a]	0.00057[a]	0.38	1.55	[1.18-2.05]	1.85	[1.30-2.62]	1.37	[0.68-2.73]	1.000
rs6057776	0.28	0.21	0.0052[a]	0.0015[a]	0.53	1.49	[1.13-1.98]	1.77	[1.25-2.52]	1.26	[0.62-2.53]	1.000
SNP01	0.28	0.21	0.0027[a]	0.00087[a]	0.39	1.53	[1.16-2.02]	1.81	[1.28-2.57]	1.35	[0.68-2.71]	1.000
rs6057777	0.45	0.43	0.63	0.64	0.75	1.06	[0.84-1.35]	1.09	[0.76-1.56]	1.07	[0.70-1.65]	0.351
rs6059173	0.29	0.21	0.0015[a]	0.00043[a]	0.37	1.57	[1.19-2.07]	1.87	[1.32-2.66]	1.38	[0.69-2.74]	1.000
rs6057778	0.28	0.21	0.0030[a]	0.00090[a]	0.45	1.53	[1.16-2.02]	1.81	[1.27-2.56]	1.31	[0.65-2.66]	1.000
rs1998149	0.28	0.21	0.0026[a]	0.00090[a]	0.38	1.53	[1.16-2.02]	1.81	[1.27-2.56]	1.37	[0.68-2.74]	1.000
rs6059177	0.28	0.21	0.0016[a]	0.00038[a]	0.44	1.57	[1.19-2.08]	1.89	[1.33-2.68]	1.32	[0.65-2.67]	1.000
rs6059178	0.28	0.21	0.0044[a]	0.0011[a]	0.55	1.50	[1.13-1.98]	1.79	[1.26-2.54]	1.24	[0.61-2.50]	0.988
rs750065	0.28	0.21	0.0030[a]	0.00092[a]	0.45	1.52	[1.15-2.01]	1.80	[1.27-2.56]	1.31	[0.65-2.65]	1.000
rs2752901	0.28	0.21	0.0020[a]	0.00056[a]	0.42	1.55	[1.18-2.05]	1.85	[1.31-2.63]	1.34	[0.66-2.70]	0.988
rs2752902	0.28	0.21	0.0010[a]	0.00064[a]	0.17	1.60	[1.21-2.12]	1.84	[1.30-2.61]	1.64	[0.80-3.33]	0.988
rs6141897	0.28	0.21	0.0030[a]	0.0010[a]	0.39	1.52	[1.15-2.01]	1.79	[1.27-2.54]	1.36	[0.68-2.72]	1.000
rs3787144	0.45	0.43	0.61	0.58	0.81	1.07	[0.84-1.36]	1.11	[0.77-1.59]	1.06	[0.68-1.63]	0.354
rs3787145	0.29	0.21	0.0022[a]	0.00056[a]	0.46	1.55	[1.17-2.04]	1.85	[1.30-2.62]	1.30	[0.64-2.62]	0.994
rs6059181	0.28	0.21	0.0026[a]	0.00090[a]	0.37	1.53	[1.16-2.03]	1.81	[1.27-2.56]	1.38	[0.68-2.76]	0.994

Table 3.17, continued

SNP ID	Cases	Controls	*P* value			Odds ratio [95%CI]						r^{2b}
	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2		vs 11		vs 22		
rs1407019	0.29	0.21	0.0029[a]	0.00077[a]	0.51	1.53	[1.16-2.03]	1.82	[1.29-2.59]	1.27	[0.62-2.59]	0.994
rs6059183	0.29	0.21	0.0020[a]	0.00063[a]	0.37	1.55	[1.17-2.05]	1.84	[1.30-2.60]	1.37	[0.69-2.75]	1.000
rs6141898	0.28	0.21	0.0025[a]	0.00090[a]	0.38	1.54	[1.17-2.04]	1.81	[1.27-2.56]	1.37	[0.67-2.81]	0.994
rs6059184	0.28	0.21	0.0024[a]	0.00059[a]	0.52	1.54	[1.17-2.04]	1.84	[1.30-2.61]	1.27	[0.62-2.57]	0.994
rs6059185	0.28	0.21	0.0020[a]	0.00050[a]	0.46	1.55	[1.18-2.06]	1.86	[1.31-2.64]	1.31	[0.65-2.64]	1.000
rs3746392	0.29	0.21	0.0020[a]	0.00044[a]	0.52	1.56	[1.18-2.06]	1.87	[1.32-2.65]	1.26	[0.62-2.57]	0.994
rs6141899	0.28	0.21	0.0041[a]	0.0012[a]	0.53	1.50	[1.14-1.98]	1.77	[1.26-2.50]	1.25	[0.62-2.53]	0.994
rs6059187	0.44	0.43	0.64	0.51	0.96	1.06	[0.83-1.34]	1.13	[0.79-1.61]	1.01	[0.66-1.56]	0.352
rs2273529	0.16	0.22	0.012	0.029	0.05	0.68	[0.50-0.92]	0.67	[0.47-0.96]	0.42	[0.18-1.00]	0.075
rs6059186	0.28	0.21	0.0017[a]	0.00049[a]	0.38	1.56	[1.18-2.06]	1.86	[1.31-2.64]	1.36	[0.68-2.73]	1.000
rs12480799	0.29	0.21	0.0017[a]	0.00049[a]	0.37	1.56	[1.18-2.07]	1.86	[1.31-2.64]	1.37	[0.68-2.75]	0.988
rs10875488	0.28	0.21	0.0040[a]	0.0012[a]	0.51	1.51	[1.14-1.99]	1.78	[1.26-2.52]	1.27	[0.62-2.58]	0.988
rs1047595	0.29	0.21	0.0022[a]	0.00069[a]	0.39	1.54	[1.17-2.04]	1.83	[1.29-2.59]	1.36	[0.67-2.72]	0.988
rs2295575	0.28	0.22	0.0064	0.0016[a]	0.66	1.47	[1.11-1.93]	1.74	[1.24-2.46]	1.17	[0.58-2.35]	0.958
rs6120186	0.28	0.21	0.0034[a]	0.00073[a]	0.61	1.52	[1.15-2.01]	1.83	[1.29-2.59]	1.21	[0.59-2.46]	0.994
rs6141900	0.28	0.21	0.0023[a]	0.0013[a]	0.23	1.54	[1.17-2.04]	1.78	[1.25-2.52]	1.53	[0.76-3.07]	0.969
rs6059197	0.28	0.21	0.0014[a]	0.00057[a]	0.27	1.58	[1.19-2.09]	1.85	[1.30-2.62]	1.48	[0.74-2.98]	0.981
rs911137	0.28	0.21	0.0032[a]	0.0018[a]	0.26	1.52	[1.15-2.01]	1.74	[1.23-2.46]	1.50	[0.74-3.03]	0.969

Table 3.17, continued

SNP ID	Cases	Controls	*P* value			Odds ratio [95%CI]						r^{2b}
	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2		vs 11		vs 22		
rs6141901	0.29	0.21	0.0021[a]	0.00095[a]	0.27	1.55	[1.17-2.04]	1.80	[1.27-2.55]	1.48	[0.74-2.94]	0.982
rs6057786	0.28	0.21	0.0022[a]	0.0012[a]	0.22	1.55	[1.17-2.04]	1.78	[1.25-2.52]	1.54	[0.77-3.09]	0.975
rs6059214	0.30	0.22	0.00073[a]	0.00016[a]	0.35	1.60	[1.22-2.11]	1.95	[1.38-2.75]	1.37	[0.70-2.69]	0.853

MAF, Minor allele frequency; CI, Confidence interval
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 447 NPC cases; 487 controls
[a] Statistically significant at the Bonferroni-corrected significance level, based on 8 independent effective tests ($P < 0.0063$)
[b] Pairwise LD coefficients (r^2) between SNPs and rs2752903

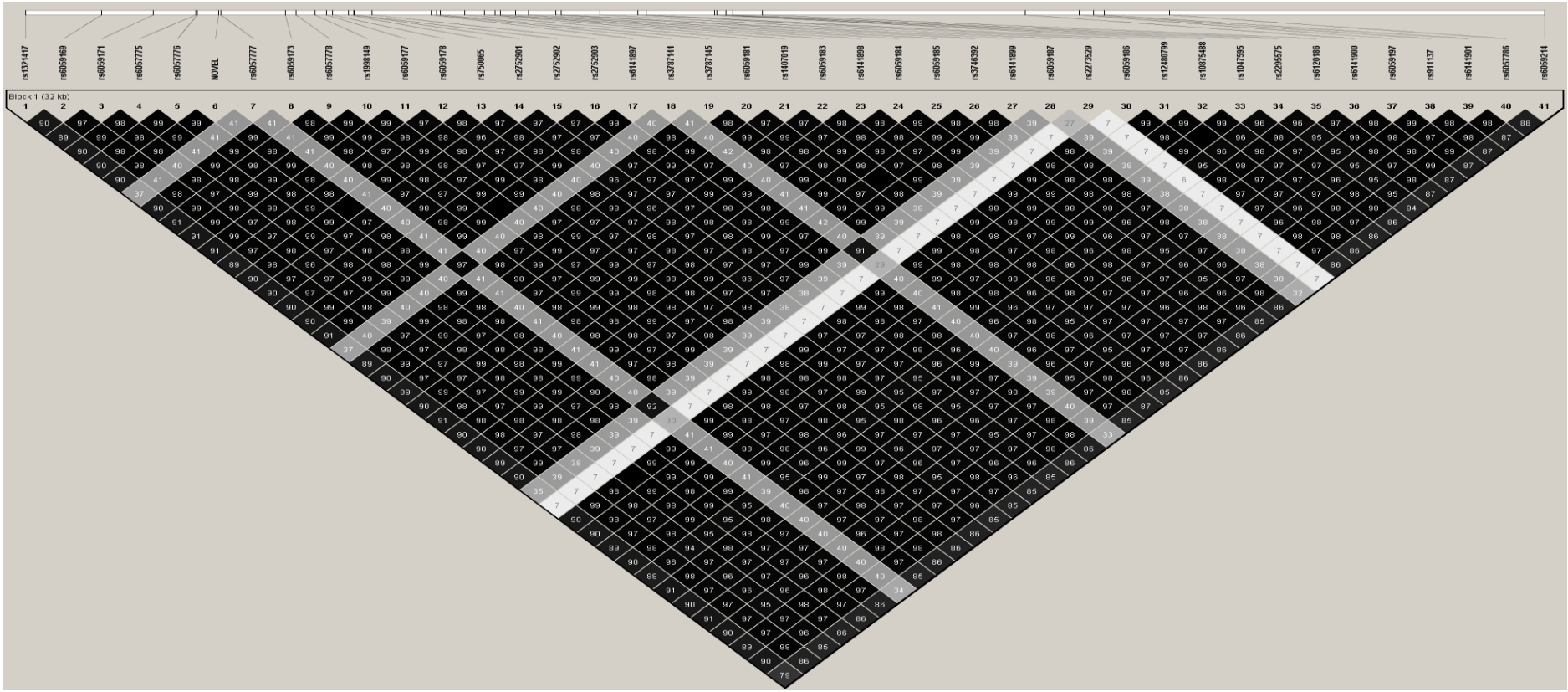


Figure 3.8: Pairwise LD pattern between SNPs in *SPLUNC1*.

Value in each diamond represents the pairwise LD coefficients (r^2) calculated using Haploview. The darker the colour, the higher is the r^2 value.

3.10.2 Screening of functional SNPs of *SPLUNC1*

a) Expression of *SPLUNC1* mRNA

The *SPLUNC1* mRNA expression was examined in 14 cell lines including four NPC cell lines (CNE1, HONE1, SUNE1 and HONE Akata), HeLa, BT549, MCF7, ZR-75-1, NCI-H520, NCI-H1781, NCI-H2170, A549, HepG2 and HuH-7. Low or no expression of SPLUNC1 was detected in most of the cell lines except for human bronchoalveolar carcinoma-derived NCI-H1781 (Figure 3.9). Therefore, NCI-H1781 was used for the following functional studies.

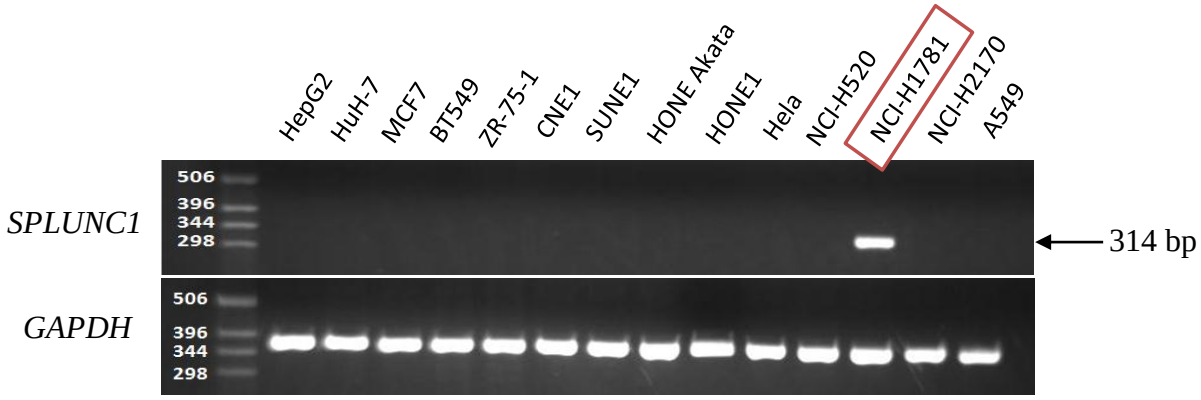


Figure 3.9: Expression of *SPLUNC1* mRNA.

SPLUNC1 mRNA expression is analysed in 14 cell lines. Expression of SPLUNC1 was detected in NCI-H1781 cell line.

b) Electrophoretic mobility shift assay (EMSA)

EMSA was performed to investigate transcription factor binding affinity of SNP alleles. We only observed a difference in the transcription factor binding affinity between two alleles of rs1407019, which is located in intron 3 of *SPLUNC1* (Chr. 20: 31289458). Two DNA-protein complexes (C1 and C2 complexes) were detected specifically to the non-susceptible allele rs1407019-G but not to the susceptible allele rs1407019-A (Figure 3.11 (a), lanes 3 and 4) and the rs1407019-G shifted bands were competed by the addition of unlabelled rs1407019-G oligonucleotide (self competitor) (Figure 3.11 (a), lane 5), suggesting that some nuclear protein(s) might interact with this genomic region.

The motif analysis by TFSEARCH database (http://mbs.cbrc.jp/research /db/ TFSEARCH.html) suggested that rs1407019 was located in the binding sequence of transcription factor specificity protein 1 (Sp1; Figure 3.10). The sequence containing rs1407019-G had higher matrix similarity to the consensus Sp1 binding motif than the sequence harbouring rs1407019-A. In the competition assay, DNA-C1 complex formation was competed by addition of an unlabelled oligonucleotide with consensus Sp1 binding motif, suggesting the binding of Sp1 or Sp1-like protein to the sequence (Figure 3.11 (a), lane 7). Furthermore, a supershifted band was observed after the addition of anti-Sp1 antibody, confirming that Sp1 interacted with the DNA sequence containing rs1407019-G (Figure 3.11 (b), lane 5). The formation of C2 complex suggested that another binding protein(s) might interact with this DNA region; however, we could not identify this binding protein(s) through the competition and supershift assays using competitors and antibodies for several transcription factors predicted by TFSEARCH database (data not shown).

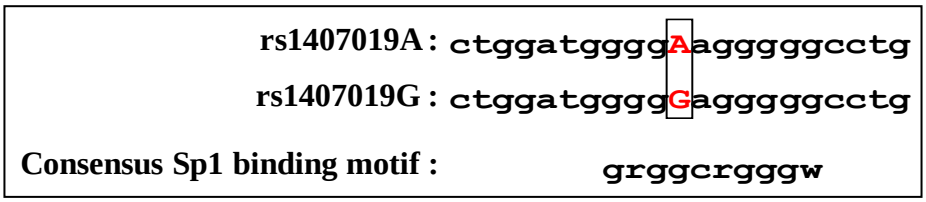


Figure 3.10: Sequence around the SNP site of rs1407019.

The motif analysis by TFSEARCH database suggested that rs1407019 was located in the binding sequence of transcription factor Specificity protein 1 (Sp1) (where r = G / A, w = A / T).

Lane	1	2	3	4	5	6	7
Labelled probes	A	G	A	G	G	G	G
Nuclear extracts	-	-	+	+	+	+	+
Unlabelled competitors	-	-	-	-	G	A	Sp1

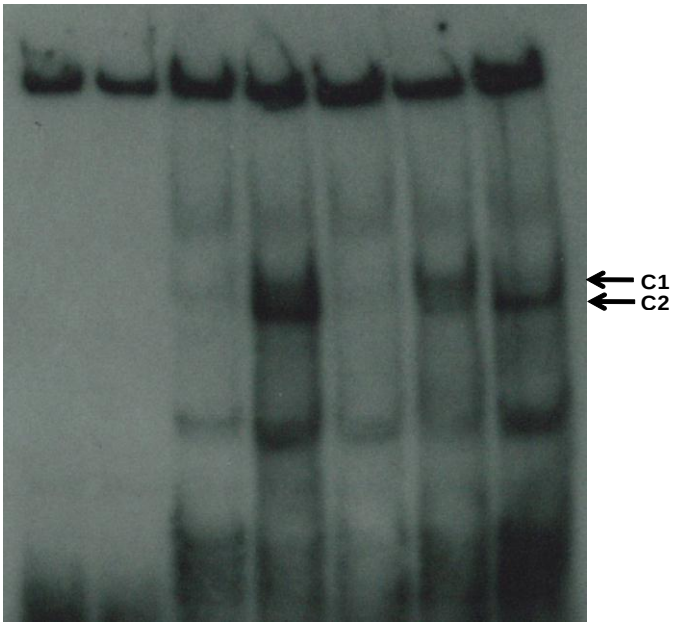
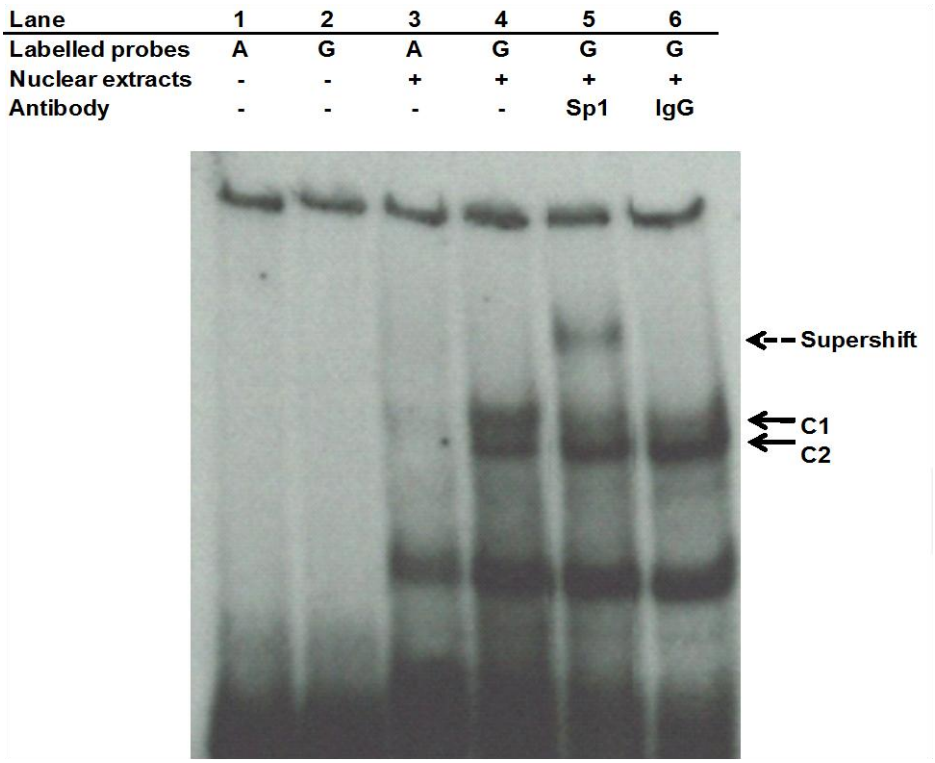


Figure 3.11: Performing EMSA on rs1407019.

(a) EMSA was performed using nuclear extracts from NCI-H1781 and oligonucleotides with rs1407019-A and rs1407019-G alleles. Two complexes (C1 and C2) were formed. Unlabelled probes in 125-fold excess as compared to the labelled probes were used for the competition experiment. The data is the representative result of three experiments.



(b) Supershift assay was performed using nuclear extracts from NCI-H1781, oligonucleotides with rs1407019-A and rs1407019-G alleles, and anti-Sp1 antibody. Supershift band was indicated by a broken arrow. The data is the representative result of three experiments.

c) Luciferase reporter assay

Luciferase reporter assay was then performed to examine the allelic difference of rs1407019 in transcriptional activity. A single copy of the 41-bp double stranded oligonucleotides which contained rs1407019-A or rs1407019-G allele of rs1407019 was then cloned into this pGL4-SV40 at the *Xho* I/*Bgl* II site.

The NCI-H1781 cells were transfected with pGL4-SV40 construct with rs1407019-A and rs1407019-G. The NCI-H1781 cells transfected with the construct containing rs1407019-G displayed a 1.6-fold greater enhancement of transcriptional activity than those with the A-allele construct (Figure 3.12). Combining the data from EMSA and luciferase reporter assay, we postulate that rs1407019 may affect the transcriptional regulation of *SPLUNC1* through transcriptional activation by Sp1.

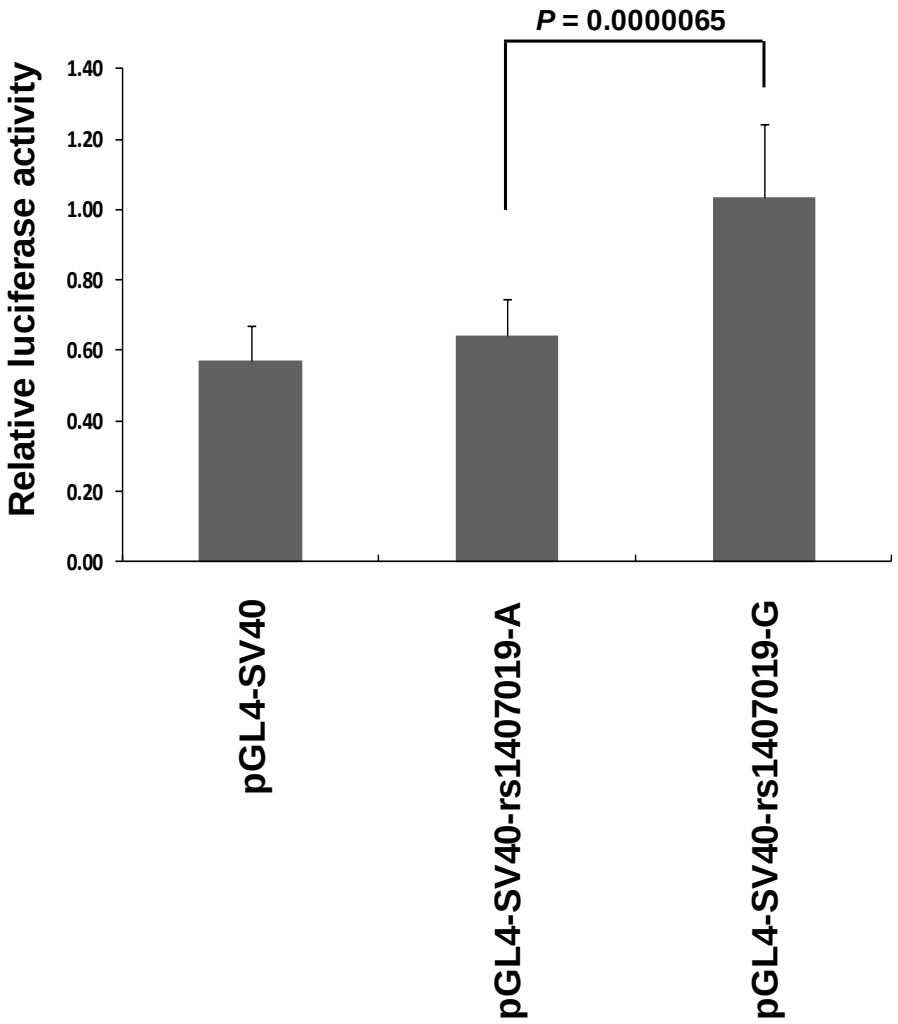


Figure 3.12: The transcriptional enhancer activity of the 41-bp sequence region around the rs1407019.

The luciferase activities in the NCI-H1781 cells transfected with A and G allele constructs were measured. The values of relative luciferase activity were expressed as mean ± SD from three independent experiments. The difference in the luciferase activities between the A and G alleles of rs1407019 were evaluated by a Student's t-test (degree of freedom = 22).

3.11 GENOME-WIDE ASSOCIATION STUDY OF NPC

A GWAS was carried out to identify NPC susceptibility genes in Malaysian Chinese. Initially, a total of 111 NPC and 260 control subjects (panel A) were examined by genome-wide SNP analysis. Overall the call rates of all subjects were 98% or higher. After genotyping, SNPs showing significant deviation from HWE in the controls ($P < 0.000001$), SNPs with call rates of less than 98%, monomorphic SNPs or those with insufficient genotyping data were excluded from further analysis. A total of 492,131 autosomal SNPs which passed the quality control filters were retained for statistical analysis. A Q-Q plot was generated to check the possible population stratification effects by comparing the distribution of observed *P*-values with expected distribution under the null hypothesis of no population stratification (Figure 3.13). This Q-Q plot (Figure 3.13) is different from the Q-Q plot reported in the paper of Ng *et al. (*2009) as more stringent quality controls (call rate > 98%, HWE > 0.000001) were applied in this analysis. The obtained genomic inflation factor was 1.09, suggesting a relatively low level of stratification in this study.

3.11.1 Validation and replication studies

To validate the GWAS finding, top 200 SNPs with the smallest *P*-values in the initial GWAS were re-genotyped by multiplex-PCR based Invader assay or direct sequencing using panel A samples. The data obtained from these assays (Illumina vs Invader assay or direct sequencing) were compared. The concordance rates of genotyping calls of these 200 SNPs in these two assays were more than 98%. We then further evaluated these 200 SNPs using an independent set of samples consisting of 335 NPC patients and 227 controls (panel B) by multiplex PCR based Invader assay for their association with NPC. The combined analysis of GWAS and replication study suggested that SNP rs2212020 is highly associated with NPC (crude $P = 0.000045$, OR = 1.80, 95% CI =

1.35-2.40, allelic-inheritance model; Table 3.18). The rs2212020 is an intronic SNP in the Integrin alpha 9 gene (*ITGA9*).

3.11.2 Fine mapping association analysis of a 40 kb LD block of *ITGA9*

Using the data of Han Chinese in the HapMap database ($r^2 > 0.8$), we found that the landmark SNP rs2212020 is located in a 40 kb LD block that covers 2.4 kb of its promoter region and up to intron 4 (Figure 3.14). Additional 19 SNPs identified in the LD block were first genotyped with 111 NPC patients and 260 healthy controls. Only rs197721 and rs149816 showed a *P*-value less than the multiple testing significance level ($P < 0.0026$, corrected against 19 SNPs; Table 3.19) was further genotyped with the panel B samples, consisting of 336 NPC patients and 227 healthy controls. The combined *P*-values of the two data sets revealed that both SNPs remained significantly associated with the increased risk of NPC after the adjustment for age and gender (Table 3.20).

Next, five SNPs (rs169188, rs169111, rs197770, rs189897 and rs197757) which are highly linked with rs197721 and rs149816 were genotyped in 447 NPC patients and 487 controls. Similarly, these SNPs showed significant association with NPC susceptibility. Among these eight associated SNPs, rs169111 showed the most significant association with NPC susceptibility. SNP rs169111 is located in the intron of *ITGA9*, About 7% of 447 NPC cases were carriers of T allele whereas only 3% of 487 controls were carriers of that ($P = 0.00012$, OR = 3.24, 95% CI = 1.78-5.91; Table 3.21).

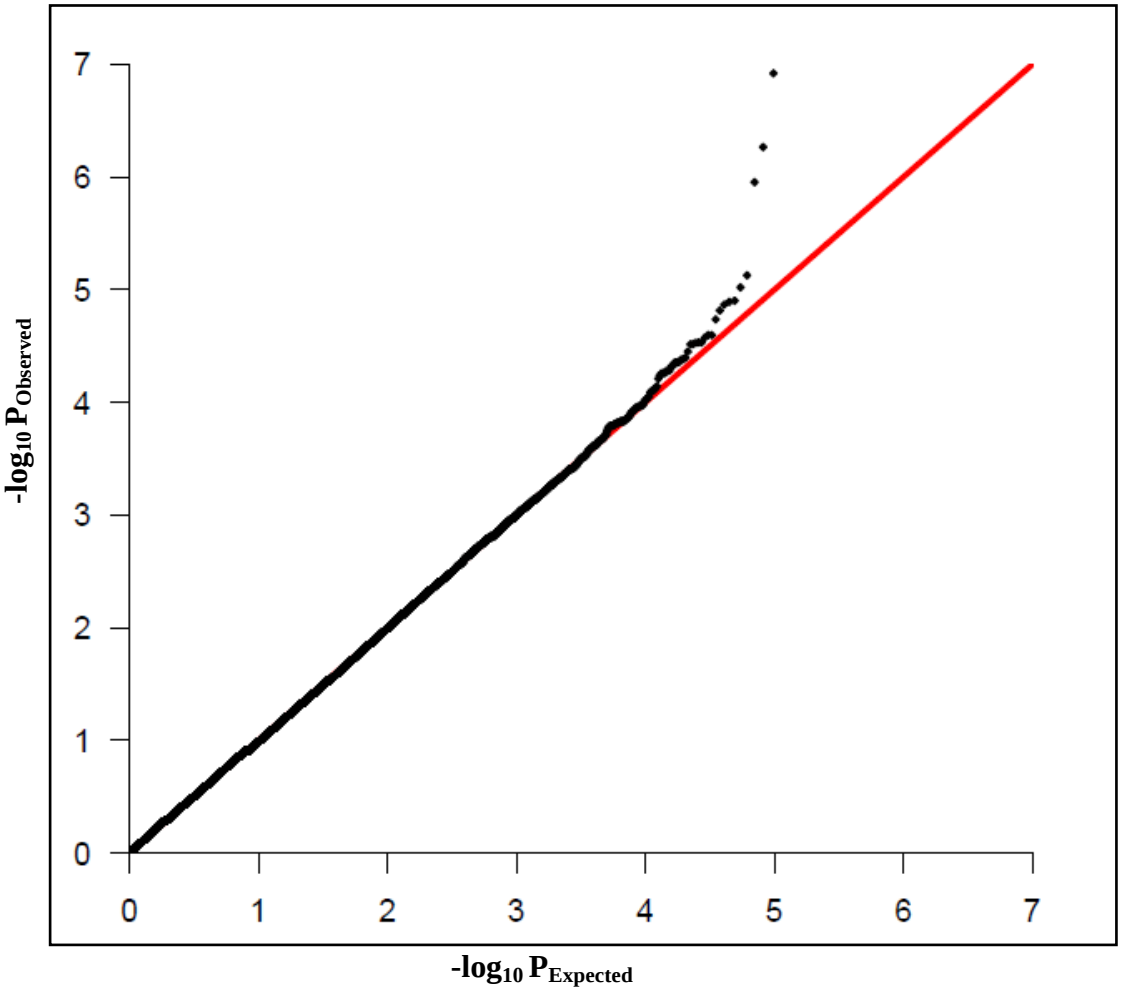


Figure 3.13: Quantile-quantile (Q-Q) *P*-value plot.

Log quantile-quantile (Q-Q) *P*-value plot showing the distribution of observed statistics by trend test for all utilized SNPs from genome-wide association study of 111 NPC patients and 260 controls of a Malaysian Chinese population (panel A). The diagonal line shows the values expected under the null hypothesis ($\lambda = 1.09$).

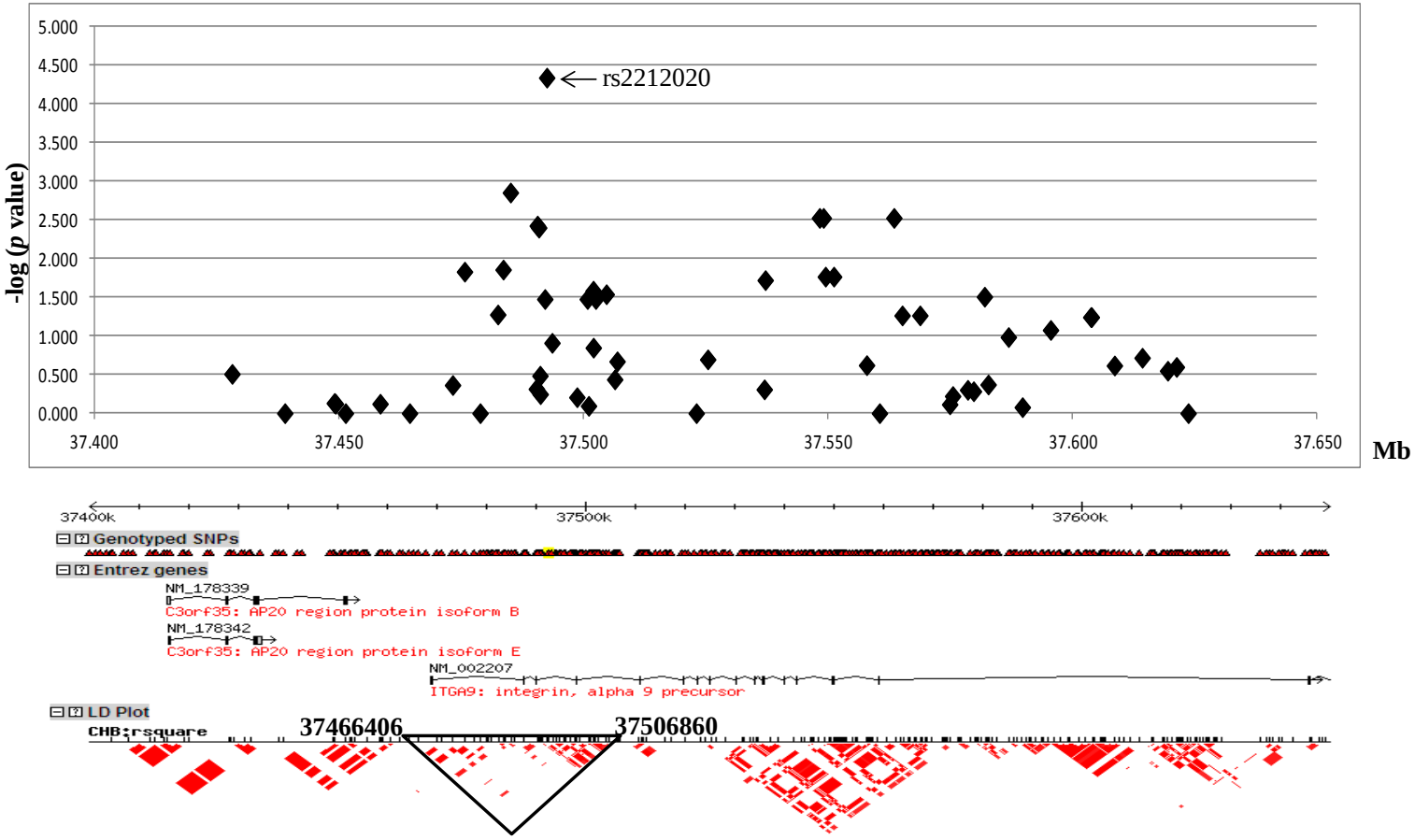


Figure 3.14: Case-control association plots, LD map and the genomic structure of an *ITGA9* region on chromosome 3p21.

Closed diamond represent –log10 (*P*-value) obtained from the genome-wide association study (Panel A). The LD map was obtained from the International HapMap Project database.

Table 3.18: Association analysis of *ITGA9* SNP rs2212020 to NPC susceptibility in 447 NPC patients and 487 controls

Gene	SNP ID	Location	Alleles[a] 1/2	Sample set[b]	Cases	Controls	*P* value			Odds ratio [95%CI]					
					MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2		vs 11		vs 22	
ITGA9 (Chr. 3p21.3)	rs2212020	Intron	C / T	panel A	0.19	0.08	0.000053[c]	0.000036[c]	0.37	2.61	[1.64-4.14]	2.99	[1.78-5.03]	2.40	[0.48-12.09]
				panel B	0.14	0.10	0.054	0.12	0.14	1.45	[1.00-2.11]	1.39	[0.92-2.10]	3.46	[0.75-15.95]
				Total	0.15	0.09	0.000045[c]	0.00014[c]	0.054	1.80	[1.35-2.40]	1.84	[1.34-2.52]	2.90	[1.03-8.20]

MAF, Minor allele frequency; CI, Confidence interval
Sample set= 447 NPC cases; 487 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] panel A: cases, N = 111, controls, N = 260
panel B: cases, N = 335, controls, N = 227
Total: cases, N= 447, controls, N = 487
[c] Statistical significant

Table 3.19: Association analysis of 19 tag-SNPs of *ITGA9* to NPC susceptibility in 111 NPC patients and 260 controls

Gene	SNP ID	Location	Alleles[a]	Cases	Controls	*P* value			Odds ratio [95%CI]		
			1/2	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2	vs 11	vs 22
ITGA9 (Chr. 3p21.3)	rs879904	Intron	T / C	0.34	0.39	0.39	0.70	0.24	0.81 [0.51-1.30]	0.88 [0.47-1.67]	0.55 [0.20-1.49]
	rs197722	Intron	G / C	0.32	0.36	0.19	0.25	0.33	0.74 [0.47-1.17]	0.69 [0.37-1.29]	0.63 [0.24-1.62]
	rs17036376	Intron	G / A	0.09	0.10	0.28	0.33	1.00	0.65 [0.30-1.41]	0.67 [0.30-1.50]	NA
	rs197721	Intron	G / T	0.09	0.03	0.00050[b]	0.00046[b]	1.00	7.04 [2.35-21.12]	7.10 [2.37-21.26]	NA
	rs149816	Intron	G / C	0.09	0.03	0.00050[b]	0.00046[b]	1.00	7.04 [2.35-21.12]	7.10 [2.37-21.26]	NA
	rs11709385	Intron	C / G	0.42	0.44	0.79	0.71	0.35	0.94 [0.61-1.46]	1.13 [0.59-2.17]	0.68 [0.30-1.54]
	rs267527	Intron	C / T	0.45	0.47	0.88	0.41	0.51	1.04 [0.66-1.62]	1.34 [0.66-2.72]	0.76 [0.35-1.68]
	rs267526	Intron	C / T	0.16	0.18	0.39	0.63	0.16	1.29 [0.72-2.31]	1.18 [0.60-2.32]	3.15 [0.63-15.7]
	rs2162683	Intron	G / A	0.09	0.06	0.15	0.14	1.00	1.91 [0.79-4.57]	1.95 [0.80-4.75]	NA
	rs197766	Intron	T / C	0.41	0.45	0.54	0.86	0.39	0.87 [0.56-1.36]	0.94 [0.48-1.86]	0.71 [0.32-1.57]
	rs391224	Intron	C / T	0.38	0.42	0.50	0.68	0.47	0.86 [0.54-1.34]	0.87 [0.45-1.69]	0.74 [0.32-1.69]
	rs17814006	Intron	G / C	0.08	0.05	0.090	0.087	1.00	2.22 [0.88-5.58]	2.29 [0.89-5.93]	NA
	rs2070478	Intron	C / T	0.27	0.32	0.52	0.86	0.25	0.86 [0.53-1.37]	0.95 [0.51-1.76]	0.52 [0.17-1.60]
	rs407930	Intron	G / T	0.47	0.52	0.93	0.87	0.99	1.02 [0.66-1.58]	1.06 [0.52-2.16]	0.99 [0.49-2.04]
	rs2019733	Intron	G / C	0.19	0.12	0.061	0.048	0.98	1.85 [0.97-3.52]	2.01 [1.01-4.01]	1.04 [0.05-22.94]

Table 3.19, continued

Gene	SNP ID	Location	Alleles[a] 1 / 2	Cases MAF	Controls MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
ITGA9 (Chr. 3p21.3)	rs1549599	Intron	A / G	0.25	0.29	0.62	0.40	0.70	0.88 [0.53-1.44]	0.76 [0.41-1.43]	1.24 [0.41-3.81]
	rs2019750	Intron	C / G	0.46	0.41	0.21	0.91	0.034	1.33 [0.85-2.08]	1.04 [0.54-2.01]	2.37 [1.07-5.25]
	rs9871277	Intron	C / T	0.07	0.05	0.98	0.95	1.00	0.99 [0.38-2.55]	1.04 [0.38-2.81]	NA

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 111 NPC cases; 260 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] Statistically significant at the Bonferroni-corrected significance level, based on 19 independent effective tests ($P < 0.0026$)

Table 3.20: Association analysis of *ITGA9* rs197721 and rs149816 to NPC susceptibility

SNP ID	Sample set	Alleles[a] 1 / 2	Cases[b] MAF	Controls[b] MAF	*P* value 1 vs 2	*P* value vs 11	*P* value vs 22	Odds ratio [95%CI] 1 vs 2	Odds ratio [95%CI] vs 11	Odds ratio [95%CI] vs 22
rs197721	panel A	G / T	0.09	0.03	0.00050[c]	0.00046[c]	1.00	7.04 [2.35-21.12]	7.10 [2.37-21.26]	NA
	panel B		0.06	0.03	0.038	0.046	1.00	2.29 [1.05-5.02]	2.29 [1.02-5.18]	NA
	Total		0.07	0.03	0.00039[c]	0.00042[c]	1.00	3.08 [1.66-5.74]	3.13 [1.66-5.89]	NA
rs149816	panel A	G / C	0.09	0.03	0.00050[c]	0.00046[c]	1.00	7.04 [2.35-21.12]	7.10 [2.37-21.26]	NA
	panel B		0.06	0.03	0.050	0.061	1.00	2.20 [1.00-4.86]	2.2 [0.96-5.02]	NA
	Total		0.07	0.03	0.00052[c]	0.00057[c]	1.00	3.03 [1.62-5.66]	3.07 [1.62-5.81]	NA

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 447 NPC cases; 487 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] panel A: cases, N = 272, controls, N = 442
panel B: cases, N = 335, controls, N = 227
Total: cases, N= 447, controls, N = 487
[c] Statistically significant at the Bonferroni-corrected significance level, based on 19 independent effective tests ($P < 0.0026$)

Table 3.21: The genotype distribution of *ITGA9* SNPs captured by rs197721 and rs149816 in Malaysian NPC samples

SNP ID	Location	Alleles[a]	Cases	Controls	*P* value			Odds ratio [95%CI]					
		1/2	MAF	MAF	1 vs 2	vs 11	vs 22	1 vs 2		vs 11		vs 22	
rs169188	Intron	G / A	0.064	0.024	0.00084[b]	0.00092[b]	1.00	3.02	[1.58-5.78]	3.07	[1.58-5.95]	NA	
rs169111	Intron	A / G	0.071	0.030	0.00012[b]	0.00013[b]	1.00	3.24	[1.78-5.91]	3.29	[1.79-6.05]	NA	
rs197770	Intron	T / C	0.080	0.034	0.00058[b]	0.00053[b]	0.14	2.65	[1.52-4.61]	2.88	[1.58-5.23]	5.62	[0.58-54.21]
rs189897	Intron	T / A	0.082	0.031	0.00021[b]	0.00026[b]	1.00	3.01	[1.68-5.40]	3.07	[1.68-5.60]	NA	
rs197757	Intron	C / T	0.080	0.034	0.00058[b]	0.00053[b]	0.14	2.65	[1.52-4.61]	2.88	[1.58-5.23]	5.62	[0.58-54.21]

MAF, Minor allele frequency; CI, Confidence interval; NA, Not applicable
P value, odds ratio and 95% CI were adjusted for age and gender
Sample set = 447 NPC cases; 487 controls
[a] Allele 1 = Major allele, Allele 2 = Minor allele
[b] Statistically significant at the Bonferroni-corrected significance level, based on 19 independent effective tests ($P < 0.0026$)

CHAPTER 4

DISCUSSION

4.1 STUDY SUBJECTS

Sample collection is indeed a daunting task for any association study. Hence, a population based association study is more viable compared to a family based study. A family based study would require sampling of nuclear families of patients, a difficult feat considering the reluctance and lack of consent from families.

In the present study, a total of 447 NPC patients and 487 healthy controls have been recruited for the candidate gene and genome-wide association studies. For candidate gene association study, a two-stage analysis was performed except for GSTs polymorphisms. These samples were collected in different stages. The initial 81 NPC patients and 147 controls were first recruited from September 2006 to January 2007, followed by a second set of 366 NPC patients and 340 controls, recruited during March 2007 until 2009. Some of the DNA amounts for second set of samples were limited as they were obtained from a shared group of researchers in the Malaysian NPC network. Therefore, the second set of samples was only used for SNPs that were significant after adjusting for multiple testing.

For association study of GSTs polymorphisms, the two-stage screening approach was not applied. Therefore, we genotyped all the samples, but excluded those samples with limited DNA amount. A total of 274 NPC patients and 442 healthy controls were examined in this association study. However, there is no additional screening for GSTs polymorphisms as no significant association was observed after genotyping those samples.

For GWAS, a total of 111 NPC patients and 260 controls were first genotyped by Illumina assay. The remaining samples (336 NPC patients and 227 controls) were

genotyped in the replication study by multiplex-PCR-based Invader assay or direct sequencing which is less expensive.

4.2 CANDIDATE GENE APPROACH AND GWAS

In this study, we adopted candidate gene and genome-wide screening (GWAS) approaches to identify NPC susceptibility genes in Malaysian Chinese population. We attempt to replicate some of the reported associations and identify the novel NPC susceptibility gene. In the candidate gene approach, SNPs in *CYP2E1*, *EPHX2* and *SPLUNC1* were associated with NPC. The association of these genes were also found to be replicated in the present GWAS at a nominal significance level (P = 0.00538 – 0.0143).

Both candidate gene and genome-wide screening approaches have their own advantages and disadvantages. GWAS is a hypothesis free approach, no prior information regarding gene function is required and it is able to discover the unanticipated genes (Hirschhorn and Daly, 2005). However, the identified genes by GWAS might have unknown function or without obvious relationship with the phenotypes (Witte, 2010), thus it would be difficult to predict the role of these genes in the pathogenesis of disease or identify gene-environmental interactions. In contrast, candidate gene approach has a higher prior probability to detect significant and true association due to the prior knowledge of biological function of the genes of interest (Tabor *et al.*, 2002; Zhu and Zhao, 2007).

Due to the larger number of genotyped SNPs in a GWAS, the criteria for statistical significance need to be more stringent compared to candidate gene approach, in order to reduce the number of false positive results. The simplest correction for multiple testing is Bonferroni correction. However, this Bonferroni correction is considered conservative as some of the genotyped SNPs for GWAS are in LD and

correlated with each other, this may inflate the type II error rate (false negative) (Pe'er *et al.*, 2008; Johnson *et al.*, 2010; Witte, 2010). Compared to candidate gene approach, GWAS requires extremely large samples size in order to have adequate statistical power to detect the associations with more modest effects (Witte, 2010). In other words, larger sample size requires more resources and money.

The genotyping arrays of GWAS might have incomplete coverage of the common variants in the genome (Spencer *et al.*, 2009). For GWAS chip, tag-SNPs identified from the International HapMap project were examined (Steemers and Gunderson, 2005). Certain SNPs might be missed due to low LD with the tag-SNPs on GWAS chip. In addition, GWAS often identify genetic variants which are located in intergenic region or non-coding sequences. The identified variant might have no function or in high LD with the functional variants (Manolio, 2010). On the other hand, a candidate approach might be able to uncover the associations that are missed by GWAS screening particularly the rare variants at lower cost.

To date, a large number of association studies have been performed to identify NPC susceptibility genes. However, most of them were failed to be replicated in the subsequent study. False positive, false negative or true positive in the association analysis might be able to explain the inconsistent results among different population studies (Kathiresan *et al.*, 2004). A false positive result is type I error, the hypothesis is rejected while there is no true association. A false positive result might arise from population stratification. Differences in the allele frequencies among patients and controls are likely to occur due to the differences in ancestry between these two groups rather than true association between gene polymorphisms and disease (Clayton *et al.*, 2005). Besides that, genotyping error may produce false positive results. Compare to homozygous genotypes, the heterozygous genotype is difficult to be assigned by many genotyping platforms and it may produce false positive results (Kathiresan *et al.*, 2004;

Marquard *et al.*, 2009). On the other hand, a false negative result is type II error, a negative association is reported while there is a true association. A false negative result might be produced by inadequate sample sizes.

In the present study, only Chinese NPC patients and Chinese controls were recruited to minimize population admixture. In addition, PCA was performed for the GWAS to investigate the population stratification. HWE test was performed for each SNP before further association analysis. HWE test can be used to detect the genotyping errors and thereby avoid the false positive results. SNPs which showed deviation from HWE were excluded from the further analysis. Besides that, multiple testing by the strict Bonferroni correction was carried out to reduce false positive results.

4.3 ASSOCIATION STUDY BETWEEN *CYP1A1, CYP2A13, CYP2E1* POLYMORPHISMS AND NPC SUSCEPTIBILITY

In the present study, we have studied the association between polymorphisms of *CYP1A1, CYP2A13* and *CYP2E1* and NPC susceptibility in Malaysian Chinese. Among these genes, only *CYP2E1* polymorphisms show significant association with increased risk of NPC. Therefore, we propose that the *CYP2E1* may play a role in the genetic susceptibility to the development of NPC in the Malaysian Chinese population.

i. *CYP1A1*

CYP1A1 involved in the metabolic activation of PAH which is known carcinogen of cigarette smoke (Bozina N *et al.*, 2009). Two polymorphism of *CYP1A1* (rs4646903 and rs1048943) has been studied extensively in relation to lung cancer (Nakachi *et al.*, 1993; Song *et al.*, 2001) and head and neck cancer (Varzim *et al.*, 2003; Marques *et al.*, 2006). These polymorphisms were suggested to have elevated enzyme activity, and thereby increased the activation of carcinogen which is expected to increase the risk of

cancer (Crofts *et al.,* 1994; Georgiadis *et al.,* 2005). To date, two studies of association between NPC susceptibility and *CYP1A1* polymorphisms have been performed in Taiwanese (Cheng *et al.,* 2003) and Cantonese nuclear families in China (Xu *et al.* 2009). Both studies reported that SNP rs4646903 was not associated with NPC susceptibility, while association of rs1048943 with NPC was observed in the Cantonese nuclear families in China (Cheng *et al*., 2003; Xu *et al.* 2009).

In our study, we examined two tag-SNPs (rs4646422 and rs1048943) selected from the HapMap in 81 NPC patients and 147 controls, but no significant association was observed. In the study by Xu *et al.* (2009), the G allele of rs1048943 is associated with the decreased risk of NPC in the low-titer group (antibody titers of VCA-IgA < 1:80) of Cantonese nuclear families in China (*P*-additive = 0.034, *P*-dominant = 0.021; Xu *et al.,* 2009). The frequency of G allele in the population reported by Xu *et al.* (0.339) was higher than ours (0.22) and HapMap CHB (0.244). The discrepancy between two studies may reflect certain differences in genetic composition of both populations. In addition, we are unable to stratify the association analysis by VCA-IgA titers due to lack of information of antibody titers of VCA-IgA. Nevertheless, the power to detect an association in our sample for *CYP1A1* was modest (between 25% and 79%), sample sizes should be increased in the future investigation to confirm the association. Besides that, variation in CYP1A1 activity might be associated with the tobacco exposure which is a potential risk factor for NPC (Wardlaw *et al*., 1998; Georgiadis *et al.,* 2005). Therefore, the potential interaction between tobacco exposure and *CYP1A1* polymorphisms in developing NPC can be evaluated in the future.

ii. *CYP2A13*

CYP2A13 has been found predominantly in the respiratory tract and is more active in the metabolism activation of a major tobacco specific carcinogen, NNK compared to CYP2A6 (Su *et al.*, 2000). Therefore, CYP2A13 might play a key role in tobacco-related tumorigenesis in the respiratory tract. Besides that, association of *CYP2A13* polymorphisms with lung cancer has been reported in Chinese and Japanese populations (Wang *et al.*, 2003; Kiyohara *et al.,* 2005). Wang *et al.* (2003) found that the variant allele of *CYP2A13* rs8192789 (Cys-257) was associated with reduced risk of smoking-related lung adenocarcinoma in a Chinese population. Zhang and co-workers reported that Cys-257 had lower enzymatic activity compared to wild type Arg-257 protein, suggesting a protective effect of variant Cys-257 against xenobiotic toxicity (Zhang *et al.*, 2002). Thus far, two association studies have been carried out to investigate the relation between *CYP2A13* polymorphisms and NPC susceptibility in the Cantonese population residing in Southern China (Jiang *et al.*, 2004; Cao *et al.*, 2011). The nonsynonymous variant (Arg257Cys; rs8192789) was examined in both studies, but they failed to detect association between NPC risk and *CYP2A13* polymorphisms (Jiang *et al.,* 2004; Cao *et al.,* 2011). Besides that, Jiang *et al.* (2004) have reported that the expression of CYP2A13 was detected in both non-cancerous nasopharynx and NPC tissues, but its expression in both tissues was not significantly different.

Five tag-SNPs of *CYP2A13* were examined in the present study, but none of them was associated with NPC susceptibility among Malaysian Chinese, it is consistent with the finding of Jiang *et al.* (2004) and Cao *et al.* (2011). Our study suggested that *CYP2A13* might not involve in the development of NPC. Nevertheless, larger sample sizes are required in the future investigation as the power to detect the association in our samples for *CYP2A13* was modest (between 13% and 42%). In addition, the potential interaction between tobacco exposure and *CYP2A13* polymorphisms in developing NPC

can be evaluated in the future as CYP2A13 is highly active in metabolizing tobacco specific carcinogen.

iii. *CYP2E1*

The association between *CYP2E1* polymorphisms and NPC susceptibility has been widely studied (Hildesheim *et al.*, 1997; Kongruttanachok *et al.*, 2001; Jia *et al.*, 2009; Guo *et al.*, 2010), and the results are summarized in Table 4.1. In the present study, five *CYP2E1* SNPs (rs2070672, rs3813865, rs3813870, rs8192772 and rs915906) which are in strong LD ($r^2 \geq 0.85$) were significantly associated with increased risk of NPC (P = 0.00081-0.0032). Our observation was in accordance with the study performed by Jia *et al.* (2009) in Southern Chinese, which reported that SNP rs3813865 was associated with an increased risk of NPC. In addition, Jia *et al.* (2009) have showed that two haplotypes are associated with NPC among individuals who had a positive smoking history and aged less than 46 years old. In contrast, our study revealed that no *CYP2E1* haplotype was associated with NPC susceptibility in Malaysian Chinese.

For *CYP2E1* rs2031920, Hildesheim *et al.* (1997) have suggested that individuals with the TT genotype are at an increased risk of developing NPC in Taiwanese. However, we are unable to replicate this result in our Malaysian Chinese. Our result was in agreement with the study by Kongruttanachok *et al.* (2001) and Guo *et al.* (2010) which reported that no association was detected between rs2031920 and NPC susceptibility among individuals of Thai origin or Chinese origin in Thailand and Southern Chinese.

In the present study, three significantly associated SNPs (rs2070672, rs3813870 and rs3813865) are located in the upstream region of *CYP2E1* (-352bp, -930bp and -1656bp of upstream, respectively) while SNP rs8192772 and rs915906 are located in the intron 2. A few studies have been performed on investigating the functional

characterization of 5' flanking region of *CYP2E1*. Ueno and Gonzalec (1990) have examined the transcriptional control of the rat hepatic *CYP2E1* and suggested that the 5' region of *CYP2E1* is positively regulated by HNF-1 in adult rats. Tambyrajah *et al.* (2004) also reported that the pig *CYP2E1* promoter can be activated by COUP-TF1 and HNF-1. In addition, Cheung *et al.* (2003) shows that CYP2E1 is expressed at lower levels in the liver of HNF-1α deficient mice, suggesting that HNF-1α may act as a positive regulator of *CYP2E1* gene.

A recent study suggested that HNF-1 binds to the -1080 to -836 region of human *CYP2E1* (Yang *et al.*, 2010), and one of the associated SNPs in our study (rs3813870) is located in this region (-930). According to the motif analysis by TFSEARCH (http://mbs.cbrc.jp/research/db/TFSEARCH.html), the susceptible allele G of rs3813870 has a similarity to the consensus binding motif of HNF-1. In addition, a recent study examined the association of *CYP2E1* polymorphisms with systemic lupus erythematosus, and they have analysed the transcription difference in different genotypes for the associated SNP rs8192772 in their study. They showed that the risk genotype (CC) has significant higher transcription compared to other two genotypes (TT and CT) (Liao *et al.*, 2011). Both rs8192772 and rs3813870 are in strong LD ($r^2 \geq 0.85$), suggesting that the risk genotype (GG) of rs3813870 might show higher CYP2E1 transcription. Hence, it is postulated that the risk genotype (GG) of rs3813870 may enhance the human CYP2E1 transcription activity by altering its ability to bind HNF-1.

CYP2E1 is important in the metabolism of variety endogenous and exogenous compounds including nitrosamines and is relevant to chemical toxicity and carcinogenesis (Nuoso *et al.*, 1992; Hou *et al.*, 2007). Therefore, overexpression of CYP2E1 may result in more procarcinogens are being changed to carcinogens and more DNA damages are produced, thereby increased the risk of developing NPC. Our study showed that *CYP2E1* polymorphisms are associated with NPC susceptibility in

Malaysian Chinese population. Further studies (e.g. EMSA, luciferase reporter assay) are needed to determine the functional consequences of the significant associated SNPs and the resulting downstream responses to provide valuable insights into the mechanism that cause susceptibility to NPC. Besides that, it is reported that nitrosamine and ethanol can induce the expression of *CYP2E1* (He *et al.*, 2002). However, our study lacked the environmental exposure data including smoking habits or food intake whereas these data might be useful in examining the interaction effect between genes polymorphism and the risk factors for NPC. Therefore, gene-environmental interaction is suggested to be analysed in the future.

Table 4.1: Case-control studies on *CYP2E1* polymorphisms and NPC risk in different populations

Gene	SNP ID	Population	References	Sample size		OR / P
				Cases	Controls	
CYP2E1	**rs3813865**	Malaysian Chinese	Present study	447	487	3.24 / 0.0017
		Southern Chinese	Jia *et al.*, 2009	755	755	2.35 / 0.0011
	rs2031920	Malaysian Chinese	Present study	81	147	1.05 / NS
		Taiwanese	Hildesheim *et al.*, 1995	50	50	7.7 / NS
		Taiwanese	Hildesheim *et al.*, 1997	364	320	2.6 / <0.05
		Thai	Kongruttanachok *et al.*, 2001	132	99	1.61 / NS
		Chinese in Thai		56	98	1.99 / NS
		Chinese	Guo *et al.*, 2010	358	629	1.21 / NS

OR, Odds ratio; NS, Non significant

4.4 ASSOCIATION STUDY BETWEEN *EPHX1* AND *EPHX2* POLYMORPHISMS AND NPC SUSCEPTIBILITY

To the best of our knowledge, this is the first study examined the association of *EPHX1* and *EPHX2* polymorphisms and NPC susceptibility. Seven tag-SNPs of *EPHX1* and ten tag-SNPs of *EPHX2* were examined in the present study. However, only SNP rs4149244 of *EPHX2* was found to be associated with NPC susceptibility in the Malaysian Chinese population, suggesting that *EPHX2* may involve in the pathogenesis of NPC.

Our study showed that the presence of T allele of rs4149244 was observed more frequently in the healthy controls compared to NPC patients, suggesting that the T allele confers a protective effect against NPC. Carriers of the T allele had an OR of 0.53 for developing NPC. SNP rs4149244 is located in the intron region of *EPHX2*, and at present it is unclear how it is related to NPC susceptibility. According to the motif analysis by TFSEARCH (http://mbs.cbrc.jp/research/db/TFSEARCH.html), the sequence harbouring allele T contains consensus binding sequence of Pbx-1 which may act as activator and enhance the gene transcription (Chang *et al.*, 2008). Hence, individual with the T allele of rs4149244 may have higher expression of *EPHX2*.

EPHX2 is a multifunctional protein involved in metabolizing endogenous fatty acid epoxides (such as arachidonic acid epoxides or EETs) which are generated by CYP 450 epoxygenases. Arachidonic acid epoxides have been shown to be involved in cell signalling, mitogenesis, apoptosis (Kroetz and Zeldin, 2002) and carcinogenesis (Jiang *et al.* 2005). Besides that, EPHX2 is suggested to be involved in the regulation of isoprenoid biosynthesis (Enayetallah and Grant, 2006). Isoprenoids are important in isoprenylation of small G-proteins, cell signaling, proliferation, apoptosis (Edwards and Ericsson, 1999) and anti-tumour activity (Burke *et al.*, 1997).

Lower expression or loss of EPHX2 might cause accumulation of certain epoxides such as arachidonic acid epoxides or reduced the synthesis of isoprenoid with anti-tumour activity, thereby increase individual's susceptibility to develop cancer (Enayetallah *et al.*, 2006). In addition, accumulation of arachidonic acid epoxides or EETs is suggested to promote the neoplatic properties of carcinoma cells and stimulate angiogenesis (Jiang *et al.*, 2005). Therefore, higher expression of *EPHX2* may decrease the risk of developing NPC.

Our finding suggested that T allele of rs4149244 of *EPHX2* may decrease the susceptibility to NPC in Malaysian Chinese population. Nevertheless, whether T allele of rs4149244 has a protective role in NPC susceptibility or it is in LD with other functional variants of *EPHX2* require further investigation. According to the data of International HapMap projects, six intronic SNPs of *EPHX2* are captured by rs4149244. These captured SNPs should be genotyped in the future. Further studies are required to replicate the results with larger cohorts in different populations to confirm the findings. The functional consequences of the significant associated SNPs can be determined by performing CHIP assay, luciferase reporter assay or EMSA. Moreover, further functional studies such as gene knockout are required to elucidate the exact mechanism on how EPHX2 involved in NPC pathogenesis.

4.5 ASSOCIATION STUDY BETWEEN *GSTM1* NULL, *GSTT1* NULL GENOTYPES AND *GSTP1* POLYMORPHISMS AND NPC SUSCEPTIBILITY

GSTs play important roles in detoxifying various carcinogens including carcinogens in tobacco smoke and deficiency of GSTs may lead to a predisposition to NPC. In the present study, deletion genotype of *GSTM1* and *GSTT1* and three SNPs of *GSTP1* including a functional variant rs1695 (Ile105Val) were genotyped in 274 NPC cases and 442 healthy controls. Our study demonstrated that there is no significant association between GSTs polymorphisms and NPC susceptibility among the Malaysian Chinese.

Numerous studies have been performed to investigate the association between *GSTM1* null, *GSTT1* null genotypes and the risk of NPC, but they show conflicting results. The summary of various studies on *GSTM1* or *GSTT1* null genotype and NPC susceptibility are shown in Table 4.2. Individuals with *GSTM1* null genotype were reported to confer a higher risk of NPC in Thai (Tiwawech *et al.*, 2005), Southern Chinese (Deng *et al.*, 2005) and Caucasian, African-American, Asian and Native American populations (Nazar-Stewart *et al.*, 1999) whereas no association has been found in the Taiwanese (Cheng *et al.*, 2003), Southern Chinese (Guo *et al.*, 2008) and Malaysian Chinese (present study).

Similarly, inconsistent results were observed for the association studies between *GSTT1* null genotype and the risk of NPC. Deng *et al.* (2005) demonstrate that *GSTT1* null genotype is associated with an increased risk of NPC in Southern Chinese. Moreover, they have suggested that carriers of both *GSTM1* and *GSTT1* null genotypes confer higher risk of developing NPC. However, lack of association was reported in Taiwanese (Cheng *et al.*, 2003), Southern Chinese (Guo *et al.*, 2008) and Malaysian Chinese (present study). Due to the inconclusive results, a meta-analysis was carried out by Zhuo *et al.* (2009) to study the association between deletion genotype of *GSTM1* and

GSTT1 and NPC risk. It was reported that *GSTM1* null but not *GSTT1* null genotype is associated with NPC susceptibility (Zhuo *et al.*, 2009).

Such conflicting results could be due to sample size variation, differences in the genetic or environmental background, and different analysis methods or sample characteristic such as distribution of gender and age. As shown in Table 4.2, sample size for the study by Nazar-Stewart *et al.* (1999), Tiwawech *et al.* (2005) and Deng *et al.* (2005) are relatively small compared to other studies (present study, Cheng *et al.*, 2003 and Guo *et al.*, 2008). Small sample size may lead to insufficient statistical power to detect the association. Differences in the genetic (e.g. ethnicities) and environmental background (e.g. diet) between the study populations might influence the effect of polymorphisms.

On the other hand, both Deng *et al.* (2005) and Guo *et al.* (2008) have examined the Southern Chinese from Guangxi. However, conflicting results were obtained for both studies, and the discrepancies between both studies may reflect differences in overall study design. The sample size of the study by Deng *et al.* (2005) is relatively small compared to Guo *et al.* (2008). Moreover, Deng *et al.* (2005) performed monoplex PCR amplification for *GSTM1* and *GSTT1* without internal control. PCR amplification without internal control may cause artefact in the detection of gene deletion, and it may result in false negative.

In the case of *GSTP1*, a functional variant rs1695 has been examined for the association with NPC risk in several studies but none of them show significant association (Cheng *et al.*, 2003; Guo *et al.*, 2010) (Table 4.3). Generally, previous studies only reported the association of GST genes individually or in combination of *GSTM1* null and *GSTT1* null genotypes on NPC risk. The evaluation of the joint effects among *GSTM1*, *GSTT1* and *GSTP1* had not been performed. *GSTM1*, *GSTT1* and

GSTP1 genes act in the same biological pathway, so there is rationale to examine their joint effects.

We have evaluated the effect of *GSTM1*, *GSTT1* and the functional variant of *GSTP1* (rs1695) in combination of two (*GSTM1* or *GSTT1* and *GSTP1*) or three genes (*GSTM1* and *GSTT1* and *GSTP1*) on NPC susceptibility, but no association was reported. In the combination analysis of these three genes, some of the combined genotypes may have small numbers of samples, leading to the insufficient statistical power. Since the sample size of this study is small, our findings are considered preliminary. The association should be replicated in larger sample sizes in order to substantiate our findings in future study.

There are several limitations in this study. Gene copy number is not considered in the analysis of *GSTM1* and *GSTT1*, as heterozygotes cannot be detected by PCR based genotyping assay. Another limitation is lack of the environmental exposure data such as smoking or alcohol consumption which were reported to be involved in the development of NPC (Sigh *et al.*, 2008). Without the exposure data, we were unable to examine the gene-environment interaction. Our data suggests that GSTs polymorphisms might not associate with NPC susceptibility among Malaysian Chinese.

Table 4.2: Case-control studies on *GSTM1*, *GSTT1* and NPC risk in different populations

Gene	Population	References	Sample size		Frequency of null genotype		Association
			Cases	Controls	Cases	Controls	
***GSTM1* null**	Malaysian Chinese	Present study	274	442	0.605	0.593	negative
	Taiwanese	Cheng *et al.*, 2003	314	337	0.551	0.501	negative
	Chinese, Guangxi	Guo *et al.*, 2008	341	590	0.598	0.556	negative
	Caucasian, African-american,Asian,Native American	Nazar-Stewart *et al.* 1999	83	142	0.542	0.444	positive
	Thai	Tiwawech *et al.*, 2005	78	145	0.641	0.510	positive
	Chinese, Guangxi	Deng *et al.*, 2005	127	207	0.614	0.459	positive
***GSTT1* null**	Malaysian Chinese	Present study	274	442	0.485	0.450	negative
	Taiwanese	Cheng *et al.*, 2003	316	336	0.506	0.518	negative
	Chinese, Guangxi	Guo *et al.*, 2008	338	585	0.458	0.460	negative
	Chinese, Guangxi	Deng *et al.*, 2005	127	207	0.622	0.411	positive
Both *GSTM1* & *GSTT1* null	Malaysian Chinese	Present study	274	442	0.299	0.282	negative
	Chinese, Guangxi	Guo *et al.*, 2008	338	584	0.293	0.252	negative
	Chinese, Guangxi	Deng *et al.*, 2005	127	207	0.394	0.217	positive

Table 4.3: Case-control studies on *GSTP1* and NPC risk in different populations

Gene	Population	References	Sample size		Minor allele frequency		Association
			Cases	Controls	Cases	Controls	
***GSTP1* - rs1695**	Malaysian Chinese	Present study	274	442	0.17	0.18	negative
	Chinese	Guo *et al.*, 2010	358	629	0.20	0.18	negative
	Taiwanese	Cheng *et al.*, 2003	314	337	0.16	0.15	negative

4.6 ASSOCIATION STUDY BETWEEN *PIGR* POLYMORPHISMS AND NPC SUSCEPTIBILITY

EBV is known as one of aetiology factors which plays a crucial role in the development of NPC (Zheng *et al.*, 1994a). It infects the B lymphocytes and epithelial cells. However, the actual mechanism of infection of epithelial cells by EBV is still obscure. Gan *et al.* (1997) and Lin *et al.* (1997) have suggested that PIGR might be exploited by EBV to enter the epithelial cells, and failure in releasing the complex of IgA-EBV increases the risk of NPC. A total of seven tag-SNPs ($r^2 \geq 0.8$) of *PIGR* gene were examined among Malaysian Chinese in this study. However, our study did not show any significant association between these SNPs and NPC susceptibility.

In a study by Hirunsatit *et al.* (2003), homozygous C allele of rs291102 (*PIGR*1739 C/T) is suggested to be associated with NPC susceptibility among individuals of Thai origin and Chinese origin in Thailand. However, no significant association between rs291102 and NPC susceptibility was observed in our study, this finding is in accordance with the study by Fan *et al.* (2005). The C allele frequencies of rs291102 in different association studies are shown in Table 4.4. The C allele frequency of rs291102 in our population (0.92) and the Cantonese-speaking Chinese in China (0.89) were comparable, which was similar to that of HapMap CHB (0.907). However, the C allele frequency in the studied population by Hirunsatit *et al.* (2003) is much lower, only 74% of Chinese and 75% of Thai carried the C allele (Table 4.4). The observed discrepancies in these studies may indicate certain differences in genetic composition of the study populations. In addition, sample size variation may contribute to the conflicting results.

In the present study, lack of association was observed between *PIGR* polymorphisms and NPC susceptibility in Malaysian Chinese.

Table 4.4: Case-control studies on rs291102 of *PIGR* and NPC risk in different populations

Gene	SNP ID	Population	References	Sample size		Frequency of C allele		Association
				Cases	Controls	Cases	Controls	
PIGR	**rs291102**	Malaysian Chinese	Present study	81	147	0.87	0.92	No association
		Chinese origin in Thai	Hirunsatit *et al.*, 2003	42	107	0.87	0.74	$P < 0.05$
		Thai	Hirunsatit *et al.*, 2003	110	104	0.84	0.75	
		Chinese	Fan *et al.*, 2005	423	332	0.89	0.89	No association

4.7 ASSOCIATION STUDY BETWEEN *LPLUNC1 AND SPLUNC1* POLYMORPHISMS AND NPC SUSCEPTIBILITY

SNPs of *SPLUNC1*, but not *LPLUNC1*, exhibited significant association with NPC. Flach *et al.* (2007) reported that LPLUNC1 is highly expressed in the duodenal mucosa of patients during the acute phase of cholera, suggesting that LPLUNC1 is possibly responsible for the innate defence in combating the infection (Flach *et al.*, 2007). This has been further supported by a family-based association study in Bangladesh population, which showed an association of a variant in the promoter region of *LPLUNC1* (rs11906665) with cholera infection (Larocque *et al.*, 2009). Since we only examined four tag-SNPs of *LPLUNC1*, it is postulated that there might be other variants of this gene associated with NPC susceptibility.

In the case of *SPLUNC1*, a previous study reported that two promoter SNPs (C-2128T and C-1888T, corresponding to rs2752903 and rs750064, respectively) showed significant association with NPC in a total of 525 Cantonese-speaking Chinese in China (He *et al.*, 2005). However, in our study, rs2752903 but not rs750064 is associated with NPC susceptibility in Malaysian Chinese. The minor allele frequency of rs750064 in our population (0.405) and the population reported by He *et al* (0.360) were comparable, which was similar to that of HapMap CHB (0.381); however, the frequency of CC genotype in the population reported by He *et al* (0.071) was much lower than ours (0.163) and HapMap CHB (0.155). Their genotype frequency in controls significantly deviated from HWE (P = 0.000019 by $\chi 2$ test) while no deviation from HWE was detected in our controls. In the present study, LD between these two SNPs in Malaysian Chinese was found to be lower (r^2 = 0.28), which was comparable to HapMap CHB (r^2 = 0.42), than that in Cantonese-speaking Chinese (r^2 = 0.73). The discrepancy between two studies may reflect certain differences in genetic composition of both populations.

Through the screening of a functional causative SNP(s), we found that an intronic SNP, rs1407019, is likely to influence the expression of *SPLUNC1* by differential binding affinity of Sp1, which mainly acts as a transcription activator (Suske, 1999; Kim *et al.*, 2010). *SPLUNC1* expression in the carriers of susceptible allele A, is expected to be lower than non susceptible allele G due to lower binding affinity of Sp1. *SPLUNC1* is proposed to be involved in the innate immune response against EBV and other viruses. SPLUNC1 protein was reported to increase apoptosis and disruption of EBV- infected B lymphocytes (Zhou *et al.*, 2008). Thus, it might play a key role in reducing the tumorigenic capability of EBV. In addition, it is reported that Sp1-mediated transcription can be activated by Rta of EBV, and that Sp1 was up-regulated in NPC cells compared to normal nasopharyngeal cells (Chang *et al.*, 2005; Su *et al.*, 2010). These results suggested that allelic difference in binding of Sp1 at the SNP site of rs1407019 may cause individual differences in the defence against EBV. Individuals with the susceptible allele A may confer an increased risk of NPC due to insufficient amount of SPLUNC1 in response to EBV infection.

Our study showed that genetic polymorphisms of *SPLUNC1* are associated with increased risk of NPC in a Malaysian Chinese population. Functional analysis suggested that rs1407019 may down regulated *SPLUNC1* expression by altering the binding site for Sp1. Further studies are still needed to clarify the detailed biological mechanisms; however, our data provides valuable insights into the mechanism that cause susceptibility to NPC.

4.8 A GWAS IDENTIFIES *ITGA9* AS NPC SUSCEPTIBILITY GENE

To the best of our knowledge, this is the first study to report *ITGA9* polymorphism is associated with the susceptibility to NPC in Malaysian Chinese through GWAS. *ITGA9* (Integrin, alpha 9) is also known as RLC, *ITGA4L* (integrin, alpha 4-like), *ALPHA-RLC* (alpha related to the development of lung cancer). It maps on chromosome 3p22-21.3, which is known to be commonly deleted in various types of carcinoma including NPC (Xiong *et al.*, 2004). Frequent loss of heterozygosity (LOH) on 3p has been reported in various types of carcinoma including small cell lung cancer, breast carcinoma, renal cell carcinoma or NPC (Senchenko *et al.*, 2004). In addition, a linkage study reported an NPC susceptibility locus on chromosome 3p21.31-21.2, indicating that genes in this region might play crucial roles in the formation of NPC (Xiong *et al.*, 2004). Many tumour suppressor candidate genes are found in this region including *RASSF1A*, *DLC1*, *FUS1*, *H37* and *SEMA3F*. *RASSF1A* which maps on chromosome 3p21.3 have been reported to be associated with the development of NPC (Lo *et al.*, 2001b). Therefore, *ITGA9* might be a novel NPC susceptibility gene in this region.

ITGA9 encodes an integrin alpha 9 (α9) subunit. Integrins are cell surface heterodimeric receptors composed of an alpha (α) and a beta (β) chain that mediate cell-cell and cell-matrix adhesion. There are 18 α-subunits and 8 β-subunits known in the mammals. These α-subunits and β-subunits may associate in different combination and form 24 integrin heterodimers (Sherpard, 2003). The integrin α9 is a 1,035 amino acids polypeptide. It contains a large N-terminal extracellular domain with seven conserved repeats of putative metal binding domains, a transmembrane segment, and a short C-terminal cytoplasmic tail. The α9 subunit associates with β1 chain to form a single integrin α9β1. This α9β1 is widely expressed including the human airway epithelial cells, and can bind to diverse ligands including tenascin, VCAM-1, osteopontin, Urokinase receptor, plasmin, angiostatin, ADAM, thrombospondin-1, L1-CAM, VEGF-

C and VEGF-D31. Integrins may also be involved in modulating cell survival, proliferation or cell migration and malignant transformation. During malignant transformation, integrin signals may enable the cancer cells to detach from neighbouring cells in order to migrate to other place. Furthermore, integrin will also enable cancer cells to re-orientate their polarity during migration and survive and proliferate in foreign microenvironments (Guo and Giancotti, 2004). Therefore, integrins may play a role in the development of cancer. A study which was performed by Anedchenko *et al.* (2008) revealed that *ITGA9* was down-regulated in the non small lung cancer.

At present, the functional relevance of α9 or integrin α9β1 on NPC development still remains unclear. Integrins are suggested to act as receptor for certain respiratory pathogens. Several integrins including α2β1 (an Echovirus receptor) and α5β5 (adenovirus receptor) have been suggested to play important roles in virus infection (Sherpard, 2003). For EBV infections, the arginine-glycine-aspartate (RGD) motif of the EBV BMRF-2 glycoprotein might interact with the α5β1 integrins to mediate the entry of EBV to epithelial cell (Hutt-Fletcher, 2007). Therefore, it is tempting to hypothesize that the α9β1 which is expressed on the surface of epithelial cells may interact with a consensus domain of EBV capsid protein for attachment and/or enter the cell.

Recently, two NPC GWAS in Taiwanese and Southern Chinese have identified several NPC susceptibility loci including *HLA-A* (Taiwanese: rs2517713 and rs2975042; *P*combined = $3.90\text{x}10^{-20}$ and $1.60\text{x}10^{-19}$ respectively, Southern Chinese: rs2860580; *P*combined = $4.88\text{x}10^{-67}$), *HLA-F* (Taiwanese: rs3129055 and rs9258122; *P*combined = $7.36\text{x}10^{-11}$ and $3.33\text{x}10^{-10}$, respectively)., *GABBR1* (Taiwanese: rs29232; *P*combined = $8.97\text{x}10^{-17}$), *TNFRSF19* (Southern Chinese: rs9510787, *P*combined = $1.53\text{x}10^{-9}$), *MDS1-EVI1* (Southern Chinese: rs6774494, *P*combined = $1.34\text{x}10^{-8}$) and *CDKN2A-CDKN2B* (Southern Chinese: rs1412829, *P*combined = $4.84\text{x}10^{-7}$) (Tze *et al.*, 2009; Bei

et al., 2010). Both studies consistently reported an association in *HLA* region. However, our GWAS failed to detect genome wide significant association in the *HLA* region (P = 0.0013 – 0.047). On the other hand, Bei *et al.* (2010) did not observe any association between *ITGA9* polymorphisms and NPC susceptibility in Southern Chinese (Table 4.5). Such conflicting results may reflect certain differences in genetic composition of the studied populations or due to small sample size in our study.

Our study suggests that *ITGA9* on chromosome 3p21 may be a genetic risk factor involved in the pathogenesis of NPC in Malaysian Chinese. This finding is of great interest as it sheds new light on the association between integrin α9β1 and NPC. Replication studies with larger sample size as well as in other independent populations should be carried out for further confirmation of the association of *ITGA9* and NPC. Further studies are needed to elucidate the causal SNP or haplotype and to provide a plausible biological mechanism for the observed association between polymorphisms and susceptibility to NPC. Furthermore, further functional studies are necessary for a better understanding of its role in the nasopharyngeal tumorigenesis.

Table 4.5: Studies on *ITGA9* polymorphisms and NPC risk in different populations

Gene	SNP ID	Population	References	Sample size		Minor allele frequency		Association
				Cases	Controls	Cases	Controls	
ITGA9	rs2212020	Malaysian Chinese	Present study	447	487	0.15	0.09	S
		Southern Chinese	Bei *et al.*, 2010	1583	1894	0.11	0.11	NS
	rs197721	Malaysian Chinese	Present study	447	487	0.07	0.03	S
		Southern Chinese	Bei *et al.*, 2010	1583	1894	0.04	0.04	NS

S, significant association; NS, non significant

4.9 PROPOSED ROLES OF THE IDENTIFIED NPC SUSCEPTIBILITY GENES IN PATHOGENESIS

The development of NPC is complex and involves multiple steps. Environmental factors (e.g. consumption of salted fish or preserved food, exposure to carcinogens), EBV infection and genetic susceptibility were suggested to be the major contributors to NPC carcinogenesis (Sriampron *et al.*, 1992; Hildesheim and Lavine, 1993; Armstrong *et al.*, 1998; Thompson and Kurzrock, 2004; Zhou *et al.*, 2007). The interaction between environmental factors and genetic susceptibility may result in impaired immune response to EBV and let the virus to infect the nasopharyngeal epithelium, leading to NPC. Our study has identified several NPC susceptibility genes including *CYP2E1*, *EPHX2*, *SPLUNC1* and *ITGA9* in Malaysian Chinese. These NPC susceptibility genes are likely to be involved in the earliest stage of NPC tumorigenesis. They may synergistically work together, along with the effect of environmental factors and EBV infection to develop NPC.

Numerous studies have suggested that smoking, consumption of salted fish and other preserved food which contain high level of nitrosamines are associated with increased risk of NPC (Ward *et al.*, 2000; Yu and Yuan, 2002) and CYP2E1 is involved in the metabolic activation of nitrosamines or other carcinogens. Higher level of CYP2E1 in nasopharynx may increase the risk of developing NPC by activating more carcinogens and leading to the formation of DNA adduct. On the other hand, lower expression of EPHX2 causes the accumulation of epoxides which may promote the neoplatic properties of carcinoma cells and stimulate angiogenesis. Accumulations of these DNA adduct may disrupt genomic integrity and cause genetic damages in the nasopharyngeal epithelial cells, resulting in low-grade pre-invasive lesions which may susceptible to EBV infection, and later facilitate the promotion of cancer development.

EBV infection may lead to the development of severe dysplasia. EBV can infect B cells and epithelial cells. ITGA9 which associated with β1 may interact with a consensus domain of EBV capsid protein for attachment and/or enter the cell. The viral genes may regulate the cellular gene expression, cell growth and differentiation, leading to formation of invasive and malignant tumour. SPLUNC1, a secreted protein, is important in the early phase of NPC as it is proposed to take part in the immune response against to EBV. It may disrupt the EBV infected cells and inhibit the oncogenicity of EBV. Indeed, SPLUNC1 was found to be down-regulated in NPC biopsy samples (Zhang *et al.*, 2003). Our study also suggested that individuals with susceptible allele A of SNP rs1407019 may have lower expression of *SPLUNC1* by altering the transcription factor binding site for Sp1. Lower expression of *SPLUNC1* may lead to loss of innate immunity, increasing the risk of developing NPC. Accumulation of sufficient genetic changes after EBV infection may result in NPC carcinogenesis.

NPC is resulted from the combined effects of genes, environmental factors, EBV infections and their interaction. Understanding the relation between these factors may help to identify the high risk subgroups in a population and provide better insight into pathogenesis of NPC. Figure 4.1 is just a proposed model of the involvement of the identified susceptibility genes in the NPC pathogenesis which involved environmental factors and EBV infections. Additional analyses can be performed to evaluate the interaction between the NPC susceptibility genes, environment, diet and EBV infection in the future.

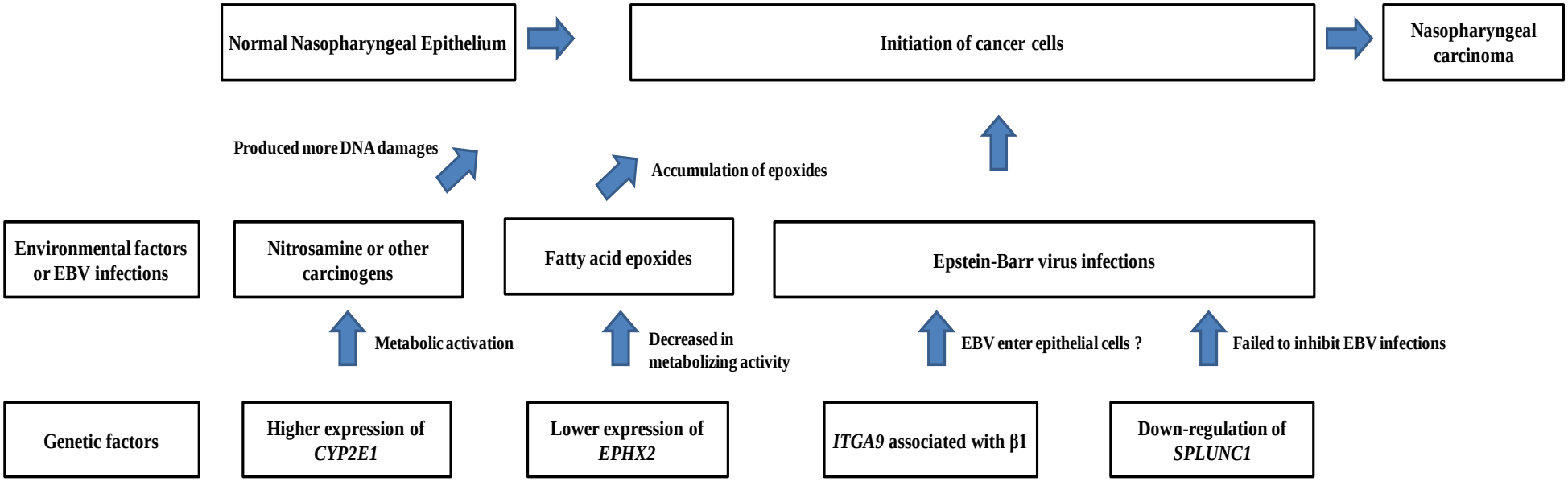


Figure 4.1: Proposed roles of the identified NPC susceptibility genes in pathogenesis.

The identified susceptibility genes may work together with environmental factors and EBV infection in the pathogenesis of NPC. Higher expression of CYP2E1 activated more carcinogens including nitrosamines and lower expression of EPHX2 accumulated more epoxides, producing more DNA damages and lead to the formation of tumours. On the other hand, *ITGA9* which associated with β1 might mediate the entry of EBV to epithelial cells. Moreover, down-regulation of *SPLUNC1* will increase the risk of developing NPC as it failed to inhibit EBV oncogenicity.

4.10 SUMMARY AND FUTURE WORKS

In the present study, candidate gene and genome-wide screening approaches were used to investigate the association between various genetic polymorphisms and NPC susceptibility in Malaysian Chinese. We have generated the genetic data including allele and genotype frequencies of numerous genes for NPC patients and healthy controls in the Malaysian Chinese population. In summary, our study further supported that *CYP2E1* and *SPLUNC1* polymorphisms are associated with NPC susceptibility and identified a functional SNP in *SPLUNC1* which may contribute to the development of NPC. In addition, we have identified two novel NPC susceptibility genes, *ITGA9* and *EPHX2*. Overall, the outcome of this project may contribute to the continuing exploration of genetic susceptibility of NPC.

Further studies would be needed to replicate the results in this study with larger sample size or different populations to confirm the findings. In addition, evaluation by meta-analysis using our results and other studies may be performed in order to increase the statistical power and overcome the limitations of individual studies. NPC is a multifactorial and polygenic disease, it might associate with combination of genes or environmental factors or EBV infection. Assessing the epistasis/gene-gene, gene-environment or gene-environment-EBV interaction may yield a larger effect size on the disease susceptibility (Hunter, 2005; Cordell, 2009), and further improve and enhance our understanding of NPC pathogenesis. Several methods (e.g. multifactor dimensionality reduction (MDR) and binary logistic regression) can be performed to detect epistasis or gene-environmental interaction.

In addition, copy number variations (CNVs), another form of common genetic variant apart from SNP, should be considered as an important issue related to the etiology of a complex disease. CNVs arise when there is duplication, insertion or deletion of chromosomal segments which are ≥ 1kb (Redon *et al.*, 2006). A recent study

have identified eight regions with CNV including six deletions (on chromosomes 3, 6, 7, 8, 19) and two duplications (on chromosomes 7 and 12) which were associated with NPC susceptibility in Taiwanese population (Tse *et al.*, 2011). Therefore information of CNVs in our GWAS can be utilized to further identify NPC susceptibility regions.

Identification of NPC susceptibility genes is an intriguing finding. Once genetic association was confirmed, further understanding of their importance and roles in the development of NPC are necessary. Further studies such as chromatin immunoprecipitation (ChIP) assay, luciferase reporter assay or EMSA can be performed to determine the functional consequences of the significant associated SNPs and how they involved in NPC susceptibility. In addition, functional studies such as gene knockout can be carried out to investigating the roles of the genes in the NPC pathogenesis. In general, the ultimate goal for a NPC research is prevention, earlier diagnosis and more effective treatment. Thus, a better understanding of the pathogenesis of NPC may lead to the prevention, improvement of diagnosis or treatment in the future.

REFERENCES:

Aden, D.P., Fogel, A., Plotkin, S., Damjanov, I. and Knowles, B.B. (1979). Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature*, *282*(5739), 615-616.

American Society of Clinical Ooncology. (2003). Nasopharyngeal cancer. Retrieve from http://www.cancer.net/patient/Cancer+Types/Nasopharyngeal+Cancer/ci.Nasopharyngeal+Cancer.printer.

Anedchenko, E.A., Dmitriev, A.A., Krasnov, G.S., Kondrat'eva, T.T., Kopantsev, E.P., Vinogradova, T.V., *et al.* (2008). Down-regulation of RBSP3/CTDSPL, NPRL2/G21, RASSF1A, ITGA9, HYAL1 and HYAL2 genes in non-small cell lung cancer. *Mol. Biol. (Mosk), 42*(6), 965-976.

Armstrong, R.W., Imrey, P.B., Lye, M.S., Armstrong, M.J., Yu, MC. and Sani S. (1998). Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. *Int. J. Cancer, 77*, 228-235.

Armstrong, R.W., Imrey, P.B., Lye, M.S., Armstrong, M.J., Yu, M.C. and Sani, S. (2000). Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles formaldehyde and heat. *Int. J. Epidemiol*., *29*, 991-998.

Balding, D.J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet., 7*(10), 781-791.

Barnes, L., Eveson, J.W., Reichart, P. and Sidransky, D. (eds). (2005). World Healthy Organization Classification of Tumours. Pathology and genetics of head and neck tumours. IARC press, Lyon.

Baujat, B., Audry, H., Bourhis, J., Chan, A.T., Onat, H., Chua, D.T., *et al.* (2006). Chemotherapy as an adjunct to radiotherapy in locally advanced nasopharyngeal carcinoma. *Cochrane. Database Syst. Rev*., (4), CD004329.

Beetham, J.K., Tian, T. and Hammock, B.D. (1993). cDNA cloning and expression of a soluble epoxide hydrolase from human liver. *Arch. Biochem. Biophys., 305*, 197–201.

Bei, J., Li, Y., Jia, W., Feng, B., Zhou, G., Chen, L., *et al.* (2010). A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat. Genet., 42*(7), 599-603.

Benhamou, S., Reinikainen, M., Bouchardy, C., Dayer, P. and Hirvonen, A. (1998). Association between lung cancer and microsomal epoxide hydrolase genotypes. *Cancer Res., 58*(23), 5291-5293.

Bingle, C.D. and Bingle, L. (2000). Characterization of the human plunc gene, a gene product with an upper airways and nasopharyngeal restricted expression pattern. *Biochim. Biophys. Acta., 1493*, 363-367.

Bingle, C.D. and Craven, C.J. (2002). PLUNC: a novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. *Hum. Mol. Genet*., *11*, 937-943.

Bingle, C.D. and Craven, C.J. (2003). Comparative analysis of the PLUNC (palate, lung and nasal epithelium clone) protein families. *Biochem. Soc. Trans.*, *31*, 806-809.

Blair, A., Stewart, P., O'Berg, M., Gaffey, W., Walrath, J., Ward, J. *et al.* (1986). Mortality among industrial workers exposed to formaldehyde. *J. Natl. Cancer Inst.*, *76*(6)**,** 1071-1084.

Borden, E.C., Hogan, T.F. and Voelkel, J.G. (1982). Comparative antiproliferative activity in vitro of natural interferons alpha and beta for diploid and transformed human cells. *Cancer Res.*, *42*(12), 4948-4953.

Bozina, N., Bradamante, V. and Lovrić, M. (2009). Genetic polymorphism of metabolic enzymes P450 (CYP) as a susceptibility factor for drug response, toxicity, and cancer risk. *Arh. Hig. Rada. Toksikol*., *60*(2), 217-242.

Braathen, R., Sandvik, A., Berntzen, G., Hammerschmidt, S., Fleckenstein, B., Sandlie, I., *et al.* (2005).Identification of a Polymeric Ig Receptor binding phage-displayed peptide that exploits epithelial transcytosis without dimeric IgA competition. *The Journal of Biological Chemistry*, *281*(11), 7075-7081.

Brennan, B. (2006). Nasopharyngeal carcinoma. *Orphanet Journal of Rare Diseases*, *1*(23), doi:10.1186/1750-1172-1-23.

Brock, S.C., McGraw, P.A., Wright, P.F. and Crowe Jr., J.E. (2002). The human polymeric immunoglobulin receptor facilitates invasion of epithelial cells by *Streptococcus pneumoniae* in a strain specific and cell type-specific manner. *Infect. Immun*, *70*(9), 5091-5095.

Brookes, A.J. (1999). The essence of SNPs. *Gene*, *234*(2), 177-186.

Burke, Y.D., Stark, M.J., Roach, S.L., Sen, S.E. and Crowell, P.L. (1997). Inhibition of pancreatic cancer growth by the dietary isoprenoids farnesol and geraniol. *Lipids*, *32*(2), 151-156.

Campbell, P. C. (1971). Chinese coolie emigration to countries within the British Empire. New York, Negro Universities Press.

Candiano, G., Bruschi, M., Pedemonte, N., Musante, L., Ravazzolo, R., Liberatori, S., *et al.* (2007). Proteomic analysis of the airway surface liquid: modulation by proinflammatory cytokines. *Am. J. Physiol. Lung Cell Mol. Physiol*., *292*(1), L185-198.

Cao, Y., Miao, X., Zeng, Y., Lin, D. and Shao, J. (2011). Association between genetic polymorphisms of CYP2A13, CYP2A6 and risk of nasopharyngeal carcinoma in southern Chinese population. *Cancer Biol.*, *1*(1).

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., *et al.* (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet*., *22*(3), 231-238.

Carlsten, C., Sagoo, G.S., Frodsham, A.J., Burke, W. and Higgins, J.P. (2008). Glutathione S-transferase M1 (GSTM1) polymorphisms and lung cancer: a literature-based systematic HuGE review and meta-analysis. *Am. J. Epidemiol.*, *167*(7), 759-774.

Catarino, R.J., Breda, E., Coelho, V., Pinto, D., Sousa, H., Lopes, C., *et al.* (2006). Association of the A870G cyclin D1 gene polymorphism with genetic susceptibility to nasopharyngeal carcinoma. *Head Neck*, *28*(7**)**, 603-608.

Casado, B., Pannell, L.K., Iadarola, P. and Baraniuk, J.N. (2005). Identification of human nasal mucous proteins using proteomics. *Proteomics*, *5*(11), 2949-2959.

Casanova, J.E., Apodaca, G. and Mostov, K.E. (1991). An autonomous signal for basolateral sorting in the cytoplasmic domain of the polymeric immunoglobulin receptor. *Cell*, *66*(1), 65-75.

Chan, A.T.C., Teo, P.M.L. and Johnson, P.J. (2002). Nasopharyngeal Carcinoma. *Ann. Oncol.*, *13*, 1007-1015.

Chan, A.T. (2005). Head and neck cancer: treatment of nasopharyngeal cancer. *Ann. Oncol.*, 16Suppl 2, ii265-8.

Chander, R. (1972). 1970 population and housing census of Malaysia: community groups. Kuala Lumpur: Department of Statistics.

Chang, C.P., Stankunas, K., Shang, C., Kao, S.C., Twu, K.Y. and Cleary, M.L. (2008). Pbx1 functions in distinct regulatory networks to pattern the great arteries and cardiac outflow tract. *Development*, *135*(21), 3577-3586.

Chang, L., Chung, J., Hong, Y., Ichimura, T., Nakao, Mitsuyoshi. and Liu, S. (2005). Activation of Sp1-mediated transcription by Rta of Epstein-Barr virus via an interaction with MCAF1. *Nucleic Acids Res.*, *33*, 6528-6539.

Chanock, S. (2001). Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis. Markers*, *17*(2), 89-98.

Chen, C.J., Liang, K.Y., Chang, Y.S., Wang, Y.F., Hsieh, T., Hsu, M.M., et al. (1990). Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency. *Anticancer Res*., *10*(2B), 547-553.

Chan, A.T., Lo, Y.M., Zee, B., Chan, L.Y., Ma, B.B., Leung, S.F., *et al.* (2002). Plasma Epstein-Barr virus DNA and residual disease after radiotherapy for undifferentiated nasopharyngeal carcinoma. *J. Natl. Cancer Inst*., *94*(21), 1614-1619.

Chen, C.L., Liu, Q., Pui, C.H., Rivera, G.K., Sandlund, J.T., Ribeiro, R., *et al.* (1997). Higher frequency of glutathione S-transferase deletions in black children with acute lymphoblastic leukemia. *Blood, 89*(5), 1701-1707.

Cheng, Y.J., Hildesheim, A., Hsu, M.M., Chen, I.H., Brinton, L.A., Levine, P.H., *et al.* (1999). Cigarette smoking, alcohol consumption and risk of nasopharyngeal carcinoma in Taiwan. *Cancer Causes Control, 10*(3), 201-207.

Cheng, Y.J., Chien, Y.C., Hildesheim, A., Hsu, M.M., Chen, I.H., Chuang, J., *et al.* (2003). No association between genetic polymorphisms of CYP1A1, GSTM1, GSTT1, GSTP1, NAT2, and nasopharyngeal carcinoma in Taiwan. *Cancer Epidemiol. Biomarkers Prev.*, *12*(2), 179-180.

Cheung, C., Akiyama, T.E., Kudo, G. and Gonzalez, F.J. (2003). Hepatic expression of cytochrome P450s in hepatocyte nuclear factor 1-alpha (HNF1alpha)-deficient mice. *Biochem. Pharmacol*., *66*(10), 2011-2020.

Cho, E.Y., Hildesheim, A., Chen, C.J., Hsu, M.M., Chen, I.H., Mittl, B.F., *et al.* (2003) Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. *Cancer Epidemiol. Biomarkers Prev.*, *12*(10), 1100-1104.

Chow, L.S., Lo, K.W., Kwong, J., To, K.F., Tsang, K.S., Lam, C.W., et al. (2004). RASSF1A is a target tumor suppressor from 3p21.3 in nasopharyngeal carcinoma. *Int. J. Cancer, 109*(6), 839-847.

Chua, D.T., Sham, J.S., Kwong, D.L. and Au, G.K. (2003). Treatment outcome after radiotherapy alone for patients with Stage I-II nasopharyngeal carcinoma. *Cancer*, *98*(1), 74-80.

Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., *et al*. (2005). Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet*., *37*(11), 1243-1246.

Cordell, H.J. (2009). Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet., 10*(6), 392-404.

Cordell, H.J. and Clayton, D.G. (2005). Genetic association studies. *Lancet., 366*(9491), 1121-1131.

Crofts, F., Taioll, E., Trachman, T., Cosma, G.N., Currie, D., Toniolo, P., *et al.* (1994). Functional significance of different human CYP1A1 genotypes. *Carcinogenesis, 15*(12), 2961-2963.

Cross, D.S., Ivacic, L.C., Stefanski, E.L. and McCarty, C.A. (2010). Population based allele frequencies of disease associated polymorphisms in the Personalized Medicine Research Project. *BMC Genet., 11,* 51.

Curado, M.P., Edwards, B., Shin, H.R., Storm, H, Ferlay, J., Heanue, M., *et al.* (eds.). (2007). Cancer Incidence in Five Continents Vol. IX. Lyon, France: IARC Scientific Publications No. 160.

Danesi, G., Zanoletti, E., Mazzoni, A. (2007). Salvage surgery for recurrent nasopharyngeal carcinoma. *Skull Base, 17*(3), 173-180.

Danko, I.M. and Chaschin, N.A. (2005).Association of CYP2E1 gene polymorphism with predisposition to cancer development. *Exp. Oncol., 27*(4), 248-256.

De Bakker, P.I., Graham, R.R., Altshuler, D., Henderson, B.E. and Haiman, C.A. (2006). Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations. *Pac. Symp. Biocomput., 11,* 478-486.

Decker, M., Arand, M. and Cronin, A. (2009). Mammalian epoxide hydrolases in xenobiotic metabolism and signalling. *Arch. Toxicol., 83*(4), 297-318.

Deng, Z., Wei, Y., Luo, W., Liao, Z. and Ma, Y. (2005). Glutathione S- transferase M1 and T1 gene deletion associated with increased susceptibility to nasopharyngeal carcinoma. *The Chinese-German Journal of Clinical Oncology*, *4*(5), 276–278.

Deschler, D.G. and Day, T. (2008). TNM staging of head and neck cancer and neck dissection classification (3rd ed.). American academy of Head and Neck Surgery.

de Waziers, I., Cugnenc, P.H., Yang, C.S., Leroux, J.P. and Beaune, P.H. (1990). Cytochrome P 450 isoenzymes, epoxide hydrolase and glutathione transferases in rat and human hepatic and extrahepatic tissues. *J. Pharmacol. Exp. Ther., 253*(1), 387-394.

Di, Y.P., Harper, R., Zhao, Y., Pahlavan, N., Finkbeiner, W. and Wu, R. (2003). Molecular cloning and characterization of spurt, a human novel gene that is retinoic acid-inducible and encodes a secretory protein specific in upper respiratory tracts. *J. Biol. Chem., 278,* 1165-1173.

Dib, J.C., Dube, M., Kelly, C., Rinaldi, M.G. and Patterson, J.E. (1996). Evaluation of pulsed-field gel electrophoresis as a typing system for Candida rugosa: comparison of karyotype and restriction fragment length polymorphisms. *J. Clin. Microbiol., 34*(6), 1494-1496.

Ding, X. and Kaminsky, L.S. (2003). Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. *Annu. Rev.Pharmacol. Toxicol*., *43*, 149-173.

Edge, S.B., Byrd, D.R., Compton, C.C., Fritz, A.G., Greene, F.L., Trotti, A., *et al.* (eds.). (2010). AJCC Cancer Staging Manual (7th ed.). New York, NY: Springer, pp 41-49.

Edwards, P.A. and Ericsson, J. (1999). Sterols and isoprenoids: signaling molecules derived from the cholesterol biosynthetic pathway. *Annu. Rev. Biochem.*, *68*,157-185.

Elsbach, P. and Weiss, J. (1998). Role of the bactericidal/permeability-increasing protein in host defence. *Curr. Opin. Immunol*., *10*(1), 45-49.

Enayetallah, A.E., French, R.A., Barber, M. and Grant, D.F. (2006). Cell-specific subcellular localization of soluble epoxide hydrolase in human tissues. *J. Histochem. Cytochem.*, *54*(3), 329-335.

Enayetallah, A.E. and Grant, D.F. (2006). Effects of human soluble epoxide hydrolase polymorphisms on isoprenoid phosphate hydrolysis. *Biochem. Biophys. Res. Commun.*, *341*(1), 254-260.

Fan, Q., Jia, W.H., Zhang, R.H., Yu, X.J., Chen, L.Z., Feng, Q. S., *et al.* (2005). Correlation to Polymeric Immunoglobulin receptor gene polymorphisms to susceptibility of nasopharyngeal carcinoma. *Chinese journal of cancer*, *24*(8), 915-918.

Fändrich, F., Degiuli, B., Vogel-Bindel, U., Arand, M. and Oesch, F. (1995). Induction of rat liver microsomal epoxide hydrolase by its endogenous substrate 16 alpha, 17 alpha-epoxyestra-1,3,5-trien-3-ol. *Xenobiotica*, *25*(3), 239-244.

Feng, B.J., Huang, W., Shugart, Y.Y., Lee, M.K., Zhang, F., Xia, J.C., *et al.*(2002). Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat. Genet.*, *31*, 395-399.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010)a. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, *127*(12), 2893-2917.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010)b. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No.10 [Internet]. Lyon, France: International Agency for Research on Cancer; Available from: http://globocan.iarc.fr.

Flach, C.F., Qadri, F., Bhuiyan, T.R., Alam, N.H., Jennische, E., Lönnroth, I., *et al.* (2007). Broad up-regulation of innate defence factors during acute cholera. *Infect. Immun.*, *75*(5), 2343-2350.

Gan, Y.J., Chodosh, J., Morgan, A. and Sixbey, J.W. (1997). Epithelial cell polarization is a determinant in the infectious outcome of immunoglobulin A-mediated entry by Epstein-Barr virus. *J. Virol.*, *71*(1), 519-526.

Ghafouri, B., Kihlström, E., Tagesson, C. and Lindahl, M. (2004). PLUNC in human nasal lavage fluid: multiple isoforms that bind to lipopolysaccharide. *Biochim. Biophys. Acta.*, *1699*(1-2), 57-63.

Geisler, S.A. and Olshan, A.F. (2001). GSTM1, GSTT1, and the risk of squamous cell carcinoma of the head and neck: A mini-Huge review. *Am. J. Epidemiol.*, *154* (2), 95-105.

Georgiadis, P., Topinka, J., Vlachodimitropoulos, D., Stoikidou, M., Gioka, M., Stephanou, G., *et al.* (2005). Interactions between CYP1A1 polymorphisms and exposure to environmental tobacco smoke in the modulation of lymphocyte bulky DNA adducts and chromosomal aberrations. *Carcinogenesis*, *26*(1), 93-101.

Goldsmith, D.B., West, T.M. and Morton, R. (2002). HLA associations with nasopharyngeal carcinoma in Southern Chinese: a meta-analysis. *Clin. Otolaryngol. Allied Sci*., *27*(1), 61-67.

Goldstein, M.N., Slotnick, I.J. and Journey, L.J. (1960). In vitro studies with HeLa cell line sensitive and resistant to actinomycin D. *Ann. N. Y. Acad. Sci.*, *89*, 474-483.

Gu, M.L. and Chu, J.Y. (2007). Construction of haplotype and haplotype block based on tag single nucleotide polymorphisms and their applications in association studies. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, *24*(6), 660-665.

Guo, W. and Giancotti, F.G. (2004). Integrin signalling during tumour progression. *Nat. Rev. Mol. Cell Biol.*, *5*(10), 816-826.

Guo, X., J. O'Brien, S., Zeng, Y., W. Nelson, G. and A. Winkler, C. (2008) GSTM1 and GSTT1 gene deletions and the risk for nasopharyngeal carcinoma in Han Chinese. *Cancer Epidemiol. Biomarkers Prev*., *17*(7), 1760-1763.

Guo, X., Zeng, Y., Deng, H., Liao, J., Zheng, Y., Li, J., *et al.* (2010). Genetic polymorphisms of CYP2E1, GSTP1, NQO1 and MPO and the Risk of Nasopharyngeal Carcinoma in a Han Chinese Population of Southern China. *BMC Res Notes*, *3*, 212.

Hadar, T., Rahima, M., Kahan, E., Sidi, J., Rakowsky, E., Sarov, B., *et al.* (1986). Significance of specific Epstein-Barr virus IgA and elevated IgG antibodies to viral capsid antigens in nasopharyngeal carcinoma patients. *J. Med. Virol., 20*(4), 329-339.

Hardy, G.H. (1908). Mendelian proportions in a mixed population. *Science, 28*(706), 49-50.

Harrison, D.J., Hubbard, A.L., MacMillan, J., Wyllie, A.H. and Smith, C.A. (1999). Microsomal epoxide hydrolase gene polymorphism and susceptibility to colon cancer. *Br. J. Cancer, 79*(1), 168-171.

Hayes, J.D. and Pulford, D.J. (1995). The Glutathione S-Transferase Supergene Family: Regulation of GST* and the Contribution of the lsoenzymes to Cancer Chemoprotection and Drug Resistance. *Biochemistry and Molecular Biology, 30*(6), 445-600.

He, Y., Zhou, G., Zhai, Y., Dong, X., Lv, L., He, F. and Yao, K. (2005). Association of PLUNC gene polymorphisms with susceptibility to nasopharyngeal carcinoma in a Chinese population. *J. Med. Genet., 42*, 172-176.

He, Z.W., Xie, L., Xu, L.G., Zhang, L., Lan, K., Ren, C.P., *et al.* (2000). Cloning of a novel gene associated with human nasopharyngeal carcinoma. *Chin. Sci. Bull., 45*, 2267-2272.

He, Z.M., Wang, S.L., Yuan, J.H. and Chen, Z.C. (2002). Expression of human cytochrome P450 2E1 gene in embryonic nasopharynx, nasopharyngeal cancer cell lines and tissue. *Ai Zheng, 21*(6), 597-600.

Henle, W., Henle, G., Ho, H.C., Burtin, P., Cachin, Y., Clifford, P., *et al.* (1970). Antibodies to Epstein-Barr virus in nasopharyngeal carcinoma, other head and neck neoplasms, and control groups. *J. Natl. Cancer. Inst., 44*(1), 225-231.

Her, C. (2001).Nasopharyngeal Cancer and the Southeast Asian Patient. *American Family Physician, 63*(9), 1776-1782.

Hildesheim, A. and Levine, P.H. (1993). Etiology of Nasopharyngeal Carcinoma: A review. *Epidemiologic Review, 15*, 466-485.

Hildesheim, A., Chen, C.J., Caporaso, N.E., Cheng, Y.J., Hoover, R.N., Hsu, M.M., *et al.* (1995). Cytochrome P4502E1 genetic polymorphisms and risk of nasopharyngeal carcinoma: results from a case-control study conducted in Taiwan. *Cancer Epidemiol. Biomarkers Prev., 4*(6), 607-610.

Hildesheim, A., Anderson, L.M., Chen, C.J., Cheng, Y.J., Brinton, L.A., Daly, A.K., *et al.* (1997). CYP2E1 genetic polymorphisms and risk of Nasopharyngeal carcinoma in Taiwan. *J. Natl. Cancer Inst., 89*, 1207-1212.

Hildesheim, A., Apple, R.J., Chen, C.J., Wang, S.S., Cheng, Y.J., Klitz, W., *et al.* (2002). Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J. Natl. Cancer Inst., 94*(23), 1780-1789.

Hirschhorn, J.N, and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet*., *6*(2), 95-108.

Hirunsatit, R., Kongruttanachok, N., Shotelersuk, K., Supiyaphun, P., Voravud, N., Sakuntabhai, A., *et al.* (2003). Polymeric immunoglobulin receptor polymorphisms and risk of nasopharyngeal cancer. *BMC Genetics*, *4*.

Hong, J. Y. and Yang, C.S. (1997). Genetic polymorphism of Cytochrome P450 as a biomarker of susceptibility to environmental toxicity. *Environmental Health Perspectives*, *105*, 759-762.

Hosgood, H.D. 3rd, Berndt, S.I, and Lan, Q. (2007). GST genotypes and lung cancer susceptibility in Asian populations with indoor air pollution exposures: a meta-analysis. *Mutat. Res., 636*(1-3), 134-143.

Hou, D.F., Wang, S.L., He, Z.M., Yang, F. and Chen, Z.C. (2007). Expression of CYP2E1 in human nasopharynx and its metabolic effect in vitro. *Mol. Cell Biochem., 298*, 93-100.

Hsu, T.M., Law, S.M., Duan, S., Neri, B.P. and Kwok, P.Y. (2001). Genotyping single-nucleotide polymorphisms by the invader assay with dual-color fluorescence polarization detection. *Clin. Chem., 47*(8), 1373-1377.

Hu, S.P., Day, N.E., Li, D.R., Luben, R.N., Cai, K.L., Ou-Yang, T., *et al.* (2005). Further evidence for an HLA-related recessive mutation in nasopharyngeal carcinoma among the Chinese. *Br. J. Cancer*, *92*(5), 967-970.

Hukkanen, J., Hakkola, J., Anttila, S., Piipari, R., Karjalainen, A., Pelkonen, O., *et al.* (1997).Detection of mRNA encoding xenobiotic-metabolizing cytochrome P450s in human bronchoalveolar macrophages and peripheral blood lymphocytes. *Mol. Carcinog., 20*(2), 224-30.

Hunter, D.J. (2005). Gene-environment interactions in human diseases. *Nat. Rev. Genet*., *6*(4), 287-298.

Hutt-Fletcher, L.M. (2007). Epstein-Barr virus entry. *J. Virol*., *81*(15), 7825-32.

IARC. (1995). Wood dust and formaldehyde. Monographs on evaluation of the carcinogenic risk in human. Volume 62. Lyon: IARC.

International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature, 437*(7063), 1299-1320.

International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature, 449*(7164), 851-61.

International HapMap Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature, 467*(7311), 52-58.

International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/

Iskandar, Y. (1992). The Malay sultanate of Malacca: a study of various aspects of Malacca in the 15th and 16th centuries in Malaysian history. Dewan Bahasa dan Pustaka, Ministry of Education Malaysia.

Imig, J.D., Zhao, X., Capdevila, J.H., Morisseau, C. and Hammock, B.D. (2002). Soluble epoxide hydrolase inhibition lowers arterial blood pressure in angiotensin II hypertension. *Hypertension, 39,* 690–694.

Iwao, K., Watanabe, T., Fujiwara, Y., Takami, K., Kodama, K., Higashiyama, M., *et al.* (2001). Isolation of a novel human lung-specific gene, lunx, a potential molecular marker for detection of micrometastasis in non-small-cell lung cancer. *Int. J. Cancer, 91,* 433-437.

Jancova, P., Anzenbacher, P. and Anzenbacherova, E. (2010).Phase II Drug metabolizing enzymes. *Biomed. Pap. Med. Fac. Univ. Palacky. Olomouc Czech Repub., 154*(2), 103–116.

Jia, W.H., Collins, A., Zeng, Y.X., Feng, B.J., Yu, X.J., Huang, L.X., *et al.* (2005). 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*(2), 248-252.

Jia, W.H., Pan, Q.H., Qin, H.D., Xu, Y.F., Shen, G.P., Chen, L., *et al.* (2009). A case-control and a family-based association study revealing an association between *CYP2E1* polymorphisms and nasopharyngeal carcinoma risk in Cantonese. *Carcinogenesis, 30*(12), 2031-2036.

Jiang, J.H., Jia, W.H., Chen, H.K., Feng, B.J., Qin, H.D., Pan, Z.G., *et al.* (2004). Genetic polymorphisms of CYP2A13 and its relationship to nasopharyngeal carcinoma in the Cantonese population. *J. Transl. Med., 2*(1), 24.

Jiang, J.G., Chen, C.L., Card, J.W., Yang, S., Chen, J.X., Fu, X.N., *et al.* (2005). Cytochrome P450 2J2 promotes the neoplastic phenotype of carcinoma cells and is up-regulated in human tumors. *Cancer Res.*, *65*, 4707-4715.

Jiang, R.C., Qin, H.D., Zeng, M.S., Huang, W., Feng, BJ., Zhang, F., *et al.* (2006). A functional variant in the transcriptional regulatory phenotype in familial nasopharyngeal carcinoma. *Cancer Res.*, *66*(2), 693-700.

Johnson, G.C. and Todd, J.A. (2000). Strategies in complex disease mapping. *Curr.Opin.Genet. Dev.*, *10*(3), 330-334.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A., *et al*. (2010). Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*, *11*, 724.

Joshua Project database for Malaysia. (2010). Malaysia. Retrieve from http://www.joshuaproject.net/countries.php?rog3=MY.

Kathiresan, S., Newton-Cheh, C., Gerszten, R.E. (2004). On the interpretation of genetic association studies. *Eur. Heart J.*, *25*(16), 1378-1381.

Khanna, R., Burrows, S.R. and Moss, D.J. (1995). Immune regulation in Epstein-Barr virus-associated diseases. *Microbiol. Rev*., *59*(3), 387–405.

Kim, H., Lee, S., Park, B., Che, L. and Lee, S. (2010). Sp1 and Sp3 mediate basal and serum-induced expression of human CENP-W. *Mol. Biol. Rep*., *37*, 3593-3600.

Kirk, B.W., Feinsod, M., Favis, R., Kliman, R.M. and Barany, F. (2002). Single nucleotide polymorphism seeking long term association with complex disease. *Nucleic Acids Res.*, *30*(15), 3295-3311.

Kiyohara, C., Takayama, K and Nakanishi, Y. (2005). CYP2A13, CYP2A6, and the risk of lung adenocarcinoma in a Japanese population. *J. Health Sci*., *51*(6), 658-666.

Kongruttanachok, N., Sukdikul, S., Setavarin, S., Kerekhjanarong, V., Supiyaphun, P., Voravud, V., *et al.* (2001). Cytochrome P450 2E1 polymorphisms and nasopharyngeal carcinoma development in Thailand: a correlative study. *BMC cancer*, *10*, 652.

Krajci, P., Gedde-Dahl, T. Jr,, Høyheim, B., Rogde, S., Olaisen, B. and Brandtzaeg, P. (1992). The gene encoding human transmembrane secretory component (locus PIGR) is linked to D1S58 on chromosome 1. *Hum. Genet*., *90*(3), 215-219.

Kroetz, D.L. and Zeldin, D.C. (2002). Cytochrome P450 pathways of arachidonic acid metabolism. *Curr. Opin. Lipidol.*, *13*(3), 273-283.

Kruglyak, L. and Nickerson, D.A. (2001). Variation is the spice of life. *Nat. Genet.*, *27*(3), 234-236

Laird, N.M. and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nat. Rev. Genet.*, *7*(5), 385-394.

Laasanen, J., Romppanen, E.L., Hiltunen, M., Helisalmi, S., Mannermaa, A., Punnonen, K., *et al.* (2002). Two exonic single nucleotide polymorphisms in the microsomal epoxide hydrolase gene are jointly associated with preeclampsia. *Eur. J. Hum. Genet.*, *10*(9), 569-573.

Lancaster, J.M., Brownlee, H.A., Bell, D.A., Futreal, P.A., Marks, J.R., Berchuck, A., et al. (1996).Microsomal epoxide hydrolase polymorphism as a risk factor for ovarian cancer. *Mol. Carcinog.*, *17*, 160–162.

Larocque, R.C., Sabeti, P., Duggal, P., Chowdhury, F., Khan, A.I., Lebrun, L.M., *et al.* (2009). A variant in long palate, lung and nasal epithelium clone 1 is associated with cholera in a Bangladeshi population. *Genes Immun*., *10*(3), 267-272.

Larsen, K., Madsen, L.B. and Bendixen, C. (2005). Porcine SPLUNC1: molecular cloning, characterization and expression analysis. *Biochim. Biophys. Acta*., *1727*(3), 220-226.

Lee, A.W., Sze, W.M., Au, J.S., Leung, S.F., Leung, T.W., Chua, D.T., *et al.* (2005). Treatment results for nasopharyngeal carcinoma in the modern era: the Hong Kong experience. *Int. J. Radiat. Oncol. Biol. Phys*., *61*(4), 1107-1116.

Lee, C.R., North, K.E., Bray, M.S., Fornage, M., Seubert, J.M., Newman, J.W., *et al.* (2006). Genetic variation in soluble epoxide hydrolase (EPHX2) and risk of coronary heart disease: The Atherosclerosis Risk in Communities (ARIC) study. *Hum. Mol. Genet*., *15*(10), 1640-1649.

Lee, K.H. and Tan, C.B. (eds.). (2000). The Chinese in Malaysia. Kuala Lumpur: Oxford University Press.

Lee, L.K., Tan, E.L., Gopala, K. and Sam, C.K. (2007). Human leukocyte class I antigen alleles A2 and A11 are not associated with nasopharyngeal carcinoma in West Malaysia. *Singapore Med. J*., *48*(7), 632-634.

Lee, Y.C., Hwang, Y.C., Chen, K.C., Lin, Y.S., Huang, D.Y., Huang, T.W., *et al.* (2007a). Effect of Epstein-Barr virus infection on global gene expression in nasopharyngeal carcinoma. *Funct. Integr. Genomics*, *7*(1), 79-93.

Leong, J.L., Fong, K.W. and Low, W.K. (1999). Factors contributing to delayed diagnosis in nasopharyngeal carcinoma. *J. Laryngol. Otol.*, *113*(7), 633-636.

Leung, S.F., Tam, J.S., Chan, A.T., Zee, B., Chan, L.Y., Huang, D.P., *et al.* (2004). Improved accuracy of detection of nasopharyngeal carcinoma by combined application of circulating Epstein-Barr virus DNA and anti-Epstein-Barr viral capsid antigen IgA antibody. *Clin. Chem*., *50*(2), 339-345.

Lewin, B. (2003). Genes VIII. USA: Prentice Hall.

Liang, H., Chen, H., Shen, Y., Fenf, Q., Jin, W., Huang, W., *et al.* (2004). A rare polymorphism of the COX7B2 gene in a Cantonese family with nasopharyngeal carcinoma. *Science in China Ser. C Life Sciences*, *47*(5), 449-453.

Li, X., Ghandri, N., Piancatelli, D., Adams, S., Chen, D., Robbins, F.M., *et al.* (2007). Associations between HLA class I alleles and the prevalence of nasopharyngeal carcinoma (NPC) among Tunisians. *J. Transl. Med.*, *5*, 22.

Liao, L.H., Zhang, H., Lai, M.P., Chen, S.L., Wu, M. and Shen, N. (2011). Single-nucleotide polymorphisms and haplotype of CYP2E1 gene associated with systemic lupus erythematosus in Chinese population. *Arthritis. Res. Ther*., *13*(1), R11.

Lim, G.C.C., Yahaya, H. and Lim, T.O. (eds.). (2003). The first report of the national cancer registry cancer incidence in Malaysia 2002. National Cancer Registry, Ministry of Health Malaysia.

Lim, G.C.C. and Yahaya, H (2004). Second Report of the National Cancer Registry. Cancer Incidence in Malaysia 2003. National Cancer Registry. Kuala Lumpur.

Lin, C.T., Lin, C.R., Tan, G.K., Chen, W., Dee, A.N. and Chan, W.Y. (1997). The mechanism of Epstein-Barr virus infection in nasopharyngeal carcinoma cells. *Am. J. Pathol.* , *150*(5), 1745-1756.

Lin, C.Y., Tsai, S.T., Jin, Y.T., Yang, M.W., Yeh, I.C. and Hsiao, J.R. (2008). Outcome of surgical management of persistent or recurrent neck mass in patients with nasopharyngeal carcinoma after radiotherapy. *Eur. Arch. Otorhinolaryngol.*, 265 Suppl 1, S69-74.

Lin, T.M., Yang, C.S., Chiou, J.F., Tu, S.M., Chen, T.Y., Tu, Y.C., *et al.* (1977). Antibodies to Epstein-Barr virus capsid antigen and early antigen in nasopharyngeal carcinoma and comparison groups. *Am. J. Epidemiol*., *106*(4), 336-339.

Lindah, M., Stahlbom, B. and Tagesson, C. (2001). Identification of a new potential airway irritation marker, palate lung nasal epithelial clone protein, in human nasal lavage fluid with two-dimensional electrophoresis and matrix-assisted laser desorption/ionization-time of flight. *Electrophoresis*, *22*, 1795-1800.

Lo, Y. M. (2001a). Quantitative analysis of Epstein-Barr virus DNA in plasma and serum: applications to tumour detection and monitoring. *Ann. N. Y. Acad. Sci*., *945*, 68-72.

Lo, K.W., Kwong, J., Hui, A.B., Chan, S.Y., To, K.F., Chan, A.S., *et al.*, (2001b). High frequency of promoter hypermethylation of RASSF1A in nasopharyngeal carcinoma. *Cancer Res*., *61*(10), 3877-3881.

Loh, L.E., Chee, T.S. and John, A.B. (1991). The anatomy of the Fossa of Rosenmuller—its possible influence on the detection of occult nasopharyngeal carcinoma. *Singapore Med. J.*, *32*(3), 154-155.

Lu, S.J., Day, N.E., Degos, L., Lepage, V., Wang, P.C., Chan, S.H. *et al.* (1990). Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature*, *346*, 470-471.

Marques, C.F., Koifman, S., Koifman, R.J., Boffetta, P., Brennan, P. and Hatagima, A. (2006). Influence of CYP1A1, CYP2E1, GSTM3 and NAT2 genetic polymorphisms in oral cancer susceptibility: results from a case-control study in Rio de Janeiro. *Oral Oncol*., *42*(6), 632-637.

Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., *et al.* (2004). Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat. Methods, 2*, 109-111.

McGlynn, K.A., Rosvold, E.A., Lustbader, E.D., Hu, Y., Clapper, M.L., Zhou, T., *et al.* (1995). Susceptibility to hepatocellular carcinoma is associated with genetic variation in the enzymatic detoxification of aflatoxin B1. *Proc. Natl. Acad. Sci. USA*, *92*, 2384–2387.

Mcllwain, C.C., Townsend, D.M. and Tew, K.D. (2006). Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene*, *25*, 1639-1648.

Manolio, T.A. (2010). Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med*., *363*(2), 166-176.

Marquard, V., Beckmann, L., Heid, I.M., Lamina, C. and Chang-Claude, J. (2009). Impact of genotyping errors on the type I error rate and the power of haplotype-based association methods. *BMC Genet*., *10*, 3.

Mooney, L.V., Bell, D.A., Santella, R.M., Van Bennekum, A.M., Ottman, R., Paik, M., *et al.* (1997). Contribution of genetic and nutritional factors to DNA damage in heavy smokers. *Carcinogenesis*, *18*, 503-509.

Morrow, C.S., Cowan, K.H., Goldsmith, M.E. (1989). Structure of the human genomic glutathione S-transferase-pi gene. *Gene*, *75*(1), 3-11.

Mostov, K.E., Friedlander, M. and Blobel, G. (1984). The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. *Nature, 308*(5954), 37-43.

Moyer, A.M., Salavaggione, O.E., Wu, T., Moon, I., Eckloff, B.W., Hildebrandt, M.A.T., *et al.* (2008). Glutathione S-transferase P1: Gene sequence variation and functional genomic studies. *Cancer Res., 68*(12), 4791-4801.

Mueller, J.C,, Lõhmussaar, E., Mägi, R., Remm, M., Bettecken, T., Lichtner, P., *et al.* (2005). Linkage disequilibrium patterns and tagSNP transferability among European populations. *Am. J. Hum. Genet., 76*(3), 387-398.

Nakachi, K., Imai, K., Hayashi, S. and Kawajiri, K. (1993). Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. *Cancer Res., 53*(13), 2994-2999.

Nakamura, Y. (2009). DNA variations in human and medical genetics: 25 years of my experience. *J. Hum. Genet., 54*(1),1-8.

Nam, J.M., McLaughlin, J.K. and Blot, W.J. (1992). Cigarette smoking, alcohol, and nasopharyngeal carcinoma: a case-control study among U.S. whites. *J. Natl. Cancer Inst., 84*(8), 619-622.

Nazar-Stewart, V., Vaughan, T.L., Burt, R.D., Chen, C., Berwick, M. and Swanson, G.M. (1999). Glutathione S-transferase M1 and susceptibility to nasopharyngeal carcinoma. *Cancer Epidemiol. Biomarkers Prev., 8*(6), 547-551.

NCBI dbSNP database, http://www. ncbi.nlm.nih.gov/projects/SNP/

NCBI BLAST tool, http://www.ncbi.nlm.nih.gov/BLAST/

Nelson, D.R. (2009). The Cytochrome P450 Homepage.*Human Genomics, 4*, 59-65.

Nickerson, D.A., Tobe, V.O. and Taylor, S.L. (1997). Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res., 23*, 1294-1296.

Niedobitek, G. (2000). Epstein-Barr virus infection in the pathogenesis of nasopharyngeal carcinoma. *J. Clin. Pathol: Mol. Pathol., 53*, 238-254.

Nonoyama, M., Huang, C.H., Pagano, J.S., Klein, G. and Singh, S. (1973). DNA of Epstein-Barr virus detected in tissue of Burkitt's lymphoma and nasopharyngeal carcinoma. *Proc. Natl. Acad. Sci. U S A, 70*(11), 3265-3268.

Norderhaug, I.N., Johansen, F.E., Schjerven, H. and Brandtzaeg, P. (1999). Regulation of the formation and external transport of secretory immunoglobulins. *Crit. Rev. Immunol., 19*(5-6), 481-508.

Nuoso, K., Thorgeirsson, S.S. and Battula, N. (1992). Stable Expression of Human Cytochrome P450IIE1 in Mammalian Cells: Metabolic Activation of Nitrosodimethylamine and Formation of Adducts with Cellular DNA. *Cancer Res., 52*, 1796-1800.

Oesch, F., Herrero, M.E., Hengstler, J.G., Lohmann, M. and Arand, M. (2000). Metabolic detoxification: implications for thresholds. *Toxicol. Pathol*., *28*(3), 382-387.

Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. (2001). A high-throughput SNP typing system for genome wide association studies. *J. Hum. Genet., 46*, 471-477.

Old, L.J., Boyse, E.A., Oettgen, H.F., Harven, E.D., Geering, G., Williamson, B., *et al.* (1966). Precipitating antibody in human serum to an antigen present in cultured Burkitt's lymphoma cells. *Proc. Natl. Acad. Sci. U S A, 56*, 1699–1704.

Olivier, M. (2005). The Invader assay for SNP genotyping. *Mutat. Res*., 573, 103-110.

Olshan, A.F., Weissler, M.C., Watson, M.A. and Bell, D.A. (2000). GSTM1, GSTT1, GSTP1, CYP1A1, and NAT1 polymorphisms, tobacco use, and the risk of head and neck cancer. *Cancer Epidemiol. Biomarkers Prev., 9*(2), 185-191.

Omar, Z.A., Ali, Z.M. and Ibrahim Tamin, N.S. (2006). Malaysian Cancer Statistics data and figure, Peninsular Malaysia. National cancer registry, Kuala Lumpur.

Omar, Z.A. and Ibrahim Tamin, N.S. (2011). National Cancer Registry Report 2007. Ministry of Health, Malaysia.

Ooi, K.G. (2004). Southeast Asia: A Historical Encyclopedia, from Angkor Wat to East Timor. California: ABC-CLIO.

Ortiz de Montellano, P.R. (2005). Cytochrome P450: structure, mechanism, and biochemistry (3rd ed.). New York: Kluwer Academic/Plenum Publishers.

Ota, K. and Hammock, B.D. (1980). Cytosolic and microsomal epoxide hydrolases: differential properties in mammalian liver. *Science*, *207*(4438), 1479-1481.

Oyama, T., Kawamoto, T., Mizoue, T., Sugio, K., Kodama, Y., Mitsudomi, T., *et al.* (1997). Cytochrome P450 2E1 polymorphism as a risk factor for lung cancer: in relation to p53 gene mutation. *Anticancer Res., 17*(1B), 583-587.

Pan, Q.H., Cao, Y., Xu, J.F., Chen, L.Z., Feng, Q.S., Zeng, Y.X., *et al.* (2007). Family-based association study of XRCC1 gene polymorphisms in nasopharyngeal carcinoma. *Zhonghua Yu Fang Yi Xue Za Zhi*, 41 Suppl, 12-16.

Pathmanathan, R., Prasad, U., Chandrika, G., Sadler, R., Flynn, K. and Raab-Traub, N. (1995). Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx.Variants of Epstein-Barr virus-infected neoplasia. *Am. J. Pathol.*, *146*(6), 1355-1367.

Parl, F.F. (2005). Glutathione S-transferase genotype and cancer risk. *Cancer Lett.*, *221*, 123-129.

Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol., 32*(4), 381-385.

Pemble, S., Schroeder, K.R., Spencer, S.R., Meyer, D.J., Hallier, E., Bolt, H.M., *et al.* (1994). Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem. J., 300* (Pt 1), 271-276.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet., 38*(8), 904-909.

Przybyla-Zawislak, B.D., Srivastava, P.K., Vazquez-Matias, J., Mohrenweiser, H.W., Maxwell, J.E., Hammock, B.D., *et al.* (2003). Polymorphisms in human soluble epoxide hydrolase. *Mol. Pharmacol.*, *64*(2), 482-490.

Pua, K.C., Khoo, A.S.B., Yap, Y.Y., Subramanian, S.K., Ong, C.A., Gopala Krishnan, *et al.* (2008). Nasopharyngeal Carcinoma Database. *Med. J. Malaysia*, *63*, 59-62.

Purcell, S., Cherny, S.S. and Sham, P.C. (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, *19*(1), 149-150.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet., 81*, 559–575.

Rebbeck, T.R. (1997). Molecular epidemiology of the human Glutathione S-Transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. *Cancer Epidemiol. Biomarkers Prev., 6*, 733-743.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., *et al.* (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444-454.

Repeat Masker program, http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker

Ribas, G., González-Neira, A., Salas, A., Milne, R.L., Vega, A., Carracedo, B., *et al.* (2006). Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum. Genet., 118*(6), 669-679.

Rickinson, A.B. and Kieff, E. (2001). Epstein-Barr virus. In: Knipe DM, Howley PM, editors. Field's virology (4th ed.). Philadelphia (Pennsylvania): Lippincott, Williams and Wilkins; p. 2575–627.

Roxo-Rosa, M., da Costa, G., Luider, T.M., Scholte, B,J,, Coelho, A.V., Amaral, M.D., *et al.* (2006). Proteomic analysis of nasal cells from cystic fibrosis patients and non-cystic fibrosis control individuals: search for novel biomarkers of cystic fibrosis lung disease. *Proteomics, 6*, 2314-2325.

Sandberg, M. and Meijer, J. (1996). Structural characterization of the human Soluble Epoxide Hydrolase gene (EPHX2). *Biochem. Biophys., 221*, 333-339.

Sarmanová, J., Benesová, K., Gut, I., Nedelcheva-Kristensen, V., Tynková, L. and Soucek, P. (2001). Genetic polymorphisms of biotransformation enzymes in patients with Hodgkin's and non-Hodgkin's lymphomas. *Hum. Mol. Genet., 10*(12), 1265-1273.

Sawyer, S.L., Mukherjee, N., Pakstis, A.J., Feuk, L., Kidd, J.R., Brookes, A.J., *et al.* (2005). Linkage disequilibrium patterns vary substantially among populations. *Eur. J. Hum. Genet., 13*(5), 677-686.

Scheetz, T.E., Zabner, J., Welsh, M.J., Coco, J., Eyestone, M. de F., Bonaldo, M., *et al.* (2004). Large-scale gene discovery in human airway epithelia reveals novel transcripts. *Physiol. Genomics, 17*, 69–77.

Seidegard, J. and Ekstrom, G. (1997). The role of human glutathione transferases and epoxide hydrolase in the metabolim of xenobiotics. *Environmental Health Perspectives, 105*, 791-799.

Senchenko, V.N., Liu, J., Loginov, W., Bazov, I., Angeloni, D., Seryogin, Y., *et al.*, (2004).Discovery of frequent homozygous deletions in chromosome 3p21.3 LUCA and AP20 regions in renal, lung and breast carcinomas. *Oncogene, 23*(34), 5719-5728.

Shao, J.Y., Zhang, Y., Li, Y.H., Gao, H.Y., Feng, H.X., Wu, Q.L., *et al.* (2004). Comparison of Epstein-Barr virus DNA level in plasma, peripheral blood cell and tumor tissue in nasopharyngeal carcinoma. *Anticancer Res., 24*(6), 4059-4066.

Shao, Y.M., Poirier, S., Ohshima, H., Malaveille, C., Zeng, Y., de Thé, G., *et al.* (1988). Epstein-Barr virus activation in Raji cells by extracts of preserved food from high risk areas for nasopharyngeal carcinoma. *Carcinogenesis, 9*(8), 1455-1457.

Sherpard, D. (2003). Functions of Pulmonary Epithelial Integrins: From Development to Disease. *Physiol. Rev., 83*, 673–686.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., *et al.* (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, *29*(1), 308-311.

Sigma DNA calculator, http://www.sigma-genosys.com/calc/DNACalc.asp

Singh, M., P. Shah, P., P. Singh, A., Ruwali, M., Mathur, N., C. Pant, M., *et al.* (2008). Association of genetic polymorphisms in glutathione S-transferases and susceptibility to head and neck cancer. *Mutat. Res.*, *638*, 184-194.

Smith, B.D., Sanders, J.L., Porubsky, P.R., Lushington, G.H., Stout, C.D. and Scott, E.E. (2007). Structure of the human lung cytochrome P450 2A13. *J. Biol. Chem.*, *282*(23), 17306-17313.

Snustad, D.P. and Simmons, M.J. (2003). Principles of genetics (3rd ed.). USA: John Wiley & Sons.

Spano, J.P., Busson, P., Atlan, D., Bourhis, J., Pignon, J.P., Esteban, C., *et al.* (2003). Nasopharyngeal carcinomas: an update. *European Journal of Cancer*, *39*, 2121-2135.

Spencer, C.C., Su, Z., Donnelly, P. and Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet*., *5*(5), e1000477.

Song, N., Tan, W., Xing, D. and Lin, D. (2001). CYP1A1 polymorphism and risk of lung cancer in relation to tobacco smoking: a case-control study in China. *Carcinogenesis*, *22*, 11-16.

Song, C., Chen, L.Z., Zhang, R.H., Yu, X.J. and Zeng, Y.X. (2006). Functional variant in the 3'-untranslated region of Toll-like receptor 4 is associated with nasopharyngeal carcinoma risk. *Cancer Biol. Ther.*, *5*(10), 1285-1291.

Sriampron, S., Vatanasapt, V., Pisani, P., Yongchaiyudha, S. and Rungpitarangsri, V. (1992). Environmental Risk Factors for Nasopharyngeal Carcinoma: A Case-Control Study in Northeastern Thailand. *Cancer Epidemiol. Biomarkers Prev.*, *1*, 345-348.

Steemers, F.J. and Gunderson, K.L. (2005). Illumina, Inc. *Pharmacogenomics*, *6*(7), 777-782.

Steemers, F.J. and Gunderson, K.L. (2007).Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J.*, *2*(1), 41-49.

Stone, K.R., Mickey, D.D., Wunderli, H., Mickey, G.H. and Paulson, D.F. (1978). Isolation of a human prostate carcinoma cell line (DU 145). *Int. J. Cancer*, *21*(3), 274-281.

Strober, W. (2001). Trypan blue exclusion test of cell viability. *Curr. Protoc. Immunol.*, Appendix 3, Appendix 3B.

Su, T., Bao, Z., Zhang, Q.Y., Smith, T.J., Hong, J.Y. and Ding, X. (2000). Human cytochrome P450 CYP2A13: predominant expression in the respiratory tract and its high efficiency metabolic activation of a tobacco-specific carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Cancer Res., 60*(18), 5074-5079.

Su, B., Xiang, B., Wang, L., Cao, L., Xiao, L., Li, X., *et al.* (2010). Profiling and comparing transcription factors activated in non-metastatic nasopharyngeal carcinoma cells. *Journal of Cellular Biochemistry*, *109*, 173-183.

Sugimura, H., Hamada, G.S., Suzuki, I., Iwase, T., Kiyokawa, E., Kino, I., *et al.* (1995). CYP1A1 and CYP2E1 polymorphism and lung cancer, case-control study in Rio de Janeiro, Brazil. *Pharmacogenetics*, 5 Spec No, S145-148.

Suh, Y. and Vijg, J. (2005). SNP discovery in associating genetic variation with human disease phenotypes. *Mutat. Res., 573*(1-2), 41-53.

Sung, Y.K., Moon, C., Yoo, J.Y., Moon, C., Pearse, D., Pevsner, J., *et al.* (2002). Plunc, a member of the secretory gland protein family, is up-regulated in nasal respiratory epithelium after olfactory bulbectomy. *J. Biol. Chem., 277*, 12762-12769.

Suske, G. (1999). The Sp-family of transcription factors. *Gene*, *238*, 291-300.

Suzina, S.A. and Hamzah, M. (2003). Clinical presentation of patients with nasopharyngeal carcinoma. *Med. J. Malaysia*, *58*(4), 539-545.

Tabor, H.K., Risch, N.J. and Myers, R.M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet., 3*(5), 391-397.

Tambyrajah, W.S., Doran, E., Wood, J.D. and McGivan, J.D. (2004). The pig CYP2E1 promoter is activated by COUP-TF1 and HNF-1 and is inhibited by androstenone. *Arch. Biochem. Biophys., 431*(2), 252-260.

Applied Biosystems. (2010). TaqMan SNP genotyping assays protocol.

Tiwawech, D., Srivatanakul, P., Karalak, A. and Ishida, T. (2005). Glutathione S-transferase M1 gene polymorphism in Thai nasopharyngeal carcinoma. *Asian Pacific Journal of Cancer Prevention*, *6*, 270-275.

Thompson, D., Stram, D., Goldgar, D. and Witte, J.S. (2003).Haplotype tagging single nucleotide polymorphisms and association studies. *Hum. Hered., 56*(1-3), 48-55.

Thompson, M.P. and Kurzrock, R. (2004).Epstein Barr virus and cancer. *Clinical Cancer Res., 10*, 803-821.

Thompson, L.D.R. (2007). Update on Nasopharyngeal carcinoma. *Head and Neck Pathol., 1*, 81-86.

Thorisson, G.A., Smith, A.V., Krishnan, L. and Stein, L.D. (2005). The International HapMap Project Web site. *Genome Res., 15*(11), 1592-1593.

Tracy, J.D. (ed). (1993). The Rise of merchant empires: long-distance trade in the early modern world 1350-1750. Cambridge ; New York : Cambridge University Press.

Tse, K.P., Su, W.H., Chang, K.P., Tsang, N.M., Yu, C.J., Tang, P., *et al.* (2009). Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. *Am. J. Hum. Genet., 85*, 194-203.

Tse, K.P., Su, W.H., Yang, M.L., Cheng, H.Y., Tsang, N.M., Chang, K.P., *et al.* (2011). A gender-specific association of CNV at 6p21.3 with NPC susceptibility. *Hum. Mol. Genet*., *20*(14), 2889-2896

Tugizov, S.M,, Berline, J.W., Palefsky, J.M. (2003). Epstein-Barr virus infection of polarized tongue and nasopharyngeal epithelial cells. *Nat. Med*., *9*(3), 307-314.

Ueno, T. and Gonzalec, T.J. (1990). Transcriptional control of the rat hepatic CYP2E1 gene. *Mol. Cell Biol., 10*(9), 4495-4505.

University College Santa Cruz (USCS), http://genome.ucsc.edu/cgi

Wang, H., Tan, W., Hao, B., Miao, X., Zhou, G., He, F., *et al.* (2003). Substantial reduction in risk of lung adenocarcinoma associated with genetic polymorphism in CYP2A13, the most active cytochrome P450 for the metabolic activation of tobacco-specific carcinogen NNK. *Cancer Res., 63*, 8057-8061.

Wang, X., Tomso, D.J., Liu, X. and Bell, D.A. (2005). Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. *Toxicol. Appl. Pharmacol., 207*, 84-90.

Ward, M.H., Pan, W.H., Cheng, Y.J., Li, F.H., Brinton, L.A., Chen, C.J., *et al.* (2000). Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. *Int. J. Cancer*, *86*(5), 603-609.

Wardlaw, S.A., Nikula, K.J., Kracho, D.A., Finchi, G.L., Thornton-Manning, J.R. and Dahl, A.R. (1998). Effect of cigarette smoke on CYP1A1, CYP1A2 and CYP2B1/2 of nasal mucosae in F344 rats. *Carcinogenesis*, *19*, 655-662.

Wei, W.I. and Kwong, D.L. (2010). Current management strategy of nasopharyngeal carcinoma. *Clin. Exp. Otorhinolaryngol*, *3*(1), 1-12.

Witte, S. (2010). Genome-wide association studies and beyond. *Annu. Rev. Public Health, 31*, 9-20.

Van der Weide, J. and Hinrichs, J.W.J. (2006). The influence of cytochrome P450 pharmacogenetics on disposition of common antidepressant and antipsychotic medications. *Clin. Biochem. Rev., 27*, 17-25.

Van Poppel, G. and Van den Berg, H. (1997).Vitamins and cancer. *Cancer Lett., 114*(1-2), 195-202.

Varzim, G., Monteiro, E., Silva, R.A., Fernandes, J. and Lopes, C. (2003). CYP1A1 and XRCC1 gene polymorphisms in SCC of the larynx. *Eur. J. Cancer Prev*., *12*(6), 495-499.

Vaughan, T.L., Shapiro, J.A., Burt, R.D., Swanson, G.M., Berwick, M., Lynch, C.F., *et al.* (1996). Nasopharyngeal cancer in a low-risk population: defining risk factors by histological type. *Cancer Epidemiol. Biomarkers Prev., 5*(8), 587-593.

Xiong, W., Zeng, Z.Y., Xia, J.H., Xia, K., Shen, S.R., Li, X.L., *et al.* (2004). A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res., 64*(6), 1972-1974.

Xu, Y.F., Pan, Q.H., Cui, C., Chen, L.Z., Feng, Q.S., Zeng, Y.X., *et al.* (2009). Association of nasopharyngeal carcinoma risk with cytochrome P450 CYP1A1 gene polymorphisms. *Zhonghua Yu Fang Yi Xue Za Zhi, 43*(7), 586-590.

Xu, Z., Kaplan, N.L. and Taylor, J.A. (2007). Tag SNP selection for candidate gene association studies using HapMap and gene resequencing data. *Eur. J. Hum. Genet., 15*(10), 1063-1070.

Yang, H., Nie, Y., Li, Y. and Wan, Y.J. (2010). Histone modification-mediated CYP2E1 gene expression and apoptosis of HepG2 cells. *Exp. Biol. Med. (Maywood)*, 235(1), 32-39.

Yang, Y.X., Yang, Y.B., Li, X.L., Peng, C., Guo, Q., Shen, S.R., *et al.* (2007). The study of LPLUNC1 gene inhibit human nasopharyngeal carcinoma cell line HNE1 growth and proliferation. *Prog. Biochem. Biophhys., 34*, 366-374.

Yuan, J.M., Wang, X.L., Xiang, Y.B., Gao, Y.T., Ross, R.K. and Yu, M.C. (2000). Non-dietary risk factors for nasopharyngeal carcinoma in Shanghai, China. *Int. J. Cancer*, *85*(3), 364-369.

Yu, M.C., Ho, J.H., Lai, S.H. and Henderson, B.E. (1986). Cantonese-style salted fish as a cause of nasopharyngeal carcinoma: report of a case-control study in Hong Kong. *Cancer Res*., *46*, 956–61.

Yu, M.C., Nichols, P.W., Zou, X.N., Estes, J. and Henderson, B.E. (1989)b. Induction of malignant nasal cavity tumours in Wistar rats fed Chinese salted fish. *Br. J. Cancer, 60*(2), 198-201.

Yu, M.C., Huang, T.B. and Henderson, B.E. (1989)a, Diet and nasopharyngeal carcinoma: a case-control study in Guangzhou, China. *Int. J. Cancer*, *43*,1077–1082.

Yu, M.C. and Yuan, J.M. (2002). Epidemiology of nasopharyngeal carcinoma. *Cancer Biol.*, *12*, 421-429.

Zeng, Y.X. and Jia, W.H. (2002). Familial nasopharyngeal carcinoma. *Semin. Cancer Biol*., *12*(6), 443-450.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S, Ohwaki, M., *et al.* (2000). The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell*, *102*, 827-837.

Zhang, B., Nie, X., Xiao, B., Xiang, J., Shen, S., Gong, J., *et al.* (2003). Identification of tissue-specific genes in nasopharyngeal epithelial tissue and differentially expressed genes in nasopharyngeal carcinoma by suppression subtractive hybridization and cDNA microarray. *Genes Chromosomes Cancer*, *38*, 80-90.

Zhang, X., Su, T., Zhang, Q.Y., Gu, J., Caggana, M., Li, H., *et al*. Genetic polymorphisms of the human CYP2A13 gene: Identification of single-nucleotide polymorphisms and functional characterization of an Arg257Cys variant. *J. Pharmacol. Exp. Ther.*, *302*(2), 416-423.

Zheng, X., Hu, L., Chen, F. and Christensson, B. (1994a). Expression of Ki67 antigen, epidermal growth factor receptor and Epstein-Barr virus-encoded latent membrane protein (LMP1) in nasopharyngeal carcinoma. *Eur. J. Cancer B. Oral Oncol.*, *30B*(5), 290-295.

Zheng, Y.M., Tuppin, P., Hubert, A., Jeannel, D., Pan, Y.J., Zeng, Y., *et al.* (1994b). Environmental and dietary risk factors for nasopharyngeal carcinoma: a case-control study in Zangwu County, Guangxi, China. *Br. J. Cancer, 69*(3), 508-14.

Zheng, M.Z., Qin, H.D., Yu, X.J., Zhang, R.H., Chen, L.Z., Feng, Q.S., *et al.* (2007). Haplotype of gene Nedd4 binding protein 2 associated with sporadic nasopharyngeal carcinoma in the Southern Chinese population. *J. Transl. Med.*, *5*, 36.

Zhou, X.F., Cui, J., Macias, V., Kajdacsy-Balla, A.A., Ye, H., Wang, J.G., *et al.* (2007). The progress on genetic analysis of nasopharyngeal carcinoma. *Comparative and functional genetics*, 13 pages, doi:10.1155/2007/57513.

Zhou, H.D., Li, X.L., Li, G.Y., Zhou, M., Liu, H.Y., Yang, Y.X., *et al.* (2008). Effect of SPLUNC1 protein on the Pseudomonas aeruginosa and Epstein-Barr virus. *Mol. Cell Biochem.*, *309*(1-2), 191-197.

Zhu, M. and Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *Int. J. Biol. Sci.*, *3*(7), 420-427.

Zhuo, X., Cai, L., Xiang, Z., Li, Q. and Zhang, X. (2009). GSTM1 and GSTT1 polymorphisms and nasopharyngeal cancer risk: an evidence-based meta-analysis. *J. Exp. Clin. Cancer Res*., *28*, 46.

APPENDICES

APPENDIX A: MATERIALS

i. Chemicals and enzymes

All the chemicals used in this study were of molecular grade and available commercially. These chemicals were obtained from Sigma-Aldrich Chemical Company, USA; GIBCO®, USA; Invitrogen, Carlsbad, CA, USA; Wako Pure Chemical Ind., Ltd., Osaka, Japan; Promega, Madison, WI, USA.

These chemicals included ethylenediamine tetraacetic acid (EDTA) disodium salt, sodium dodecyl sulphate (SDS), sodium chloride, magnesium chloride ($MgCl_2$), hydrochloric acid, tris base, tris-hydrochloride (HCL), phenol, chloroform, Proteinase-K, boric acid, bromophenol blue, ethidium bromide, absolute ethanol and other common inorganic chemicals and solvents. Restriction enzymes including *Xho* I, *Bgl* II and *Hind* III were purchased from TaKaRa BIO Incorporated, Japan. All the antibodies used for electrophoretic mobility shift assay (EMSA) were obtained from Santa Cruz Biotechnology, Santa Cruz, CA, USA.

Custom primers (synthetic oligonucleotides), invader probes, oligonucleotides used for EMSA and luciferase assay were made to order from Sigma-Aldrich Chemical Co., USA or Invitrogen, Carlsbad, CA, USA. All the reagents for TaqMan genotyping assay were obtained from Applied Biosystems, Foster City, CA, USA.

All the reagents used for polymerase chain reaction (PCR), Invader genotyping assay, sequencing, and the solutions used for nuclear extraction (lysis buffer and nuclear extract buffer) were supplied by the Laboratory for Pharmacogenetics, Center for Genomic Medicine of RIKEN, Japan. TaKaRa Ex-Taq polymerase was supplied by TaKaRa BIO Incorporated, Japan. GeneRuler™ Lambda DNA/HindIII DNA ladder and

1kb DNA ladder used in gel electrophoresis was purchased from Fermentas, Canada and Invitrogen, Carlsbad, CA. respectively.

ii. Kits

The following are the kits used in the present study:

DIG Gel Shift Kit, 2nd Generation (Roche Diagnostics, Mannheim, Germany)

Dual-luciferase reporter assay system (Promega, Madison, WI)

QIAmp DNA blood mini kit (Qiagen, Valencia, CA, USA)

QIAfileter midi and maxi kits (Qiagen, Valencia, CA, USA)

QIAGEN MinElute gel extraction kit (Qiagen, Valencia, CA, USA)

RNeasy plus Mini Kit (Qiagen, Valencia, CA, USA)

SuperScriptIII First Strand Synthesis System (Invitrogen, Carlsbad, CA, USA)

BigDye® terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA)

iii. Instruments

The following are the instruments used in the present study:

Applied Biosystems 7900HT Fast Real time PCR system (Applied Biosystems, Foster City, CA, USA)

Applied Biosystems Prism 3700 DNA analyzer (Applied Biosystems, Foster City, CA, USA)

Applied Biosystems thermal cycler (Applied Biosystems, Foster City, CA, USA)

BioDoc-It® Imaging System (UVP, Upland, CA, USA)

Multimek 96 (Beckman Coulter, Inc., Fullerton, CA, USA)

Fuji Medical Film Processing system Cepros SV (Fuji Photo Film Co., Ltd., Japan)

Luminometer VICTOR Light 1420-060 (Perkin Elmer, USA)

Nanodrop 1000 (ND-1000, Thermo Scientific, USA)

Nucleofector system (Amaxa, Köln, Germany)

Ultraviolet Crosslinkers CL-1000 (UVP, Upland, CA, USA)

APPENDIX B: BUFFER AND CHEMICAL PREPARATIONS

1. DNA extraction

1X Red cell lysis buffer

320 mM Sucrose	54.77 g
75 % Triton-X-100	3.75 ml
5mM Magnesium chloride (MgCl2)	2.5 ml
1mM Tris-HCl; pH 7.5	5 ml
Distilled water (dH2O) to	500 ml

10 X Proteinase K buffer

0.375M Sodium chloride (NaCl)	10 ml
0.12M EDTA	10 ml
Total	20 ml

20% (w/v) Sodium Dodecyl Sulphate (SDS)

SDS	20 g
Distilled water (dH2O) to	100 ml

6M Sodium chloride (NaCl)

NaCl	35.064 g
Distilled water to (dH2O)	100 ml

Phenol-chloroform (1:1; w/v)

Phenol	250 ml
Chloroform	250 ml
Total	500 ml

70% Ethanol	
Ethanol (100%)	70 ml
Distilled water(dH2O)	30 ml
Total	100 ml

TE buffer; pH 8.0	
10mM Tris-HCl	50 ml
1mM EDTA	5 ml
Distilled water(dH2O) to	500 ml

2. Sequencing

80% Ethanol	
Ethanol (100%)	80 ml
Distilled water(dH2O)	20 ml

70% Ethanol	
Ethanol (100%)	70 ml
Distilled water(dH2O)	30 ml
Total	100 ml

3. Agarose gel electrophoresis

5X TBE	
Tris base	54 g
Boric acid	27.5 g
EDTA	3.72 g
Distilled water (dH2O) to	1 L

2.0% agarose gel	
Agarose	2.0 g
1 X TBE	100ml

1.5% agarose gel	
Agarose	1.5 g
1 X TBE	100ml

0.8% agarose gel	
Agarose	0.8 g
1 X TBE	100ml

APPENDIX C: PRIMER SEQUENCES OF RE-SEQUENCING *CYP2E1*

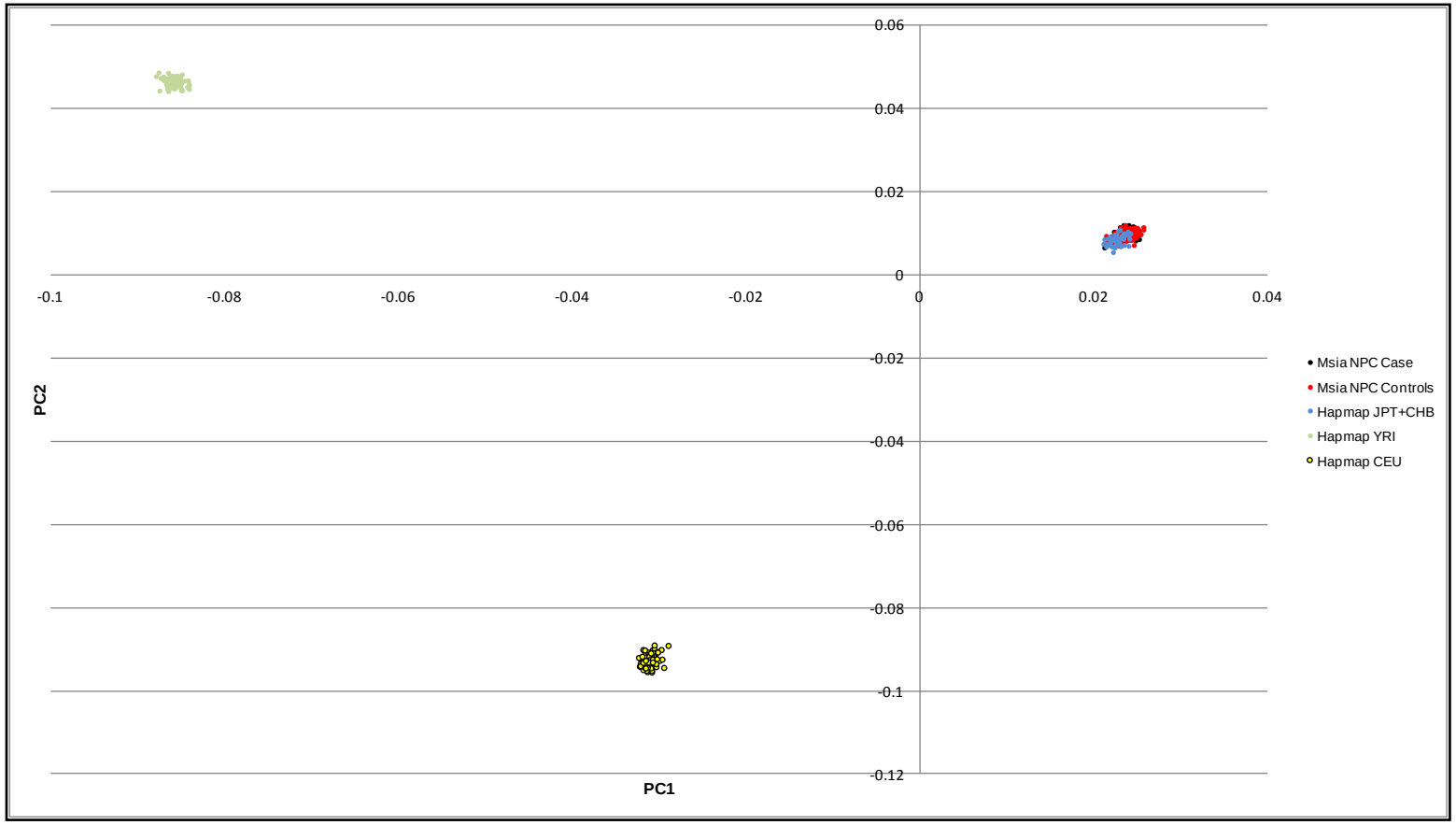
no	Primer	Sequence
1	CY2E1_F1	AGAGGACTCAGACCACAGC
	CY2E1_R1	CTCCACATTGACTAGCTTCTTC
	CY2E1_Fi1	CCACACAGGACAACAGG
	CY2E1_Ri1	ATGTAGACTCGACTGGCTCAC
2	CY2E1_F2	TCTTCATTCTAACCACACACAC
	CY2E1_R2	AAGAGGTTCCCGATGATG
	CY2E1_Fi2	GTGGTGATTCAGGTACTACTGG
	CY2E1_Ri2	GGAACTCCAGCCACATC
3	CYP2E1_F3	CCGTTTCCACAGGATTG
	CYP2E1_R3	TTCCAGCTGCTCTGCTC
	CYP2E1_Fi3	AAATGTGTCGCCTGTGG
	CYP2E1_Ri3	CTTTGGCAAACACAGGC
4	CYP2E1_F4	GGTCCAGAACCTTGATTCC
	CYP2E1_R4	TCTGGAGACAGGAAGCG
	CYP2E1_Fi4	AAGCGCTGCTGGACTAC
	CYP2E1_Ri4	TAATCCTGATCTCATCCTTCG
5	CYP2E1_F5	ACACTGCTCGCTTCAGG
	CYP2E1_R5	ATCTGTGTTGATGTGAATTCC
	CYP2E1_Fi5	CTTTGATGACGAGGCTCC
	CYP2E1_Ri5	AGATTTGTTGGCCCATTC
6	CYP2E1_F6	ATTAACCATCTCAGTCCTTGTC
	CYP2E1_R6	TCCTCCCGAGAGTATGAAC
	CYP2E1_Fi6	CTCAACATGACCACGTCTG
	CYP2E1_Ri6	CACATGAGATGAGGTGGG
7	CYP2E1_F7	GCAGAGGCAGCAAAGTC
	CYP2E1_R7	GTTCAGATGCAGTCTAGGAGTC
	CYP2E1_Fi7	GGAAGACTCAGCAGAAAGG
	CYP2E1_Ri7	GGTCCTCACAGTCTGCAC
8	CYP2E1_F8	GACCTACCTGGAAGGACATC
	CYP2E1_R8	AAACCTGAATTTGTGGCTG
	CYP2E1_Fi8	ACTGAGCAGGTGGAGGAG
	CYP2E1_Ri8	GCCTTAGAAACTTCTCATCATTG
9	CYP2E1_F9	GGTGTTTGATTAGACAGCCC
	CYP2E1_R9	CTTGGTCATAGCTCAGGC
	CYP2E1_Fi9	CATGATGTGGGCTCTCC
	CYP2E1_Ri9	ATGCTGAAGCCAAGAGTG
10	CYP2E1_F10	CTTCTTGTATCCTTGTCTTGTCC
	CYP2E1_R10	CTTTCACACCTCCTCCAG
	CYP2E1_Fi10	AGCGCTTGTACACAATGG
	CYP2E1_Ri10	GTGTCCCTTCAGTCACTTG

no	Primer	Sequence
11	CYP2E1_F11	GCGTCTTCCTCAAATGC
	CYP2E1_R11	GCAGGAGGATGACTTGAG
	CYP2E1_Fi11	ACATGTGATGGATCCAGG
	CYP2E1_Ri11	TACAATGAGCCATGATTGC
12	CYP2E1_F12	ACCTCAGAGGCACAGGG
	CYP2E1_R12	ATGTTTATGGAGCATCCTAAGC
	CYP2E1_Fi12	AGCCAGCCAAAGTCTACC
	CYP2E1_Ri12	AAATGTTATGTACAGTACCCAGG
13	CYP2E1_F13	TTATCAATGCAAAGGTTACCC
	CYP2E1_R13	TGTCAATTTCTTCATGGAGC
	CYP2E1_Fi13	TGTGGATGGATGGAAGG
	CYP2E1_Ri13	ACCCTCCTATCCATCCATC
14	CYP2E1_F14	TGGATGGATGGAGGATG
	CYP2E1_R14	TCCTCAAGGGAAGGTACTG
	CYP2E1_Fi14	ATACTGCATCTCCAGGAGTG
	CYP2E1_Ri14	TGAAAGCCCTAATCTTCTCAC
15	CYP2E1_F15	GTTTCTCCTAGGGCACAGTC
	CYP2E1_R15	TCTCCAGCACACACTCG
	CYP2E1_Fi15	GCCATATTACCCAATTCTCC
	CYP2E1_Ri15	CACAACAATCCCTAAGATGAG
16	CYP2E1_F16	CATCACTGATGAGCACTGG
	CYP2E1_R16	ACCTCAATATGATTCTGACACTG
	CYP2E1_Fi16	GTTTGGCTGTATCCCACC
	CYP2E1_Ri16	CAAATACTCATATGCAAAGAAAGG
17	CYP2E1_F17	ATTACAGGCACCCACCAC
	CYP2E1_R17	TTCAGTGAGACTCAAGGGC
	CYP2E1_Fi17	CAGTAATCCTTTCCTAACGTCAG
	CYP2E1_Ri17	AATCTGTGGCTGTACTCATTG
18	CYP2E1_F18	TATTCCCTGTACTTGTTAGCTCTG
	CYP2E1_R18	AGCTTCCTGAATTGCTCAC

APPENDIX D: PRIMER SEQUENCES FOR RE-SEQUENCING *SPLUNC1*

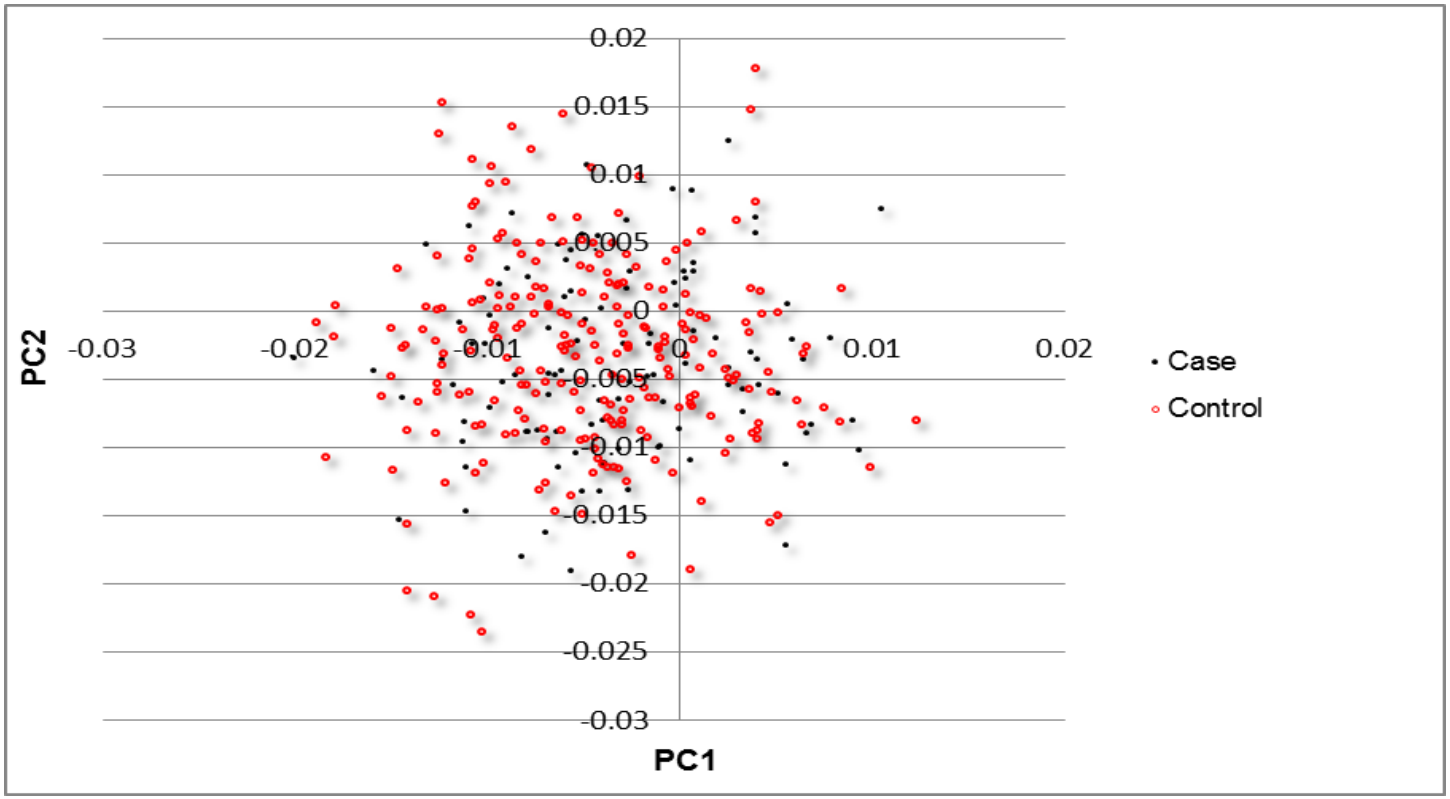
no	Primer	Sequence
1	SPLUNC1_F1	CCTTGCCTTATCTGCCTC
	SPLUNC1_R1	CCTCCCAACACTTGCTG
	SPLUNC1_Fi1	GCTGAACAGACTCCAAGG
	SPLUNC1_Ri1	TGGCATCCTGGTCTCTATC
2	SPLUNC1_F2	AGGTGGAGGTTGTCATGAG
	SPLUNC1_R2	GGAGCCTGAGATCCCTG
	SPLUNC1_Fi2	AGAGAGCTCCTGACGTGC
	SPLUNC1_Ri2	TCCAGTTGGAGGATTCTTTC
3	SPLUNC1_F3b	AACACATGCTTCCTCGTC
	SPLUNC1_R3b	CAGATTTATTCTCCAATCGG
	SPLUNC1_Fi3	GATGGAGCTTTAGCTGAGTC
	SPLUNC1_Ri3	AGCAGAGTTGTCCAGAGC
4	SPLUNC1_F4	AAAGGACTGAACAGGGATG
	SPLUNC1_R4	CTTCAGGATGTCCAGGAG
	SPLUNC1_Fi4	ACAGGTGCCCATGTCTTAG
	SPLUNC1_Ri4	GGCATGGGTATCAGAGAG
5	SPLUNC1_F5	CCTGAGGATCTGGAATTGTC
	SPLUNC1_R5	ACCATCCATCCATCATCAC
	SPLUNC1_Fi5	GGCTCTCACTCTACTATCCTAACTC
	SPLUNC1_Ri5	TGTTCTAAGGGATTTCACAGG
6	SPLUNC1_F6	GAAAGTGTATTGTGAGGACTGG
	SPLUNC1_R6	GCCAAGAATGCTAGCCTC
	SPLUNC1_Fi6	AAGGCTCTTAGCTGGCATAC
	SPLUNC1_Ri6	CTGGAACAGATCTGACACAC
7	SPLUNC1_F7	CATTTGCTGCTATGCTAAGTG
	SPLUNC1_R7	AGCAGTGAACACACAGGTATG
	SPLUNC1_Fi7	GGAGCGTGGAAAGATGG
	SPLUNC1_Ri7	TCTATCATCTATCTATCCATCCATC
8	SPLUNC1_F8	CACATGTGTCATCTCATTGG
	SPLUNC1_R8	TGTCCAAGCCTCTGAGAAC
	SPLUNC1_Fi8	GGCAACGTAAGTAGGCAAG
	SPLUNC1_Ri8	GACAAATAGATTGATGAATGGAAG
9	SPLUNC1_F9	GGATTGCTTAGAATCACTTTCAC
	SPLUNC1_R9	ACAGGAGGAGGCTGAGAC
	SPLUNC1_Fi9	AAGCATGCTGAACACAGTC
	SPLUNC1_Ri9	CTGTAGTCCGTGGATCAGC
10	SPLUNC1_F10	GACTCACCACACATAAGTACACG
	SPLUNC1_R10	CATTCCCACCAGCAGTATAAG
	SPLUNC1_Fi10	CTGAATAGTCTTGAGCAGGTC
	SPLUNC1_Ri10	TACCTGGAGTATTATATCTACCCTG

no	Primer	Sequence
11	SPLUNC1_F11	GGTGACAAAGTGAGACCCTATC
	SPLUNC1_R11	AGTGAACACTTTGCACCAG
	SPLUNC1_Fi11	GCAATTATCCTAAGTGAATCAATG
	SPLUNC1_Ri11	AGATAGGTCTGGTGTGTATTACTTC
12	SPLUNC1_F12	ATGAACTGATTCCGACAGAC
	SPLUNC1_R12	TCTCCCTATGTTGCCCAG
13	SPLUNC1_F13	ACTCTAGCTCTGAAGGGTGC
	SPLUNC1_R13	ATACCCAGCACCTAGATCAG
	SPLUNC1_Fi13	TGGCTCAATGAGTGGATG
	SPLUNC1_Ri13	CCATGTTTAACTCTAGGAGATGC
14	SPLUNC1_F14	GGCATGTGTTGCATCTC
	SPLUNC1_R14	GATAAGATGCTGGATGTAAGCTC
	SPLUNC1_Fi14	CCTGGGTGATAGAGCAAG
	SPLUNC1_Ri14	TTCCCATGAACCCATCACATGTG
15	SPLUNC1_F15	ATGGATTCCCACTCTCCTC
	SPLUNC1_R15	GTCATGGAGCTAAGGACATC
	SPLUNC1_Fi15	TCCCTAATCCGGAATGC
	SPLUNC1_Ri15	AGTGAACCAGGTAGCCTTG
16	SPLUNC1_F16b	GAGCTTTATAACTTCCATGCC
	SPLUNC1_R16b	TCGCCTAGGCTGAACTG
	SPLUNC1_Fi16	TGGCACACACCCATAATC
	SPLUNC1_Ri16	TTCTTAGTCTTGCTCTGTCTCC
17	SPLUNC1_F17b	CAGGTAGACCAATTGGGC
	SPLUNC1_R17b	TCAATCAAGAGATCCATTTCTG
	SPLUNC1_Fi17	AGCAAGTGCAAAGATTAATCTC
	SPLUNC1_Ri17	GATGGCCTTGGAAGCTC
18	SPLUNC1_F18	TAGCTGGGTTTACAGGTGC
	SPLUNC1_R18	GCCTGGATCCACCTAGAC
	SPLUNC1_Fi18	TGGTAGTGACAGAGAACAAACAG
	SPLUNC1_Ri18	GAGTCTTACCACATGTATAGTCAGC
19	SPLUNC1_F19	GAACAGAATCCCAGGCAC
	SPLUNC1_R19	CTGTCTGAGTCCCTGCTAATC
	SPLUNC1_Fi19	GGCATTTGAAATGGTAAAGTATAG
	SPLUNC1_Ri19	CCTGGTTCAAATCCTACGAC



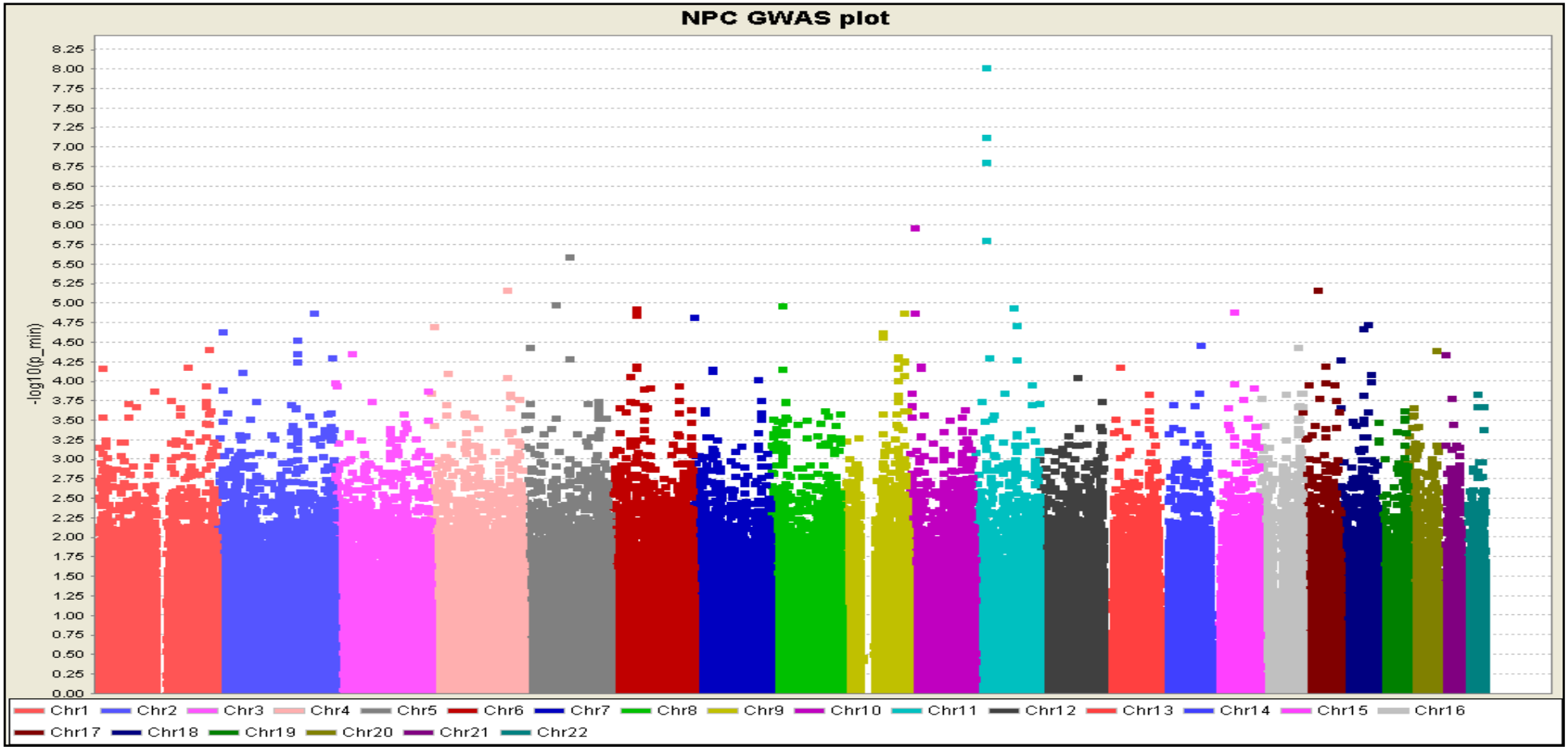
Appendix E: Principal component analysis (PCA) plot of samples enrolled in the GWAS of NPC.

NPC patients and controls in the panel A were plotted based on eigenvectors 1 and 2 obtained from the PCA, along with the European (CEU), African (YRI), and Japanese (JPT) + Chinese (CHB) individuals obtained from the phase II HapMap database.



Appendix F: Principal component analysis (PCA) plot of samples enrolled in the GWAS of NPC.

Only NPC patients and controls in the panel A were plotted based on eigenvectors 1 and 2 obtained from the PCA.



Appendix G: **Summary results of GWAS, based on chromosome.**

Manhattan plot for the GWAS of NPC indicating −log10 *P*-value of the genotyped SNPs plotted against their respective positions on each chromosome.

Journal of Human Genetics (2009) 54, 392–397

www.nature.com/jhg

ORIGINAL ARTICLE

A genome-wide association study identifies *ITGA9* conferring risk of nasopharyngeal carcinoma

Ching Ching Ng[1,2], Poh Yin Yew[1,2], Suat Moi Puah[1,2], Gopala Krishnan[3], Lee Fah Yap[4], Soo Hwang Teo[4], Paul Vey Hong Lim[5], Selvaratnam Govindaraju[6], Kananathan Ratnavelu[6], Choon Kook Sam[7], Atsushi Takahashi[8], Michiaki Kubo[9], Naoyuki Kamatani[8], Yusuke Nakamura[1,10] and Taisei Mushiroda[11]

To identify a gene(s) susceptible to nasopharyngeal carcinoma (NPC), we carried out a genome-wide association study (GWAS) through genotyping of more than 500 000 tag single-nucleotide polymorphisms (SNPs), using an initial sample set of 111 unrelated NPC patients and 260 controls of a Malaysian Chinese population. We further evaluated the top 200 SNPs showing the smallest *P*-values, using a replication sample set that consisted of 168 cases and 252 controls. The combined analysis of the two sets of samples found an SNP in intron 3 of the *ITGA9* (integrin-α 9) gene, rs2212020, to be strongly associated with NPC ($P=8.27\times10^{-7}$, odds ratio (OR)=2.24, 95% confidence intervals (CI)=1.59–3.15). The gene is located at 3p21 which is commonly deleted in NPC cells. We subsequently genotyped additional 19 tag SNPs within a 40-kb linkage disequilibrium (LD) block surrounding this landmark SNP. Among them, SNP rs189897 showed the strongest association with a *P*-value of 6.85×10^{-8} (OR=3.18, 95% CI=1.94–5.21), suggesting that a genetic variation(s) in *ITGA9* may influence susceptibility to NPC in the Malaysian Chinese population.
Journal of Human Genetics (2009) **54,** 392–397; doi:10.1038/jhg.2009.49; published online 29 May 2009

Keywords: genome-wide association study; GWAS; integrin-α 9; ITGA9; nasopharyngeal carcinoma; NPC; SNP

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a malignancy of the head and neck regions developed from epithelial cells that cover the surface and line the nasopharynx.[1] Its incidence is low in most parts of the world, but it is one of the most common cancers in some geographical regions, including Southeast Asian countries, with a high incidence of 15–50 per 100 000. Intermediate incidence rates ranging from 15 to 20 per 100 000 persons have also been reported in Alaskan Eskimos and in the Mediterranean basin including Southern Italy, Greece and Turkey.[2] In Peninsular Malaysia, NPC ranked the second among men and the twelfth among women in terms of cancer incidence, in the year 2003. The Chinese had the highest incidence compared with that in other ethnic groups, namely, Malays and Indians.[3] The age-standardized incidence in Chinese men and women was 17.0 and 6.6 per 100 000 individuals, respectively, in 2003–2005 in Peninsular Malaysia.[4] A high risk of NPC has also been observed among indigenous groups, particularly in Bidayuh in Sarawak in East Malaysia.[5]

Similar to most types of cancer, NPC is a complex disease, and its etiology is also considered to be caused by interactions of multiple factors including an Epstein–Barr virus (EBV) chronic infection, environmental factors and genetic susceptibility. Latent EBV infection is observed in nearly all NPC tumor cells, and EBV DNA is detected in all undifferentiated NPC cells, suggesting that EBV has a critical role in the pathogenesis of NPC.[2,6] Significant environmental factors, including exposure to various carcinogens such as nitrosamine and polycyclic hydrocarbon, and dietary factors, including a high consumption of salted preserved foods such as salted fish and salted eggs, have been suggested to increase the risk of NPC in Malaysian Chinese.[7]

NPC shows remarkable differences in ethnic and geographical distributions. Strong evidence from familial aggregation of NPC, migrant studies and case–control association studies of specific markers in certain ethnic groups suggests that genetic factors contribute significantly to NPC development.[8–10] Most genetic linkage analyses conducted among Chinese show that HLA alleles, A2, B14,

[1]Laboratory for International Alliance, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Japan; [2]Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia; [3]Department of Otorhinolaryngology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; [4]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Selangor, Malaysia; [5]Tung Shin Hospital, Kuala Lumpur, Malaysia; [6]NCI Cancer Hospital, Negeri Sembilan, Malaysia; [7]National Institute of Education, Nanyang Technological University, Singapore; [8]Laboratory for Statistical Analysis, RIKEN Center for Genomic Medicine, Minato-ku, Tokyo, Japan; [9]Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Japan; [10]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan and [11]Laboratory for Pharmacogenetics, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Japan
Correspondence: Dr Y Nakamura, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan.
E-mail: yusuke@ims.u-tokyo.ac.jp or Dr CC Ng, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia.
E-mail: ccng@um.edu.my
Received 20 February 2009; revised 31 March 2009; accepted 24 April 2009; published online 29 May 2009

B46, AW19 and BW17, are associated with an increased risk of NPC.[11,12] Deletion of chromosome 3p was the most common genetic alteration reported in NPC.[13,14] A recent study has linked a susceptibility locus to a 13.6-cM region on chromosome 3p21.31–21.2 to familial NPC. Investigation on this region has identified many candidate tumor suppressor genes, for example, *DLC1*, *RASSF1A*, *CACNA2D2* and *FUS1*[15]. In addition, Feng *et al.*[16] have also provided evidence of susceptibility loci for NPC on chromosome 4p15.1–q12 from a whole-genome scan in families at high risk of NPC from Guangdong Province, China.

Many recent studies have attempted to search for NPC susceptibility genes through a case–control candidate-gene approach. The possible association of genetic polymorphisms in enzymes that are involved in carcinogen metabolism, such as GSTM1 (glutathione *S*-transferase M1), CYP2E1 (cytochrome P450 2E1) and CYP2A6 (cytochrome P450 2A6), with NPC in Taiwanese and Thais have been shown.[17,18] GSTM1 is a phase II enzyme known to detoxify several carcinogens, including those found in tobacco smoke, and homozygous deletions of *GSTM1* are associated with an increased risk of NPC.[19] The CYP2E1 and CYP2A6 enzymes are known to activate nitrosamines and related carcinogens, and are possibly involved in the development of this disease. Certain variants of the *CYP2E1* and *CYP2A6* genes are thought to be more highly expressed than others and thus induce higher levels of cellular damage.

Association with NPC development has also been reported for genetic variations in DNA repair enzymes, *XRCC1* (X-ray repair cross complementing group 1)[20] and *hOGG1* (8-oxoguanine glycosylase 1).[21] In addition to them, single-nucleotide polymorphisms (SNPs) in some other genes, namely, *PLUNC* (palate, lung and nasal epithelial clone),[22] *MMPs* (matrix metalloproteinases),[23] *CCND1* (Cyclin D1),[24] TLR 4 and TLR 10 (toll-like receptor 4, 10),[25,26] may also influence susceptibility to NPC. Although a number of genetic variants have been reported to have associations with NPC susceptibility, many of them could not be replicated in subsequent studies in other populations. The genetic factors involved in determination of an individual's susceptibility to NPC are still puzzling, and we still do not understand the interactions between NPC susceptibility genes and other risk factors.

SNPs are the most abundant DNA sequence variations. The construction of a large body of SNP information through the International HapMap project[27,28] and rapid technological advances enabled us to perform genome-wide association studies (GWAS) routinely for identifying the genetic determinants for many complex diseases. In this study, we aimed to identify multiple and novel susceptibility loci of NPC through population-based case–control GWAS, using high-throughput SNP genotyping technologies.

MATERIALS AND METHODS

Participants

All the individuals who participated in this SNP association study were unrelated Chinese from different states of Peninsular Malaysia. For an initial screening, a total of 111 NPC patients and 260 healthy volunteers (Panel A) were recruited from the University Malaya Medical Centre (UMMC) and from the NCI Cancer Hospital. For a replication study, a second set of samples, consisting of 168 NPC patients and 252 control individuals, were later recruited from UMMC, the NCI Cancer Hospital and the Tung Shin Hospital. Of the 279 cases, 68.5% were male and their ages ranged from 14 to 79 years. Of the 512 controls, 70.5% were male, and their ages ranged from 18 to 60 years. All participants gave their written informed consent. The study was approved by both the ethical committees of the Yokohama Institute, The Institutes of Physical and Chemical Research (RIKEN), Yokohama, Japan and UMMC.

Genotyping of SNPs

Genomic DNA was extracted from peripheral blood leukocytes using a conventional method. A genome-wide analysis on Panel A samples was conducted using Illumina HumanHap550v3 Genotyping BeadChip, according to the manufacturer's protocols (San Diego, CA, USA). To validate the Illumina BeadChip genotyping results, we performed genotyping using multiplex-PCR-based Invader assays (Third Wave Technologies, Madison, WI, USA)[29] or by direct sequencing of PCR products using a 96-capillary 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) for the top 200 SNPs showing the smallest *P*-values, and compared the data obtained by the two platforms. SNPs that passed through the validation process were further evaluated using the second set of samples. The call rates for the landmark SNP in the GW scan, rs2212020 of *ITGA9,* were 0.99 and 1.00 in cases and controls, respectively. The *P*-values for the Hardy–Weinberg Equilibrium (HWE) test in cases and controls (Panel A) were 0.761 and 0.223, respectively.

To further analyze SNPs within the 40-kb linkage disequilibrium (LD) region including the landmark SNPs (rs2212020), 19 tSNPs (squared correlation coefficient between the two SNPs (r^2) >0.8, minor allele frequency (MAF) of >0.05) were selected from the International HapMap project database (http://www.hapmap.org/index.html.en) and genotyped on 279 cases and 512 controls by multiplex-PCR-based invader or TaqMan assay (Applied Biosystems). The LD plot was annotated from the International HapMap Project database.

Statistical analyses

We carried out statistical analyses for association and HWE. For association, the allele and genotype distributions in cases and controls were compared and evaluated in allelic, dominant- and recessive-inheritance models by two-tailed Fisher's exact test. Statistical analyses were carried out for the data obtained from both Panel A and a combined set of samples. SNPs were sorted according to the lowest *P*-value in a combined set of samples in one of these models.

RESULTS

Case–control GWA genotyping

We conducted a case–control whole-genome association study to identify genes susceptible to NPC. We first genotyped 111 NPC and 260 control individuals (Panel A) by means of a genome-wide SNP analysis. More than 500 000 SNPs derived from the International HapMap project were examined. The overall call rates of all individuals were 0.98 or higher. Of the 554 496 SNPs genotyped, 21 448 SNPs had no genotyping data and were removed from the study and 533 048 SNPs on autosomal chromosomes were further analyzed. We first generated a quantile–quantile (Q–Q) plot to inspect possible population stratification effects by comparing the distribution of the observed *P*-values with expected distribution under the null hypothesis of no population stratification. The Q–Q *P*-value plot showed no evidence of population stratification of observed statistics versus expected statistics (Figure 1). The observed *P*-values matched the expected *P*-values under the null distribution over the range of $1<-\log10\ (P)<5$. There was a departure of distribution at the extreme tail with low *P*-values at $-\log10\ (P)$ of >5, suggesting that the associations of these SNPs with NPC are likely to be true rather than the population stratification.

Validated genotyping

To validate the genotyping results of the Illumina assay, we re-genotyped all individuals in Panel A for the top 200 SNPs showing the smallest *P*-values in the initial GWA study by multiplex-PCR based Invader assay or direct sequencing. We compared the genotype frequencies and allele frequencies for each of the 200 SNPs obtained from the two assays (Illumina vs Invader assay or direct sequencing). When the concordant rates of genotyping calls in these two assays were less than 98%, we did not perform an additional analysis for such

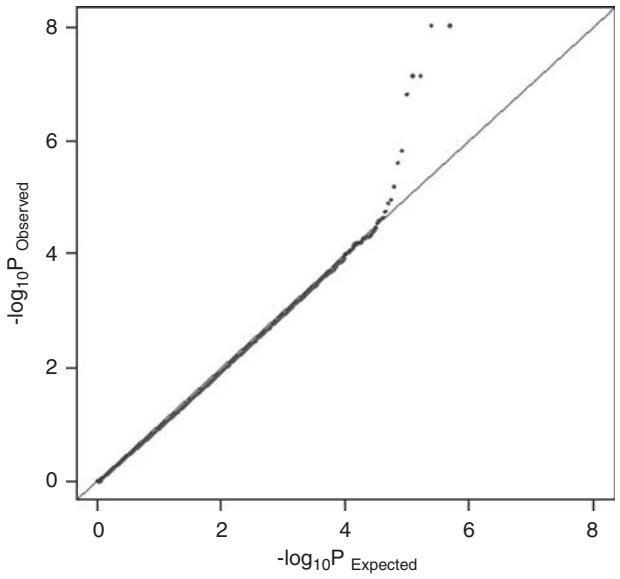


Figure 1 Log quantile–quantile (Q–Q) *P*-value plot showing the distribution of observed statistics by allelic test for all utilized 533 048 single-nucleotide polymorphisms from a genome-wide association study of 111 NPC patients and 260 controls of a Malaysian Chinese population (Panel A). The diagonal line shows the values expected under the null hypothesis.

SNPs. Therefore, the possibility of less accurate genotyping data affecting the GWAS analysis in this study was extremely low.

Replicated genotyping

After confirmation of the GWAS data, we further evaluated the top 200 SNPs showing the smallest *P*-values in an independent set of samples consisting of 168 NPC patients and 252 controls by multiplex PCR-based Invader assay or direct sequencing for their association with NPC. The association of NPC and each SNP was analyzed under the three genetic models (allelic, recessive- and dominant-inherited models) using the combined set of cases (n=279) and controls (n=512). Statistical analysis (Fisher's exact test) of the top 200 SNPs in the combined set identified 10 SNPs with *P*-values ranging from 10^{-5} to 10^{-7}, showing possible associations with NPC. Of these, one SNP, rs2212020 ($P=8.27\times10^{-7}$), was associated with NPC with an odds ratio (OR) of 2.24 (95% confidence interval (CI), 1.59–3.15) (Table 1). The rs2212020 is an intronic SNP in the *ITGA9* (Integrin-α 9) gene.

SNPs on ITGA9 gene associated with NPC susceptibility

The *ITGA9* gene contains 28 exons and spans approximately 367.5 kb. LD blocks were plotted using the data of Han Chinese in the International HapMap Project database ($r^2>0.8$) (Figure 2), and we found that the LD block including the landmark SNP is located within one gene, *ITGA9*. To further define a genomic region of interest, we analyzed 58 SNPs that cover an approximately 0.25-Mb genomic region, including a part of *ITGA9* and *C3orf35,* but found no SNP to be significantly associated with NPC (Figure 2).

The landmark SNP is located in intron 3 of the *ITGA9* gene and is included in a 40-kb LD block that covers 2.4 kb of their promoter region and up to intron 4 (Figure 2). From the Han Chinese genotype data in the International HapMap Project database, 19 tSNPs with an MAF of >0.05 were identified within this 40-kb LD block. We subsequently genotyped these 19 tSNPs for 279 cases and 512 controls and found two SNPs, rs197721 and rs149816 in intron 1, revealing some level of associations ($P<10^{-5}$) (Table 1). We then genotyped five SNPs completely linked with rs197721 and rs149816 ($r^2>0.8$), and confirmed their strong association with NPC ($P<10^{-6}$) (Table 1). Among the eight associated SNPs identified, rs189897 revealed the most significant association with NPC ($P=6.85\times10^{-8}$, OR=3.18, 95% CI=1.94–5.21).

DISCUSSION

We report here the association of SNPs in *ITGA9* with the susceptibility to NPC in Malaysian Chinese through a GWA analysis. Although no study has so far addressed the functional consequences of the SNPs we identified, one particular SNP or some haplotype may function as an enhancer and cause considerable difference in the expression and activities of *ITGA9*, which in turn determine individual susceptibility to NPC. We also cannot exclude the possibility that other unidentified candidate functional polymorphisms are present in *ITGA9*.

ITGA9 is also known as ITGA4L (integrin-α 4-like), ALPHA-RLC (alpha related to the development of lung cancer). The gene is located at the chromosomal 3p22–21.3 segment, which is known to be commonly deleted in various types of carcinoma including NPC. Loss of heterozygosity on 3p is observed in almost all primary NPC.[14,30] A linkage study also mapped an NPC susceptibility locus to chromosome 3p21.31–21.2, indicating that the genes in this region are crucial for the formation of NPC.[15] Hence, *ITGA9* might be a novel NPC susceptibility gene.

ITGA9 encodes an integrin, α9 subunit. Integrins constitute a superfamily of integral membrane glycoproteins bound to extracellular matrix ligands, cell surface ligands and soluble ligands, which mediate cell–cell and cell–matrix adhesion. Each integrin consists of an alpha (α) and a beta (β) chain. In humans, 18 α-subunits and 8 β-subunits are known, forming 24 integrin heterodimers.[31] In addition to mediating cell adhesion, signals from integrins are now known to modulate cell behaviors including proliferation, survival or apoptosis, maintenance of polarity, shape, motility, haptotaxis, gene expression, differentiation and malignant transformation.[31–33]

The integrin α9 is a 1035 amino acids polypeptide that contains a large N-terminal extracellular domain with seven conserved repeats, a transmembrane segment and a short C-terminal cytoplasmic tail. The α9 subunit associates with the β1 chain only to form a single integrin, α9β1, which is widely expressed, including in the human airway epithelial cells, and binds to various diverse ligands such as tenascin, VCAM-1, osteopontin, uPAR, plasmin, angiostatin, ADAM, thrombospondin-1, VEGF-C and VEGF-D.[31]

At present, the functional relevance of α9 on NPC development is still unclear. The risk of NPC associated with multiple variants in *ITGA9* may be under the influence of other genetic factors and/or environmental factors. Herpes viruses have been shown to exploit integrin for cell entry and infection. Their outer envelope glycoproteins contain a consensus motif that can bind to integrin to promote cell entry and infection.[34] The apparent molecular mechanisms by which EBV invades and infects epithelial cells are still largely unknown, although a surface molecule, which is antigenically related to the CD21 receptor, has been described and may serve as a receptor for virus internalization.[35,36] Alternatively, it has been suggested that EBV may enter nasopharyngeal cells through IgA-mediated endocytosis.[37] Therefore, it is tempting to hypothesize that a consensus domain of the EBV capsid protein may interact with the α9β1 expressed on the surface of epithelial cells for attachment and/or cell entry. Ligation of α9β1 by EBV may subsequently elicit potent

Table 1 Relationship between SNPs in *ITGA9* and NPC susceptibility in Malaysian Chinese

					Case				*Control*					P-*value (Fisher's exact test)*			
					Genotype				*Genotype*								
SNP ID	*Position*		*Alleles*[a]	*Panel*	*11*	*12*	*22*	*MAF*	*11*	*12*	*22*	*MAF*	*Risk allele*	*Allele 1 vs allele2*	*Genotype 11 vs 12+22*	*Genotype 11+12 vs 22*	*Odds ratio (95% CI)*
rs169188	37482453	Intron 1	G>A	A	95	15	1	0.077	247	11	0	0.021	A	0.000598	0.00107	0.301	3.78 (1.69–8.44)
				Combined	235	32	5	0.077	485	22	0	0.022		0.000000316	0.00000572	0.00507	3.47 (2.00–6.02)
rs197721	37485040	Intron 1	C>A	A	94	16	1	0.081	244	13	0	0.025	A	0.000916	0.00167	0.302	3.39 (1.59–7.26)
				Combined	238	33	5	0.078	482	25	0	0.025		0.00000145	0.0000218	0.00531	3.08 (1.82–5.22)
rs149816	37485770	Intron 1	C>G	A	94	16	1	0.081	244	13	0	0.025	G	0.000916	0.00167	0.302	3.39 (1.59–7.26)
				Combined	238	33	5	0.078	482	25	0	0.025		0.00000145	0.0000218	0.00531	3.08 (1.82–5.22)
rs169111	37490537	Intron 3	A>G	A	93	16	1	0.082	243	14	0	0.027	G	0.00152	0.00238	0.300	3.17 (1.50–6.69)
				Combined	231	35	5	0.083	480	27	0	0.027		0.000000898	0.0000124	0.00501	3.08 (1.84–5.14)
rs197770	37490831	Intron 3	T>C	A	94	15	1	0.077	244	13	0	0.025	C	0.00172	0.00312	0.300	3.19 (1.48–6.90)
				Combined	234	32	5	0.077	482	24	0	0.024		0.00000110	0.0000173	0.00504	3.18 (1.86–5.43)
rs2212020	37492466	Intron 3	C>T	A	72	35	3	0.186	221	36	3	0.081	T	0.0000531	0.0000363	0.368	2.99 (1.78–5.03)
				Combined	190	76	11	0.177	425	81	6	0.091		0.000000827	0.00000434	0.0121	2.24 (1.59–3.15)
rs189897	37493549	Intron 3	A>T	A	90	19	2	0.104	240	16	0	0.031	A	0.000123	0.000379	0.0909	3.50 (1.75–7.01)
				Combined	226	37	7	0.094	473	29	0	0.029		0.0000000685	0.00000334	0.000608	3.18 (1.94–5.21)
rs197757	37497832	Intron 3	C>T	A	89	19	2	0.105	238	17	0	0.033	T	0.000205	0.000581	0.0902	3.30 (1.67–6.55)
				Combined	224	37	7	0.095	472	30	0	0.030		0.000000104	0.00000470	0.000588	3.09 (1.89–5.05)

Abbreviations: CI, confidence interval; MAF, minor allele frequency; NPC, nasopharyngeal carcinoma; SNP, single-nucleotide polymorphism.
Panel A: 111 NPC cases, 260 controls. Combined: 279 NPC cases, 512 controls. Major alleles in cases and controls were defined as allele 1.
[a]Major allele >minor allele.

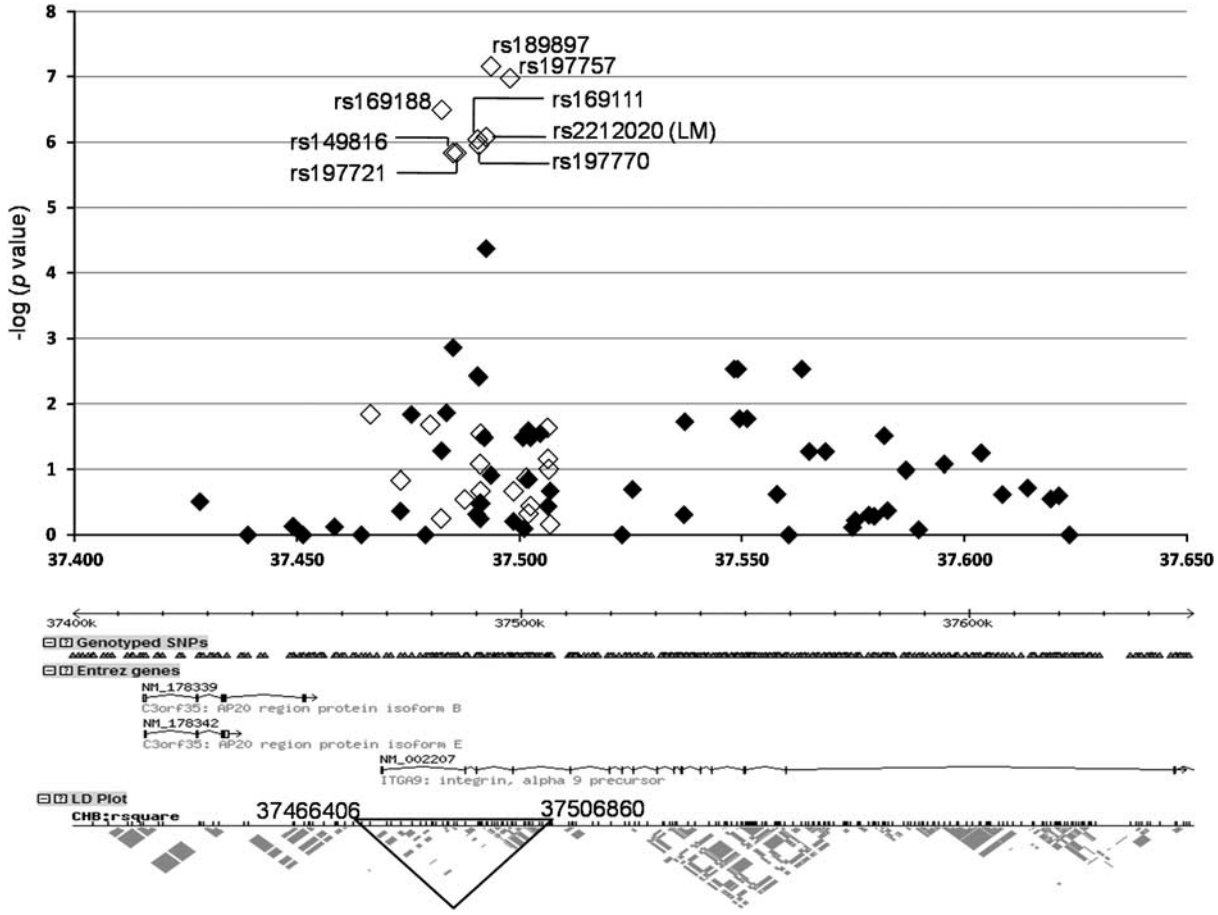


Figure 2 Case–control association plots, linkage disequilibrium (LD) map and the genomic structure of an *ITGA9* region on chromosome 3p21. Closed and open diamond represent $-\log_{10}$ (*P*-value) obtained from the genome-wide association study (Panel A) and fine mapping (combined samples), respectively. Pairwise LD (r^2) was based on the genotype data of Han Chinese in the International HapMap Project database.

signaling responses that could promote cell proliferation and viral pathogenesis.

This study was carried out on a relatively small initial sample containing 279 cases and 512 controls. Although our approach did not have sufficient statistical power to identify all the genetic determinants, it should have the power to identify a gene(s) having a relatively large effect such as *ITGA9*. To our knowledge, this was the first report of the GWA approach to identify genetic risk factors of NPC. Replication studies with a larger number of samples as well as in other independent populations need to be carried out for further confirmation of the association of *ITGA9* and NPC.

In conclusion, we suggest that *ITGA9* on chromosome 3p21 may be a genetic risk factor for NPC development in Malaysian Chinese. Our finding is of great interest as it sheds new light on the association between integrin α9β1 and NPC. Further studies are needed to elucidate the causal SNP or haplotype and to provide a plausible biological mechanism for the observed association between polymorphisms and susceptibility to NPC. The identification and characterization of NPC susceptibility genes help us to understand the pathogenesis of NPC and could possibly lead to an improvement of treatment in the future.

ACKNOWLEDGEMENTS

We thank all the participants in this study, Dr Shashinder Singh, Dr Anura Michelle Manuel and staffs of the Department of Otorhinolaryngology, Dr Veera Sekaran Nadarajah from the Department of Pathology, staffs of the Blood Bank, UMMC and Ms Maliza Nawawi from the NCI Cancer Hospital. In Malaysia, this study was supported by the Research University Grant (FQ018/2007A) from the University of Malaya.

1 Brennan, B. M. Nasopharyngeal carcinoma. *Orphanet J. Rare Dis.* **1,** 23 (2006).
2 Zhou, X. F., Cui, J., Macias, V., Kajdacsy-Balla, A. A., Ye, H., Wang, J. G. *et al.* The progress on genetic analysis of nasopharyngeal carcinoma. *Comp Func Genomics* Article ID 57513, 13 pages (2007).
3 Lim, G. C. C. & Halimah, Y. (eds) *Second Report of the National Cancer Registry. Cancer Incidence in Malaysia 2003* (National Cancer Registry, Kuala Lumpur, 2004).
4 Abu Bakar, S. in *Cancer Incidence in Peninsular Malaysia 2003–2005* (eds. Lim, G.C.C., Rampal, S. & Halimah, Y.) 71–73 (National Cancer Registry, Kuala Lumpur, 2008).
5 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).
6 Raab-Traub, N. in *Epstein-Barr Virus* (ed. Robertson, E.S.) 73 (Caister Academic Press, England, 2005).
7 Armstrong, R. W., Imrey, P. B., Lye, M. S., Armstrong, M. J., Yu, M. C. & Sani, S. Nasopharyngeal carcinoma in Malaysian Chinese: salted fish and other dietary exposures. *Int. J. Cancer* **77,** 228–235 (1998).
8 Jeannel, D., Bouvier, G. & Huber, A. Nasopharyngeal carcinoma: an epidemiological approach to carcinogenesis. *Cancer Surv.* **33,** 125–155 (1999).
9 McDermott, A. L., Dutt, S. N. & Watkinson, J. C. The aetiology of nasopharyngeal carcinoma. *Clin. Otolaryngol.* **26,** 82–92 (2001).
10 Hildesheim, A. & Levine, P. H. Etiology of nasopharyngeal carcinoma: a review. *Epidemiol. Rev.* **15,** 466–485 (1993).

11 Goldsmith, D. B., West, T. M. & Morton, R. HLA associations with nasopharyngeal carcinoma in Southern Chinese: a meta-analysis. *Clin. Otolaryngol.* **27,** 61–67 (2002).
12 Lieboitz, D. Nasopharyngeal carcinoma. The Epstein–Barr virus association. *Semin. Oncol.* **21,** 382–397 (1994).
13 Hui, AB. T., Lo, K. W., Leung, S. F., Teo, P., Fung, M. K. F., To, K. F. *et al.* Detection of recurrent chromosomal gains and losses in primary nasopharyngeal carcinoma by comparative genomic hybridization. *Int. J. Cancer* **82,** 498–503 (1999).
14 Lo, K. W., Teo, P. M., Hui, A. B., To, K. F., Tsang, Y. S., Chan, S. Y. *et al.* High-resolution allelotype of microdissected primary nasopharyngeal carcinoma. *Cancer Res.* **60,** 3348–3353 (2000).
15 Xiong, W., Zeng, Z. Y., Xia, J. H., Xia, K., Shen, S. R., Li, X. L. *et al.* A susceptibility locus at chromosome 3p21 linked to familial nasopharyngeal carcinoma. *Cancer Res.* **64,** 1972–1974 (2004).
16 Feng, B. J., Huang, W., Shugart, Y. Y., Lee, M. K., Zhang, F., Xia, J. C. *et al.* Genome-wide scan for familial nasopharyngeal carcinoma reveals evidence of linkage to chromosome 4. *Nat. Genet.* **31,** 395–399 (2002).
17 Hildesheim, A., Anderson, L. M., Chen, C. J., Cheng, Y. J., Brinton, L. A., Daly, A. K. *et al.* CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. *J. Natl Cancer Inst.* **89,** 1207–1212 (1997).
18 Tiwawech, D., Srivatanakul, P., Karalak, A. & Ishida, T. Cytochrome P450 2A6 polymorphism in nasopharyngeal carcinoma. *Cancer Lett.* **241,** 135–141 (2006).
19 Deng, Z. L., Wei, Y. P., Luo, W., Liao, Z. L. & Ma, Y. Glutathione *S*-transferase M1 and T1 gene deletion associated with increased susceptibility to nasopharyngeal carcinoma. *The Chinese-German Journal of Clinical Oncology* **4,** 276–278 (2005).
20 Cho, E. Y., Hildesheim, A., Chen, C. J., Hsu, M. M., Chen, I. H., Mittl, B. F. *et al.* Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. *Cancer Epidemiol. Biomarkers Prev.* **12,** 1100–1104 (2003).
21 Cao, Y., Miao, X. P., Huang, M. Y., Deng, L., Hu, L. F., Ernberg, I. *et al.* Polymorphisms of XRCC1 genes and risk of nasopharyngeal carcinoma in the Cantonese population. *BMC Cancer* **6,** 167 (2006).
22 He, Y., Zhou, G., Zhai, Y., Dong, X., Lv, L., He, F. *et al.* Association of PLUNC gene polymorphisms with susceptibility to nasopharyngeal carcinoma in a Chinese population. *J. Med. Genet.* **42,** 172–176 (2005).
23 Nasr, H. B., Mestiri, S., Chahed, K., Bouaouina, N., Gabbouj, S., Jalbout, M. *et al.* Matrix metalloproteinase-1 (-1607) 1G/2G and -9 (-1562) C/T promoter polymorphisms: susceptibility and prognostic implications in nasopharyngeal carcinomas. *Clinica. Chimica. Acta.* **384,** 57–63 (2007).
24 Catarino, R. J., Breda, E., Coelho, V., Pinto, D., Sousa, H., Lopes, C. *et al.* Association of the A870G cyclin D1 gene polymorphism with genetic susceptibility to nasopharyngeal carcinoma. *Head Neck* **28,** 603–608 (2006).
25 Song, C., Chen, L. Z., Zhang, R. H., Yu, X. J. & Zeng, Y. X. Functional variant in the 3 - untranslated region of toll-like receptor 4 is associated with nasopharyngeal carcinoma risk. *Cancer Biol. Ther.* **5,** 1285–1291 (2006).
26 Zhou, X. X., Jia, W. H., Shen, G. P., Qin, H. D., Yu, X. J., Chen, L. Z. *et al.* Sequence variants in toll-like receptor 10 are associated with nasopharyngeal carcinoma risk. *Cancer Epidemiol. Biomarkers Prev.* **15,** 862–866 (2006).
27 International HapMap Consortium. A haplotype map of the human genome. *Nature* **437,** 1229–1320 (2005).
28 International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–862 (2007).
29 Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. & Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46,** 471–477 (2001).
30 Chan, A. S. C., To, K. F., Lo, K. W., Mak, K. F., Pak, W., Chiu, B. *et al.* High frequency of chromosome 3p deletion in histologically normal nasopharyngeal epithelia from southern Chinese. *Cancer Res.* **60,** 5365–5370 (2000).
31 Takada, Y., Ye, X. J. & Scott, S. The integrins. *Genome Biol.* **8,** 215 (2007).
32 Sheppard, D. Functions of pulmonary epithelial integrins: from development to disease. *Physiol. Rev.* **83,** 673–686 (2003).
33 Guo, W. & Giancotti, F. G. Integrin signalling during tumour progression. *Nat. Rev. Mol. Cell Biol.* **5,** 816–826 (2004).
34 Stewart, P. L. & Nemerow, G. R. Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol.* **15,** 500–507 (2007).
35 Young, L., Dawson, C., Brown, K. & Rickinson, A. Identification of human epithelial cell surface protein sharing an epitope with the C3d/Epstein-Barr virus receptor molecule of B lymphocytes. *Int. J. Cancer* **43,** 786–794 (1989).
36 Birkenbach, M., Tong, X., Bradbury, L. E., Tedder, T. F. & Kieff, E. Characterization of an Epstein-Barr virus receptor on human epithelial cells. *J. Exp. Med.* **176,** 1405–1414 (1992).
37 Lin, C. T., Lin, C. R., Tan, G. K., Chen, W., Dee, A. N. & Chan, W. Y. The mechanism of Epstein-Barr virus infection in nasopharyngeal carcinoma cells. *Am. J. Pathol.* **150,** 1745–1756 (1997).

Identification of a Functional Variant in *SPLUNC1* Associated With Nasopharyngeal Carcinoma Susceptibility Among Malaysian Chinese

Poh-Yin Yew,[1,2] Taisei Mushiroda,[3] Kazuma Kiyotani,[3] Gopala Krishnan Govindasamy,[4] Lee-Fah Yap,[5] Soo-Hwang Teo,[5] Paul Vey-Hong Lim,[6] Selvaratnam Govindaraju,[7] Kananathan Ratnavelu,[7] Choon-Kook Sam,[8] Yoke-Yeow Yap,[9] Alan Soo-Beng Khoo,[10] Kin-Choo Pua,[11]Yusuke Nakamura[1,12] and The Malaysian NPC Study Group,[13] Ching-Ching Ng[1,2]*

[1]*Laboratory for International Alliance, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Japan*
[2]*Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia*
[3]*Laboratory for Pharmacogenetics, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Japan*
[4]*Department of Otorhinolaryngology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia*
[5]*Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, Selangor, Malaysia*
[6]*Tung Shin Hospital, Jalan Pudu, Kuala Lumpur, Malaysia*
[7]*NCI Cancer Hospital, Nilai, Negeri Sembilan, Malaysia*
[8]*National Institute of Education, Nanyang Technological University, Singapore*
[9]*Faculty of Medicine and Health Sciences, Clinical Campus Universiti Putra Malaysia, Jalan Pahang, Kuala Lumpur, Malaysia*
[10]*Molecular Pathology Unit, Cancer Research Centre, Institute for Medical Research, Jalan Pahang, Kuala Lumpur, Malaysia*
[11]*Department of Otorhinolaryngology, Hospital Pulau Pinang, Jalan Residensi, Penang, Malaysia*
[12]*Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan*
[13]*Malaysian Nasopharyngeal Carcinoma study group: Hospital Pulau Pinang, Hospital Kuala Lumpur/UPM, University of Malaya, Institute of Medical Research, Cancer Research Initiatives Foundation, Sarawak General Hospital/UNIMAS, Queen Elizabeth Hospital and Hospital USM Kubang Kerian*[†]

Nasopharyngeal carcinoma (NPC) is a multifactorial and polygenic disease with high incidence in Asian countries. Epstein–Barr virus infection, environmental and genetic factors are believed to be involved in the tumorigenesis of NPC. The association of single nucleotide polymorphisms (SNPs) in *LPLUNC1* and *SPLUNC1* genes with NPC was investigated by performing a two-stage case control association study in a Malaysian Chinese population. The initial screening consisted of 81 NPC patients and 147 healthy controls while the replication study consisted of 366 NPC patients and 340 healthy controls. The combined analysis showed that a SNP (rs2752903) of *SPLUNC1* was significantly associated with the risk of NPC (combined $P = 0.00032$, odds ratio = 1.62, 95% confidence interval = 1.25–2.11). In the subsequent dense fine mapping of *SPLUNC1* locus, 36 SNPs in strong linkage disequilibrium with rs2752903 ($r^2 \geq 0.85$) were associated with NPC susceptibility. Screening of these variants by electrophoretic mobility shift and luciferase reporter assays showed that rs1407019 located in intron 3 ($r^2 = 0.994$ with rs2752903) caused allelic difference in the binding of specificity protein 1 (Sp1) transcription factor and affected luciferase activity. This SNP may consequently alter the expression of *SPLUNC1* in the epithelial cells. In summary, our study suggested that rs1407019

Additional Supporting Information may be found in the online version of this article.

Abbreviations: NPC, nasopharyngeal carcinoma; EBV, Epstein–Barr virus; HLA, human leukocyte antigens; PLUNC, palate, lung, nasal epithelium clone; CYP2E1, cytochrome P450 2E1; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; MAF, minor allele frequency; HWE, Hardy–Weinberg equilibrium; EMSA, electrophoretic mobility shift assay; DIG, digoxigenin-11-ddUTP; Sp1, specificity protein 1; CI, confidence interval.

[†]Malaysian Nasopharyngeal Carcinoma Study Group—Hospital Pulau Pinang: K.C. Pua (Project Leader), S. Subathra, Y.S. Ee, M. Goh, A.H. Rahmah, C.H. Fong, N. Punithavathi, L.M. Ong. Hospital Kuala Lumpur/UPM: Y.Y. Yap, B.D. Dipak, R. Deepak, F.N. Lau, P. Shalini, M.A. Atiliyana. University of Malaya: G. Gopala Krishnan, M. Anura, A.Z. Bustam, L.M. Looi, P. Shahfinaz, O. Hashim, C.C. Ng, O. Rahmat, J. Amin. Institute for Medical Research: A.S.B. Khoo, (Programme Leader), N.M. Kumaran. Cancer Research Initiatives Foundation: S.H. Teo, L.F. Yap. Sarawak General Hospital/UNIMAS: S.K. Subramaniam, T.S. Tiong, U.H. Sim, T.W. Tharumalingam, D. Norlida, M. Zulkarnaen, W.H. Lai. Queen Elizabeth Hospital: C.A. Ong, S. Bokari, C.L. Lum, A. Vivekananda, A. Othman, D. Jayendran, A. Kam. Hospital USM Kubang Kerian: S. Hassan, B. Biswal, H. Nik Fariza, H.A. Mubassir, A.H. Suzina Sheikh.

*Correspondence to: Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Received 6 June 2011; Revised 3 November 2011; Accepted 17 November 2011

DOI 10.1002/mc.21857

Published online in Wiley Online Library (wileyonlinelibrary.com).

in intronic enhancer of *SPLUNC1* is associated with NPC susceptibility in which its A allele confers an increased risk of NPC in the Malaysian Chinese population.

Key words: head and neck; single nucleotide polymorphism; genetic susceptibility

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a head and neck malignancy with striking racial and geographical distribution. It is rare in most countries (approximately 1 per 100,000); however, the highest incidence is observed in Southern China (20–30 per 100,000 people) [1]. Chinese emigrants also showed higher incidence of NPC in their countries of migration [1–4]. According to the report of Malaysia Cancer Statistic 2006, NPC is the third most common cancer among male and fifteenth among female in Peninsular Malaysia. The Chinese had a higher incidence compared to Malay and Indian population [3]. The high incidence of NPC in certain ethnic group suggested that genetic factors play a role in the contribution to the development of NPC in addition to environmental factors such as exposure to food carcinogen or chemical carcinogens, and Epstein–Barr virus (EBV) infection [2,4–8].

Identification of susceptible genes to NPC is important for improving our understanding of biological and etiologic mechanism involved in the development of NPC. Several studies have reported the association between NPC and polymorphisms in the genes which are involved in the activation of carcinogens and in the defence against xenobiotics, including cytochrome P450 2E1 (*CYP2E1)*, human leukocyte antigens (*HLA*) and palate, lung, nasal epithelium clone (*PLUNC*) genes [7,9–12]. Besides that, recent genome-wide association studies in the Taiwanese, Southern Chinese, and Malaysian Chinese populations have identified several NPC susceptibility loci including *HLA-A*, *HLA-F*, *GABBR1*, *TNFRSF19*, *MDS1-EVI1*, *CDKN2A-CDKN2B*, and *ITGA9* [13–15].

The PLUNC family may play a significant role in host defence in the mouth, nose, and upper airways [16,17]. The gene cluster of PLUNC family is located on chromosome 20q11.2. Members of the PLUNC family are subdivided into two groups based on their size, that is, ''short'' protein and ''long'' protein. ''Short'' protein comprises SPLUNC1, SPLUNC2, and SPLUNC3, while ''Long'' protein comprises LPLUNC1, LPLUNC2, LPLUNC3, and LPLUNC4 [16–19]. It was reported that SPLUNC1 and LPLUNC1 were down-regulated in the NPC biopsies [20], suggesting that these proteins might be involved in the development of NPC.

LPLUNC1 is highly expressed in normal adult and fetal nasopharyngeal epithelial tissue. LPLUNC1 was reported to suppress the growth and proliferation of human NPC cell line, thus it might play a role in inhibiting the development of NPC [21]. However, no association study of *LPLUNC1* polymorphism and NPC susceptibility has yet been conducted. SPLUNC1 (also known as *LUNX*, *SPURT*, or *NASG*) is most prominently expressed in nose, salivary glands, and upper respiratory tract [16,17]. The *SPLUNC1* mRNA was found to be highly expressed in the normal nasal septum epithelium, salivary gland, and tracheobronchial epithelial cells [16]. A previous study revealed that two promoter single nucleotide polymorphisms (SNPs) of *SPLUNC1* gene (C-2128T and C-1888T; rs2752903 and rs750064, respectively) were associated with susceptibility to NPC in the Cantonese-speaking Chinese in China [12]. However, the functional importance of these two promoter SNPs are still unknown. Therefore, the aim of this study was to investigate the associations of SNPs in *LPLUNC1* and *SPLUNC1* with NPC susceptibility among Malaysian Chinese and to elucidate the functional significance of the associated SNPs.

MATERIALS AND METHODS

Subjects

All the subjects who participated in this study were unrelated Malaysian Chinese. The patients were pathologically diagnosed as NPC, while the healthy controls were consisted of a group of blood donors. For an initial screening, 81 NPC patients from University of Malaya Medical Center (UMMC) and 147 healthy controls from Blood Bank of UMMC were recruited. For a replication study, independent 366 NPC patients were recruited later from UMMC, NCI Cancer Hospital, Tung Shin Hospital, and two NPC referring hospitals of the Malaysian NPC study group, which are Hospital Pulau Pinang and Hospital Kuala Lumpur/Universiti Putra Malaysia (UPM) as well as 340 controls from Blood Bank of UMMC. The Malaysian NPC study group consists of clinicians and research scientists which recruits NPC patients from five major NPC referring hospitals across Malaysia: Hospital Pulau Pinang, Hospital Kuala Lumpur/Universiti Putra Malaysia (UPM), Sarawak General Hospital (Hospital Umum Sarawak)/ Universiti Malaysia Sarawak, Queen Elizabeth Hospital in Sabah, UMMC, and Hospital Universiti Sains Malaysia (USM) in Kubang Kerian, Kelantan.

Of the 447 cases, 75% were males and their median age was 53 (ranged from 14 to 86). Of the 487 controls, 77% were males and their median age was 37 (ranged from 19 to 60). Written informed consent was obtained from all the study subjects.

Ethical approvals for conducting the study were obtained from the Medical Ethics Committee of UMMC and Ethical Committees of the Yokohama Institute, The Institutes of Physical and Chemical Research (RIKEN), Yokohama, Japan.

Selection and Genotyping of SNPs

We selected the tag-SNPs that encompass the linkage disequilibrium (LD) block containing *SPLUNC1* and *LPLUNC1* genes using the HapMap phase II CHB data (www.hapmap.org) with a pairwise r^2 of > 0.80 and minor allele frequency (MAF) of ≥ 0.1. Four SNPs of *SPLUNC1* were genotyped with commercially available TaqMan pre-designed probes and primers by means of GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. The four SNPs in *LPLUNC1* were genotyped by multiplex PCR-based Invader assay as described previously (Third Wave Technologies, Madison, WI) [22]. The genotypes were determined by ABI PRISM 7900HT (Applied Biosystems). In the subsequent fine mapping of the *SPLUNC1* locus, SNP screening was performed by direct sequencing of genomic DNA from 48 healthy controls using ABI3730*xl* sequencer (Applied Biosystems). The identified SNPs were then genotyped by multiplex PCR-based Invader assay in all the samples [22]. For quality control, random samples were genotyped twice to verify the genotyping consistency, and concordant results were obtained. All the SNPs had a call rate of more than 96% and no deviation from Hardy–Weinberg equilibrium (HWE) ($P \geq 0.01$) was observed.

Cell Culture and Reverse Transcription (RT)-PCR

Hela, BT549, MCF7, ZR-75-1, NCI-H520, NCI-H1781, NCI-H2170, A549, HepG2 and HuH-7 cell lines were purchased from the American Type Culture Collection (ATCC, Rockville, MD) while the NPC cell lines (CNE1, HONE1, SUNE1, HONE Akata) were kindly gifted by various laboratories. The cells were maintained in RPMI-1640 or Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum at 37°C in 5% CO_2.

Total RNA was isolated using an RNeasy plus Mini Kit (Qiagen, Valencia, CA). cDNA was synthesized from total RNA (1 μg) using SuperScriptIII First Strand Synthesis System (Invitrogen, Carlsbad, CA). *SPLUNC1* mRNA expression was determined by semi-quantitative PCR using forward primer: 5′-GCCTGAACAACATCATTGAC-3′ and reverse primer: 5′-TGTCCAGAAGACCTT-GAATG-3′. *GAPDH* was used as an internal control.

Electrophoretic Mobility Shift Assay (EMSA)

Nuclear extracts were prepared from NCI-H1781 cells. EMSA was performed using digoxigenin-11-ddUTP (DIG) Gel Shift Kit, 2nd Generation (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. Oligonucleotides (41 base pairs) were designed corresponding to genomic sequences that surrounded the SNPs of interest, and were labelled with DIG. We incubated the nuclear extracts with the DIG-labelled probes for 20 min at 4°C. Protein-DNA complexes were separated by electrophoresis on a 7.5% polyacrylamide gel with 0.5 × Tris-borate EDTA buffer, followed by transfer to a nylon membrane and detection by chemiluminescent signal detection system. For the competition assay, we pre-incubated the nuclear extracts with a 125-fold excess of unlabelled oligonucleotides before adding the DIG-labelled probes. In the supershift assay, the nuclear extracts were pre-incubated with monoclonal antibody specific to specificity protein 1 (Sp1) (E-3, Santa Cruz Biotechnology, Santa Cruz, CA) for 30 min at 4°C and then incubated with DIG-labelled probe for another 20 min at 4°C. Each experiment was independently repeated three times.

Luciferase Reporter Assay

The pGL4-SV40 was constructed by inserting the SV40 promoter into the pGL4 luciferase vector (pGL4.10 [luc2]) (Promega, Madison, WI) at the *Bgl* II/*Hin*d III site. A single copy of the 41-bp double stranded oligonucleotides which contain A or G allele of rs1407019 was then cloned into *Xho* I/*Bgl* II site of this pGL4-SV40. The NCI-H1781 cells were transfected with these constructs and pGL4-SV40-hRluc (an internal control for transfection efficacy) using nucleofector system (Amaxa, Köln, Germany). After 48 h, cells were lysed and luciferase activity was measured using a Dual-Luciferase reporter assay system (Promega). Each experiment was independently repeated three times and each sample was studied in quadruplicate.

Statistical Analyses

A two-stage approach was carried out for the association analysis. A SNP that showed a *P*-value less than a significance level in eight multiple tests (0.00625) in the first screening was further genotyped in the second stage. Haploview 4.1 (Daly laboratory at the Broad Institute, Cambridge, MA) was used to calculate the r^2 correlation coefficient between SNPs. We performed single SNP association analysis using a Fisher's exact test under allelic, dominant-inheritance and recessive-inheritance models. The association between rs2752903 and NPC risk was then adjusted for age and sex by a logistic regression analysis using PLINK (v1.07) [23]. The difference of luciferase reporter activities between two alleles was analyzed by Student's *t*-test.

RESULTS

Association Between SNPs in *PLUNC* and Risk of NPC

In the first screening, a total of eight haplotype-tagging SNPs (tag-SNPs), four each of *LPLUNC1* and S*PLUNC1* were genotyped in 81 NPC patients and 147 healthy controls. Only one promoter SNP of *SPLUNC1* (rs2752903) showed *P*-value of less than a significance level in eight multiple tests ($P < 0.00625$; Table 1), indicating that there is a significant association of this SNP with the susceptibility to NPC.

Next, this SNP was further genotyped using the second set of samples, consisting of 366 NPC patients and 340 healthy controls, and the association was successfully replicated ($P = 0.040$; Table 2). The combined *P*-values of the two data sets revealed that the rs2752903 remained significantly associated with NPC susceptibility after the adjustment for multiple testing by the strict Bonferroni correction ($P = 0.00032$; Table 2). The frequency of individuals with at least one risk C allele (CC, TC) in NPC patients (47.6%) was higher compared to controls (35.9%) (Odds ratio = 1.62, 95% confidence interval (CI) = 1.25–2.11 for dominant-inheritance model). Potential confounding factors were also analyzed. The median age in cases was significantly higher than that in controls ($P < 0.0001$). The gender distribution was not significantly different between two groups ($P = 0.62$). Even after adjusting for age and sex in a logistic regression analysis, the association of rs2752903 with NPC susceptibility still remained significant (adjusted $P = 0.00097$, odds ratio = 1.79, 95% CI = 1.27–2.52 for dominant-inheritance model; Table 3), suggesting that rs2752903 genotype is an independent risk factor for NPC.

To search whether additional SNPs in *SPLUNC1* are associated with NPC susceptibility, we re-sequenced a 17-kb LD block including the whole gene of *SPLUNC1*. We identified a total of 40 SNPs including one novel SNP (SNP01); however, no SNP was found within the exon region of *SPLUNC1*. Genotyping of these 40 SNPs in all cases and controls revealed that 36 SNPs, in high LD with rs2752903, are significantly associated with NPC ($P = 0.000044$–0.00056; Table 4).

Screening of Functional SNPs Using Electrophoretic Mobility Shift and Luciferase Reporter Assays

The *SPLUNC1* mRNA expression was examined in 14 cell lines including four NPC cell lines (CNE1, HONE1, SUNE1, and HONE Akata), Hela, BT549, MCF7, ZR-75-1, NCI-H520, NCI-H1781, NCI-H2170, A549, HepG2, and HuH-7. Low or no expression of *SPLUNC1* was detected in most of the cell lines except for human bronchoalveolar carcinoma-derived NCI-H1781 (Supplementary

Table 1. Relationship Between tag-SNPs in *LPLUNC1* and *SPLUNC1* and NPC Susceptibility in Malaysian Chinese patients

Gene	SNP ID	Location	Alleles (1/2)	Cases[a]				Controls[a]				*P*-value			Odds ratio (95%CI)	
				11[b]	12	22	MAF	11[b]	12	22	MAF	1 vs. 2	vs. 11	vs. 22	vs. 11	vs. 22
LPLUNC1	rs1884882	Intergenic	T/C	55	22	4	0.19	79	59	9	0.26	0.066	0.049	0.78	1.82 (1.03–3.21)	1.26 (0.37-4.21)
	rs4911314	Intronic	C/G	52	26	3	0.20	80	62	5	0.24	0.29	0.16	1.00	1.50 (0.86–2.62)	1.09 (0.25–4.69)
	rs1999663	Non_synonymous_coding	C/G	45	31	5	0.25	81	55	11	0.26	0.91	1.00	0.79	1.02 (0.59–1.76)	1.23 (0.41–3.67)
	rs6087476	Intergenic	C/T	32	34	15	0.40	57	68	22	0.38	0.84	1.00	0.57	1.03 (0.59–1.80)	1.29 (0.63–2.66)
SPLUNC1	rs927159	Intergenic	C/T	58	21	2	0.15	85	53	9	0.24	0.031	0.046	0.34	1.84 (1.03–3.30)	2.58 (0.54–12.22)
	rs2752903	5′-near gene	T/C	39	34	8	0.31	104	39	4	0.16	0.00024[c]	0.0010[c]	0.029	2.60 (1.48–4.57)	3.92 (1.14–13.44)
	rs750064	5′-near gene	T/C	24	38	19	0.47	52	71	24	0.40	0.20	0.46	0.22	1.30 (0.72–2.33)	1.57 (0.80–3.08)
	rs1570034	Intergenic	C/T	23	40	18	0.47	49	75	23	0.41	0.24	0.46	0.28	1.26 (0.70–2.28)	1.54 (0.77–3.06)

MAF, minor allele frequency; CI, confidence interval.
[a]Cases, $N = 81$; Controls, $N = 147$.
[b]Major allele in controls was defined as allele 1.
[c]Statistically significant after Bonferroni correction, based on 8 independent effective tests ($P < 0.00625$).

Table 2. Relationship Between rs2752903 and NPC Susceptibility in Exploratory and Replication Studies

Gene	SNP ID	Location	Alleles (1/2)	Sample set	Cases[a]				Controls[a]				*P*-value			Odds ratio (95%CI)	
					11[b]	12	22	MAF	11[b]	12	22	MAF	1 vs. 2	vs. 11	vs. 22	vs. 11	vs. 22
SPLUNC1	rs2752903	5′-near gene	T/C	1st	39	34	8	0.31	104	39	4	0.16	0.00024[c]	0.0010[c]	0.029	2.60 (1.48–4.57)	3.92 (1.14–13.44)
				2nd	194	143	27	0.27	208	105	27	0.23	0.13	0.040	0.89	1.38 (1.02–1.86)	1.08 (0.62–1.88)
				1st+2nd	233	177	35	0.28	312	144	31	0.21	0.00097[c]	0.00032[c]	0.44	1.62 (1.25–2.11)	1.26 (0.76–2.07)

MAF, minor allele frequency; CI, confidence interval.
[a]1st: Cases, $N = 81$; Controls, $N = 147$, 2nd: Cases, $N = 366$; Controls, $N = 340$, 1st+2nd: Cases, $N = 447$; Controls, $N = 487$.
[b]Major allele in controls was defined as allele 1.
[c]Statistically significant after Bonferroni correction, based on 8 independent effective tests ($P < 0.00625$).

Table 3. Multivariate Logistic Regression Analysis for NPC Susceptibility

Variables	Estimated regression coefficient (SE)	Odds ratio (95%CI)	*P*-value
rs2752903*C	0.58 (0.18)	1.79 (1.27–2.52)	0.00097
Age	0.15 (0.0099)	1.16 (1.14–1.19)	<0.0001
Sex	0.095 (0.19)	1.10 (0.76–1.60)	0.62

Materials: Figure S1). Therefore, NCI-H1781 was used for the following functional studies.

To identify the causative SNP(s), screening of allelic differences in transcription factor binding using EMSA was performed (data not shown). We observed a difference in the transcription factor binding ability between two alleles for only one SNP, rs1407019, which is located in intron 3. Two DNA-protein complexes (C1 and C2 complexes) were detected specifically to the non-susceptible allele G of rs1407019 but not to the susceptible allele A (Figure 1B) and the G-allele shifted bands were disappeared only by the addition of unlabelled G-allele oligonucleotide (self competitor), suggesting that some nuclear protein(s) might interact with this genomic region.

The motif analysis by TFSEARCH database (http://mbs.cbrc.jp/research/db/TFSEARCH.html) suggested that rs1407019 was located in the binding sequence of transcription factor Sp1 (Figure 1A). The sequence containing G-allele had higher matrix similarity to the consensus Sp1 binding motif than the sequence harboring A-allele. In the competition assay, DNA-C1 complex formation was competed by addition of an unlabelled oligonucleotide with consensus Sp1 binding motif, suggesting the binding of Sp1 or Sp1-like protein to the sequence (Figure 1B). Furthermore, a supershifted band was observed after the addition of anti-Sp1 antibody, indicating that Sp1 interacted with the DNA sequence containing rs1407019 G-allele (Figure 1C). The formation of C2 complex suggested that another binding protein(s) might interact with this DNA region; however, we could not identify this binding protein(s) through the competition and supershift assays using competitors and antibodies for several transcription factors predicted by TFSEARCH database (data not shown).

Luciferase reporter assay was then performed to examine the allelic difference of rs1407019 in transcriptional activity. The NCI-H1781 cells transfected with the construct containing G-allele of rs1407019 displayed a 1.6-fold greater enhancement of transcriptional activity than those with

Table 4. Relationship Between Additional SNPs of *SPLUNC1* and NPC Susceptibility

Gene	SNP ID	Chromosome position (bp)	Gene position	Alleles (1/2)	Cases[a]				Controls[a]				*P*-value			Odds ratio (95% CI)		LD Information(r^2)[c]
					11[b]	12	22	MAF	11[b]	12	22	MAF	1 vs. 2	vs. 11	vs. 22	vs. 11	vs. 22	
SPLUNC1	rs1321417	31279634	Intergenic	C/T	232	173	27	0.26	318	142	26	0.20	0.0015	0.00035	0.57	1.63 (1.25–2.13)	1.18 (0.68–2.05)	0.933
	rs6059169	31281265	Intergenic	A/C	223	173	36	0.28	313	142	31	0.21	0.00027	0.00010	0.31	1.70 (1.30–2.21)	1.33 (0.81–2.20)	0.994
	rs6059171	31282365	Intergenic	G/A	220	173	36	0.29	310	143	31	0.21	0.00027	0.000098	0.26	1.69 (1.30–2.21)	1.34 (0.81–2.20)	0.981
	rs6057775	31283289	Intergenic	T/G	219	173	36	0.29	311	143	31	0.21	0.00022	0.00010	0.25	1.71 (1.31–2.22)	1.34 (0.82–2.22)	1.000
	rs6057776	31283311	Intergenic	A/G	219	169	35	0.28	309	141	31	0.21	0.00045	0.00015	0.31	1.67 (1.28–2.19)	1.31 (0.79–2.16)	1.000
	SNP01	31283777	Intergenic	C/T	222	174	36	0.28	310	143	31	0.21	0.00033	0.00013	0.31	1.69 (1.29–2.20)	1.33 (0.81–2.19)	1.000
	rs6057777	31283814	Intergenic	A/C	130	216	86	0.45	157	236	92	0.43	0.51	0.48	0.74	1.11 (0.84–1.47)	1.06 (0.77–1.47)	0.351
	rs6059173	31285204	Intergenic	G/A	219	174	37	0.29	310	141	31	0.21	0.00014	0.000054	0.26	1.74 (1.33–2.26)	1.37 (0.83–2.25)	1.000
	rs6057778	31285431	Intergenic	G/C	223	175	34	0.28	312	144	31	0.21	0.00060	0.00017	0.44	1.67 (1.28–2.18)	1.26 (0.76–2.08)	1.000
	rs1998149	31285831	5′-near gene	C/A	223	173	36	0.28	312	144	31	0.21	0.00041	0.00017	0.26	1.67 (1.28–2.18)	1.34 (0.81–2.20)	1.000
	rs6059177	31286081	5′-near gene	A/C	222	175	35	0.28	312	143	31	0.21	0.00033	0.00010	0.37	1.70 (1.30–2.21)	1.29 (0.78–2.14)	1.000
	rs6059178	31286208	5′-near gene	G/C	222	172	35	0.28	311	143	31	0.21	0.00048	0.00017	0.31	1.67 (1.28–2.17)	1.30 (0.79–2.15)	0.988
	rs750065	31286548	5′-near gene	T/A	222	173	35	0.28	311	144	31	0.21	0.00049	0.00017	0.31	1.67 (1.28–2.17)	1.30 (0.79–2.15)	1.000
	rs2752901	31286660	5′-near gene	G/A	223	173	35	0.28	314	142	31	0.21	0.00032	0.000099	0.31	1.69 (1.30–2.21)	1.30 (0.79–2.15)	0.988
	rs2752902	31286685	5′-near gene	T/C	222	170	36	0.28	311	146	29	0.21	0.00031	0.00022	0.16	1.65 (1.27–2.15)	1.45 (0.87–2.40)	0.988
	rs6141897	31288326	Intronic	T/G	221	173	36	0.28	311	144	31	0.21	0.00033	0.00013	0.26	1.68 (1.29–2.19)	1.34 (0.81–2.21)	1.000
	rs3787144	31288430	Intronic	A/G	125	223	83	0.45	157	236	91	0.43	0.42	0.28	0.87	1.18 (0.89–1.56)	1.03 (0.74–1.43)	0.354
	rs3787145	31288513	Intronic	C/T	218	177	35	0.29	309	143	31	0.21	0.00022	0.000056	0.37	1.73 (1.32–2.25)	1.29 (0.78–2.13)	0.994
	rs6059181	31289034	Intronic	A/G	223	173	36	0.28	312	145	30	0.21	0.00033	0.00017	0.25	1.67 (1.28–2.18)	1.38 (0.84–2.29)	0.994
	rs1407019	31289458	Intronic	G/A	217	178	34	0.29	308	144	31	0.21	0.00033	0.000073	0.44	1.72 (1.32–2.24)	1.26 (0.76–2.08)	0.994
	rs6059183	31289688	Intronic	T/C	220	174	36	0.29	312	144	31	0.21	0.00022	0.000099	0.25	1.70 (1.31–2.22)	1.34 (0.82–2.21)	1.000
	rs6141898	31289791	Intronic	G/T	223	175	34	0.28	312	145	30	0.21	0.00049	0.00017	0.36	1.67 (1.28–2.18)	1.30 (0.78–2.16)	0.994
	rs6059184	31290126	Intronic	C/T	221	177	34	0.28	310	143	31	0.21	0.00041	0.00010	0.44	1.70 (1.31–2.22)	1.25 (0.75–2.07)	0.994
	rs6059185	31290393	Intronic	T/C	221	176	35	0.28	311	143	31	0.21	0.00027	0.000077	0.37	1.71 (1.31–2.22)	1.29 (0.78–2.13)	1.000
	rs3746392	31290991	Intronic	G/A	217	180	34	0.29	308	144	31	0.21	0.00028	0.000044	0.44	1.74 (1.33–2.26)	1.25 (0.75–2.07)	0.994
	rs6141899	31291100	Intronic	A/G	233	179	35	0.28	311	145	31	0.21	0.0010	0.00033	0.44	1.62 (1.25–2.11)	1.25 (0.76–2.06)	0.994
	rs6059187	31291926	Intronic	A/G	140	223	84	0.44	160	236	90	0.43	0.71	0.62	0.93	1.08 (0.82–1.42)	1.02 (0.73–1.42)	0.352
	rs2273529	31292744	Intronic	G/A	302	114	12	0.16	299	160	26	0.22	0.0019	0.0051	0.067	1.49 (1.13–1.97)	1.96 (0.98–3.94)	0.075
	rs6059186	31292915	Intronic	T/C	221	171	36	0.28	311	143	31	0.21	0.00039	0.00016	0.25	1.67 (1.28–2.18)	1.34 (0.82–2.22)	1.000
	rs12480799	31294373	Intronic	T/C	218	175	36	0.29	309	145	31	0.21	0.00027	0.00010	0.26	1.70 (1.30–2.21)	1.34 (0.81–2.21)	0.988
	rs10875488	31294434	Intronic	A/G	222	175	34	0.28	309	145	31	0.21	0.00074	0.00022	0.44	1.65 (1.27–2.15)	1.25 (0.76–2.08)	0.988
	rs1047595	31294616	3′-UTR	T/C	221	175	36	0.29	309	145	31	0.21	0.00035	0.00013	0.31	1.68 (1.29–2.18)	1.33 (0.81–2.19)	0.988
	rs2295575	31294769	3′-UTR	C/A	231	181	35	0.28	307	147	33	0.22	0.0022	0.00056	0.61	1.59 (1.23–2.07)	1.17 (0.71–1.92)	0.958
	rs6120186	31295406	3′-near gene	A/G	222	175	34	0.28	311	145	31	0.21	0.00061	0.00017	0.44	1.66 (1.28–2.17)	1.26 (0.76–2.09)	0.994
	rs6141900	31301024	Intergenic	C/A	224	171	37	0.28	311	146	29	0.21	0.00027	0.00022	0.16	1.65 (1.27–2.15)	1.48 (0.89–2.44)	0.969
	rs6059197	31302181	Intergenic	A/G	222	172	36	0.28	310	145	30	0.21	0.00040	0.00017	0.25	1.66 (1.27–2.16)	1.39 (0.84–2.29)	0.981
	rs911137	31302481	Intergenic	G/A	223	172	36	0.28	309	149	29	0.21	0.00050	0.00038	0.20	1.62 (1.24–2.11)	1.44 (0.87–2.39)	0.969
	rs6141901	31302719	Intergenic	G/T	221	173	37	0.29	310	147	30	0.21	0.00028	0.00017	0.16	1.66 (1.28–2.17)	1.43 (0.87–2.36)	0.982
	rs6057786	31304113	Intergenic	A/G	223	172	37	0.28	310	148	29	0.21	0.00028	0.00023	0.16	1.64 (1.26–2.14)	1.48 (0.89–2.45)	0.975
	rs6059214	31312132	Intergenic	A/C	221	185	40	0.30	305	151	31	0.22	0.00011	0.000072	0.14	1.71 (1.31–2.22)	1.45 (0.89–2.36)	0.853

MAF, minor allele frequency; CI, confidence interval.
[a]Cases, $N = 447$; Controls, $N = 487$.
[b]Major allele in controls was defined as allele 1.
[c]LD with rs2752903.

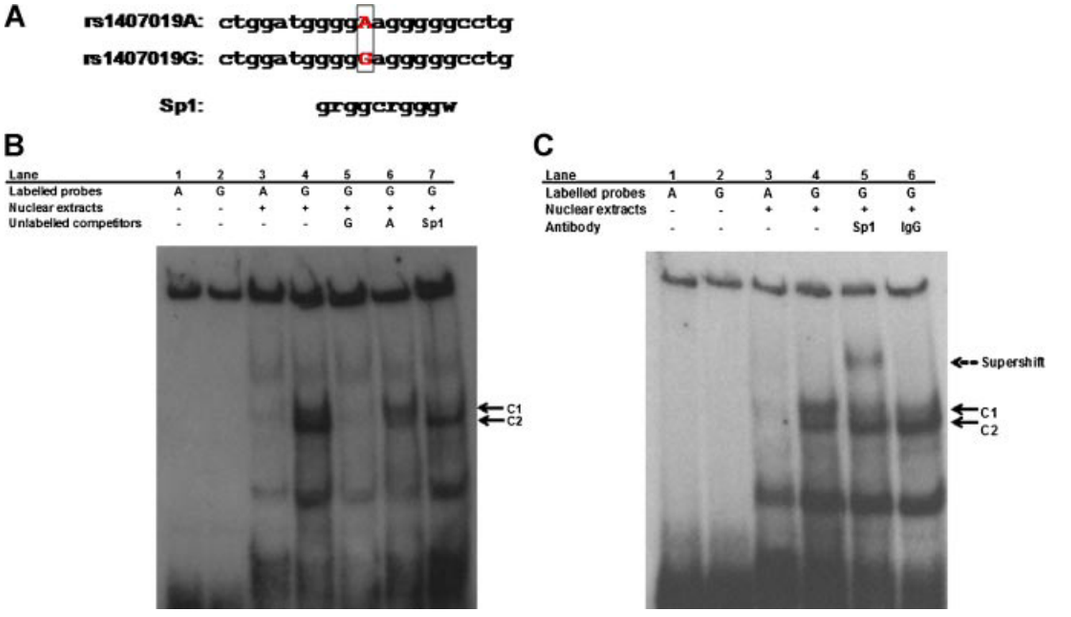


Figure 1. Allelic difference in binding of Sp1 to the SNP site of rs1407019 in intron 3 of *SPLUNC1*. (A) Sequence around the SNP site of rs1407019. (B) EMSA was performed using nuclear extracts from NCI-H1781 and oligonucleotides with A and G alleles. Two complexes (C1 and C2) were formed. Unlabelled probes in 125-fold excess as compared to the labelled probes were used for the competition experiment. The data is the representative result of three experiments. (C) Supershift assay was performed using nuclear extracts from NCI-H1781, oligonucleotides with A and G alleles, and anti-Sp1 antibody. Supershift band was indicated by a broken arrow. The data is the representative result of three experiments.

the A-allele construct (Figure 2). Combining the data from EMSA and luciferase reporter assay, it is suggested that rs1407019 may affect the transcriptional regulation of *SPLUNC1* through transcriptional activation by Sp1.

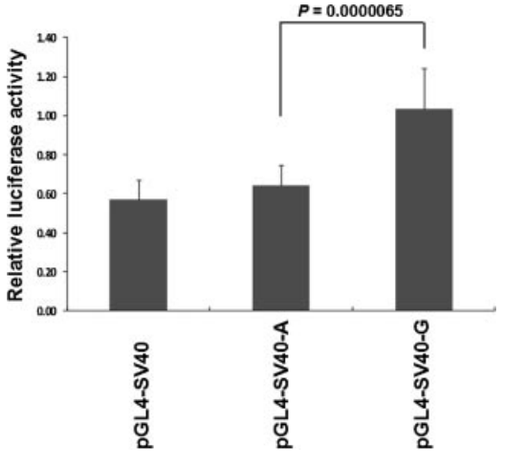


Figure 2. The transcriptional enhancer activity of the 41-bp sequence region around the rs1407019. The luciferase activities in the NCI-H1781 cells transfected with A and G allele constructs were measured. The values of relative luciferase activity were expressed as mean ± SD from three independent experiments performed in quadruplicate. The difference in the luciferase activities between the A and G alleles of rs1407019 were evaluated by a Student's *t*-test.

DISCUSSION

We investigated the association of *LPLUNC1* and *SPLUNC1* genes with the risk of NPC in Malaysian Chinese by performing case-control association study. We found that SNPs of *SPLUNC1*, but not *LPLUNC1*, exhibited significant association with NPC. Flach et al. [24] reported that *LPLUNC1* is highly expressed in the duodenal mucosa of patients during the acute phase of cholera, suggesting that *LPLUNC1* is possibly responsible for the innate defence in combating the infection. This has been further supported by a family-based association study in Bangladesh population, which showed an association of a variant in the promoter region of *LPLUNC1* with cholera infection [25]. In the present study, since we only examined four tag-SNPs of *LPLUNC1*, it is postulated that there might be other variants of this gene which are associated with NPC susceptibility.

In the case of *SPLUNC1*, a previous study reported that two promoter SNPs (C-2128T and C-1888T, corresponding to rs2752903, and rs750064, respectively) showed significant association with NPC in a total of 525 Cantonese-speaking Chinese in China [12]. However, in our study, rs2752903 but not rs750064 is associated with NPC susceptibility in Malaysian Chinese. The MAF of rs750064 in our population (0.405) and the population reported by He et al. (0.360) were comparable,

which was similar to that of HapMap CHB (0.381); however, the frequency of CC genotype in the population reported by He et al. (0.071) was much lower than ours (0.163) and HapMap CHB (0.155). Their genotype frequency in controls was significantly deviated from HWE ($P = 0.000019$ by χ^2-test) while no deviation from HWE was detected in our controls. In the present study, LD between these two SNPs in Malaysian Chinese was found to be lower ($r^2 = 0.28$), which was comparable to HapMap CHB ($r^2 = 0.42$), than that in Cantonese-speaking Chinese ($r^2 = 0.73$). The discrepancy between two studies may reflect certain differences in genetic composition of both populations.

Through the screening of a functional causative SNP(s), we found that an intronic SNP, rs1407019, is likely to influence the expression of *SPLUNC1* by differential binding affinity of Sp1, which mainly acts as a transcription activator [26–28]. *SPLUNC1* expression in the carriers of susceptible allele A, is expected to be lower than non-susceptible allele G due to lower binding affinity of Sp1. SPLUNC1 is proposed to be involved in the innate immune response against EBV and other viruses [19]. In addition, it is reported that Sp1-mediated transcription can be activated by Rta of EBV, and that Sp1 was up-regulated in NPC cells compared to normal nasopharyngeal cells [29,30]. These results suggested that allelic difference in binding of Sp1 at the SNP site of rs1407019 may cause individual differences in the defence against EBV. Individuals with the susceptible allele A may confer an increased risk of NPC due to insufficient amount of SPLUNC1 in response to EBV infection.

In summary, our study shows that genetic polymorphisms of *SPLUNC1* confer risk of NPC in a Malaysian Chinese population. Functional analysis suggests that rs1407019 may affect *SPLUNC1* expression by altering the binding site for Sp1. Further studies are still needed to clarify the detailed biological mechanisms; however, our data provides valuable insights into the mechanism that cause susceptibility to NPC.

ACKNOWLEDGMENTS

The authors would like to thank the Director of General of Health of Malaysia for his permission to publish this article, and the Director of the IMR for her support. The authors thank Tai Mei Chee, Yuka Kikuchi, and Mitsunori Yahata for their constructive suggestions and technical assistance. The authors also thank all the voluntary participants in this project; clinicians and staff at the University of Malaya Medical Centre, NCI Cancer Hospital, Tung Shin Hospital and the Malaysian NPC study group for their help in procurement of the samples; and all the members in Laboratory for Pharmacogenetics, RIKEN Center for Genomic Medicine for their contributions in completion of this study. This study was supported by the Ministry of Health Malaysia (JPP-IMR No. 06-062) and University of Malaya (PS160/2008B and RG042/11BIO).

REFERENCES

1. Curado MP, Edwards B, Shin HR, et al. editors. Cancer incidence in five continents Vol. IX. Lyon, France: IARC Scientific Publications No. 160; 2007.
2. Spano JP, Busson P, Atlan D, et al. Nasopharyngeal carcinomas: An update. Eur J Cancer 2003;39:2121–2135.
3. Omar ZA, Mohd Ali Z, Tamin NSI, editors. Malaysian cancer statistics—Data and figure Peninsular Malaysia 2006. Malaysia: National Cancer Registry, Ministry of Health Malaysia; 2006.
4. Armstrong RW, Imrey PB, Lye MS, Armstrong MJ, Yu MC, Sani S. Nasopharyngeal carcinoma in Malaysian Chinese: Occupational exposures to particles formaldehyde and heat. Int J Epidemiol 2000;29:991–998.
5. Ward MH, Pan WH, Cheng YJ, et al. Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. Int J Cancer 2000;86:603–609.
6. 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;77:228–235.
7. Zhou XF, Cui J, Macias V, et al. The progress on genetic analysis of nasopharyngeal carcinoma. Comp Funct Genomics 2007; 57513.
8. Hildesheim A, Levine PH. Etiology of nasopharyngeal carcinoma: A review. Epidemiol Rev 1993;15:466–485.
9. Hildesheim A, Chen CJ, Caporaso NE, et al. Cytochrome P4502E1 genetic polymorphisms and risk of nasopharyngeal carcinoma: Results from a case–control study conducted in Taiwan. Cancer Epidemiol Biomarkers Prev 1995;4:607–610.
10. Hildesheim A, Anderson LM, Chen CJ, et al. CYP2E1 genetic polymorphisms and risk of nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst 1997;89:1207–1212.
11. Kongruttanachok N, Sukdikul S, Setavarin S, et al. Cytochrome P450 2E1 polymorphisms and nasopharyngeal carcinoma development in Thailand: A correlative study. BMC Cancer 2001;1:4.
12. He Y, Zhou G, Zhai Y, et al. Association of PLUNC gene polymorphisms with susceptibility to nasopharyngeal carcinoma in a Chinese population. J Med Genet 2005; 42:172–176.
13. Tse KP, Su WH, Chang KP, et al. Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. Am J Hum Genet 2009;85:194–203.
14. Bei JX, Li Y, Jia WH, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat Genet 2010;42:599–603.
15. Ng CC, Yew PY, Puah SM, et al. A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. J Hum Genet 2009;54:392–397.
16. Bingle CD, Bingle L. Characterization of the human plunc gene, a gene product with an upper airways and nasopharyngeal restricted expression pattern. Biochim Biophys Acta 2000;1493:363–367.
17. Bingle CD, Craven CJ. PLUNC: A novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. Hum Mol Genet 2002;11:937–943.
18. He ZW, Xie L, Xu LG, et al. Cloning of a novel gene associated with human nasopharyngeal carcinoma. Chin Sci Bull 2000;45:2267–2272.
19. Bingle CD, Craven CJ. Comparative analysis of the PLUNC (palate, lung and nasal epithelium clone) protein families. Biochem Soc Trans 2003;31:806–809.

20. Zhang B, Nie X, Xiao B, et al. Identification of tissue-specific genes in nasopharyngeal epithelial tissue and differentially expressed genes in nasopharyngeal carcinoma by suppression subtractive hybridization and cDNA microarray. Genes Chromosomes Cancer 2003;38: 80–90.
21. Yang YX, Yang YB, Li XL, et al. The study of LPLUNC1 gene inhibit human nasopharyngeal carcinoma cell line HNE1 growth and proliferation. Prog Biochem Biophhys 2007;34:366–374.
22. Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y. A high-throughput SNP typing system for genome-wide association studies. J Hum Genet 2001;46: 471–477.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
24. Flach CF, Qadri F, Bhuiyan TR, et al. Broad up-regulation of innate defence factors during acute cholera. Infect Immun 2007;75:2343–2350.
25. Larocque RC, Sabeti P, Duggal P, et al. A variant in long palate, lung and nasal epithelium clone 1 is associated with cholera in a Bangladeshi population. Genes Immun 2009;10:267–272.
26. Suske G. The Sp-family of transcription factors. Gene 1999;238:291–300.
27. Kim H, Lee S, Park B, Che L, Lee S. Sp1 and Sp3 mediate basal and serum-induced expression of human CENP-W. Mol Biol Rep 2010;37:3593–3600.
28. Ihn H, Trojanowska M. Sp3 is a transcriptional activator of the human α2(I) collagen gene. Nucleic Acids Res 1997; 25:3712–3717.
29. Chang L, Chung J, Hong Y, Ichimura T, Nakao M, Liu S. Activation of Sp1-mediated transcription by Rta of Epstein–Barr virus via an interaction with MCAF1. Nucleic Acids Res 2005;33:6528–6539.
30. Su B, Xiang B, Wang L, et al. Profiling and comparing transcription factors activated in non-metastatic and metastatic nasopharyngeal carcinoma cells. J Cell Biochem 2010;109:173–183.