RESEARCH ARTICLE

Identification of Genetic and Non-genetic Risk Factors for Nasopharyngeal Carcinoma in a Southeast Asian Population

Nikman Adli Nor Hashim[1], Nurul Hanis Ramzi[1&], Sharmila Velapasamy[1&], Livy Alex[1&], Jagdish Kaur Chahil[1&], Say Hean Lye[1&], Khamsigan Munretnam[1&], Mohd Roslan Haron[2], Lian Wee Ler[1*]

Abstract

Background: Nasopharyngeal carcinoma (NPC) is endemic in Southern Chinese and Southeast Asian populations. Geographical and ethnic clustering of the cancer is due to genetic, environmental, and lifestyle risk factors. This case-control study aimed to identify or confirm both genetic and non-genetic risk factors for NPC in one of the endemic countries, Malaysia. Materials and Methods: A panel of 768 single-nucleotide polymorphisms (SNPs) previously associated with various cancers and known non-genetic risk factors for NPC were selected and analyzed for their associations with NPC in a case-control study. Results: Statistical analysis identified 40 SNPs associated with NPC risk in our population, including 5 documented previously by genome-wide association studies (GWAS) and other case-control studies; the associations of the remaining 35 SNPs with NPC were novel. In addition, consistent with previous studies, exposure to occupational hazards, overconsumption of salt-cured foods, red meat, as well as low intake of fruits and vegetables were also associated with NPC risk. Conclusions: In short, this study confirmed and/or identified genetic, environmental and dietary risk factors associated with NPC susceptibility in a Southeast Asian population.

Keywords: SNP - nasopharyngeal carcinoma - association study - cancer risk - Southeast Asia

Asian Pacific J Cancer Prev, **13 (12)**, 6005-6010

Introduction

Nasopharyngeal carcinoma (NPC) constitutes 75-95% of the cancer cases of nasopharynx in low-risk populations and almost all those in high-risk populations (Whelan and Ferlay 1992). Globally, nasopharyngeal carcinoma is considered a relatively rare disease, having an age-standardized incidence rate in both sexes of less than 1 in 100,000 persons per year. This accounts for merely ~0.7% of the cancer burden across the globe (Jemal et al., 2011). However, there is a clustering of NPC in Southern Chinese and Southeast Asian populations (Clifford 1970; Vokes et al., 1997; Yu and Yuan, 2002; Yoshizaki et al., 2012). Located in Southeast Asia, Malaysia has one of the highest incidence rates of NPC in the world, together with the other two Southeast Asian countries Indonesia and Singapore (Whelan and Ferlay, 1992). In Malaysia, NPC is the fifth most common cancer nationwide (4.5% of all cancer cases) and has an age-standardized incidence of 8.5 and 2.6 per 100,000 males and females respectively (Zainal, 2006). NPC is the most prevalent cancer among young male adults (aged 15-49) and co-dominant with colorectal, lung, prostate and liver cancers in the older age group (≥50 years old).

Findings from epidemiological studies suggest that genetic predispositions, environmental risk factors and Epstein-Barr virus (EBV) infection may have important roles in the development of NPC (Zheng et al., 1994; Zhang et al., 2004; Yoshizaki et al., 2012). However, the near ubiquity of EBV infection and other environmental risk factors cannot fully explain the geographical and ethnic clustering of NPC incidences, suggesting a strong genetic link to NPC carcinogenesis (Serraino et al., 2005). The studies of NPC genetic predisposition, such as genome-wide association study (GWAS), have reported some early successes (Ng et al., 2009; Tse et al., 2009; Bei et al., 2010).

All epidemiological studies on the associations of genetic and/or non-genetic factors with NPC have provided insights into the etiology of NPC, and will be beneficial to the construction of a cancer risk prediction model. Risk prediction has a huge potential in bringing benefits to the public, as individuals at high risk to developing cancer may be provided with strategies for intervention and prevention (Spitz et al., 2007). Many statistical models, built on both genetic and non-genetic

*[1]Molecular Research and Services Laboratory, INFOVALLEY® Life Sciences Sdn. Mines Resort City, Selangor, [2]Department of Radiotherapy and Oncology, Hospital Sultan Ismail, Johor Bahru, Malaysia [&]Equal contributors *For correspondence: drlivy@infovalley.net.my*

risk factors, had been developed previously to assess and manage the risks of cancers, particularly breast, colorectal, ovarian and prostate cancers (Taplin et al., 1990; Hartge et al., 1994; Eastham et al., 1999; Rockhill et al., 2001; Selvachandran et al., 2002; Imperiale et al., 2003; Tice et al., 2005). However, such a tool developed for any head and neck cancers, including NPC, remains scarce (Jiang and Liu, 2009; Bosch et al., 2011; Jin et al., 2012). This may be due to the fact that globally NPC is an uncommon form of cancer compared to other cancer types, in spite of its geographical and ethnic clustering. In addition, the genetic markers associated with NPC have not yet been fully explored and catalogued.

In the present study, a case-control association study was conducted with the specific aims of identifying and confirming both genetic and non-genetic risk factors associated with NPC in a Southeast Asian population in Malaysia.

Materials and Methods

Study population

NPC and control subjects were recruited from four different hospitals in Malaysia: Hospital Putrajaya (Putrajaya), Hospital Sultan Ibrahim (Johor Bharu), Hospital Kuala Lumpur (Kuala Lumpur), and Beacon International Specialist Centre (Selangor). For controls, only those who have never been diagnosed with any cancer and those without a family history of NPC were included. Demographic data and information about the known clinical risk factors for NPC were collected by face-to-face interview.

The risk factors included personal and family history of cancers, exposure to occupational hazards (daily or near daily exposure to ionizing radiation, heavy metals, fume, wood dust, and volatile chemicals), cigarette smoking, alcohol drinking, and dietary intakes of fruits and vegetables, red meat, and salt-cured food (Chang and Adami, 2006). In accordance with Malaysian dietary habit, foodstuff was quantified as the average proportion of total daily food intake over the past 10 years, instead of the commonly used portion sizes. Salt-cured food was measured as the average number of meals per month that included salt-cured food over the past 10 years. A regular consumer of salt-cured food was defined as those who included the foodstuff at least once a month.

Informed consent was obtained from all the subjects and the study protocol was conformed to national ethics guidelines and was approved by Medical Review & Ethics Committee (MREC), under the Ministry of Health, Malaysia (Registration ID: NMRR-10-652-6473).

DNA isolation and processing

Blood samples (1-5 ml) were collected at the time of subject recruitment. Genomic DNA was extracted using QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA, USA). The quantity and purity of the isolated DNA were determined using a Smart SpecTM plus spectrophotometer (Bio-rad, Inc), and the integrity of the DNA was assessed by 0.8% agarose gel electrophoresis followed by visual inspection.

SNP selection and genotyping

A literature search was used to identify a total of 768 candidate SNPs previously reported to associate with various cancer types in East, Southeast and South Asian populations. The cancer types were breast, cervical, colorectal, gastric, liver, lung, NPC, oral, ovarian, prostate and thyroid cancers, as well as leukemia. Genotyping was conducted using Illumina Golden Gate Genotyping Assay Platform (Illumina Inc., San Diego, USA) according to the manufacturer's protocols and recommendations. Custom genotyping probes were designed and submitted to Illumina Inc. for quality control and assessment using Assay Design Tool (ADT) (Illumina Inc., San Diego, USA). All 768 SNPs achieved designability scores of 0.5 or 1.0. Beadchips were scanned using Bead Array Reader System (Illumina Inc., San Diego, USA). Raw data generated from the scan were deciphered and quality-checked using Genome Studio software (version 2011.1, Illumina Inc., San Diego, USA).

Statistical analysis

The average call rate per sample was >99% and the average call rate per SNP was 93%. SNPs with call rate <93% were excluded from further analysis. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using STATA version 10 (StataCorp, Texas). Statistical significance was defined as p value <0.05. All SNPs were tested for deviation from Hardy-Weinberg equilibrium (HWE) using the chi-square (χ^2) test; and SNPs deviated from HWE (PHWE<0.05) were excluded. The association of SNPs with NPC susceptibility was analyzed further according to a best-fitted dominant and recessive genetic models as previously described (Lerman, 1996). JLIN (Java Linkage disequilibrium plotter) software version 1.6.0 was used for the analysis of linkage disequilibrium (LD) (Carter et al., 2006).

Results

The 96 NPC cases and controls were matched according to the age, ethnicity and gender (Table 1). None of these controls had a family history of NPC or personal history of any cancer. The mean age of the controls was 49.1 years (±10.6), and the mean age of the cancer subjects was 45.8 years (±12.7).

A total of 40 SNPs were identified to be significantly associated with NPC risk, attaining statistical significance

Table 1. Demographic Characteristics of Cancer Cases and Controls

		Case		Control	
		No.	%	No.	%
Ethnicity:	Chinese	24	50.0	26	54.2
	Malay	24	50.0	22	45.8
Gender:	Male	39	81.3	39	81.3
	Female	9	18.7	9	18.7
Average Age (±SD)*		45.8 (±12.7)		49.1 (±10.6)	p=0.168
Family History of Cancers**:					
	Yes (NPC)	7	14.6	0	0
	Yes (Other Cancers)	15	31.3	3	6.3
	No	26	54.2	45	93.7

*SD, standard deviation; **includes first and second degree relatives

Table 2. SNPs Associated with NPC Susceptibility Best-Fitted

	SNP Name	Gene	Genotype	Case		Control		P value	OR	95% CI		HWE	Effect
				No.	%	No.	%,						
Dominant Genetic Model:	rs1346044	WRN	AG+GG	13	27.1	5	10.4	0.036	3.19	1.04	9.83	0.70	Risk
			AA	35	72.9	43	89.6	Reference					
	rs16896923	HCG9	AG+GG	12	25.9	23	47.9	0.02	0.36	0.15	0.86	0.89	Protective
			AA	36	75	25	52.1	Reference					
	rs2236722	CYP19A1	AG+GG	2	4.2	8	16.7	0.045	0.22	0.04	1.08	0.53	Protective
			AA	46	95.8	40	83.3	Reference					
	rs243842	MMP2	AG+GG	28	58.3	18	37.5	0.041	2.33	1.03	5.29	0.23	Risk
			AA	20	41.7	30	62.5	Reference					
	rs3842759	**INS-IGF2**	**AT+TT**	**24**	**54.6**	**12**	**26.7**	**0.007**	**3.30**	**1.36**	**8.02**	**0.30**	**Risk**
			AA	**20**	**45.4**	**33**	**73.3**	**Reference**					
	rs6869366	TMEM167A	AC+CC	11	22.9	4	8.3	0.049	3.27	0.96	11.1	0.76	Risk
			AA	37	77.1	44	91.7	Reference					
	rs872071	IRF4	AG+GG	32	66.7	22	45.8	0.04	2.36	1.04	5.40	0.07	Risk
			AA	16	33.3	26	54.2	Reference					
	rs1137100	LEPR	AG+AA	12	25	22	45.8	0.033	0.39	0.17	0.94	0.44	Protective
			GG	36	75	26	54.2	Reference					
	rs1867277	**FOXE1**	**AG+AA**	**4**	**9.5**	**17**	**35.4**	**0.004**	**0.19**	**0.06**	**0.63**	**0.52**	**Protective**
			GG	**38**	**90.5**	**31**	**64.6**	**Reference**					
	rs2236225	MTHFD1	AG+AA	1	2.1	6	12.5	0.05	0.15	0.02	1.29	0.64	Protective
			GG	47	97.9	42	87.5	Reference					
	rs243844	MMP2	AG+AA	28	58.3	18	37.5	0.041	2.33	1.03	5.29	0.23	Risk
			GG	20	41.7	30	62.5	Reference					
	rs243845	MMP2	AG+AA	28	58.3	18	37.5	0.041	2.33	1.03	5.29	0.23	Risk
			GG	20	41.7	30	62.5	Reference					
	rs3025039	VEGFA	AG+AA	20	41.7	10	20.8	0.028	2.71	1.10	6.69	0.60	Risk
			GG	28	58.3	38	79.2	Reference					
	rs723147	4p13	AG+AA	19	39.6	8	16.7	0.013	3.28	1.26	8.51	0.53	Risk
			GG	29	60.4	40	83.3	Reference					
	rs724165	ADCY4	AG+AA	17	35.4	8	16.7	0.036	2.74	1.05	7.18	0.33	Risk
			GG	31	64.6	40	83.3	Reference					
Recessive Genetic Model:	rs243839	MMP2	GG	2	4.2	10	20.8	0.014	0.17	0.03	0.80	0.83	Protective
			AA+AG	46	95.8	38	79.2	Reference					
	rs2517713	**HLA-A**	**CC**	**1**	**2.1**	**11**	**22.9**	**0.002**	**0.07**	**0.01**	**0.58**	**0.11**	**Protective**
			AA+AC	**47**	**97.9**	**37**	**77.1**	**Reference**					
	rs29232	**GABBR1**	**GG**	**2**	**4.2**	**16**	**33.3**	**<0.001**	**0.09**	**0.02**	**0.41**	**0.42**	**Protective**
			AA+AG	**46**	**95.8**	**32**	**66.7**	**Reference**					
	rs304270	**RAD51C**	**GG**	**20**	**41.7**	**8**	**16.7**	**0.007**	**3.57**	**1.38**	**9.25**	**0.49**	**Risk**
			AA+AG	**28**	**58.3**	**40**	**83.3**	**Reference**					
	rs324013	STAT6	GG	20	41.7	11	22.9	0.049	2.40	0.99	5.82	0.79	Risk
			AA+AG	28	58.3	37	77.1	Reference					
	rs3792796	**GPX3**	**GG**	**29**	**64.4**	**13**	**28.3**	**<0.001**	**4.60**	**1.90**	**11.2**	**0.73**	**Risk**
			CC+CG	**16**	**35.6**	**33**	**71.7**	**Reference**					
	rs532841	DLC1	GG	23	48.9	14	29.2	0.048	2.33	1.00	5.42	0.08	Risk
			AA+AG	24	51.1	34	70.8	Reference					
	rs560191	TP53BP1	GG	23	47.9	13	27.1	0.035	2.48	1.06	5.81	0.53	Risk
			CC+CG	25	52.1	35	72.9	Reference					
	rs569143	MRE11A	GG	20	41.7	10	20.8	0.028	2.71	1.10	6.69	0.63	Risk
			CC+CG	28	58.3	38	79.2	Reference					
	rs6496724	BLM	CC	10	20.8	3	6.2	0.037	3.95	1.01	15.4	0.19	Risk
			AA+AC	38	79.2	45	93.8	Reference					
	rs719293	NRXN1	TT	7	14.6	18	37.5	0.011	0.29	0.11	0.77	0.19	Protective
			AA+AT	41	85.4	30	62.5	Reference					
	rs865094	**MMP2**	**GG**	**1**	**2.1**	**9**	**18.8**	**0.008**	**0.09**	**0.01**	**0.76**	**0.52**	**Protective**
			AA+AG	**47**	**97.9**	**39**	**81.2**	**Reference**					
	rs11132383	**KLKB1**	**AA**	**5**	**10.4**	**16**	**33.3**	**0.007**	**0.23**	**0.08**	**0.70**	**0.27**	**Protective**
			GG+AG	**43**	**89.6**	**32**	**66.7**	**Reference**					
	rs11225395	MMP8	AA	1	2.1	8	16.7	0.014	0.11	0.01	0.89	0.77	Protective
			GG+AG	47	97.9	40	83.3	Reference					
	rs1799796	**KLC1**	**AA**	**24**	**51.1**	**10**	**20.8**	**0.002**	**3.97**	**1.61**	**9.77**	**0.56**	**Risk**
			GG+AG	**23**	**48.9**	**38**	**79.2**	**Reference**					
	rs1800975	XPA	AA	6	13.3	17	35.4	0.014	0.28	0.10	0.80	0.17	Protective
			GG+AG	39	86.7	31	64.6	Reference					
	rs2046210	ESR1	AA	1	2.1	8	16.7	0.014	0.11	0.01	0.89	0.21	Protective
			GG+AG	47	97.9	40	83.3	Reference					
	rs2070593	GPX3	AA	11	22.9	21	43.8	0.03	0.38	0.16	0.92	0.51	Protective
			GG+AG	37	77.1	27	56.2	Reference					
	rs2602141	TP53BP1	AA	23	47.9	12	25	0.02	2.76	1.16	6.55	0.37	Risk
			CC+AC	25	52.1	36	75	Reference					
	rs2736100	TERT	AA	9	19.1	20	41.7	0.017	0.33	0.13	0.84	0.44	Protective
			CC+AC	38	80.9	28	58.3	Reference					
	rs3129055	HLA-F	AA	10	20.8	21	43.8	0.016	0.34	0.14	0.83	0.31	Protective
			GG+AG	38	79.2	27	56.2	Reference					
	rs36686	**B3GNT3**	**AA**	**19**	**39.6**	**7**	**14.6**	**0.006**	**3.84**	**1.43**	**10.3**	**0.30**	**Risk**
			GG+AG	**29**	**60.4**	**41**	**85.4**	**Reference**					
	rs5009448	**HCG9**	**AA**	**1**	**2.1**	**9**	**18.8**	**0.008**	**0.09**	**0.01**	**0.76**	**0.17**	**Protective**
			GG+AG	**47**	**97.9**	**39**	**81.2**	**Reference**					
	rs727479	CYP19A1	AA	26	54.2	15	31.2	0.023	2.60	1.13	5.98	0.66	Risk
			CC+AC	22	45.8	33	68.8	Reference					
	rs8036601	BLM	AA	7	14.6	1	2.1	0.027	8.02	0.95	68.0	0.07	Risk
			GG+AG	**41**	**85.4**	**47**	**97.9**	**Reference**					

*SNPs previously reported by GWAS are underscored; SNPs with p value < 0.01 are in boldface

Table 3. Clinical Risk Factors for NPC

Risk Factor		Case	Control	OR (95% CI)	P-value
Occupational hazards	Yes	18	1	28.3	<0.0001
	No	28	44	(3.91 -1204)	
Cigarette smoking	Yes	20	17	1.22	0.6403
	No	28	29	(0.49-3.04)	
Alcohol drinking	Yes	10	10	0.97	0.9567
	No	37	36	(0.32-2.96)	
Fruits & Vegetables*	<50%	24	3	14.3	<0.0001
	≥50%	24	43	(3.67-79.6)	
Red meat*	≥50%	23	6	6.13	0.0003
	<50%	25	40	(2.02-20.7)	
Salt-cured food	≥1	41	29	3.43	0.0129
(meal/month)	<1	7	17	(1.15-11.0)	

*Measured as a percentage of daily food consumption

at p value <0.05. All of the 40 SNPs did not deviate from HWE. Genetic and functional information of these SNPs were detailed in Supplemental Table 1. Of the 40 SNPs, 15 were best-fitted with a dominant genetic model (Table 2), while 25 were best-fitted with a recessive genetic model (Table 3). Furthermore, the OR of these 40 SNPs ranged from 0.07-8.02 (Table 2), with 22 SNPs conferring risk (OR>1) and the remaining 18 having a protective effect (OR<1). Of the 40 SNPs, 11 attained a higher statistical significance of p value<0.01 level (Table 2 boldface). Next, we analyzed LD among the 40 SNPs identified. Three SNPs located in MMP2 gene in chromosome 16 (rs243842, rs243844 and rs243845) were found to be in perfect LD (r^2=1.000), while 2 SNPs in TP53BP1 gene in chromosome 15 (rs2602141 and rs560191, data not shown) were in high LD (r^2=0.979, data not shown). All the remaining SNP pairs did not show significant linkage (r^2<0.8, data not shown).

Several known non-genetic risk factors for cancers were analyzed for their associations with NPC risk. Our data demonstrated that four risk factors were significantly associated with NPC, including exposure to occupational hazards (OR=28.3, 95%CI=3.91-1204, p value<0.0001), low dietary intake of fruits and vegetables (OR=14.3, 95%CI=3.67-79.6, p value<0.0001), high red meat diet (OR=6.13, 95%CI=2.02-20.7, p value=0.0003), and regular consumption of salt-cured food (OR=3.43, 95%CI =1.15-11.0, p value=0.0129) (Table 3). However, the associations of cigarette smoking and alcohol consumption with NPC did not reach statistical significance (Table 3).

Discussion

As compared with other major cancer types, the study of genetic predisposition to NPC is lacking, even in the disease endemic areas. In the present study, we demonstrated the association of 40 SNPs with NPC in a Southeast Asian population. A search of the catalog of published GWAS revealed that out of the 11 documented NPC-associated SNPs (rs1412829, rs1572072, rs189897, rs2517713, rs28421666, rs2860580, rs2894207, rs29232, rs3129055, rs6774494, and rs9510787), 3 SNPs (rs29232, rs2517713, and rs3129055) were statistically significant in the present study (Supplemental Table S1) (Ng et al., 2009; Tse et al., 2009; Bei et al., 2010; Hindorff et al., 2011). Another SNP rs189897 was in our panel of screened SNP, but the associated did not reach statistical significance (data not shown). The remaining 7 SNPs from GWAS database were, however, not included in our original 768-SNP panel and thus their associations cannot be ascertained. In addition to the 3 SNPs identified by GWAS, the associations of rs16896923 and rs5009448 with NPC have also been reported previously in a Han Chinese population (Tse et al., 2009). Interestingly, all these 5 previously reported, NPC-associated SNPs are located in chromosomal region 6p21.3, and the associations of rs2517713, rs29232 and rs5009448 with NPC reached a higher statistical significance of p value<0.01 (Table 1). The mapping of chromosome 6p21.3 had been previously achieved by positional cloning approach (Lu et al., 2003), and the associations of a number of variants located within 6p21.3 with NPC risk had been described in a Taiwanese population (Lu et al., 2005).

Besides these 5 SNPs, to our knowledge, this study is the first to document the associations of the remaining 35 SNPs with NPC susceptibility. Among them, the associations of 8 SNPs (rs3842759, rs1867277, rs304270, rs3792796, rs865094, rs11132383, rs1799796, and rs36686) reached a higher significance level of p value<0.01 (Table 2). Furthermore, three SNPs had been described to be associated with breast cancer (rs2046210), lung cancer (rs2736100), and chronic lymphocytic leukemia (rs872071) in three separate GWAS (Di Bernardo et al., 2008; Landi et al., 2009; Zheng et al., 2009; Hindorff et al., 2011).

The known environmental and lifestyle risk factors for NPC include Epstein-Barr virus infection, occupational exposure to wood dust, consumption of salt-cured foods, cigarette smoking and alcohol drinking; and a diet high in fruits and vegetables may reduce the risk of NPC (Armstrong et al., 1983; Yu et al. 1986; 1988; Armstrong et al., 2000; Yuan et al., 2000; Chang and Adami 2006; Ekburanawat et al., 2010). Our data were in agreement with these previous reports, with the exception of cigarette smoking and alcohol drinking, both of which did not attain statistical significance (Table 3). However, in addition to wood dust, the occupational hazards registered in this study also included ionizing radiation, heavy metals, fume, and volatile chemicals. Due to the small sample size, the contribution of each individual type of hazard cannot be determined. It will be interesting to investigate the contribution of individual occupational hazard using a larger sample size in the future.

The major limitation of the present study is its sample size. A moderate-sized study usually loses its power as allele frequency and effect size decrease. Nevertheless, selecting disease cases and controls based on a family history of the disease can considerably increase the power of the case–control association study, by which the inclusion of cases with affected relatives decreases the required sample size and thus the cost of such studies (Peng et al., 2010). Among the NPC subjects in this study, 45.9% reported having a family history of cancers, including 14.6% having a family history of NPC. In contrast, among the control subjects, only 6.3% have a family history of any cancer, and none of which was that of NPC (Table 1). Hence, the negative impact of the small sample size was minimized in this study as case and

control subjects were selected according to their family history of cancers. This was exemplified by the fact that known environmental and dietary risk factors for NPC, as well as 3 out of 4 previously reported GWAS SNPs in the panel were successfully identified in this study.

Small study sample size poses a challenge to performing racial stratification analysis. The study subjects of this study consisted of Malaysians of self-identified ethnic Malay and Chinese (Table 1). However, ethnic self-identification does not necessarily translate into racial and genetic differences at the population level. A previous genetic study has demonstrated high similarity between the two ethnic groups and to other East Asian populations (Teo et al., 2009).

In NPC endemic areas, such as Malaysia, a risk prediction method will prove to be useful for the identification of high risk individuals and risk management. Such a tool can only be realized if both the genetic and non-genetic risk factors for NPC are identified and studied. It is hoped that the list of risk factors identified here will not only lead to the better understanding of NPC, but hopefully will also gear research towards the development of better NPC risk prediction and management approaches.

Acknowledgements

We are grateful to all of the study subjects for their generous contributions to this research project. We would like to extend our gratitude to all of the medical doctors, nurses, and other staff at the participating medical centers for their assistance in the recruitment of study subjects. This work was fully funded by INFOVALLEY® Life Sciences Sdn. Bhd. We declare that this study was in compliant with Malaysia's national ethics guidelines and was approved by Medical Review & Ethics Committee (MREC), under the Ministry of Health, Malaysia (Registration ID: NMRR-10-652-6473). The authors declare no conflict of interest.

References

Armstrong RW, Armstrong MJ, Yu MC, Henderson BE (1983). Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese. *Cancer Res*, 43, 2967-70

Armstrong RW, Imrey PB, Lye MS, et al (2000). Nasopharyngeal carcinoma in Malaysian Chinese: occupational exposures to particles, formaldehyde and heat. Int J Epidemiol, 29, 991-8

Bei JX, Li Y, Jia WH, et al (2010). A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat Genet, 42, 599-603.

Bosch DJ, Pultrum BB, de Bock GH, et al (2011). Comparison of different risk-adjustment models in assessing short-term surgical outcome after transthoracic esophagectomy in patients with esophageal cancer. Am J Surg, **202**, 303-9.

Carter KW, McCaskie PA, Palmer LJ (2006). JLIN: a java based linkage disequilibrium plotter. *BMC Bioinformatics*, **7**, 60.

Chang ET, Adami H-O (2006). The Enigmatic Epidemiology of Nasopharyngeal Carcinoma. *Cancer Epidemiology Biomarkers and Prev*, **15**, 1765-77.

Clifford P (1970). A review on the epidemiology of nasopharyngeal carcinoma. *J Int Du Cancer*, **5**, 287-309.

Di Bernardo MC, Crowther-Swanepoel D, Broderick P, et al (2008). A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet*, **40**, 1204-10.

Eastham JA, May R, Robertson JL, Sartor O, Kattan MW (1999). Development of a nomogram that predicts the probability of a positive prostate biopsy in men with an abnormal digital rectal examination and a prostate-specific antigen between 0 and 4 ng/mL. *Urology*, **54**, 709-13.

Ekburanawat W, Ekpanyaskul C, Brennan P, et al (2010). Evaluation of non-viral risk factors for nasopharyngeal carcinoma in Thailand: results from a case-control study. *Asian Pac J Cancer Prev*, **11**, 929-32.

Hartge P, Whittemore AS, Itnyre J, McGowan L, Cramer D (1994). Rates and risks of ovarian cancer in subgroups of white women in the United States. The collaborative ovarian cancer group. *Obstetrics Gynecol*, **84**, 760-4.

Hindorff LA, MacArthur J, Wise A, et al (2011). A Catalog of Published Genome-Wide Association Studies.

Imperiale TF, Wagner DR, Lin CY, et al (2003). Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. *Ann Internal Med*, **139**, 959-65.

Jemal A, Bray F, Center MM, et al (2011). Global cancer statistics. *CA: A Cancer J Clinicians*, **61**, 69-90.

Jiang SQ, Liu Q (2009). Application of logistic regression in combination with multiple diagnostic tests for auxiliary diagnosis of nasopharyngeal carcinoma. *Chinese J Cancer*, **28**, 177-80.

Jin Y, Cai XY, Cai YC, et al (2012). To build a prognostic score model containing indispensible tumour markers for metastatic nasopharyngeal carcinoma in an epidemic area. *Eur J Cancer*, **48**, 882-8.

Landi MT, Chatterjee N, Yu K, et al (2009). A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet*, **85**, 679-91.

Lerman J (1996). Study design in clinical research: sample size estimation and power analysis. Canadian journal of anaesthesia. *J Canadien d'anesthesie*, **43**, 184-91.

Lu C-C, Chen J-C, Jin Y-T, et al (2003). Genetic susceptibility to nasopharyngeal carcinoma within the HLA-A locus in Taiwanese. *Int J Cancer*, **103**, 745-51.

Lu C-C, Chen J-C, Tsai S-T, et al (2005). Nasopharyngeal carcinoma–susceptibility locus is localized to a 132 kb segment containing HLA-A using high-resolution microsatellite mapping. *Int J Cancer*, **115**, 742-6.

Ng CC, Yew PY, Puah SM, et al (2009). A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. *J Hum Genet*, **54**, 392-7.

Peng B, Li B, Han Y, Amos CI (2010). Power analysis for case-control association studies of samples with known family histories. *Human Genetics*, **127**, 699-704.

Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA (2001). Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Nat Cancer Inst*, **93**, 358-66

Selvachandran SN, Hodder RJ, Ballal MS, Jones P, Cade D (2002). Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. *Lancet*, **360**, 278-83

Serraino D, Piselli P, Angeletti C, et al (2005). Infection with Epstein-Barr virus and cancer: an epidemiological review. *J Biol Regul Homeost Agents*, **19**, 63-70

Spitz MR, Hong WK, Amos CI, et al (2007). A risk model for prediction of lung cancer. *J Natl Cancer Inst*, **99**, 715-26.

Taplin SH, Thompson RS, Schnitzer F, Anderman C, Immanuel V (1990). Revisions in the risk-based breast cancer screening

program at group health cooperative. *Cancer*, **66**, 812-8.
Teo YY, Sim X, Ong RT, et al (2009). Singapore genome variation project: a haplotype map of three Southeast Asian populations. *Genome Res*, **19**, 2154-62.
Tice JA, Cummings SR, Ziv E, Kerlikowske K (2005). Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat*, **94**, 115-22.
Tse K-P, Su W-H, Chang K-P, et al (2009). Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. *Am J Human Genetics*, **85**, 194-203.
Vokes EE, Liebowitz DN, Weichselbaum RR (1997). Nasopharyngeal carcinoma. *Lancet*, **350**, 1087-91.
Whelan SL, Ferlay J (1992). Cancer incidence in five continents. Age-specific and standardized incidence rates. *IARC Scientific Publications*, **120**, 178-861
Wijnen JT, Vasen HFA, Khan PM, et al (1998). Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *New Engl J Med,* **339**, 511-8.
Yoshizaki T, Ito M, Murono S, et al (2012). Current understanding and management of nasopharyngeal carcinoma. *Auris Nasus Larynx*, **39**, 137-44.
Yu MC, Ho JHC, Lai S-H, Henderson BE (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 MC, Mo C-C, Chong W-X, Yeh F-S, Henderson BE (1988). Preserved foods and nasopharyngeal carcinoma: a case-control study in Guangxi, China. *Cancer Res*, **48**, 1954-9.
Yu MC, Yuan JM (2002) Epidemiology of nasopharyngeal carcinoma. *Sem Cancer Biol*, **12**, 421-9
Yuan J-M, Wang X-L, Xiang Y-B, et al (2000). Preserved foods in relation to risk of nasopharyngeal carcinoma in Shanghai, China. *Int J Cancer*, **85**, 358-63.
Zainal AO, Zainudin MA, Nor Saleha IT (2006). Malaysia Cancer Statistics - Data and Figure Peninsular Malaysia 2006. National Cancer Registry, Ministry of Health Malaysia
Zhang XS, Wang HH, Hu LF, et al (2004). V-val subtype of Epstein-Barr virus nuclear antigen 1 preferentially exists in biopsies of nasopharyngeal carcinoma. *Cancer Letters*, **211**, 11-8.
Zheng W, Long J, Gao YT, et al (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet*, **41**, 324-8.
Zheng YM, Tuppin P, Hubert A, et al (1994). Environmental and dietary risk factors for nasopharyngeal carcinoma: a case-control study in Zangwu County, Guangxi, China. *Br J Cancer*, **69**, 508-14.