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.

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 poly-cyclic 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,

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 suscept-ibility 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 suscept-ibility 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 popula-tions. 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 con-struction of a large body of SNP information through the Interna-tional 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 correla-tion 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 indivi-duals 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 hypoth-esis 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
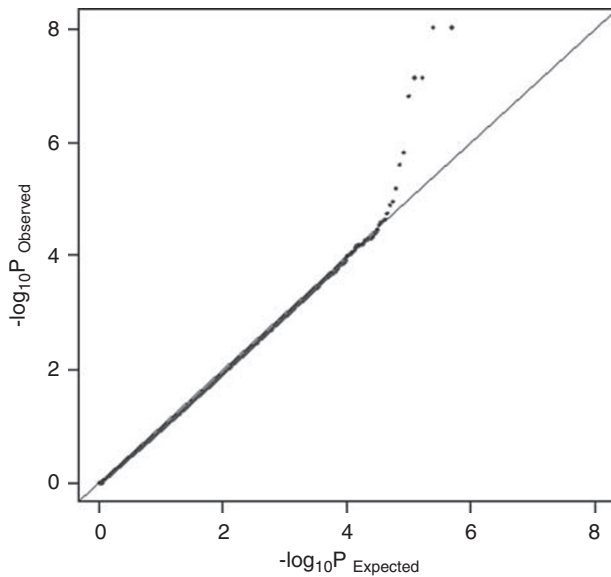
Figure 1 Log quantile–quantile (Q–Q) *P*-value plot showing the distribution of observed statistics by allelic test for all utilized 533 048 single-nucleotide polymorphisms from a genome-wide association study of 111 NPC patients and 260 controls of a Malaysian Chinese population (Panel A). The diagonal 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

SNP ID	Position		Alleles[a]	Panel	Case Genotype 11	12	22	MAF	Control Genotype 11	12	22	MAF	Risk allele	P-value (Fisher's exact test) 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.
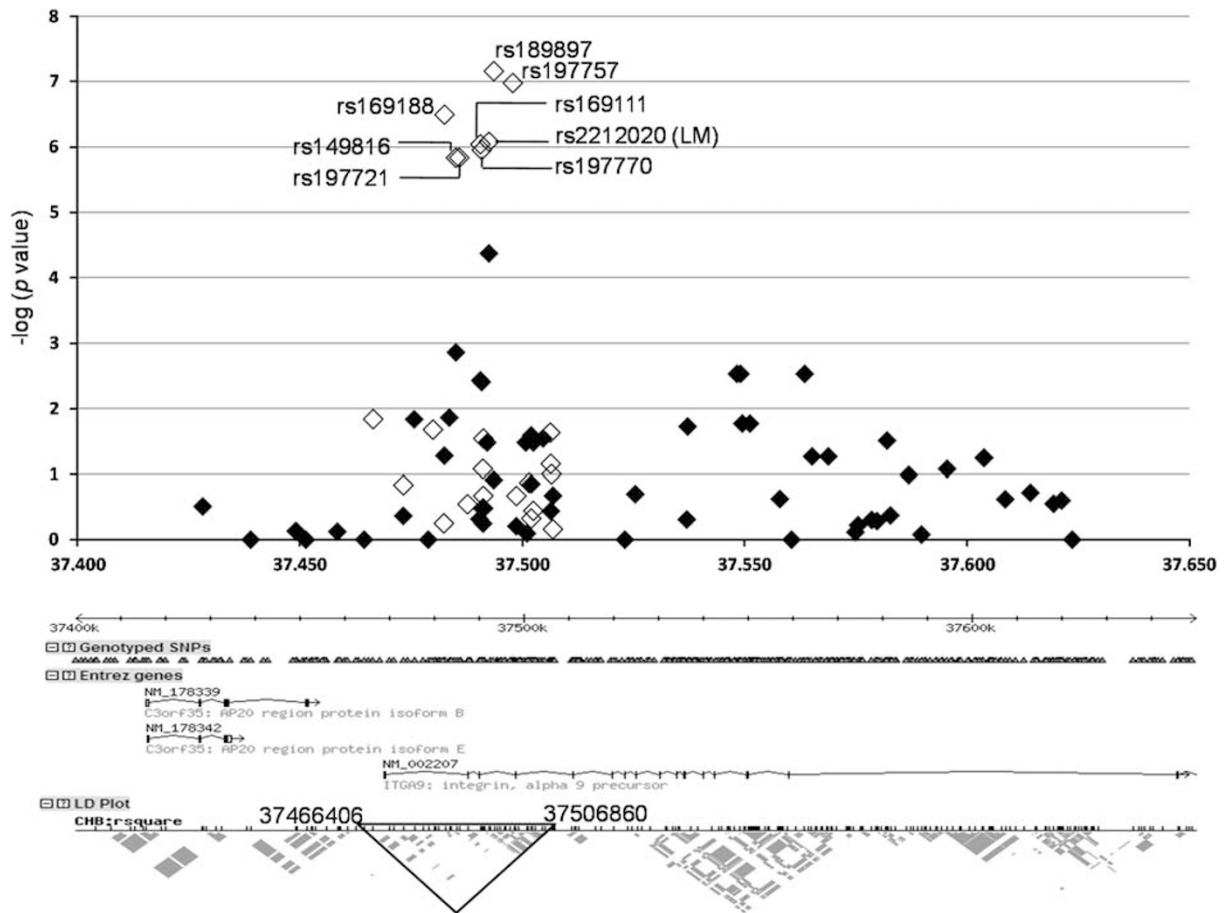
Figure 2 Case–control association plots, linkage disequilibrium (LD) map and the genomic structure of an *ITGA9* region on chromosome 3p21. Closed and open diamond represent $-\log_{10}$ (*P*-value) obtained from the genome-wide association study (Panel A) and fine mapping (combined samples), respectively. 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

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).