IJC
International Journal of Cancer

HLA-A SNPs and amino acid variants are associated with nasopharyngeal carcinoma in Malaysian Chinese

Yoon-Ming Chin[1,2], Taisei Mushiroda[3], Atsushi Takahashi[4], Michiaki Kubo[5], Gopala Krishnan[6], Lee-Fah Yap[7,8], Soo-Hwang Teo[8], Paul Vey-Hong Lim[9], Yoke-Yeow Yap[10,11], Kin-Choo Pua[12], Naoyuki Kamatani[4], Yusuke Nakamura[13], Choon-Kook Sam[14], Alan Soo-Beng Khoo[15], The Malaysian NPC Study Group and Ching-Ching Ng[1,2]

[1] Laboratory for International Alliance, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[2] Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia
[3] Laboratory for Pharmacogenetics, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[4] Laboratory for Statistical Analysis, RIKEN Center for Genomic Medicine, Minato-ku, Tokyo 108–8639, Japan
[5] Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
[6] Department of Otorhinolaryngology, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia
[7] Department of Oral Biology and Biomedical Sciences and Oral Cancer Research and Coordinating Centre, Faculty of Dentistry, University of Malaya, Kuala Lumpur 50603, Malaysia
[8] Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya 47500, Selangor, Malaysia
[9] Tung Shin Hospital, Jalan Pudu, Kuala Lumpur, Malaysia
[10] Department of Surgery, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang 43400, Malaysia
[11] Department of Otorhinolaryngology, Hospital Kuala Lumpur, Kuala Lumpur 50586, Malaysia
[12] Department of Otorhinolaryngology, Hospital Pulau Pinang, Penang 10990, Malaysia
[13] Department of Medicine, University of Chicago, Chicago, IL 60637
[14] Natural Sciences and Science Education AG, National Institute of Education, Nanyang Technological University, Singapore 637616, Singapore
[15] Molecular Pathology Unit, Cancer Research Centre, Institute for Medical Research, Kuala Lumpur 50588, Malaysia

Key words: nasopharyngeal carcinoma, *HLA-A*, GWAS, imputation, Malaysian Chinese

Abbreviations: λ_{gc}: lambda genomic control inflation factor; 3′ UTR: 3′-untranslated region; 5′-UTR: 5′ untranslated region; 95% CI: 95% confidence interval; ASR: age-standardized rate; dbMHC: database for major histocompatibility complex; EA: early antigen; EBV: Epstein–Barr virus; eQTL: expression quantitative trait loci, GWAS: genome wide association study; HKL: Kuala Lumpur General Hospital; HPP: Penang General Hospital; HUS: Hospital University Sarawak; HWE: Hardy–Weinberg equilibrium; IBS: identity-by-state; IgA: immunoglobulin A; ImmPort: Immunology Database and Analysis Portal; LD: linkage disequilibrium; MAF: minor allele frequency; MHC: major histocompatibility complex; NCBI: National Center for Biotechnology Information; NPC: nasopharyngeal carcinoma; OR: odds ratio; PCA: principal component analysis; PROVEAN: Protein Variation Effect Analyzer; Q–Q plot: quantile–quantile plot; QES: Queen Elizabeth Hospital Sabah; r^2: Pearson's correlation coefficient; RIKEN: The Institutes of Physical and Chemical Research; SIFT: Sorting Intolerant from Tolerant; SNP: single nucleotide polymorphism; UMMC: University Malaya Medical Centre; VCA: viral capsid antigen

Additional Supporting Information may be found in the online version of this article.

The Malaysian Nasopharyngeal Carcinoma Study Group—Hospital Pulau Pinang: K.C. Pua (Project Leader), S. Subathra, N. Punithavati, B.S. Tan, Y.S. Ee, L.M. Ong, R.A. Hamid, M. Goh, J.C.T. Quah, J. Lim; Hospital Kuala Lumpur/Universiti Putra Malaysia: Y.Y. Yap, B.D. Dipak, R. Deepak, F.N. Lau, P.V. Kam, S. Shri Devi; Queen Elizabeth Hospital: C.A. Ong, C.L. Lum, Ahmad N.A., Halimuddin S., M. Somasundran, A. Kam, M. Wodjin; Sarawak General Hospital/ Universiti Malaysia Sarawak: S.K. Subramaniam, T.S. Tiong, T.Y. Tan, U.H. Sim, T.W. Tharumalingam, D. Norlida, M. Zulkarnaen, W.H. Lai; University of Malaya: G. Gopala Krishnan, C.C. Ng, A.Z. Bustam, S. Marniza, P. Shahfinaz, O. Hashim, S. Shamshinder, N. Prepageran, L.M. Looi, O. Rahmat, J. Amin, J. Maznan; Hospital Universiti Sains Malaysia: S. Hassan, B. Biswal; Cancer Research Initiatives Foundation: S.H. Teo, L.F. Yap; Institute for Medical Research: A.S.B. Khoo (Program Leader), A. Munirah, A. Subasri, L.P. Tan, N.M. Kumaran, M.S. Nurul Ashikin, M.S. Nursyazwani, B. Norhasimah, R. Sasela Devi, S. Shri Devi, C.Y. Koh

Disclosure: None declared.

Author contribution: Y.-M.C. designed and performed the experiments, analysed data and wrote the paper. T.M. and N.C.C. designed and supervised the project. N.K. and Y.N. supervised the project. A.T. performed statistical analysis of GWAS data and M.K. performed GWAS genotyping. G.K., L.-F.Y., S.-H.T., P.V.-H.L.,Y.-Y.Y., A.S.-B.K., K.-C.P., C.-K.S., and T.M.N.S.G. contributed samples and reviewed the manuscript.

Grant sponsor: University of Malaya, Malaysia (UM-MOHE High Impact Research Grant); **Grant number:** UM.C/625/1/1HIR/282 MOHE/CHAN-02(H-50001-A000023)

DOI: 10.1002/ijc.29035

History: Received 28 Nov 2013; Accepted 15 May 2014; Online 20 June 2014

Correspondence to: Ching-Ching Ng, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia. Fax: +60-379-675908, E-mail: ccng@um.edu.my

Nasopharyngeal carcinoma (NPC) arises from the mucosal epithelium of the nasopharynx and is constantly associated with Epstein–Barr virus type 1 (EBV-1) infection. We carried out a genome-wide association study (GWAS) of 575,247 autosomal SNPs in 184 NPC patients and 236 healthy controls of Malaysian Chinese ethnicity. Potential association signals were replicated in a separate cohort of 260 NPC patients and 245 healthy controls. We confirmed the association of *HLA-A* to NPC with the strongest signal detected in rs3869062 ($p = 1.73 \times 10^{-9}$). *HLA-A* fine mapping revealed associations in the amino acid variants as well as its corresponding SNPs in the antigen peptide binding groove ($p_{\text{HLA-A-aa-site-99}} = 3.79 \times 10^{-8}$, $p_{\text{rs1136697}} = 3.79 \times 10^{-8}$) and T-cell receptor binding site ($p_{\text{HLA-A-aa-site-145}} = 1.41 \times 10^{-4}$, $p_{\text{rs1059520}} = 1.41 \times 10^{-4}$) of the *HLA-A*. We also detected strong association signals in the 5′-UTR region with predicted active promoter states ($p_{\text{rs41545520}} = 7.91 \times 10^{-8}$). SNP rs41545520 is a potential binding site for repressor ATF3, with increased binding affinity for rs41545520-G correlated with reduced *HLA-A* expression. Multivariate logistic regression diminished the effects of *HLA-A* amino acid variants and SNPs, indicating a correlation with the effects of *HLA-A*11:01*, and to a lesser extent *HLA-A*02:07*. We report the strong genetic influence of *HLA-A* on NPC susceptibility in the Malaysian Chinese.

What's new?

Certain variants of the *HLA-A* gene are linked to either resistance or susceptibility in nasopharyngeal carcinoma (NPC). But which variants are most strongly associated with effects in NPC remains unclear. Here, high resolution fine-mapping of the *HLA-A* region was used to better understand the effects of variants on peptide loading or *HLA-A* expression in a Malaysian Chinese population. Variants showing potential epigenetic, peptide-loading function and T-cell immune response were correlated with the effects of *HLA-A*11:01*, a protective *HLA-A* allele. Most other *HLA-A* variants did not appear to possess any potential function.

Nasopharyngeal carcinoma (NPC) [OMIM 161550] is an epithelial squamous cell carcinoma arising from the mucosal lining of the nasopharynx. NPC has a distinct geographical distribution, highly endemic to southern China, Hong Kong, Taiwan, Northern Africa, Alaska, and Southeast Asia with annual incidence ranging from 2.7 to 26.9 cases per 100,000 in males and a significantly lower incidence of 0.9–10.1 cases per 100,000 in females.[1] In Malaysia, NPC is the fourth most common cancer among Malaysians and third most common among men.[2] NPC incidence in males (age-standardized rate per 100,000 people, ASR = 6.3) is double that of females (ASR = 2.3). The disease is particularly prevalent in the Chinese (ASR = 10.9), followed by the Malays (ASR = 3), and Indians (ASR = 1.1). A high risk of NPC has also been observed among indigenous groups in East Malaysia, particularly the Bidayuh in Sarawak.[3]

NPC is a multifactorial disease, suggesting a possible interaction between genetic and environmental factors in disease susceptibility. Elevated levels of immunoglobulin antibodies (IgA) to Epstein–Barr virus (EBV) viral capsid antigen (EBV-IgA/VCA) and early antigen (EBV-IgA/EA) are observed in NPC patients, implicating EBV's role in NPC development.[4] Lifestyle trends such as smoking, alcohol intake, consumption of salted fish or preserved food and occupational hazard all contribute to increased risk of NPC.[5,6] Many candidate genes have been suggested in NPC susceptibility,[7–10] in particular the major histocompatibility complex (MHC) region, with great focus on the antigen presenting molecule *HLA-A* gene.[11] The association of the *HLA-A* with NPC susceptibility has been well-documented in linkage studies, candidate gene approaches and GWAS studies.[12–16] The most recent GWAS identified strong association signals in *HLA-A* amino acid variants.[14] *HLA-A* allele studies have also demonstrated both NPC resistance and NPC susceptible associations.[17,18] However, with the existence of multiallelic variants as well as benign and damaging variants in the *HLA-A* region, a high resolution genotyping as well as *in silico* approach is needed to resolve association signals corresponding to damaging variants and the correlation of these variants to the effects of the *HLA-A* alleles.

We present an imputation-based approach to methodically fine-map and resolve association signals from LD-proxy associations of the *HLA-A*. Imputation is an attractive approach considering the extensive LD of the MHC region due to a relatively low recombination rate.[19] Our strategy employs the following methods: (1) A genome-wide sweep of 712,717 SNPs on the Illumina Human OmniExpress platform to identify NPC associated loci; (2) imputation with HapMap, 1000 Genomes datasets and previously reported GWAS data[20] to fine map the *HLA-A* region; (3) high-resolution molecular *HLA-A* allele genotyping to identify *HLA-A* alleles as well as its corresponding SNP genotypes; (4) *in silico* prediction and amino acid variant analysis to identify functional associations.

Material and Methods

Study cohort

NPC patients were recruited from University Malaya Medical Centre (UMMC), Tung Shin Hospital, Kuala Lumpur General Hospital (HKL), Penang General Hospital (HPP), Nilai Cancer Institute Hospital (NCI), Hospital University Sarawak