Evaluation of Rare and Common Variants from Suspected Familial or Sporadic Nasopharyngeal Carcinoma (NPC) Susceptibility Genes in Sporadic NPC

Zhiwei Liu1, Alisa M. Goldstein1, Wan-Lun Hsu2, Kelly J. Yu1, Yin-Chu Chien2, Jenq-Yuh Ko3, James Jer-Min Jian4, Yung-An Tsou5, Yi-Shing Leu6, Li-Jen Liao7, Yen-Liang Chang8, Cheng-Ping Wang3, Jia-Shing Wu4, Chun-Hung Hua5, Jehn-Chuan Lee6, Tsung-Lin Yang3, Chuhsing Kate Hsiao9, Ming-Shiang Wu10, Ming-Hsui Tsai5, Kuei-Kang Huang11, Kai Yu1, Kristie Jones1, Bin Zhu1, Meredith Yeager1, Guoqin Yu1, Pei-Jen Lou3, Chien-Jen Chen2, and Allan Hildesheim1; for the GEV-NPC group

Abstract

Background: Genetic susceptibility is associated with nasopharyngeal carcinoma (NPC). We previously identified rare variants potentially involved in familial NPC and common variants significantly associated with sporadic NPC.

Methods: We conducted targeted gene sequencing of 20 genes [16 identified from the study of multiplex families, three identified from a pooled analysis of NPC genome-wide association study (GWAS), and one identified from both studies] among 819 NPC cases and 938 controls from two case-control studies in Taiwan (independent from previous studies). A targeted, multiplex PCR primer panel was designed using the custom Ion AmpliSeq Designer v4.2 targeting the regions of the selected genes. Gene-based and single-variant tests were conducted.

Results: We found that NPC was associated with combined common and rare variants in CDKN2A/2B (P = 1.3 × 10^{-5}), BRD2 (P = 1.6 × 10^{-3}), TNFRSF19 (P = 4.0 × 10^{-3}), and CLPTM1L/TERT (P = 5.4 × 10^{-3}). Such associations were likely driven by common variants within these genes, based on gene-based analyses evaluating common variants and rare variants separately (e.g., for common variants of CDKN2A/2B, P = 4.6 × 10^{-1}; for rare variants, P = 0.04). We also observed a suggestive association with rare variants in HNRNPU (P = 3.8 × 10^{-3}) for NPC risk. In addition, we validated four previously reported NPC risk-associated SNPs.

Conclusions: Our findings confirm previously reported associated variants and suggest that some common variants in genes previously linked to familial NPC are associated with the development of sporadic NPC.

Impact: NPC-associated genes, including CLPTM1L/TERT, BRD2, and HNRNPU, suggest a role for telomere length maintenance in NPC etiology.

Introduction

While Epstein–Barr virus (EBV) infection is a necessary cause for virtually all undifferentiated nasopharyngeal carcinoma (NPC), only a small proportion of infected individuals develop NPC (1, 2). Inherited genetic susceptibility is hypothesized to be an important risk factor for NPC among EBV-infected individuals, and agnostic and targeted studies of sporadic and familial NPC support this notion (3, 4).

We recently completed a whole-exome sequencing study among 251 individuals from 97 multiplex NPC families from Taiwan in which novel, rare variants were implicated in the development of familial NPC (5). These variants are located in genes potentially involved in notch signaling (NOTCH1, CDKN2A/2B, BRD2, TNFRSF19, and CLPTM1L/TERT).

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

P.-J. Lou, C.-J. Chen, and A. Hildesheim contributed equally to this article.

Corresponding Author: Zhiwei Liu, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892. Phone: 240-276-6726; Fax: 240-276-7806; E-mail: zhiwei.liu@nih.gov

Cancer Epidemiol Biomarkers Prev 2019;28:1682-6
doi: 10.1158/1055-9965.EPI-19-0007
©2019 American Association for Cancer Research.

1Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. 2Department of Genetic Research, Academia Sinica, Taipei, Taiwan. 3Department of Otorhinolaryngology, China Medical University Hospital, Taichung, Taiwan. 4Department of Otorhinolaryngology, National Taiwan University Hospital, Taichung, Taiwan. 5Department of Otolaryngology, Mackay Memorial Hospital, Taipei, Taiwan. 6Department of Otolaryngology, Far Eastern Memorial Hospital, New Taipei City, Taiwan. 7Department of Otolaryngology, Cathay General Hospital, Taipei, Taiwan. 8Graduate Institute of Epidemiology and Preventative Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan. 9Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. 10Department of Oncology, Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan. 11Department of Radiation Oncology, Koo Foundation Sun Yat-Sen Cancer Center, Taipei, Taiwan.
DLL3, LFNG, MAML1, MFNG, PSEN2), magnesium transport (NIPAL1), EBV entry into epithelial cells (ITGβ6), modulation of EBV (BCL2L12, NEDD4L, LAMC2), telomere biology (CLPTM1L, BRD2, HRNPU), DNA repair (PRKDC, MLH1), or modulation of cAMP signaling (RAPGEF3). Separate findings from a meta-analysis of sporadic NPC noted significant associations for common polymorphisms within genes involved in telomere biology (CLPTM1L/TERT), apoptosis (MECOM, TNFRSF19), and cell-cycle regulation (CDKN2A/2B; ref. 6). Taken together, these studies of familial and sporadic NPC point to a set of candidate genes that might be important in NPC pathogenesis.

To follow up on these findings, we performed targeted gene sequencing of 20 genes/gene regions in an independent set of 1,757 NPC cases and controls from two studies in Taiwan. In this study, we aimed (i) to evaluate whether variability in genes involved in familial NPC might also be associated with sporadic NPC and (ii) to determine whether additional polymorphisms in genes previously associated with sporadic NPC could be identified.

Materials and Methods

Study population

The two case–control studies in Taiwan have been described previously (ref. 7; W.L. Hsu et al., submitted for publication). In brief, the first case–control study, conducted between 1991 and 1994, recruited incident NPC cases from the Taipei metropolitan area and age, sex, and residential area matched controls. Of 1,850 eligible cases and 372 controls, 369 cases (98%) and 320 controls (86%) agreed to participate. The second case–control study, conducted between 2010 and 2014, recruited prevalent and incident NPC cases diagnosed after January 1, 2007, from northern and central Taiwan and age and sex frequency–matched controls from the same geographic regions. Of 1,850 eligible cases and 1,885 controls, 1,600 cases (86%) and 1,804 controls (96%) agreed to participate.

For the current study, we included all individuals with DNA available from the first case–control study (241 cases and 234 controls), and all individuals from the initial phase of the second case–control study (578 cases and 704 controls), as samples were not available for the remaining participants. For cases and controls, separately, we noted a similar distribution of age and sex for those who were included and excluded from each study (P > 0.05). In total, we included 819 NPC cases and 938 controls. Characteristics of the NPC cases and controls are presented in Supplementary Table S1. Of 819 NPC cases, 256 were prevalent (i.e., recruited after treatment was initiated/completed) and 563 were incident (i.e., recruited at the time of diagnosis and prior to treatment initiation). Written informed consent was obtained from all participants, and both studies were approved by human subject committees in Taiwan and the United States.

Targeted sequencing pipeline

DNA was extracted from the peripheral blood lymphocytes using the QIAamp DNA Blood Mini Kit (Qiagen). The quantity and quality of the genomic DNA were assessed by Nanodrop 1000 (Thermo Fisher Scientific) and Qubit (Life Technologies), respectively. A targeted, multiplex PCR primer panel was designed using the custom Ion AmpliSeq Designer v4.2 (Life Technologies) targeting the coding regions of the 20 selected genes/gene regions. Sample DNA (30 ng) was amplified using this custom AmpliSeq primer pool, and libraries were prepared following the manufacturer’s Ion AmpliSeq Library Preparation protocol (Life Technologies).

Raw sequencing reads generated by the Ion Torrent sequencer were quality and adaptor trimmed by Ion Torrent Suite and then aligned to the hg19 reference sequence by TMAP (https://github.com/ontorrent/TS/tree/master/Analysis/TMAP) using default parameters. Resulting BAM files were aligned using Genome Analysis Toolkit (GATK) UnifiedGenotyper v3.1; ref. 8). Variant annotation was done by snpEff, SnpSift (http://snpeff.sourceforge.net/), and Annovar (9).

On average, we achieved a read length of 224 bp (with a minimum of 198 bp) across the samples, with an average of mean depth of 1,113 (interquartile range: 849–1,296). Insertions and deletions (INDEL) were not called because of the potential high false positive rate of INDELS calling for the data generated from the Ion Torrent sequencer. A total of 1,966 variants were identified. Genotypes with read depth (DP) < 8 or genotype quality (GQ) < 20 were considered as ambiguous calls and therefore were considered as missing values. Variants missing in more than 5% of the samples in cases and controls were excluded (N = 120). We excluded multiallelic variants (172 SNPs) and SNPs with Hardy–Weinberg equilibrium P < 10^{-6} (N = 2). Six candidate SNPs, which were significantly associated with NPC in a previous genome-wide association study (GWAS) pooled analysis (6), were interrogated with the Ion Torrent panel, as compared with the variant calling data, no discrepancy was observed.

Statistical analysis

We first focused on genes previously identified from a study of NPC multiplex families (N = 17) to test the hypothesis that genes involved in familial NPC are also associated with sporadic NPC. We then focused on genes associated with sporadic NPC from a previously reported NPC GWAS (N = 4; 1 overlapped with the list from the family study) both to confirm previous findings and to determine whether there is support for an association with NPC for additional variants within those genes.

For gene-based analyses, we used the Sequence Kernel Association combined sum test (SKAT-O) to estimate the effect of common variants and combined effect of common (minor allele frequency, MAF, ≥1% among all study subjects) and rare (MAF < 1% among all study subjects) variants with NPC risk (10). To evaluate the association between rare variants alone and NPC risk, we used the SKAT optimal adjusted test (SKAT-O) that encompasses SKAT and the burden test. All SKAT analyses were performed using default weights, adjusting for sex and age groups (in 10-year category). For single-variant analyses (restricted to common variants), associations between SNPs and NPC were analyzed by logistic regression under a log additive model, adjusting for sex, age, and sex groups (in 10-year category).

We considered a gene-based association significant if the nominal P value was < 2.5 × 10^{-3}, corresponding to a Bonferroni correction for 20 genes evaluated. For the four SNPs with previously reported genome-wide significance (P < 5 × 10^{-8}; ref. 6), we considered P value of <0.05 as significant. For other SNPs,
we considered a single-variant association significant if the nominal $P$ value was $< 2.7 \times 10^{-3}$, corresponding to a Bonferroni correction for 187 tests (total number of common variants evaluated). Sensitivity analysis was conducted among incident cases and all controls.

On the basis of the current sample size (~800 cases and ~900 controls), after correction for multiple tests (significant level = $2.7 \times 10^{-3}$), we had at least 80% power to detect an OR as low as 3.4, 1.9, and 1.6, corresponding to MAF of 1%, 5%, and 10%, respectively. Because of a modest sample size, we have a limited power to identify disease association for rare variants.

**Results**

**Number of variants**

From the selected 20 genes/gene regions, we identified a total of 1,670 variants among 1,757 participants (819 NPC; 938 controls), of which 187 were common (MAF $\geq$ 1%).

**Gene-based analyses**

The association between the combined effect of rare and common variants and NPC risk is summarized in Table 1. For genes previously identified from a study of multiplex families and a meta-analysis of NPC GWAS, the most significant association was observed for BRD2 located on chromosome 6 ($P_{\text{SKAT-C}} = 0.0016$). A suggestive association was observed for CLPTM1L/TERT located on chromosome 5 ($P_{\text{SKAT-C}} = 0.0054$). In analyses restricted to rare variants (MAF < 1%), however, no evidence for significant associations for these two genes was noted, suggesting that the observed association was more likely explained by common variants within these gene regions. We observed a suggestive association with NPC for rare variants in HNRNPU located on chromosome 1 ($P_{\text{SKAT-O}} = 0.0038$).

For genes previously identified from a meta-analysis of NPC GWAS, the most significant association was observed for CDKN2A/2B located on chromosome 9 ($P_{\text{SKAT-C}} = 1.3 \times 10^{-5}$), followed by a suggestive association for TNFRSF19 located on chromosome 13 ($P_{\text{SKAT-C}} = 0.0040$). Again, no evidence for significant associations for these two genes was noted in analyses restricted to rare variants.

**SNP-based analyses**

The association between individual common SNPs and NPC risk is summarized in Table 2 and Supplementary Tables S2 and S3. For SNPs from genes previously identified from a multiplex family study, suggestive evidence in support of an association with NPC was observed for rs78231671 ($P = 4.7 \times 10^{-4}$, Table 2), an intronic SNP in the PRDRC gene region, and for rs76146382 ($P = 5.8 \times 10^{-4}$) in the BRD2 gene region.

We next evaluated the four SNPs that were previously reported to be significantly associated with NPC in a meta-analysis (6). All four SNPs, located within the CLPTM1L/TERT, CDKN2A/2B, TNFRSF19, and MECOM genes/gene regions, were significantly associated with NPC, and the direction of the association was the same as previously reported ($P < 0.05$, Table 2). Of an additional 29 SNPs evaluated within these four genes (Supplementary Table S2), 10 were in suggestive linkage disequilibrium (LD) with the primary $a$ priori SNPs evaluated ($r^2 > 0.20$ with the primary $a$ priori SNP, Supplementary Fig. S1A–S1D). Of the remaining 19 SNPs, we found suggestive evidence for an association with NPC for one SNP within CLPTM1L/TERT, rs13167280 ($P = 0.012$, Supplementary Table S2), and one SNP within the MECOM region, rs17466625 ($P = 0.011$).

The association between other common SNPs and NPC risk is summarized in Supplementary Table S3. Associations of similar...
magnitude were observed in analysis among incident cases and controls (Supplementary Tables S4 and S5).

Discussion

In this targeted gene–sequencing study, we extended our findings from a study of NPC multiplex families aimed at identifying genes associated with familial NPC risk by showing that two genes, CLPTM1L/TERT and BRD2, are also associated with sporadic NPC risk in Taiwan. A suggestive association was observed for HNRNPU. In addition, we confirmed results from a recent meta-GWAS of NPC by showing that four previously reported SNPs in genes CDKN2A/2B, MECOM, TNFRSF19, and CLPTM1L/TERT loci also conferred risk for sporadic NPC in this Taiwanese sample.

The biological implication for genes CDKN2A/2B, TNFRSF19, and CLPTM1L/TERT in NPC pathogenesis have been discussed previously (6, 11). These genes are involved in cell growth control (CDKN2A/2B and TNFRSF19) and telomere length maintenance (CLPTM1L/TERT). Both BRD2 and HNRNPU are also involved in telomere biology (12, 13). BRD2 is one of the host chromatin factors interacting with the Kaposi sarcoma–associated herpesvirus latency–associated nuclear antigen 1 that affects chromatin structure and maintenance of the telomere repeat (14). Of note, the most significantly associated SNP within BRD2 was rs76146382. This SNP is in LD with an extended HLA haplotype known to be associated with NPC. HLA-A*3303–R*5801–DRB1*0301 (r2 > 0.6; refs. 7, 15). It remains to be determined whether the association with BRD2 observed in this study is explained by or independent of this previously reported HLA association. The role of HNRNPU in NPC pathogenesis is not clear. Overexpression of HNRNPU has been associated with telomere shortening (12). In addition, HNRNPU shows a strong homology with sequence of EBV nuclear antigen 1 (EBNA1), which is important for NPC development (16). Whether such similarity would have a biological implication is unclear. Taken together, although other mechanisms merit further investigation, our findings further support a role for telomere maintenance in NPC pathogenesis.

Variants in other genes previously linked to familial NPC were not associated with the risk of sporadic NPC. This finding could be due to the fact that genes involved in familial NPC pathogenesis differ from genes in sporadic NPC pathogenesis or to false positive findings in our previous familial NPC study. In addition, false negative findings cannot be ruled out given the modest sample size and power of our study to identify disease associations for rare variants that occur with very low frequency in the population and account for a small proportion of NPC cases. More studies are needed to confirm our findings.

In summary, variants in several genes previously linked to familial NPC were associated with the risk of sporadic NPC, and the risk was more likely to be driven by common rather than rare variants within these genes. We also confirmed previously reported associations from NPC GWAS for SNPs in four genes. Our findings highlight the important role for telomere length maintenance in NPC pathogenesis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The funding organization played no role in the study design, collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript.

Authors’ Contributions


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): W.-L. Hsu, K.J. Yu, Y.-C. Chien, J.-Y. Ko, J.-M. Jian, Y.-A. Tsou, L.-J. Liao, C.-P. Wang, C.-H. Hua, T.-L. Yang, C.K. Hsiao


Acknowledgments

This research was supported by the NCI Intramural Research Program and research grants from Academia Sinica, Taipei, Taiwan.

The GENV-NPC Study Group includes Genomics Research Center, Academia Sinica, Taipei, Taiwan (Wan-Lun Hsu, Yin-Chu Chien, Chien-Jen Chen); Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD (Allan Hildesheim, Zhwei Liu, Kelly J. Yu); Department of Otolaryngology, National Taiwan University Hospital and National Taiwan University Hospital (Wan-Lun Hsu, K.J. Yu, Y.-C. Chien, J.-Y. Ko, J.-M. Jian, Y.-A. Tsou, L.-J. Liao, C.-P. Wang, C.-H. Hua, T.-L. Yang, C.K. Hsiao).
References


Evaluation of Rare and Common Variants from Suspected Familial or Sporadic Nasopharyngeal Carcinoma (NPC) Susceptibility Genes in Sporadic NPC

Zhiwei Liu, Alisa M. Goldstein, Wan-Lun Hsu, et al.