Title: A large-scale population-based study reveals that gp42-IgG antibody is protective against Epstein-Barr virus-associated nasopharyngeal carcinoma

Authors: Xiang-Wei Kong[1, 2†], Guo-Long Bu[1†], Hua Chen[1, 13†], Yu-Hua Huang[1, 12†], Zhiwei Liu[6], Yin-Feng Kang[1, 7], Yan-Cheng Li[3,4], Xia Yu[5], Biao-Hua Wu[5], Zi-Qian Li[1], Xin-Chun Chen[1], Shang-Hang Xie[1], Dong-Feng Lin[1], Tong Li[1], Shu-Mei Yan[1, 12], Run-Kun Han[1], Nan Huang[1], Qian-Yu Wang[11], Yan Li[1, 12], Ao Zhang[1] , Qian Zhong[1], Xiao-Ming Huang[2], Weimin Ye[8, 9, 10], Ming-Fang Ji[5*], Yong-Lin Cai[3, 4*], Su-Mei Cao[1*], Mu-Sheng Zeng[1*]

Affiliations:

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center (SYSUCC), Guangzhou, Guangdong, China.

[2]Department of Otorhinolaryngology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

[3]Guangxi Health Commission Key Laboratory of Molecular Epidemiology of Nasopharyngeal Carcinoma, Wuzhou Red Cross Hospital, Wuzhou, Guangxi, China.

[4]Department of Preventive Medicine, Wuzhou Cancer Center, Wuzhou, Guangxi, China.

[5]Cancer Research Institute of Zhongshan City, Zhongshan City People's Hospital, Zhongshan, Guangdong, China.

[6]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA.

[7]Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong, China.

[8]Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Fuzhou, Fujian, China.

[9]Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Fujian Medical University, Fuzhou, Fujian, China.

[10]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[11]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing, China.

[12]Department of pathology, Sun Yat-sen University Cancer Center (SYSUCC), Guangzhou, Guangdong, China.

[13]Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, China.

[#]Address correspondence to Prof. Mu-Sheng Zeng： Sun Yat-sen University Cancer Center, 651 Dongfengdong Avenue, Guangzhou, Guangdong, China. Mail stop code: 510060. Phone number: +86-20-87343192. Email: zengmsh@sysucc.org.cn (M.-S.Z.).

[#]Address correspondence to Prof. Su-Mei Cao: Sun Yat-sen University Cancer Center, 651 Dongfengdong Avenue, Guangzhou, Guangdong, China. Mail stop code: 510060. Phone number: +86-20-87345685. Email: caosm@sysucc.org.cn (S.-M.C.).

[#]Address correspondence to Prof. Yong-Lin Cai: Wuzhou Red Cross Hospital, #3-1, Xinxing Yi Road, Wuzhou, Guangxi, China. Mail stop code: 543002. Phone number: +86-774-3849595. Email: cylzen@163.com.

[#]Address correspondence to Prof. Ming-Fang Ji, Cancer Research Institute of Zhongshan City, Zhongshan City People's Hospital, Zhongshan, Guangdong, China. Mail stop code: 528404. Phone number: +86-760-89880417. Email: jmftbh@sina.com.

[†]X.-W.K., G.-L.B., H.C., and Y.-H.H. were co-first authors.

46 The authors have declared that no conflict of interest exists.

47

Abstract

Background

Epstein-Barr virus (EBV) is associated with nasopharyngeal carcinoma (NPC), but the existence of NPC protective antibody against EBV-associated antigens remains inconclusive.

Methods

NPC cases and matched controls were identified from prospective cohorts comprising 75,481 participants in southern China. ELISA and conditional logistic regression were applied to assess effects of gp42-IgG on NPC. The expression of HLA-II, the gp42 receptor, in nasopharyngeal atypical dysplasia and its impact on EBV infecting epithelial cells were evaluated.

Findings

gp42-IgG titers were significantly lower in NPC cases compared to controls across various follow-up years before NPC diagnosis (P<0.05). Individuals in the highest quartile of gp42-IgG titers had a 71% NPC risk reduction comparing to those in the lowest quartile (odds ratios $[OR]_{Q4vsQ1}$=0.29, 95% confidence intervals [CIs]=0·15 to 0·55, P<0.001). Each unit antibody titer increase was associated with 34% lower risk of NPC (OR=0.66, 95% CI=0.54 to 0.81, P_{trend}<0.001). Their protective effect was observed in cases diagnosed ≥5 years, 1-5 years and <1 year after blood collection (P<0.05). HLA-II expression was detected in 13 of 27 nasopharyngeal atypical dysplasia and its overexpression substantially promoted epithelial-cell-origin EBV infection.

Conclusion

Elevated EBV gp42-IgG titers can reduce NPC risk, indicating gp42 as a potential EBV prophylactic vaccine design target.

Trial registration

NCT00941538, NCT02501980, ChiCTR2000028776, ChiCTR2100041628.

Funding

Noncommunicable Chronic Diseases-National Science and Technology Major Project, National Natural Science Foundation of China, Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program, Central Financial Transfer Payment Projects of the Chinese Government, Cancer Research Grant of Zhongshan City.

Main text

Introduction

The incidence of nasopharyngeal carcinoma (NPC) demonstrates a striking geographical variation (1). In endemic regions such as southern China, the age-adjusted incidence can reach 23/100,000 in males and 7/100,000 in females, which is 10-fold higher than the global incidence (2). The regional variation implies a complex interplay of genetic predisposition, environmental risk factors, and notably, Epstein-Barr virus (EBV) infection (1, 3).

In endemic regions, where over 95% of NPC cases can be attributed to EBV infection, there is a great opportunity for prevention interventions via EBV vaccination to reduce NPC burden. However, a major obstacle emerges when selecting an optimal target for vaccine design, considering as many as five surface glycoproteins with definitive functions that EBV expresses (4). Glycoproteins gB, gHgL, and gp350 have been spotlighted in the development of EBV prophylactic vaccines due to their pivotal roles in EBV-host interactions (4), availability of potent monoclonal antibodies against them (5-8), and encouraging results from preliminary lab-based trials (9). Nonetheless, initial studies assessing neutralizing antibodies targeting gp350 in three cohorts with modest sample sizes have shown inconsistent results, with one showing an association with reduced NPC risk (10), while the other two found associations with increased NPC risk (11, 12). In our previous nested case-control study from a community-based NPC screening cohort, we also tested antibodies against gp42, gHgL, gB, and gp350, as well as B cell and epithelial cell neutralizing antibodies. Our results suggest a potential protective role for antibodies against gp42, with gp42-IgG showing a stronger correlation with epithelial and B cell neutralization competence compared to gp42-IgA. However, the evidence remains inconclusive, partly due to the limited sample size (20 cases and 40 controls) (13). It is worth noting that vaccines aiming at gp42 have been less studied (14).

To follow-up on our initial findings, here, we conducted a larger nested case-control study utilizing samples from 129 NPC cases and 387 matched controls, which were collected prospectively across three

independent cohorts in southern China independent from our previous study (13). Our findings confirmed the inverse association between increased levels of anti-gp42 antibodies and NPC risk. In addition, we sought to explore the physical and molecular basis by examining the presence of HLA-II, the only known receptor for gp42, in nasopharyngeal atypical dysplasia samples. This analysis showed that overexpression of HLA-II in epithelial cells promotes EBV infection. Taken together, our findings suggest that IgG antibodies against gp42 play a protective role against NPC in endemic regions, thereby marking gp42 as a potential prophylactic vaccine design target to prevent NPC in high-risk populations.

Results

Study population

Based on three prospective community-based cohorts in Sihui, Wuzhou and Zhongshan cohorts with a total 75,481 individuals, we conducted a nested case-control study to assess whether elevated levels of gp42-IgG are associated with lower risks of NPC. Blood samples from 129 NPC cases and 387 individually matched controls at 1:3 ratio by age, sex, blood collection time and region were obtained with a median of 1.3 years (IQR=0.3-2.9 years) before NPC diagnosis (Figure 1). Characteristics of the NPC cases and controls included in the analysis were shown in Table 1. A majority of participants were male (76.0%) and at ages of 50-59 years at baseline (44·2%), and NPC cases were more likely to have a higher EBV-antibody based risk score compared to controls. A longer follow-up duration was observed for those in the Zhongshan cohort (P<0.001) (Supplemental Figure 1). All samples were independent from our previous study (13).

Elevated gp42-IgG titers in disease-free controls comparing with incident NPC cases

To accurately quantify serum levels of gp42-IgG and make comparison in different batches of samples, we developed an ELISA detection system and assessed the linearity and precision using serial dilution of monoclonal human gp42 antibody in house. The ELISA showed excellent linearity (goodness of fit of least square regression R^2=0·95) over a broad dynamic range of 0.46–111 ng/mL (Supplemental Figure 2A). We demonstrated the reproducibility by repeating measurements on the three concentrations of monoclonal gp42 antibody standards in triple and across 11 batches. Results showed CVs of less than 20% (Supplemental Figure 2B; Supplemental Table 1), with the corresponding Inter-ICC of 0.97 (95%CI: 0.92-0.97) and Intra-ICC of 0.92 (95%CI: 0.43-0.99) For the randomly selected 10% serum samples (n=52), the results showed a 5% CV for the duplicated testing samples.

Notably, gp42-IgG titers were significantly lower in NPC cases compared to the controls in the overall cohort (P<0.001), and in subgroups with follow-up ≤1 year (P=0.008), 1-5 years (P=0.007), and ≥5 years

(P=0.027) before NPC diagnosis (Figure 2). However, the overall distribution of gp42-IgG titers remained consistent across sexes (P=0.174), age groups (P=0.320) (Supplemental Figure 3). To evaluate the associations of gp42-IgG titer with NPC risk score, we analyzed the gp42-IgG titers at various EBV antibody NPC risk scores in controls by different follow-up durations. Result showed that the controls with low-risk score were more likely to have higher gp42-IgG titers compared to those with medium- and high-risk scores in the overall cohort (P<0.001), and the subgroups with ≤1 year (P=0.051), 1-5 years (P<0.001), and ≥5 years (P=0.212) follow-up before NPC diagnosis (Supplemental Figure 4).

The risk of NPC is the highest in individuals with the lowest gp42-IgG titers

We categorized individuals into quartiles based on gp42-IgG levels in controls and explored the association between gp42-IgG levels and NPC risks using conditional logistic regression. We found that in the overall cohort, elevated levels of gp42-IgG in the highest quartile reduced 71% NPC risk (odds ratios $[OR]_{Q4vsQ1}$=0.29, 95% CI_{Q4vsQ1}=0·15 to 0·55) and the risks of NPC decreased as the levels of gp42-IgG increased (P_{trend}<0.001). When stratified by length of follow-up, the inverse association with dose-response trend persisted among cases diagnosed ≤1, 1-5, and ≥5 years after blood collection, with the corresponding OR_{Q4vsQ1} of 0.33, 0.33, and 0.05, respectively and all the P_{trends}<0.05 (Table 2). The inverse associations could be consistently observed among the subgroups with different sexes, ages, cohorts, and EBV-antibody NPC risk scores, with all the $P_{heterogenity}$≥0.05 (Supplemental Figure 5).

Expression of HLA-II, the sole identified gp42 receptor, is detectable in nasopharyngeal atypical dysplasia

The expression of HLA-II is notably detectable in NPC tumor cells (15); however, it remains unclear whether the expression of HLA-II precedes tumor occurrence. To address this hypothesis, we collected 27 nasopharyngeal premalignant tissues from Sun Yat-sen University Cancer Center, China, where samples underwent examination by two experienced pathologists independently. Among the 27 premalignant

157 samples, 20 were positive for EBERs, 13 were positive for HLA-DRA, and 12 were positive for both
158 HLA-DRA and EBERs (Supplemental Table 2; Figure 3; Supplemental Figure 6).

159 **HLA-II facilitates epithelial-origin EBV infection of epithelial cells**

160 To further investigate the functional implications of HLA-II expression in epithelial cells, we established
161 stable HEK293 and NP69 cell lines overexpressing HLA-DRA/DRB. The overexpression of HLA-DR
162 had a notable impact on enhancing the infection of CNE2-EBV, the representative of EBV originating
163 from epithelial cells (Supplemental Figure 7, A and C; Supplemental Figure 8, A and C). However, this
164 effect was only marginal in promoting the infection of Akata-EBV, the representative of EBV originating
165 from B cells (Supplemental Figure 7, B and D; Supplemental Figure 8, B and D).

166 To further validate the role of gp42 in mediating CNE2-EBV infection, we purified soluble gp42 and
167 HLA-DRA/DRB. Treating cells with soluble gp42 and HLA-DRA/DRB led to a dose-dependent
168 inhibition of EBV infection (Supplemental Figure 9, A and B). Interestingly, the inhibitory effects were
169 countered by the overexpression of HLA-DRA/DRB (Supplemental Figure 7, E and F; Supplemental
170 Figure 9, C and D). Overexpression of HLA-DRA/DRB was confirmed through western blot analysis
171 (Supplemental Figure 7G). These results strongly suggest that gp42 plays an important role in promoting
172 the infection of EBV originating from epithelial cells into HLA-DR-expressing epithelial cells.
173

Discussion

We have conducted the largest, prospective epidemiological study to date, focusing on NPC-associated protective antibodies against EBV based on three mass screening cohorts. Our findings establish the protective capacity of gp42-IgG against NPC in endemic regions. Moreover, our research has elucidated the underlying mechanistic foundations of these protective effects. The identification of HLA-II, the sole receptor for gp42, in nasopharyngeal atypical dysplasia constitutes a novel discovery. Our experiments have further demonstrated that the overexpression of HLA-II amplifies the infectivity of EBV within epithelial cells. Our results on the interplay between gp42-IgG, HLA-II, and EBV infection offer valuable insights for the mechanisms of the observed gp42-IgG protective effects against NPC.

Previous studies have shown that the elevation of antibodies against EBV EA, EBNA1, and VCA are associated with a higher risk of NPC, which leads to utilization the EBV-associated antibodies as the screening markers for NPC in endemic regions (16). In comparison with these markers, the precise roles of antibodies against glycoproteins in the development of NPC remain inconclusive. Although gB, gHgL, and gp350 play essential roles in EBV infection cells, antibodies targeting these components do not exhibit significant protective effects against NPC development, as demonstrated in previous studies (11-13). In this study, which encompassed 129 incident NPC cases and 387 matched controls, we established statistically significant reduction in NPC risks among cases overall and diagnosed ≤1 year, 1-5 years, ≥5 years after blood collection with dose-response trends. Moreover, the protective effects of gp42-IgG can be observed across varying sexes, age groups, and cohorts, indicating the robustness of the protective influence of gp42-IgG against NPC. However, in the previous report, gp42-IgG levels were nominally higher in controls than in those who developed NPC more than 5 years post-collection, but lower in controls than in those who developed NPC within 5 years (12). This discrepancy could be due to the differences in antigenicity and measurement methods. In terms of antigenicity, our purified gp42 can substantially block EBV entry in epithelial cells (Supplemental Figures 7F, 9B, 9D), which was not tested previously (12), and our purified gp42 was used to identify two potent EBV-neutralizing antibodies (17).

In terms of measurement methods, we applied ELISA versus the high-throughput Luminex assay in the previous report (12).

Notably, gp42 is recognized as a crucial molecule that dictates the host preference of EBV (18). A higher presence of gp42, released through epithelial cell lytic infection, results in a preference for B cell infection due to the expression of HLA-II by B cells rather than epithelial cells under physiological conditions. Our research shows that the overexpression of HLA-II only minimally promotes the infection of B-cell-origin EBV in epithelial cells. This observation prompted us to speculate whether, in physiological settings, while antibodies against gB, gHgL, and gp350 confer a foundational level of humoral protection against EBV infection and reactivation, the continuous exposure of epithelial-cell-origin EBV in nasopharyngeal niches might warrant higher titers of gp42 antibodies. Such elevated titers could counteract the susceptibility of nasopharyngeal premalignant lesions to epithelial-cell-origin EBV infection, thereby diminishing the risk of NPC development.

Direct evidence concerning the origin of EBV driving NPC remains elusive. Nevertheless, recent studies indicate persistent exposure to epithelial-cell-origin EBV in nasopharyngeal niches. First, with regard to Cp methylation percentage of the EBV genome, methylation levels are notably lower in lytic infection than in latent infection (19); furthermore, evidence suggests that EBV DNA can be detected in nasopharyngeal swabs as early as three years before NPC development (20). When compared with healthy controls, nasopharyngeal swabs from NPC patients exhibit significantly higher Cp methylation percentages of EBV DNA (21). Second, considering gp42, the quantity of surface-bound gp42 on EBV serves as an indirect indicator of its origin. HLA-II expressed in B cells can capture and reduce the amount of surface-bound gp42 on released EBV (18). Evidence demonstrates that saliva-origin EBV possesses significantly higher levels of gp42 when compared with EBV originating from LCLs (22). Third, with respect to EBV release rate, studies have proposed that the primary source of EBV in saliva is epithelial lytic infection, as the release rate far surpasses that of the maximum release rate from the pharyngeal lymphatic ring (23). Lastly, in immune-competent healthy populations, EBV lytic infection

can be detected in the parotid gland and the tongue's edge (24, 25). Collectively, these findings suggest constant exposure of epithelium-origin EBV in nasopharyngeal niches. Consequently, elevated levels of antibodies against gp42 hold the potential to reduce the likelihood of EBV infection within these niches. In our recent study, we have identified two human monoclonal antibodies against gp42, named 2C1 and 2B7, bearing strong neutralizing effect against epithelial-cell-origin EBV infecting epithelia, the NOK cell line (IC50=0.091 and 0.022 μg/mL, respectively), which provide another strong evidence on the NPC protective effects of gp42-IgG (17).

Though we have provided compelling evidence on the NPC protective role of gp42-IgG, some limitations require further exploration. First, the median observation period is 1.3 years, indicating a possible reverse causality, but our results still demonstrated a significant NPC-risk reduction in the individuals with ≥5 years follow-up period after blood collection should alleviate the concern. Second, the protective effects on gp42-IgG in less prevalent regions remain elusive, and the correlation between gp42-IgG and other NPC risk factors, such as genetic predisposition and environmental factors, requires further exploration (1, 3). Finally, while we have proposed that HLA-II expression in premalignant nasopharyngeal dysplasia may account for the protective role of gp42-IgG, it is still unclear how gp42-IgG levels relate to HLA-II expression levels in nasopharyngeal tissue, when and how HLA-II expression is initiated, and how EBV infection in HLA-II-expressing premalignant nasopharyngeal dysplasia leads to malignant transformation.

In conclusion, our proposal posits that the expression of HLA-II in nasopharyngeal atypical dysplasia predisposes EBV infection originating from the nasal cavity, pharynx, and oral cavity. This susceptibility can be efficiently countered by high levels of gp42-IgG, decreasing the risk of NPC development. Accordingly, prioritizing gp42-targeting EBV vaccines in high-risk populations appears warranted to mitigate NPC risk.

Methods

Sex as a biological variable

Our study examined male and female, and similar findings were reported for both sexes.

Prospective NPC screening cohorts

This study was nested within three large-scale population-based screening trials, conducted in Sihui city and Wuzhou city for NPC screening, and Zhongshan city for liver cancer screening in southern China (20, 26). Between 2014 and 2018, 30,038 local residences aged 30-69 years were recruited in Sihui city and followed up until 2023; between 2018 and 2021, 27,477 residences aged 30-69 years were recruited in Wuzhou city and followed-up until 2023; and in 2012, 17,966 residences aged 35-64 years were recruited in Zhongshan city until follow-up 2022. In Sihui and Wuzhou cohorts, baseline serum was tested for the screening markers of EBV VCA/EBNA1-IgA by ELISA with commercial detection kit (EUROIMMUN AG, Lübeck, Germany, and Zhongshan Bio-Tech Company, Zhongshan, China, respectively). NPC risk scores of each participant are determined by the risk prediction algorithm (Logit P=-3·934+2·203×VCA-IgA+4·797×EBNA1-IgA) and stratified by the P scores (Low risk, P<0·65; Medium risk, 0·65≤P<0·98; High risk, P≥0·98) (26). For individuals with low and medium NPC risk scores, EBV antibody retesting is conducted in every 5- and 1-years intervals, while the high-risk individuals are referred to nasopharyngeal endoscopy and pathological biopsy will be performed if suspicious lesions are found. In all the three cohorts, the baseline blood components (plasma, serum, and white blood cells) were separated within 4-6 hours post collection and stored at -80 ^{0}C for future research. The preserved baseline serum samples were used for gp42-IgG testing for the whole individuals in the nested case-control study, and for VCA-IgA, and EBNA1-IgA testing in Zhongshan cohort. NPC cases were identified through local cancer and death registration system in Zhongshan cohort, as well as in the Sihui and the Wuzhou cohorts.

The median follow-up period were 1·2 years (IQR=0.3-2.7 years), 0·6 years (IQR=0.3-1.5 years), and 3·6 years (IQR=2.2-5.2 years) in the Sihui cohort, the Wuzhou cohort, and the Zhongshan cohort and 81, 27, and 34 incident NPC cases were identified, correspondingly. However, the serum of 13 cases in Zhongshan cohort are unavailable. Thus, a total of 129 incident NPC cases and 387 controls were matched with a 1:3 ratio by sex, age (<40, 40 to 50, 50 to 60, ≥60 years old), sample collection date (±6 months), and regions. Race and ethnicity were not considered a variable in this study.

Soluble protein purification

Soluble EBV (strain M81) glycoproteins gp42 and HLA-DR were purified by 293F protein expression system cultured with chemically defined medium (UP1000, Union Bio) under CO_2 supplied shaking incubator. Transmembrane and cytoplasmic parts of them were truncated for the expression of soluble proteins, signal peptides were optimized for higher yield, and the His-tag was appended to the C-terminus. Six to seven days post transfection, supernatant was collected and purified with Ni Sepharose excel (17371202, Cytiva). Eluate was further purified by Superdex 200 Increase 10/300 GL (28990944, Cytiva). Protein concentration was determined by the BCA method (23225, Pierce). Protein was aliquoted, frozen with liquid N_2, and stored at -80℃ for further use.

Enzyme-link immunosorbent assay (ELISA)

One hundred nanogram of soluble proteins were added to 96-well plates with high binding surface followed by overnight incubation at 4 ℃. Supernatant was discarded and 5% BSA in PBST (T stands for 0·1% Tween-20) was added to block non-specific binding followed by 1 hour incubation at 37 ℃. Supernatant was discarded. Plates were washed 1 time with PBST. Serums were diluted to 60 folds with 5% BSA in PBST, and then added to 96-well plates followed by 1 hour incubation at 37 ℃. Plates were washed 3 times with PBST. Anti-human IgG was diluted at a ratio of 1:5,000 with 5% BSA in PBST and added to each well followed by 1 hour incubation at 37 ℃. Plates were washed 3 times with PBST. TMB (PA107-02, TIANGEN) substances were added to each well. Reactions were developed at room temperature for 10 mins and stopped by adding 1:12 diluted HCl. Read plates at OD450 and OD630. In

each 96-well plate, two wells were set to be negative controls (NCs), where no serum was added; seven wells contain serially diluted monoclonal gp42-IgG. Then, we derive rOD from the following equations: ((OD450-OD630)-average (OD450-OD630 of NCs))/ ((OD450-OD630 of the highest concentration monoclonal anti-gp42)-average (OD450-OD630 of NCs)).

Randomization and masking

Using a random number generator, samples from cases and controls by cohort were randomly assigned to each ELISA plate in a 1:3 allocation. The testing operators were blinded to the serum origination from cases or controls and the data were pooled by independent data monitoring staff for statistical analyses.

Linearity and precision assessment

Linearity was assessed using serial dilution of standard monoclonal human gp42 antibody in house. After serial three-fold dilution of the antibody with the initial concentration of 1,000 ng/mL, 8 samples ranging from 0.46 ng/mL to 1,000 ng/mL were tested in triplicate and simulated (17, 27). To assess within plate and across plate variation, three concentrations of gp42-IgG standards (1,000 ng/mL, 37 ng/mL, 12.3 ng/mL) were tested in triple in one plate and across 11 batches. Inter- and intra-plate variances and agreements were assessed using the coefficient of variation (CV) and intraclass correlation coefficients (ICC) (28). In addition, 10% of the serum samples were randomly selected for duplicate testing in each plate to further evaluate the precision of gp42-IgG. A standardized value (rOD)—the mean of duplicate testing, were used for statistical analysis.

EBV virion preparation

In this study, Akata strain EBV virions were prepared from CNE2-EBV cell strains and Akata-EBV cell strains, where the former and the latter represent epithelium- and lymphocyte-origin EBV. Preparation of these virions followed previous reports (15). Here, we briefly described the virion preparation process.

For epithelium-origin EBV preparation, CNE2-EBV cell strains were cultured in RPMI 1640 (C11875500BT, Gibco) supplied with 10% fetal bovine serum (FSP500, ExCell Bio). When cells reached

80% confluency, TPA (50601ES03, YEASEN) and NaB (B5887-1G, Sigma-Aldrich) were added to induce EBV lytic replication. Twelve hours post induction, medium was replaced with fresh RPMI 1640 supplied with 10% FBS.

As for lymphocyte-origin EBV preparation, Akata-EBV cell strains were cultured in RPMI 1640 supplied with 5% FBS. Aliquot cells to the density of 2×10^6/mL, then goat-anti-human IgG (H0111-6-100ML, Tianfun Xinqu Zhenglong Biochem. Lab) was added to induce EBV lytic replication. Six-hours post induction, medium was replaced with fresh RPMI 1640 supplied with 5% FBS. For both preparation, three days post medium replacement, supernatant was collected, filtered, and centrifuged. Pellets which contained the virions were resuspended with RPMI 1640. Resuspended virions were frozen in liquid N_2 and then stored at -80 ℃ for further use.

EBV in vitro infection analysis

Epithelial cell lines were seeded in 96-well plates. Twenty-four hours later, virions with or without indicated concentration of sHLA-DR or sgp42 were prepared, incubated at 37 ℃ for an hour, and then added to cells. Twenty-four to 48 hours later, flow cytometry was applied to evaluate green fluorescence. For IC50 fitting, the percentage of fluorescent positive cells were normalized by cells infected with virions without soluble proteins and dilution ratios were log10 transformed before non-linear regression was applied. IC50s were fitted by Prism 9.

In situ hybridization (ISH) and immunohistochemistry (IHC)

The cases of atypical dysplasia in the nasopharynx were identified through the pathological station system in Sun Yat-sen University Cancer Center (SYSUCC) between January 2009 and April 2020. A total of 27 cases of atypical dysplasia in nasopharynx with adequate formalin-fixed and paraffin-embedded (FFPE) specimens were collected for subsequent testing. The EBV probe in situ hybridization kit (ISH-7001, ZSGB-BIO) was used to detect EBERs in TMA slides, following the manufacturer's protocol. Immunohistochemistry was performed to analyze HLA-II expression using a benchmark ULTRA

automatic immunostaining device, with anti-HLA-DRA1 (diluted 1:600, ab92511-100ul; Abcam) as the primary antibody. The results were evaluated by two experienced pathologists who were blinded to the clinical data.

Antibody

Antibodies used in this study include the followings: anti-HLA-DRA1 (ab92511, Abcam), anti-HLA-DRB1 (ab133578, Abcam), anti-α-tubulin (2125S, CST), anti-rabbit IgG-HRP (31460, Thermo Fisher), anti-human IgG-HRP (ab6759, Abcam).

Cell lines

HEK293, and NP69 were maintained in a humidified atmosphere at 37 ℃ with 5% CO2. HEK293 was cultured in DMEM (11965092, Gibco) supplemented with 10% FBS (FSP500, ExCell Bio). NP69 was cultured in Keratinocyte SFM (1X) (17005042, Gibco). Details on the establishment of the HLA-II overexpressing cell lines refer to our previous study (17).

Statistics

For differences analysis of detected antibodies and neutralizing antibody titers, Wilcoxon or Kruskal-Wallis tests were applied for two or more stratifications of the characteristics. The continuous gp42-IgG titers were discretized according to the cutoffs based on the overall control groups, allowing for direct comparisons across different stratifications. Conditional logistic regression was applied for nested data analysis to estimate ORs and 95% CIs for the association between antibody and NPC risk stratified by time between baseline sampling and NPC diagnosis (<1 year, 1-5 years, and ≥5 years). No covariate is used to adjust OR. Categorical variables were treated as continuous variables in P_{trend} and $P_{\text{heterogenity}}$ analyses. In terms of EBV infection rate differences under various circumstances, t-test is adopted since where residues complied normal distribution. All statistical analysis was done by R (4.3.1) and R packages (forestplot_3.1.1, survminer_0.4.9, survival_3.5-520). A *P* value less than 0.05 was considered significant.

Study approval

This study has been conducted in accordance with the Declaration of Helsinki and the trail protocols were approved by the Ethics Review Committee of Sun Yat-sen University Cancer Center (YP2009051, B2020-362-01), Wuzhou Red Cross Hospital (LL2017-19), and Zhongshan City People's Hospital (ZSKY2012(02)). This study is registered at Clinical Trials.gov (NCT00941538, NCT02501980) and Chinese Clinical Trial Registry (ChiCTR2000028776, ChiCTR2100041628). All participants signed the informed consent upon recruitment.

Data availability

Data of this study is available upon reasonable request from the corresponding authors.

Author contributions

M.-S.Z., S.-M.C., and X.-W.K. were responsible for conceptualization. M.-S.Z., S.-M.C., Y.-L.C., M.-F.J., and X.-W.K. were responsible for methodology. X.-W.K., G.-L.B., Y.-H.H., Y.-F.K., Y.-C.L., X.Y., B.-H.W., Z.-Q.L., X.-C.C., S.-H.X., D.-F.L., T.L., S.-M.Y., R.-K.H., N.H., Q.-Y.W., Y.L., and A.Z. were responsible for investigation. X.-W.K., and H.C. were responsible for visualization. M.-S.Z., S.-M.C., Y.-L.C., and M.-F.J. were responsible for funding acquisition. M.-S.Z., S.-M.C., Y.-L.C., and M.-F.J. supervised the study. X.-W.K. wrote the original draft. M.-S.Z., S.-M.C., Y.-L.C., M.-F.J., Z.-W.L., Q.Z., X.-M.H., and W.-M.Y. reviewed and revised the manuscript. The order of co-first authors was assigned by the contribution to this study.

Acknowledgments

This study was supported by Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501003), National Natural Science Foundation of China (82030046, 82073625, 81860601, 82373655), Local Innovative and Research Teams Project of Guangdong Pearl River Talents

Program (2019BT02Y198), and Central Financial Transfer Payment Projects of the Chinese Government, Cancer Research Grant of Zhongshan City.

List of supplementary materials

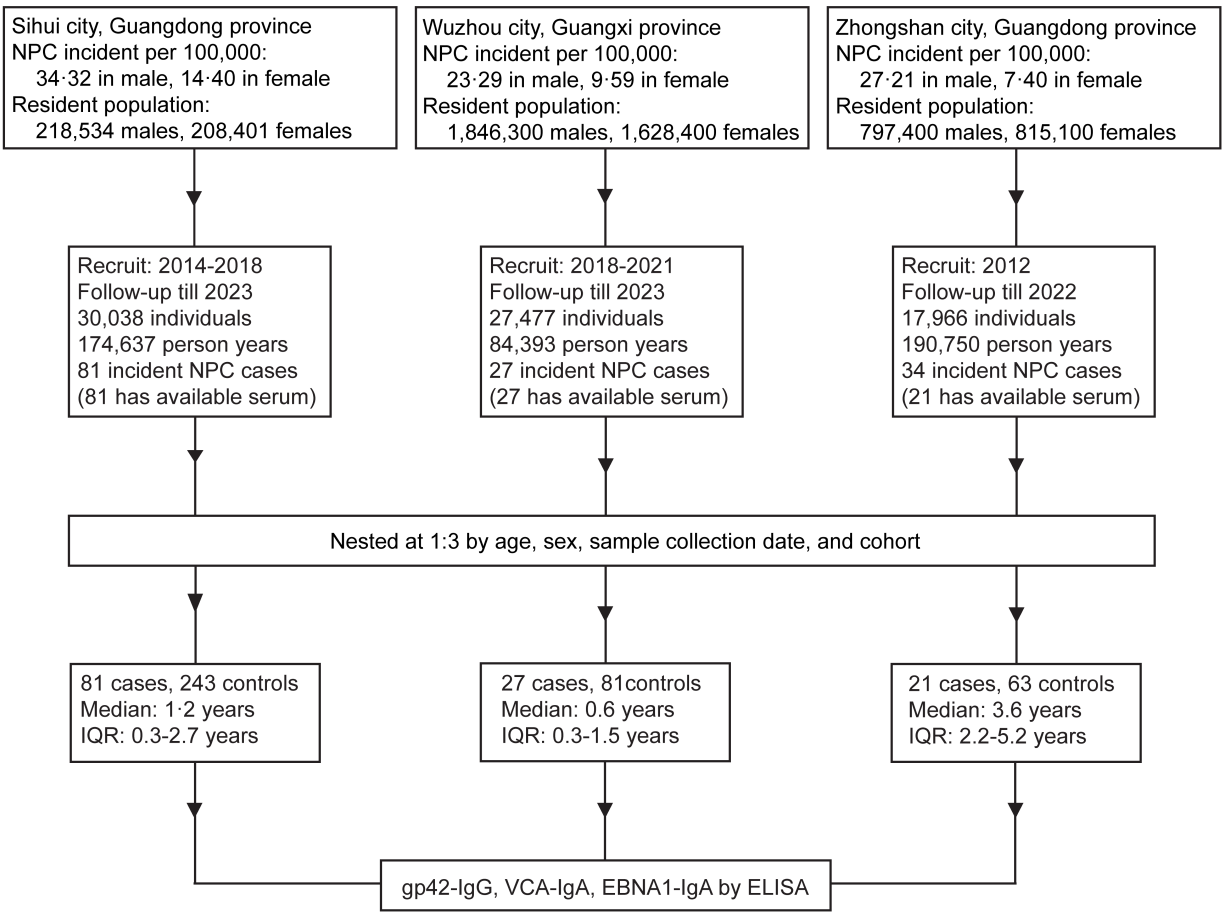
Supplemental Figure 1 to 9

Supplemental Table 1 and 2

References

1. Chang ET, Ye W, Zeng YX, and Adami HO. The Evolving Epidemiology of Nasopharyngeal Carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2021;30(6):1035-47.
2. National Cancer Center *2019 China Cancer Registry Annual Report.* PEOPLE'S MEDICAL PUBLISHING HOUSE Co., LTD; 2021.
3. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, and Ma J. Nasopharyngeal carcinoma. *Lancet.* 2019;394(10192):64-80.
4. Hutt-Fletcher LM. EBV glycoproteins: where are we now? *Future Virol.* 2015;10(10):1155-62.
5. Haque T, Johannessen I, Dombagoda D, Sengupta C, Burns DM, Bird P, et al. A mouse monoclonal antibody against Epstein-Barr virus envelope glycoprotein 350 prevents infection both in vitro and in vivo. *J Infect Dis.* 2006;194(5):584-7.
6. Zhang X, Hong J, Zhong L, Wu Q, Zhang S, Zhu Q, et al. Protective anti-gB neutralizing antibodies targeting two vulnerable sites for EBV-cell membrane fusion. *Proc Natl Acad Sci U S A.* 2022;119(32):e2202371119.
7. Zhu QY, Shan S, Yu J, Peng SY, Sun C, Zuo Y, et al. A potent and protective human neutralizing antibody targeting a novel vulnerable site of Epstein-Barr virus. *Nat Commun.* 2021;12(1):6624.
8. Snijder J, Ortego MS, Weidle C, Stuart AB, Gray MD, McElrath MJ, et al. An Antibody Targeting the Fusion Machinery Neutralizes Dual-Tropic Infection and Defines a Site of Vulnerability on Epstein-Barr Virus. *Immunity.* 2018;48(4):799-811 e9.
9. Cohen JI. Epstein-barr virus vaccines. *Clin Transl Immunology.* 2015;4(1):e32.
10. Coghill AE, Bu W, Nguyen H, Hsu WL, Yu KJ, Lou PJ, et al. High Levels of Antibody that Neutralize B-cell Infection of Epstein-Barr Virus and that Bind EBV gp350 Are Associated with a Lower Risk of Nasopharyngeal Carcinoma. *Clin Cancer Res.* 2016;22(14):3451-7.
11. Coghill AE, Bu W, Hsu WL, Nguyen H, Yu KJ, Chien YC, et al. Evaluation of Total and IgA-Specific Antibody Targeting Epstein-Barr Virus Glycoprotein 350 and Nasopharyngeal Carcinoma Risk. *J Infect Dis.* 2018;218(6):886-91.
12. Coghill AE, McGuire A, Sinha S, Homad L, Sinha I, Sholukh A, et al. Epstein-Barr Virus Glycoprotein Antibody Titers and Risk of Nasopharyngeal Carcinoma. *Open Forum Infect Dis.* 2022;9(12):ofac635.
13. Zhu QY, Kong XW, Sun C, Xie SH, Hildesheim A, Cao SM, et al. Association between Antibody Responses to Epstein-Barr Virus Glycoproteins, Neutralization of Infectivity, and the Risk of Nasopharyngeal Carcinoma. *mSphere.* 2020;5(6).
14. Hong J, Wei D, Wu Q, Zhong L, Chen K, Huang Y, et al. Antibody Generation and Immunogenicity Analysis of EBV gp42 N-Terminal Region. *Viruses.* 2021;13(12).
15. Jin S, Li R, Chen MY, Yu C, Tang LQ, Liu YM, et al. Single-cell transcriptomic analysis defines the interplay between tumor cells, viral infection, and the microenvironment in nasopharyngeal carcinoma. *Cell Res.* 2020;30(11):950-65.

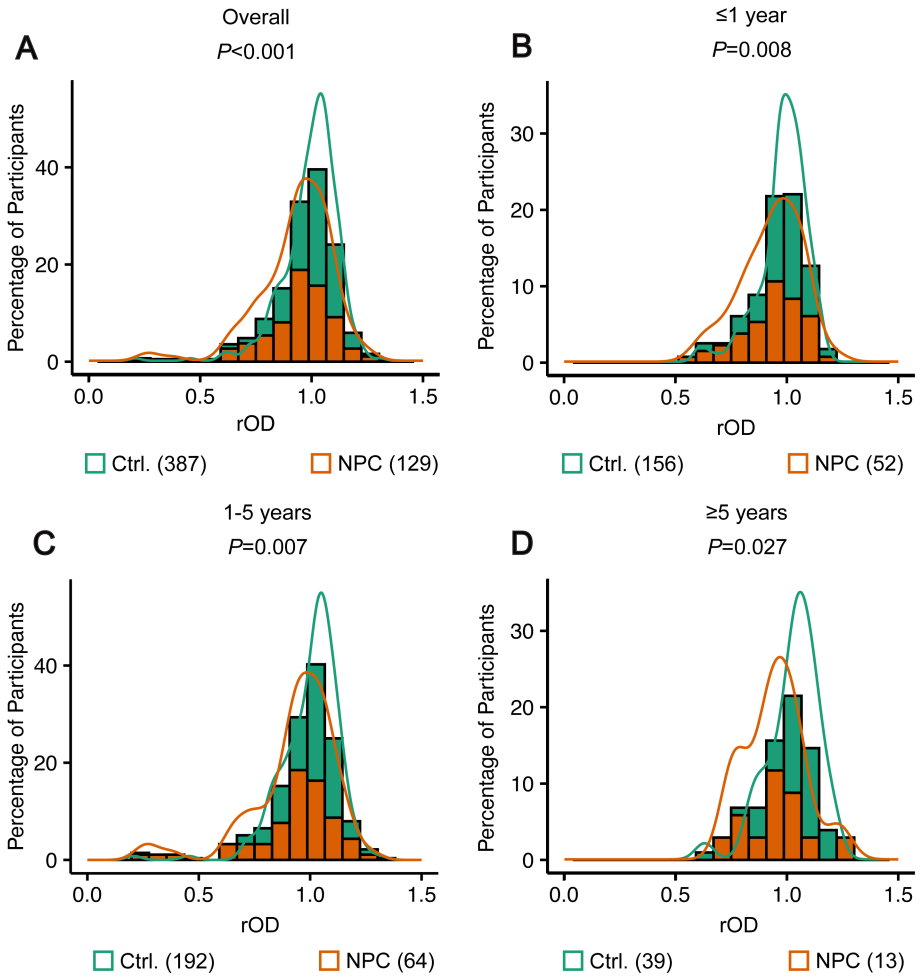
16. Tan LP, Tan GW, Sivanesan VM, Goh SL, Ng XJ, Lim CS, et al. Systematic comparison of plasma EBV DNA, anti-EBV antibodies and miRNA levels for early detection and prognosis of nasopharyngeal carcinoma. *Int J Cancer.* 2020;146(8):2336-47.
17. Zhao GX, Fang XY, Bu GL, Chen SJ, Sun C, Li T, et al. Potent human monoclonal antibodies targeting Epstein-Barr virus gp42 reveal vulnerable sites for virus infection. *Cell Rep Med.* 2024;5(5):101573.
18. Borza CM, and Hutt-Fletcher LM. Alternate replication in B cells and epithelial cells switches tropism of Epstein-Barr virus. *Nat Med.* 2002;8(6):594-9.
19. Woellmer A, and Hammerschmidt W. Epstein-Barr virus and host cell methylation: regulation of latency, replication and virus reactivation. *Curr Opin Virol.* 2013;3(3):260-5.
20. Chen GH, Liu Z, Yu KJ, Coghill AE, Chen XX, Xie SH, et al. Utility of Epstein-Barr Virus DNA in Nasopharynx Swabs as a Reflex Test to Triage Seropositive Individuals in Nasopharyngeal Carcinoma Screening Programs. *Clin Chem.* 2022;68(7):953-62.
21. Zheng XH, Wang RZ, Li XZ, Zhou T, Zhang JB, Zhang PF, et al. Detection of methylation status of Epstein-Barr virus DNA C promoter in the diagnosis of nasopharyngeal carcinoma. *Cancer Sci.* 2020;111(2):592-600.
22. Jiang R, Scott RS, and Hutt-Fletcher LM. Epstein-Barr virus shed in saliva is high in B-cell-tropic glycoprotein gp42. *J Virol.* 2006;80(14):7281-3.
23. Hadinoto V, Shapiro M, Sun CC, and Thorley-Lawson DA. The dynamics of EBV shedding implicate a central role for epithelial cells in amplifying viral output. *PLoS Pathog.* 2009;5(7):e1000496.
24. Frangou P, Buettner M, and Niedobitek G. Epstein-Barr virus (EBV) infection in epithelial cells in vivo: rare detection of EBV replication in tongue mucosa but not in salivary glands. *J Infect Dis.* 2005;191(2):238-42.
25. Wolf H, Haus M, and Wilmes E. Persistence of Epstein-Barr virus in the parotid gland. *J Virol.* 1984;51(3):795-8.
26. Liu Z, Ji MF, Huang QH, Fang F, Liu Q, Jia WH, et al. Two Epstein-Barr virus-related serologic antibody tests in nasopharyngeal carcinoma screening: results from the initial phase of a cluster randomized controlled trial in Southern China. *Am J Epidemiol.* 2013;177(3):242-50.
27. Tholen DW, Kroll M, and Astles JR. *Evaluation of the linearity of quantitative measurement procedures a statistical approach; approved guideline.* Wayne, Pa., U.S.A: CLSI; 2003.
28. Andreasson U, Perret-Liaudet A, van Waalwijk van Doorn LJ, Blennow K, Chiasserini D, Engelborghs S, et al. A Practical Guide to Immunoassay Method Validation. *Front Neurol.* 2015;6:179.



464

465 **Figure 1. A Flow chart of study design.** A total 75,481 individuals from three independent prospective
466 cohorts established in three NPC epidemic regions (Sihui city, Guangdong province; Wuzhou city,
467 Guangxi province; Zhongshan city, Guangdong province) in southern China were recruited 2014 to 2018,
468 2018 to 2021, and 2012 years. A total of 129 serum-available incident NPC cases were identified until
469 2023 and each case matched to three controls from individual cohort by age, sex, sample collection dates.
470 Titers of gp42-IgG, VCA-IgA, and EBNA1-IgA were measured by ELISA.

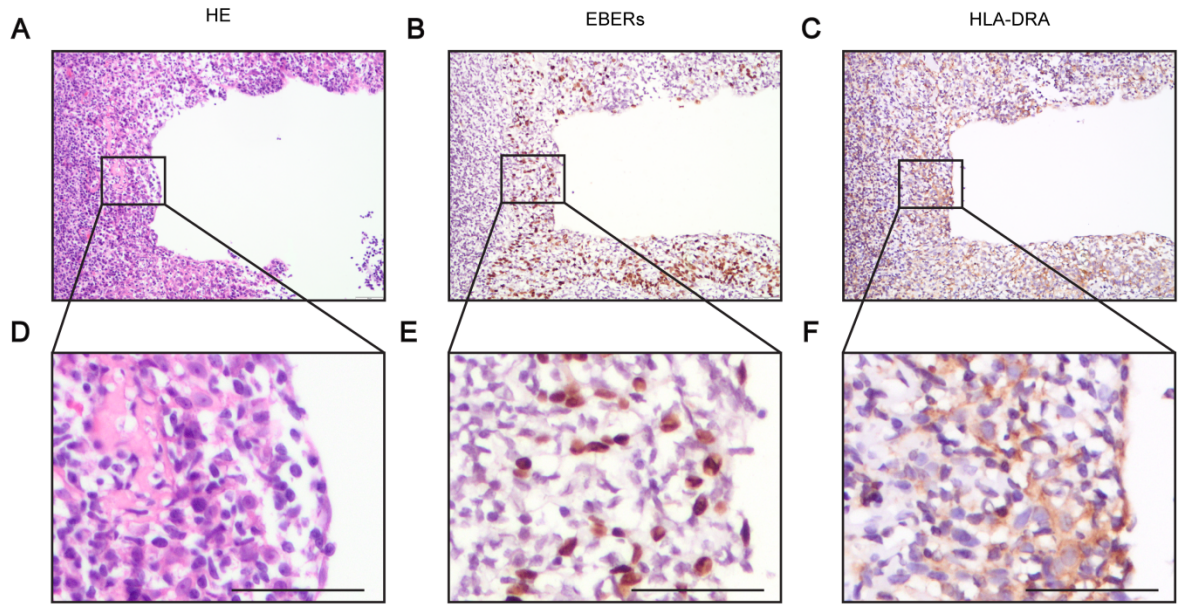
471



472

473 **Figure 2. Analysis of gp42-IgG levels in the nested case-control study.** Distributions of gp42-IgG
474 levels in (**A**) the overall cohort, (**B**) the cohort with follow-up duration ≤1 year, (**C**) the cohort with
475 follow-up duration 1-5 years, (**D**) the cohort with follow-up duration ≥5 years. rOD is a standardized
476 OD450-OD630 value, which was consistent across testing batches (Refer to Methods for details). The
477 lines at the top of each histogram represent kernel density estimations. Wilcoxon tests were used to
478 calculate *P* values, which are shown under the subtitles. Numbers in the legends refer to the numbers of
479 participants in each group.

480



481

482 **Figure 3. HLA-II expression in nasopharyngeal atypical dysplasia.** (**A and D**) HE staining. (**B and E**)
483 In situ hybridization to detect EBERs. (**C and F**) Immunohistochemistry to detect HLA-DRA. Shown
484 was a representative out of 27 samples. **A-C** were shot at 20×, **D-F** were zoomed in from the black square
485 region, and scale bars at lower right represent 50 μm.

486

Table 1. Demographic characteristics of individuals in the nested case-control study.

	Follow-up periods before NPC onset							
	Overall		≤1 year		1-5 years		≥5 years	
Characteristic	Control	Case	Control	Case	Control	Case	Control	Case
Sex								
Male	294 (76·0%)	98 (76·0%)	123 (78.8%)	41 (78.8%)	132 (68.8%)	44 (68.8%)	39 (100.0%)	13 (100.0%)
Female	93 (24·0%)	31 (24·0%)	33 (21.2%)	11 (21.2%)	60 (31.2%)	20 (31.2%)	0 (0.0%)	0 (0.0%)
Age, years								
<40	36 (9·3%)	12 (9·3%)	15 (9.6%)	5 (9.6%)	15 (7.8%)	5 (7.8%)	6 (15.4%)	2 (15.4%)
40-50	108 (27·9%)	36 (27·9%)	45 (28.8%)	15 (28.8%)	57 (29.7%)	19 (29.7%)	6 (15.4%)	2 (15.4%)
50-60	171 (44·2%)	57 (44·2%)	75 (48.1%)	25 (48.1%)	75 (39.1%)	25 (39.1%)	21 (53.8%)	7 (53.8%)
≥60	72 (18·6%)	24 (18·6%)	21 (13.5%)	7 (13.5%)	45 (23.4%)	15 (23.4%)	6 (15.4%)	2 (15.4%)
Location								
Sihui	243 (62·8%)	81 (62·8%)	105 (67.3%)	35 (67.3%)	120 (62.5%)	40 (62.5%)	18 (46.2%)	6 (46.2%)
Wuzhou	81 (20·9%)	27 (20·9%)	45 (28.8%)	15 (28.8%)	36 (18.8%)	12 (18.8%)	0 (0.0%)	0 (0.0%)
Zhongshan	63 (16·3%)	21 (16·3%)	6 (3.8%)	2 (3.8%)	36 (18.8%)	12 (18.8%)	21 (53.8%)	7 (53.8%)
NPC risk score								
Low	305 (78·8%)	21 (16·3%)	123 (78.8%)	5 (9.6%)	148 (77.1%)	11 (17.2%)	34 (87.2%)	5 (38.5%)
Medium	59 (15·2%)	23 (17·8%)	24 (15.4%)	2 (3.8%)	32 (16.7%)	18 (28.1%)	3 (7.7%)	3 (23.1%)
High	23 (5·9%)	85 (65·9%)	9 (5.8%)	45 (86.5%)	12 (6.2%)	35 (54.7%)	2 (5.1%)	5 (38.5%)

NPC, nasopharyngeal carcinoma; NPC risk score (P value) of each participant is determined by a risk prediction algorithm (Logit P=-3·934+2·203×VCA-IgA+4·797×EBNA1-IgA) and the risks of NPC are stratified as: low risk, P<0·65; medium risk, 0·65≤P<0·98; high risk, P≥0·98.

Table 2. Conditional logistic regression analysis reveals nasopharyngeal carcinoma risk associated with gp42-IgG level across different follow-up periods.

Follow-up periods	Cases (%)	Controls (%)	ORs (95% CIs)	*P* values
Overall				
Quartile 1 (<0.94)	56 (43.4)	97 (25.1)	1	
Quartile 2 (0.94-1.02)	28 (21.7)	96 (24.8)	0.43 (0.24-0.76)	0.004
Quartile 3 (1.02-1.07)	24 (18.6)	97 (25.1)	0.36 (0.20-0.65)	0.001
Quartile 4 (≥1.07)	21 (16.3)	97 (25.1)	0.29 (0.15-0.55)	<0.001
Per unit increase			0.66 (0.54-0.81)	<0.001
≤1 year				
Quartile 1 (<0.94)	24 (46.2)	37 (23.7)	1	
Quartile 2 (0.94-1.02)	11 (21.2)	56 (35.9)	0.25 (0.11-0.61)	0.002
Quartile 3 (1.02-1.07)	9 (17.3)	35 (22.4)	0.30 (0.11-0.81)	0.018
Quartile 4 (≥1.07)	8 (15.4)	28 (17.9)	0.33 (0.11-0.99)	0.048
Per unit increase			0.67 (0.47-0.96)	0.030
1-5 years				
Quartile 1 (<0.94)	25 (39.1)	52 (27.1)	1	
Quartile 2 (0.94-1.02)	15 (23.4)	33 (17.2)	0.84 (0.36-1.98)	0.696
Quartile 3 (1.02-1.07)	13 (20.3)	53 (27.6)	0.46 (0.21-1.03)	0.058
Quartile 4 (≥1.07)	11 (17.2)	54 (28.1)	0.33 (0.13-0.84)	0.020
Per unit increase			0.68 (0.51-0.90)	0.007
≥5 years				
Quartile 1 (<0.94)	7 (53.8)	8 (20.5)	1	
Quartile 2 (0.94-1.02)	2 (15.4)	7 (17.9)	0.32 (0.04-2.40)	0.266
Quartile 3 (1.02-1.07)	2 (15.4)	9 (23.1)	0.21 (0.03-1.69)	0.141
Quartile 4 (≥1.07)	2 (15.4)	15 (38.5)	0.05 (0.00-0.89)	0.042
Per unit increase			0.40 (0.18-0.90)	0.027

The quartile cutoffs based on the overall control groups, allowing for direct comparisons across different stratifications. In the trend analysis, the values of discretized gp42-IgG were treated as continuous variable.