RESEARCH ARTICLE

WILEY

Association between HLA alleles and Epstein–Barr virus Zta-IgA serological status in healthy males from southern China

Lei-Lei Yuan[1,2] | Chang-Mi Deng[2] | Wen-Qiong Xue[2] | Yong-Qiao He[2] | Tong-Min Wang[2] | Jiang-Bo Zhang[2] | Da-Wei Yang[1,2] | Ting Zhou[2] | Zi-Yi Wu[2] | Ying Liao[2] | Mei-Qi Zheng[2] | Dan-Hua Li[2] | Lian-Jing Cao[2] | Yi-Jing Jia[1,2] | Wen-Li Zhang[2] | Ruo-Wen Xiao[2] | Lu-Ting Luo[1,2] | Xia-Ting Tong[1,2] | Yan-Xia Wu[2] | Jing-Wen Huang[1,2] | Wei-Hua Jia[1,2]

[1]School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]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, Guangzhou, China

Correspondence
Wei-Hua Jia, School of Public Health, Sun Yat-Sen University, Guangzhou 510060, China.
Email: jiawh@sysucc.org.cn

Abstract

Background: Nasopharyngeal carcinoma (NPC), an Epstein–Barr virus (EBV) associated cancer, exhibits an extremely high incidence in southern Chinese. Given that human leukocyte antigen (HLA) plays critical roles in antigen presentation and relates to NPC susceptibility, it is speculated that certain HLA variants may affect EBV reactivation, which is a key pathogenic factor of NPC. Therefore, we attempted to identify HLA alleles associated with the indicator of EBV reactivation, Zta-IgA, in healthy males from NPC endemic area.

Methods: HLA alleles of 1078 healthy males in southern China from the 21-RCCP study were imputed using genome-wide single nucleotide polymorphism data. EBV Zta-IgA in blood samples were measured using an enzyme-linked immunosorbent assay. Multiple logistic regression analysis was used to evaluate the effect of HLA allele on Zta-IgA serological status and its potential joint association with smoking. The binding affinity for Zta-peptide was predicted using NetMHCIIpan 4.0.

Results: HLA-DRB1*09:01 was found to be associated with a higher risk of Zta-IgA seropositivity (odds ratio = 1.80, 95% confidence interval = 1.32–2.45; $p = 1.82 \times 10^{-4}$). Compared with non-smokers without HLA-DRB1*09:01, the effect size increased to 2.19- and 3.70-fold for the light and heavy smokers carrying HLA-DRB1*09:01, respectively. Furthermore, HLA-DRB1*09:01 showed a stronger binding affinity to Zta peptide than other HLA-DRB1 alleles.

Conclusions: Our study highlighted the pivotal role of genetic HLA variants in EBV reactivation and the etiology of NPC. Smokers with HLA-DRB1*09:01 have a significantly higher risk of being Zta-IgA seropositive, which indicates the necessity of smoking cessation in certain high-risk populations and also provide clues for further research on the etiology of NPC.

Lei-Lei Yuan, Chang-Mi Deng and Wen-Qiong Xue contributed equally to this work.

----

1 | INTRODUCTION

Nasopharyngeal carcinoma (NPC) exhibits an extremely high ethnic and geographic disparity, with southern Chinese, especially in Guangdong, Guangxi and Hong Kong, showing an almost 50-fold higher incidence than that in northern China or Caucasians. Males have two to three times the risk of NPC compared to females.[1–3] The individual genetic susceptibility may convey higher risk of NPC through its main effect,[4,5] and interactions with EBV infection and certain environmental risk factors.[6] Much effort has been made to investigate the genetic susceptibility of NPC. Previous studies, including our genome-wide association study (GWAS) in Guangdong,[7,8] have found consistently strong association signals in human leukocyte antigen (HLA) region located on chromosome 6. In addition, it is reported that both HLA class I and class II alleles (eg. HLA-A2, -A11, -B13, -B46, -B58, -C03, -C04, -C07, -DRB1*03, -DRB1*09 and -DRB1*12) were associated with NPC in a case–control study design.[9–14] Considering these genes are mainly responsible for antigen processing and presentation, it is speculated that some high-risk HLA alleles may have an impact on host immune surveillance against EBV, resulting in the loss of virus–host homeostasis and an increased level of EBV reactivation.[15] Exploring the association between HLA alleles and potential precancerous indicators of NPC in disease-free subjects from NPC endemic areas may help uncover actual causal factors.

After primary infection, which usually occurred in early human life, EBV stays dormant for most of the time,[15–17] whereas it can be reactivated by certain factors such as tobacco smoking.[18] Antibodies against lytic EBV proteins including capsid antigen (VCA), BZLF1 transcription activator protein (Zta) and early antigen (EA) have shown great potential to stratify NPC cases and controls (area under the curve = 0.90–0.93).[19,20] Individuals with elevated levels of immunoglobulin (Ig)A antibodies against lytic EBV proteins have a 20- to 30-fold higher risk of NPC.[21] Longitudinal cohort studies have revealed that the elevated IgA level precedes the occurrence of NPC by 1–15 years.[21,22] Our former study conducted in northern and southern China revealed that the seropositive rate of EBV Zta-IgA in males from NPC endemic areas (21-RCCP and North populations in Guangdong) was 2.6 times of that in males from non-endemic areas (Shanxi), after adjusting for age and education.[23] Thus, EBV reactivation probably plays a pivotal role in the initiation and early stage of NPC. The most important protein triggering EBV latent-lytic switch is Zta encoded by BZLF1,[15,17] which is transcribed during immediate-early stage of the lytic cycle. Zta directly interacts with histone acetylating complexes, and initiates a cascade of early-lytic gene transcription of EBV. It also plays a critical role in lytic EBV DNA replication.[24] Therefore, it is noteworthy to identify the HLA alleles associated with Zta-IgA, as an indicator for EBV reactivation, in NPC endemic populations.

We conducted the present study focusing on the associations between HLA alleles and Zta-IgA in our male population with a higher level of EBV reactivation in Guangdong (21-RCCP).[23] Four-digit HLA alleles were imputed using high-dense single nucleotide polymorphism (SNP) array data from our former GWAS study.[8,25] We investigated the potential associations between all available HLA alleles and serum Zta-IgA. Because smoking could induce EBV reactivation, we also examined the possible synergic effect of HLA alleles and smoking on Zta-IgA.

2 | MATERIALS AND METHODS

2.1 | Study population and data collection

In total, 1498 healthy males were recruited from physical examination centers in 21 municipalities of Guangdong Province, during 2005–2007 (21-RCCP). Details about the design, methods and subjects have been reported previously.[8,23,26] All participants were Chinese Cantonese, lived in Guangdong province at the time of the study, and had no history of malignancy or immunological diseases. Each subject was interviewed face-to-face by well-trained interviewers to obtain demographic characteristics (age, education level), family history of cancer, lifestyle behaviors (smoking and alcohol consumption). The structured questionnaire was mainly designed based on the questionnaires issued by the University of Arizona Cancer Center (https://cancercenter.arizona.edu/researchers/shared-resources/behavioral-measurement-and-interventions). For cigarette smoking, participants were divided into ever smokers and never smokers. Ever smokers were defined as having smoked at least 100 cigarettes during their lifetime. Detailed information on smoking was collected, including the amount of cigarettes consumption per day and the duration of exposure. Pack-year was used as the cumulative measurement in the analysis. Participants consuming < 20 and ≥ 20 pack-years were assigned as light and heavy smokers, respectively. The study was reviewed and approved by the Human Ethics Approval Committee at Sun Yat-sen University Cancer Center. Informed consent was obtained from all participants.

2.2 | Serologic test for EBV Zta-IgA

As stated previously,[23,26] a 5–10-ml blood sample was collected from each subject. Zta-IgA serostatus was measured with a commercially available enzyme-linked immunosorbent assay (ELISA) kit (Zhongshan Bio-technology Co., Ltd, Zhongshan city, China). Zta was produced with purified recombinant peptides specified by BZLF1 (36 + 38-kDa fine doublet). All samples were detected by the same technicians in the same laboratory of the Sun Yat-Sen University Cancer Center to limit potential biases. The seropositive/seronegative status of Zta-IgA

was determined based on the ELISA optical density value in accordance with the manufacturer's instructions. To evaluate the reproducibility for the assay, we randomly selected 46 samples to conduct a test–retest assay for Zta-IgA and calculated intraclass correlation coefficients (ICCs) and the agreement percentage. The ICCs of Zta-IgA were 0.70 and the agreement percentage of Zta-IgA was 82.6%.

2.3 | Genotyping and imputation of HLA alleles

Of all the 1498 samples, genetic data of 1080 samples were available and retrieved. HLA allele genotyping was performed on an Infinium Global Screening Array and a Human610-Quad BeadChip (Illumina, San Diego, CA, USA). Following the quality control and imputation procedures described previously,[13,25] samples were removed with genotyping call rates < 95%, duplicates or relatives, and population outliers detected by principal component analysis. Two samples were removed in the process. SNPs were removed with call rates < 95%, minor allele frequency (MAF) < 0.01 and deviation from Hardy–Weinberg equilibrium in controls ($p < 10^{-7}$). Quality control filtering were conducted using PLINK, version 1.09.[27] Imputation were performed within the extracted entire HLA region on chromosome 6 (29–34 Mb; NCBI Build 37) using SNP2HLA.[28] We imputed four-digit classical HLA alleles in three HLA class I genes (HLA-A, HLA-B and HLA-C) and five HLA class II genes (HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1 and HLA-DPB1) based on the HAN Chinese reference panel.[29] Post-imputation quality control was performed to removed variants with poor imputation quality (INFO score < 0.5, MAF < 0.005, call rate < 95%). We compared the frequencies of imputed alleles with the previously reported data in Southern Chinese population[30,31] and with that in the online HLA Frequency Database (http://allelefrequencies.net). The alleles with abnormal frequencies (discrepancy greater than 0.15) were removed. In the end, 1078 samples with 25 four-digit alleles were included in the subsequent analysis.

2.4 | Statistical analysis

Comparisons of demographic characteristics, family history, smoking status and alcohol consumption between Zta-IgA seropositive and seronegative groups were performed using a chi-squared test for categorical variables and a *t* test for continuous variables, respectively. Under additive assumption, the effect size (odd ratio [OR]) and *p* value of HLA dosage allele (0, 1 and 2) was calculated via logistic regression models adjusting for age and smoking status using PLINK, version 1.09. Because 25 four-digit HLA alleles were included in our study, the significance threshold was set as 2.00×10^{-3} ($0.05/25 = 2.00 \times 10^{-3}$) by applying Bonferroni correction. Logistic regression analyses were used to examine the effect of HLA alleles (Present versus Absent) and smoking (never, light and heavy group) on Zta-IgA seropositivity using SPSS, version 20.0 (IBM Corp., Armonk, NY, USA). *p* < 0.05 (two-sided) was considered tatistically significant.

2.5 | *In silico* binding affinity predictions using NetMHCIIpan 4.0

We employed NetMHCIIpan 4.0[32] to predict and to compare the binding affinity against protein Zta [Human gammaherpesvirus 4] (YP_401673.1) among HLA alleles. Based on the NN-align, a neural network-based approach combined with peptide sequence representation, the method is trained on an extensive dataset of over 500.000 measurements of Binding Affinity (BA) and Eluted Ligand mass spectrometry (EL), covering common HLA class II molecules. Using NetMHCIIpan 4.0, a sliding window of 1-mer overlapping 15-mer peptides based on the full-length amino acid sequences of the Zta antigen (245 amino acid) was screened to identify peptides with strong and weak binding affinities. Binding affinities for each peptide are provided as the score of EL (Score_EL). Compared with a set of random natural peptides, the percentile rank value of the score of EL (%Rank_EL) was used to reflect the strength of binding affinities. The threshold for strong binding peptides (%Rank_EL) is 2% and the threshold for weak binding peptides (%Rank_EL) is 10%. Peptides with %Rank_EL > 10% were filtered. The resulting sequence motifs were visualized using WebLogo.[33]

3 | RESULTS

3.1 | Basic characteristic of the study population

In total, 1078 healthy males were included in the analyses with a mean age of 49.79 years. The positive rate of Zta-IgA was 15.4% ($n = 166$). Almost 70% ($n = 750$) of these individuals had ever smoked. Other characteristics of the participants are presented in Table 1.

The EBV Zta-IgA seropositive individuals were significantly older than the seronegative individuals (52.79 ± 10.57 versus 49.24 ± 10.89, $p < 0.001$). A significantly higher proportion of ever-smokers were found in the Zta-IgA seropositive group, with a proportion of 77.7% compared to 68.1% in the Zta-IgA seronegative group ($p = 0.017$). No significant differences were observed for other characteristics, including education levels, alcohol consumption, family history of cancer and family history of NPC between the seropositive and seronegative groups (Table 1).

3.2 | HLA-DRB1*09:01 is related to EBV Zta-IgA serological status in healthy males from southern China

After genotyping and imputation, 25 four-digit HLA alleles of eight genes were ultimately included in the subsequent analysis. In the process of quality control, some common HLA-A, HLA-B and HLA-C alleles in the Southern Chinese population, such as HLA-A*02:07, HLA-A*11:01, HLA-B*46:01 and HLA-C*01:02, were filtered as a result of poor imputation quality or abnormal frequencies. Thus, in our study population, the frequencies of the most common alleles for each HLA gene were: for HLA-A: HLA-A*26:01 (0.019); for HLA-B: HLA-B*13:01 (0.079) and HLA-B*58:01 (0.079); for HLA-C:

TABLE 1 Characteristics of 1078 healthy males in an NPC endemic area

Variables	All, n (%)	EBV Zta-IgA seropositive, n (%)	EBV Zta-IgA seronegative, n (%)	p value[a]
Total	1,078	166 (15.4)	912 (84.6)	
Age (mean ± SD)	49.79 ± 10.91	52.79 ± 10.57	49.24 ± 10.89	< 0.001
Education level				0.092
Primary school or less	137 (12.8)	27 (16.5)	110 (12.1)	
High school	608 (56.6)	97 (59.1)	511 (56.2)	
University or more	329 (30.6)	40 (24.4)	289 (31.8)	
Smoking status				0.017
never	328 (30.4)	37 (22.3)	291 (31.9)	
ever	750 (69.6)	129 (77.7)	621 (68.1)	
Alcohol consumption				0.436
never	464 (43.0)	77 (46.4)	387 (42.4)	
<1 drink per week	395 (36.6)	61 (36.7)	334 (36.6)	
≥1 drink per week	219 (20.3)	28 (16.9)	191 (20.9)	
Family history of NPC				0.448
no	992 (95.8)	153 (94.4)	839 (96.1)	
yes	43 (4.2)	9 (5.6)	34 (3.9)	
Family history of cancer				0.656
no	759 (73.3)	116 (71.6)	643 (73.7)	
yes	276 (26.7)	46 (28.4)	230 (26.3)	

[a]A t test and chi-squared test were used to compare the mean of continuous factors and proportions of categorical factors between the EBV Zta-IgA seronegative group and the EBV Zta-IgA seropositive group.

HLA-C*07:02 (0.228); for HLA-DPA1: HLA-DPA1*02:01 (0.077); for HLA-DPB1: HLA-DPB1*21:01 (0.053); for HLA-DQA1: HLA-DQA1*06:01 (0.134); for HLA-DQB1: HLA-DQB1*05:02 (0.180); for HLA-DRB1: HLA-DRB1*09:01 (0.131).

When compared between Zta-IgA seropositive and negative males, HLA-DRB1*09:01 showed much higher frequencies in the seropositive group compared to the seronegative group (0.202 versus 0.125; $p = 1.58 \times 10^{-4}$) and HLA-DPB1*21:01 showed much lower frequencies in the seropositive group compared to the seronegative group (0.021 versus 0.054; $p = 0.011$).

After adjusting for age and smoking, we found that only HLA-DRB1*09:01 was significantly associated with a higher risk of Zta-IgA seropositivity (OR = 1.80, 95% confidence interval [CI] = 1.32–2.45; $p = 1.82 \times 10^{-4}$) when applying Bonferroni correction. Other 24 HLA alleles were not found to be associated with Zta-IgA serological status (Table 2).

3.3 | HLA-DRB1*09:01 enhanced the effect of smoking on Zta-IgA seropositivity

To investigate the potential joint association of smoking and HLA-DRB1*09:01 with Zta-IgA serological status, ORs were calculated for heavy and light-smokers (≥ 20 and < 20 pack-years) with or without HLA-DRB1*09:01 compared to the never-smokers without

HLA-DRB1*09:01, which was set as the reference group. For those without HLA-DRB1*09:01, heavy smokers (≥ 20 pack-years) had a moderate risk of Zta-IgA seropositivity (OR = 1.72, 95% CI = 1.02–2.90; $p = 0.041$), whereas the OR for light smokers (< 20 packy-years) was not significant (OR = 1.19, 95% CI = 0.67–2.09; $p = 0.554$) compared to the never-smokers without HLA-DRB1*09:01.

However, when smoking was combined with potential risk allele HLA-DRB1*09:01, the effect size increased almost two-fold, especially for heavy smokers. The heavy smokers carrying HLA-DRB1*09:01 had a 3.70-fold increased risk of Zta-IgA seropositivity (95% CI = 2.01–6.80; $p < 0.001$) compared to never-smokers without HLA-DRB1*09:01, and light smokers carrying HLA-DRB1*09:01 also had a 2.19-fold increased risk of Zta-IgA seropositivity (95% CI = 1.15–4.15; $p = 0.017$) (Figure 1).

3.4 | Zta protein showed a stronger binding affinity for HLA-DRB1*09:01 in silico predictions

Given the function of HLA-DR in EBV infection, we predicted and compared the binding affinity for Zta protein among three successfully imputed HLA-DR alleles (HLA-DRB1*09:01, HLA-DRB1*12:02 and HLA-DRB1*13:02). Table 3 shows the peptides predicted to have

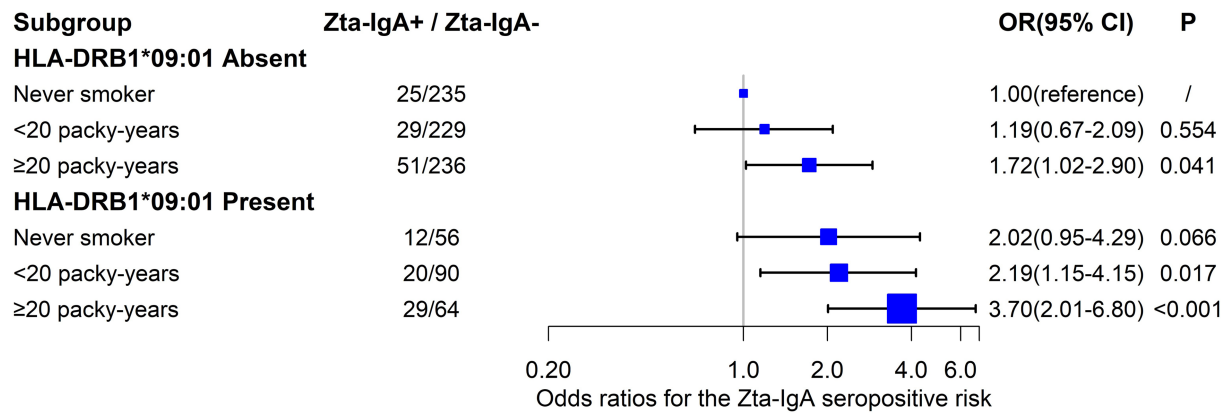**TABLE 2** Association between HLA alleles and EBV Zta-IgA serological status

HLA alleles	Allele frequencies			OR (95% CI)[a]	p value
	Total (n = 1078)	Zta-IgA seropositive (n = 166)	Zta-IgA seronegative (n = 912)		
HLA-A					
26:01	0.019	0.030	0.018	1.85 (0.88–3.88)	0.104
HLA-B					
13:01	0.079	0.066	0.076	0.88 (0.54–1.43)	0.610
58:01	0.079	0.054	0.081	0.68 (0.41–1.13)	0.134
07:05	0.022	0.012	0.026	0.48 (0.17–1.34)	0.158
13:02	0.021	0.009	0.025	0.38 (0.11–1.23)	0.106
HLA-C					
07:02	0.228	0.259	0.226	1.16 (0.88–1.54)	0.290
08:01	0.115	0.090	0.113	0.81 (0.54–1.22)	0.309
03:02	0.068	0.054	0.069	0.79 (0.48–1.31)	0.365
15:02	0.035	0.027	0.033	0.75 (0.36–1.53)	0.422
04:03	0.022	0.015	0.026	0.60 (0.23–1.55)	0.292
HLA-DPA1					
02:01	0.077	0.051	0.081	0.67 (0.40–1.11)	0.116
HLA-DPB1					
21:01	0.053	0.021	0.054	0.38 (0.17–0.82)	0.014
HLA-DQA1					
06:01	0.134	0.142	0.130	1.11 (0.78–1.56)	0.568
01:04	0.086	0.066	0.088	0.74 (0.47–1.18)	0.205
05:01	0.064	0.057	0.067	0.84 (0.51–1.39)	0.492
HLA-DQB1					
05:02	0.180	0.178	0.173	1.00 (0.74–1.36)	0.986
06:01	0.121	0.090	0.130	0.66 (0.44–1.00)	0.048
02:01	0.065	0.060	0.068	0.88 (0.54–1.43)	0.599
03:02	0.059	0.066	0.058	1.13 (0.70–1.81)	0.620
05:03	0.054	0.036	0.056	0.62 (0.34–1.14)	0.122
05:01	0.033	0.027	0.036	0.76 (0.37–1.53)	0.437
06:09	0.012	0.012	0.013	0.96 (0.33–2.84)	0.947
HLA-DRB1					
09:01	0.131	0.202	0.125	1.80 (1.32–2.45)	**1.82×10^{-4}**
12:02	0.130	0.130	0.127	1.03 (0.72–1.48)	0.869
13:02	0.019	0.018	0.020	1.00 (0.41–2.42)	0.996

[a]Odds ratios (ORs) and 95% confidence intervals (CIs) of HLA dosage alleles were calculated using logistic regression adjusting for two covariates (age, smoking status) under additive assumption.
Bold indicates significant p values after Bonferroni correction for multiple testing ($0.05/25 = 2.00 \times 10^{-3}$).

strong binding affinities for HLA-DR alleles, as well as how these 9-mer peptides cores were selected. Among the total 231 peptides of the Zta protein, there were twelve peptides of Zta protein exhibiting potential strong binding affinity (%Rank_EL ≤ 2%) for HLA-DRB1*09:01 in the prediction analysis. By contrast, only three peptides for HLA-DRB1*12:02 and one peptide for HLA-DRB1*13:02 showed potentially strong binding affinity. The complete results of the prediction analysis showing all Zta-peptides with any binding affinity (%Rank_EL ≤ 10%) among the total 231 peptides for HLA-DRB1 alleles is provided in the Supporting information (Table S1). Furthermore, WebLogo plots (see Supporting information, Figure S1) showed that the sequence motif for 9-mer core peptides that bind preferentially to HLA-DRB1*09:01, which was particularly rich in hydrophobic and polar AAs, presents a significantly different binding motif pattern compared to those of HLA-DRB1*12:02 or HLA-DRB1*13:02.

Subgroup	Zta-IgA+ / Zta-IgA-		OR(95% CI)	P
HLA-DRB1*09:01 Absent				
Never smoker	25/235		1.00(reference)	/
<20 packy-years	29/229		1.19(0.67-2.09)	0.554
≥20 packy-years	51/236		1.72(1.02-2.90)	0.041
HLA-DRB1*09:01 Present				
Never smoker	12/56		2.02(0.95-4.29)	0.066
<20 packy-years	20/90		2.19(1.15-4.15)	0.017
≥20 packy-years	29/64		3.70(2.01-6.80)	<0.001

Odds ratios for the Zta-IgA seropositive risk

FIGURE 1 The effect of smoking and HLA-DRB1*09:01 on EBV Zta-IgA serological status. We divided smoking categories by cumulative amount (never, < 20, ≥ 20 pack-years) for the analysis. Never smokers without HLA-DRB1*09:01 were defined as the reference group. Odds ratios (ORs) for the Zta-IgA seropositive group (Zta-IgA+) versus Zta-IgA seronegative group (Zta-IgA-) among different carrier status of HLA-DRB1*09:01 and smoking status were calculated using logistic regression models. Test was adjusted for age. The test was two-sided and $p < 0.05$ was considered statistically significanr. Squares indicate study-specific odds ratios; horizontal lines indicate study-specific confidence intervals (CIs); the solid vertical line indicates an odds ratio of 1.0

TABLE 3 The EBV Zta-peptides with predicted strong binding affinities for the HLA-DRB1 alleles

HLA allele	Pos[a]	Zta-peptide	Zta-peptide Core	Score _EL[b]	%rank _EL[c]	Bind Level[d]
HLA-DRB1*09:01						
	197	LQH**YREVAAAKS**SEN	YREVAAAKS	0.78	0.29	Strong
	196	LLQH**YREVAAAKS**SE	YREVAAAKS	0.73	0.47	Strong
	198	QH**YREVAAAKS**SEND	YREVAAAKS	0.70	0.60	Strong
	83	ENA**YQAYAAPQL**FPV	YQAYAAPQL	0.67	0.77	Strong
	82	PENA**YQAYAAPQL**FP	YQAYAAPQL	0.67	0.78	Strong
	60	QLTA**YHVSTAPTG**SW	YHVSTAPTG	0.67	0.80	Strong
	61	LTA**YHVSTAPTG**SWF	YHVSTAPTG	0.66	0.84	Strong
	59	GQLTA**YHVSTAPTG**S	YHVSTAPTG	0.66	0.87	Strong
	195	QLLQH**YREVAAAKS**S	YREVAAAKS	0.61	1.19	Strong
	62	TA**YHVSTAPTG**SWFS	YHVSTAPTG	0.55	1.67	Strong
	81	APENA**YQAYAAPQL**F	YQAYAAPQL	0.54	1.80	Strong
	84	NA**YQAYAAPQL**FPVS	YQAYAAPQL	0.53	1.86	Strong
HLA-DRB1*12:02						
	174	ELE**IKRYKNRVA**SRK	IKRYKNRVA	0.36	1.40	Strong
	190	RAK**FKQLLQHYR**EVA	FKQLLQHYR	0.35	1.48	Strong
	189	CRAK**FKQLLQHYR**EV	FKQLLQHYR	0.32	1.98	Strong
HLA-DRB1*13:02						
	174	ELE**IKRYKNRVA**SRK	IKRYKNRVA	0.49	1.17	Strong

[a]Pos, peptide position (starting from 0).
[b]Score_EL, raw eluted ligand likelihood prediction score.
[c]%Rank EL, rank of the predicted eluted ligand likelihood score compared to a set of random natural peptides.
[d]Binding level: Strong = strong binding. The threshold for strong binding peptides (%Rank_EL) is 2%. Peptides with %Rank_EL > 2% are not shown.
Bold indicates the 9-mer core of EBV Zta-peptide showing strong binding affinity for the corresponding HLA-DRB1 allele.

4 | DISCUSSION

Because EBV reactivation probably plays pivotal roles in NPC etiology, Zta-IgA, an indicator of EBV reactivation, may present a risk for NPC in high-endemic areas. In case–control studies in NPC endemic areas, such as Zhongshan City and Taiwan Province, consistent evidence has been observed indicating that increased Zta-IgA seropositive rates are associated with NPC.[20,34,35] Previous studies have provided evidence for genetic variants in the HLA region that may affect host immune surveillance against EBV being associated with

the risk of NPC.[8,23] Therefore, association analyses exploring potential HLA alleles affecting Zta-IgA in a high-risk population of NPC were more conducive for revealing the pathogenesis of NPC. The present study was conducted in males from Guangdong, China, where the incidences of NPC were the highest worldwide.[2] Consistently, the seropositive rate of Zta-IgA is 15.4% in our study subjects, and much higher than thre rate of 6.3% in NPC non-endemic areas.[23] We found an HLA class II allele, HLA-DRB1*09:01, was significantly associated with Zta-IgA seropositivity in Guangdong healthy males. In addition, the known association between smoking and Zta-IgA tended to be enhanced by HLA-DRB1*09:01 because smokers with HLA-DRB1*09:01 exhibit a 119–270% higher risk of Zta-IgA seropositivity compared to never-smokers without HLA-DRB1*09:01. Our further epitope prediction bioinformatics analysis showed that epitopes in HLA-DRB1*09:01 had a stronger binding affinity for peptides of Zta.

HLA-DRB1*09 has been reported as being related to NPC in Asian populations as a risk allele (OR = 1.37).[14,36] This is consistent with our results, indicating that HLA-DRB1*09 alleles may contribute to the risk of NPC via its positive impact on EBV reactivation. As a cell surface receptor binding to T cell receptor, HLA-DR is involved in presenting peptides from outside the cell to $CD4^+$ T lymphocytes (helper T cells), which in turn stimulate B cells to produce antibodies.[37] Furthermore, HLA-DR is also considered to be the co-factor of EBV infection of B lymphocytes, which assists EBV in infecting cells by binding with EBV-specific glycoprotein gp42.[38] Our epitope prediction analysis proposed that HLA-DRB1*09:01 might have more potential to bind Zta antigen and thereby elevate the efficiency of cell surface antigen presentation to T cell receptors. We considered that this would sequentially induce a higher level of IgA response. Further laboratory experiments are still required to confirm the HLA-DRB1*09:01 restricted EBV epitopes.

We did not observe any significantly associations between other NPC-associated HLA alleles such as HLA-A*26:01 and Zta-IgA serological status in our study. This may partly be a result of their low frequency. Nevertheless, two genome-wide association studies focusing on seroreactivity to EBV consistently found the strongest signals in the HLA class II region, especially in the HLA-DRB region.[37,39] In addition, previous studies focusing on other EBV-associated diseases such as Hodgkin's lymphoma and multiple sclerosis in Europe and America provided similar results, with anti-EBV IgG being significantly related to HLA-DRB1 alleles.[37,40,41] However, Zta-IgA, representing EBV reactivation and risk of NPC, had seldom been studied. Determination of the Zta-IgA related HLA allele may help to identify high-risk groups and provide clues for the etiology of NPC in Southern China.

In our results, ever-smokers were more likely to be Zta-IgA seropositive than never smokers, which was consistent with previous reports.[23] Experimental studies also confirmed that cigarette smoke extract could promote EBV reactivation and replication, possibly as a result of its function in inflammation and the cell-mediated immune response.[26,42–44] Although there were no statistically significant interactions between HLA-DRB1*09:01 and smoking on Zta-IgA seropositivity, our study provided evidence for a joint association of HLA-DRB1*09:01 and cigarette smoking with Zta-IgA seropositivity.

In comparison with never-smokers without HLA-DRB1*09:01, the effect of smoking on Zta-IgA serological status increased when carrying HLA-DRB1*09:01, especially for the heavy smokers (OR = 3.70). These findings suggest that some HLA genetic variations might aggravate the effect of cigarette smoking on Zta-IgA seropositivity. Smoking affects both innate and adaptive immune responses by aggravating pathological immune responses or weakening defense immunity.[45] Interactions between HLA alleles and smoking have been observed in EBV-associated diseases such as multiple sclerosis.[46,47] It has been found that smoking triggers HLA-DR shared epitope-specific immune reactions in rheumatoid arthritis, indicating that the effect of smoking in immune response may be regulated by HLA alleles.[48,49] Further functional research is still warranted aiming to uncover the underlying biological mechanism of the synergistic effect of HLA alleles and smoking on EBV activation.

Because HLA alleles in the present study were genotyped from GeneChips (Illumina, San Diego, CA, USA) data, we were hardly able to evaluate extremely rare variants. In addition, the imputation was performed on two GeneChips using the HAN reference panel,[29] resulting in some common alleles in the Southern Chinese population, such as HLA-A*11:01, HLA-A*02:07, HLA-B*46:01 and HLA-C*01:02, having to be excluded after quality control. However, the imputation was conducted based on the most common reference panel of the HAN population and under strict quality control. Thus, the four-digit HLA alleles imputed were reliable in our study subjects. Moreover, multicenter studies with a large sample size are still warranted to explore the effect of HLA alleles, especially those rare HLA alleles with respect to Zta-IgA seropositivity in NPC endemic areas.

In summary, we found that an NPC risk allele, HLA-DRB1*09:01, is an important genetic factor for EBV reactivation and may play a role in NPC development. This risk allele could exaggerate the effect of smoking with respect to EBV reactivation, as represented by an increased seropositive rate of Zta-IgA. The strong binding affinity to Zta peptide confirmed the specificity of HLA-DRB1*09:01 in the process of Zta antigen presentation. Our study highlighted the important roles of HLA variations regarding the etiology of NPC in southern China. From a public health perspective, determination of this high-risk HLA allele may help to identify individuals at risk of NPC. Future approaches for inhibiting EBV reactivation may serve as a method of preventing NPC in individuals with a higher potential to reactivate EBV. The synergy of HLA-DRB1*09:01 and smoking also indicates the necessity of smoking cessation in certain high-risk groups of NPC and provides clues for further research on the etiology of NPC.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

W-HJ was responsible for study conceptualization. L-LY, C-MD, W-QX, Y-QH and T-MW were responsible for the methodology. L-LY and C-MD were responsible for software. L-LY, C-MD, Y-JJ, W-QX, Y-QH and T-MW were responsible for study validation. L-LY and C-MD were responsible for the formal analysis. Y-JJ, TZ, Z-YW, YL, M-QZ, L-JC, W-LZ, D-HL, R-WX, L-TL, X-TT, Y-XW and J-WH were responsible for the investigation. Y-QH and T-MW were responsible for data curation. L-LY, C-MD and W-QX were responsible for writing – original draft preparation. W-HJ, W-QX, Y-QH, T-MW, and D-WY were responsible for writing – review and editing. L-LY was responsible for visualization. J-BZ was responsible for supervision. W-QX and Y-QH were responsible for project administration. All authors have read and approved the final version of the manuscript sumbitted for publication.

DATA AVAILABILITY STATEMENT

All of the data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Wei-Hua Jia https://orcid.org/0000-0002-0528-8715

REFERENCES

1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. 2016;66(2):115-132. https://doi.org/10.3322/caac.21338
2. Tang LL, Chen WQ, Xue WQ, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett*. 2016;374(1):22-30. https://doi.org/10.1016/j.canlet.2016.01.040
3. Young LS, Rickinson AB. Epstein-Barr virus: 40 years on. *Nat Rev Cancer*. 2004;4(10):757-768. https://doi.org/10.1038/nrc1452
4. Bei JX, Jia WH, Zeng YX. Familial and large-scale case-control studies identify genes associated with nasopharyngeal carcinoma. *Semin Cancer Biol*. 2012;22(2):96-106. https://doi.org/10.1016/j.semcancer.2012.01.012
5. Hildesheim A, Wang CP. Genetic predisposition factors and nasopharyngeal carcinoma risk: a review of epidemiological association studies, 2000-2011: Rosetta Stone for NPC: genetics, viral infection, and other environmental factors. *Semin Cancer Biol*. 2012;22(2):107-116. https://doi.org/10.1016/j.semcancer.2012.01.007
6. Chatterjee K, Das P, Chattopadhyay NR, Mal S, Choudhuri T. The interplay between Epstein-Bar virus (EBV) with the p53 and its homologs during EBV associated malignancies. *Heliyon*. 2019;5(11):e02624. https://doi.org/10.1016/j.heliyon.2019.e02624
7. Lu SJ, Day NE, Degos L, et al. Linkage of a nasopharyngeal carcinoma susceptibility locus to the HLA region. *Nature*. 1990;346(6283):470-471. https://doi.org/10.1038/346470a0
8. Bei JX, Li Y, Jia WH, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet*. 2010;42(7):599-603. https://doi.org/10.1038/ng.601
9. Hildesheim A, Apple RJ, Chen CJ, et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. *J Natl Cancer Inst*. 2002;94(23):1780-1789. https://doi.org/10.1093/jnci/94.23.1780
10. Tse KP, Su WH, Chang KP, et al. Genome-wide association study reveals multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. *Am J Hum Genet*. 2009;85(2):194-203. https://doi.org/10.1016/j.ajhg.2009.07.007
11. Tang M, Lautenberger JA, Gao X, et al. The principal genetic determinants for nasopharyngeal carcinoma in China involve the HLA class I antigen recognition groove. *PLoS Genet*. 2012;8(11):e1003103. https://doi.org/10.1371/journal.pgen.1003103
12. Chin YM, Mushiroda T, Takahashi A, et al. HLA-A SNPs and amino acid variants are associated with nasopharyngeal carcinoma in Malaysian Chinese. *Int J Cancer*. 2015;136(3):678-687. https://doi.org/10.1002/ijc.29035
13. Wang TM, Zhou T, He YQ, et al. Fine-mapping of HLA class I and class II genes identified two independent novel variants associated with nasopharyngeal carcinoma susceptibility. *Cancer Med*. 2018;7(12):6308-6316. https://doi.org/10.1002/cam4.1838
14. Yao K, Yang S, Shen J, Zhang R, Li L. HLA-DRB1 allele polymorphism and nasopharyngeal carcinoma risk: a meta-analysis. *Eur Arch Otorhinolaryngol*. 2017;274(1):297-303. https://doi.org/10.1007/s00405-016-4264-2
15. Li H, Liu S, Hu J, et al. Epstein-Barr virus lytic reactivation regulation and its pathogenic role in carcinogenesis. *Int J Biol Sci*. 2016;12(11):1309-1318. https://doi.org/10.7150/ijbs.16564
16. Babcock GJ, Decker LL, Volk M, Thorley-Lawson DA. EBV persistence in memory B cells in vivo. *Immunity*. 1998;9(3):395-404. https://doi.org/10.1016/s1074-7613(00)80622-6
17. Murata T. Regulation of Epstein-Barr virus reactivation from latency. *Microbiol Immunol*. 2014;58(6):307-317. https://doi.org/10.1111/1348-0421.12155
18. He YQ, Liao XY, Xue WQ, et al. Association between environmental factors and oral Epstein-Barr virus DNA Loads: a multicenter cross-sectional study in China. *J Infect Dis*. 2019;219(3):400-409. https://doi.org/10.1093/infdis/jiy542
19. Feng Y, Xia W, He G, et al. Accuracy evaluation and comparison of 14 diagnostic markers for nasopharyngeal carcinoma: a meta-analysis. *Front Oncol*. 2020;10:1779. https://doi.org/10.3389/fonc.2020.01779
20. Coghill AE, Pfeiffer RM, Proietti C, et al. Identification of a novel, EBV-based antibody risk stratification signature for early detection of nasopharyngeal carcinoma in Taiwan. *Clin Cancer Res*. 2018;24(6):1305-1314. https://doi.org/10.1158/1078-0432.Ccr-17-1929
21. Cao SM, Liu Z, Jia WH, et al. Fluctuations of epstein-barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *Plos One*. 2011;6(4):e19100. https://doi.org/10.1371/journal.pone.0019100
22. Chien YC, Chen JY, Liu MY, et al. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N Engl J Med*. 2001;345(26):1877-1882. https://doi.org/10.1056/NEJMoa011610
23. He YQ, Xue WQ, Xu FH, et al. The relationship between environmental factors and the profile of Epstein-Barr virus antibodies in the lytic and latent infection periods in healthy populations from endemic and non-endemic nasopharyngeal carcinoma areas in China. *EBioMedicine*. 2018;30:184-191. https://doi.org/10.1016/j.ebiom.2018.02.019
24. Kenney SC, Mertz JE. Regulation of the latent-lytic switch in Epstein-Barr virus. *Semin Cancer Biol*. 2014;26:60-68. https://doi.org/10.1016/j.semcancer.2014.01.002

25. Dai J, Lv J, Zhu M, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med*. 2019;7(10):881-891. https://doi.org/10.1016/s2213-2600(19)30144-4

26. Xu F-H, Xiong D, Xu Y-F, et al. An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein–Barr virus activation. *JNCI: J Natl Cancer Inst*. 2012;104(18):1396-1410. https://doi.org/10.1093/jnci/djs320

27. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. https://doi.org/10.1086/519795

28. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *Plos One*. 2013;8(6):e64683. https://doi.org/10.1371/journal.pone.0064683

29. Zhou F, Cao H, Zuo X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat Genet*. 2016;48(7):740-746. https://doi.org/10.1038/ng.3576

30. Ning L, Ko JM, Yu VZ, et al. Nasopharyngeal carcinoma MHC region deep sequencing identifies HLA and novel non-HLA TRIM31 and TRIM39 loci. *Commun Biol*. 2020;3(1):759. https://doi.org/10.1038/s42003-020-01487-y

31. Tang MZ, Cai YL, Zheng YM, Zeng Y. Association between human leukocyte antigen and nasopharyngeal-carcinoma. *Yi Chuan = Hereditas*. 2012;34(12):1505-1512.

32. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*. 2020;48(W1):W449-w454. https://doi.org/10.1093/nar/gkaa379

33. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188-1190. https://doi.org/10.1101/gr.849004

34. Liu W, Li H, Sheng H, et al. A randomized controlled trial on evaluation of plasma Epstein-Barr virus biomarker for early diagnosis in patients with nasopharyngeal carcinoma. *Adv Ther*. 2020;37(10):4280-4290. https://doi.org/10.1007/s12325-020-01461-4

35. Guo X, Li T, Li F, et al. Intermittent abortive reactivation of Epstein-Barr virus during the progression of nasopharyngeal cancer as indicated by elevated antibody levels. *Oral Oncol*. 2019;93:85-90. https://doi.org/10.1016/j.oraloncology.2019.04.024

36. Wang R, Hu Y, Yindom LM, et al. Association analysis between HLA-A, -B, -C, -DRB1, and -DQB1 with nasopharyngeal carcinoma among a Han population in Northwestern China. *Hum Immunol*. 2014;75(3):197-202. https://doi.org/10.1016/j.humimm.2013.12.015

37. Rubicz R, Yolken R, Drigalenko E, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet*. 2013;9(1):e1003147. https://doi.org/10.1371/journal.pgen.1003147

38. Wang X, Kenyon WJ, Li Q, Müllberg J, Hutt-Fletcher LM. Epstein-Barr virus uses different complexes of glycoproteins gH and gL to infect B lymphocytes and epithelial cells. *J Virol*. 1998;72(7):5552-5558. https://doi.org/10.1128/jvi.72.7.5552-5558.1998

39. Kachuri L, Francis SS, Morrison ML, et al. The landscape of host genetic factors involved in immune response to common viral infections. *Genome Med*. 2020;12(1):93. https://doi.org/10.1186/s13073-020-00790-x

40. Sundqvist E, Sundstrom P, Linden M, et al. Epstein-Barr virus and multiple sclerosis: interaction with HLA. *Genes Immun*. 2012;13(1):14-20. https://doi.org/10.1038/gene.2011.42

41. Wergeland S, Myhr KM, Loken-Amsrud KI, et al. Vitamin D, HLA-DRB1 and Epstein-Barr virus antibody levels in a prospective cohort of multiple sclerosis patients. *Eur J Neurol*. 2016;23(6):1064-1070. https://doi.org/10.1111/ene.12986

42. Yang QY, He YQ, Xue WQ, et al. Association between serum cotinine level and serological markers of Epstein-Barr Virus in healthy subjects in South China where nasopharyngeal carcinoma is endemic. *Front Oncol*. 2019;9:865. https://doi.org/10.3389/fonc.2019.00865

43. Robbins CS, Franco F, Mouded M, Cernadas M, Shapiro SD. Cigarette smoke exposure impairs dendritic cell maturation and T cell proliferation in thoracic lymph nodes of mice. *J Immunol*. 2008;180(10):6623-6628. https://doi.org/10.4049/jimmunol.180.10.6623

44. Arnson Y, Shoenfeld Y, Amital H. Effects of tobacco smoke on immunity, inflammation and autoimmunity. *J Autoimmun*. 2010;34(3):J258-J265. https://doi.org/10.1016/j.jaut.2009.12.003

45. Sopori M. Effects of cigarette smoke on the immune system. *Nat Rev Immunol*. 2002;2(5):372-377. https://doi.org/10.1038/nri803

46. Hedstrom AK, Katsoulis M, Hossjer O, et al. The interaction between smoking and HLA genes in multiple sclerosis: replication and refinement. *Eur J Epidemiol*. 2017;32(10):909-919. https://doi.org/10.1007/s10654-017-0250-2

47. Hedström AK, Bomfim IL, Barcellos LF, et al. Interaction between passive smoking and two HLA genes with regard to multiple sclerosis risk. *Int J Epidemiol*. 2014;43(6):1791-1798. https://doi.org/10.1093/ije/dyu195

48. Klareskog L, Stolt P, Lundberg K, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum*. 2006;54(1):38-46. https://doi.org/10.1002/art.21575

49. Hill JA, Southwood S, Sette A, Jevnikar AM, Bell DA, Cairns E. Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. *J Immunol*. 2003;171(2):538-541. https://doi.org/10.4049/jimmunol.171.2.538

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.