

**Supplemental information**

**Host genetic variants, Epstein-Barr virus  
subtypes, and the risk of nasopharyngeal  
carcinoma: Assessment of interaction and mediation**

**Miao Xu, Ruimei Feng, Zhonghua Liu, Xiang Zhou, Yanhong Chen, Yulu Cao, Linda Valeri, Zilin Li, Zhiwei Liu, Su-Mei Cao, Qing Liu, Shang-Hang Xie, Ellen T. Chang, Wei-Hua Jia, Jincheng Shen, Youyuan Yao, Yong-Lin Cai, Yuming Zheng, Zhe Zhang, Guangwu Huang, Ingemar Ernberg, Minzhong Tang, Weimin Ye, Hans-Olov Adami, Yi-Xin Zeng, and Xihong Lin**

**Table S1. Association between two host genetic variants per risk allele and nasopharyngeal carcinoma, related to Table 2.**

	<b>Allele frequency in cases N (%)</b>	<b>Allele frequency in controls N (%)</b>	<b>OR per risk allele* (95% CI)</b>	<b>P value</b>
<b>Original study in Guangdong</b>				
<b>rs2860580 (risk allele = G)</b>				
A	273 (23.9)	496 (35.6)	Reference	6.00E-10
G	871 (76.1)	896 (64.4)	1.79 (1.49, 2.15)	
<b>rs2894207 (risk allele = T)</b>				
C	155 (13.5)	299 (21.5)	Reference	2.74E-06
T	989 (86.5)	1093 (78.5)	1.71 (1.37, 2.14)	
<b>Replication study in Guangxi</b>				
<b>rs2860580 (risk allele = G)</b>				
A	240 (24.1)	547 (33.1)	Reference	3.62E-07
G	754 (75.9)	1105 (66.9)	1.62 (1.34, 1.95)	
<b>rs2894207 (risk allele = T)</b>				
C	120 (12.1)	300 (18.2)	Reference	3.64E-05
T	874 (87.9)	1352 (81.8)	1.63 (1.29, 2.06)	
<b>Pooled study</b>				
<b>rs2860580 (risk allele = G)</b>				
A	513 (24.0)	1043 (34.3)	Reference	3.76E-15
G	1625 (76.0)	2001 (65.7)	1.68 (1.47, 1.91)	
<b>rs2894207 (risk allele = T)</b>				
C	275 (12.9)	599 (19.7)	Reference	7.48E-10
T	1863 (87.1)	2445 (80.3)	1.65 (1.40, 1.93)	

Abbreviation: OR, odds ratio; CI, confidence interval.

\* The OR per risk allele was estimated with the additive genetic model using logistic regression and adjusted for age at interview, sex and smoking joint status, education level, salt-preserved fish consumption in 2000-2002, nasopharyngeal carcinoma history among first-degree relatives, rural or urban area of residence, current occupation, and environmental exposure.

**Table S2. Association between host genetic variants, rs2860580 and rs2894207, and their joint status and EBV variant 163364 †, related to Table 3.**

	OR *	95% CI *	P value
<b>rs2860580 and EBV 163364</b>			
Original study	1.46	1.16, 1.84	0.002
Replication study	1.55	1.24, 1.92	1.0E-4
Pooled study	1.48	1.27, 1.74	8.7E-7
<b>rs2894207 and EBV 163364</b>			
Original study	1.62	1.27, 2.07	1.0E-4
Replication study	1.55	1.22, 1.98	4.0E-4
Pooled study	1.54	1.30, 1.83	5.2E-7
<b>Joint status of host SNPs and EBV 163364</b>			
Original study	1.74	1.36, 2.23	1.2E-5
Replication study	1.68	1.34, 2.11	7.4E-6
Pooled study	1.67	1.42, 1.97	1.0E-9

Abbreviation: OR, odds ratio; CI, confidence interval.

\* Adjusted for age at interview, sex and smoking joint status, education level, residential area, salt-preserved fish consumption in 2000-2002, nasopharyngeal carcinoma history among first-degree relatives, current occupation, and environmental exposure.

† Coding in logistic regression for rs2860580 (0=AA/AG; 1=GG), rs2894207 (0=CC/CT; 1=TT), the joint status of two host SNPs (low risk, 0 = AA/AG for rs2860580 or CC/CT for rs2894207; high risk, 1 = GG for rs2860580 and TT for rs2894207) and EBV variant 163364 (0=C, 1=CT/T).

**Table S3. Direct and indirect effects on nasopharyngeal carcinoma between per risk allele of the two host SNPs and EBV variant 163364 under the additive genetic model, related to Table 3.**

		<b>OR *</b>	<b>95% CI *</b>	<b>P value</b>
<b>rs2860580 and EBV 163364</b>				
<b>Original study</b>	Natural direct effect	1.76	1.44, 2.15	3.9E-8
	Natural indirect effect	1.01	0.93, 1.09	0.883
	Marginal total effect	1.77	1.43, 2.18	1.3E-7
<b>Replication study</b>	Natural direct effect	1.46	1.18, 1.80	4.0E-4
	Natural indirect effect	1.12	1.02, 1.24	0.023
	Marginal total effect	1.64	1.30, 2.07	3.0E-5
<b>Pooled study</b>	Natural direct effect	1.58	1.37, 1.82	2.4E-10
	Natural indirect effect	1.05	0.98, 1.12	0.151
	Marginal total effect	1.66	1.42, 1.94	1.2E-10
<b>rs2894270 and EBV 163364</b>				
<b>Original study</b>	Natural direct effect	1.53	1.20, 1.96	7.1E-4
	Natural indirect effect	1.15	1.03, 1.30	0.017
	Marginal total effect	1.77	1.35, 2.31	3.2E-5
<b>Replication study</b>	Natural direct effect	1.48	1.15, 1.90	0.002
	Natural indirect effect	1.09	0.98, 1.22	0.128
	Marginal total effect	1.61	1.23, 2.10	4.9E-4
<b>Pooled study</b>	Natural direct effect	1.49	1.25, 1.77	6.6E-6
	Natural indirect effect	1.12	1.03, 1.21	0.008
	Marginal total effect	1.66	1.38, 2.01	9.3E-8

Abbreviation: OR, odds ratio; CI, confidence interval.

\* Adjusted for age at interview, sex and smoking joint status, education level, salt-preserved fish consumption in 2000-2002, nasopharyngeal carcinoma history among first-degree relatives, rural or urban area of residence, current occupation, and environmental exposure.

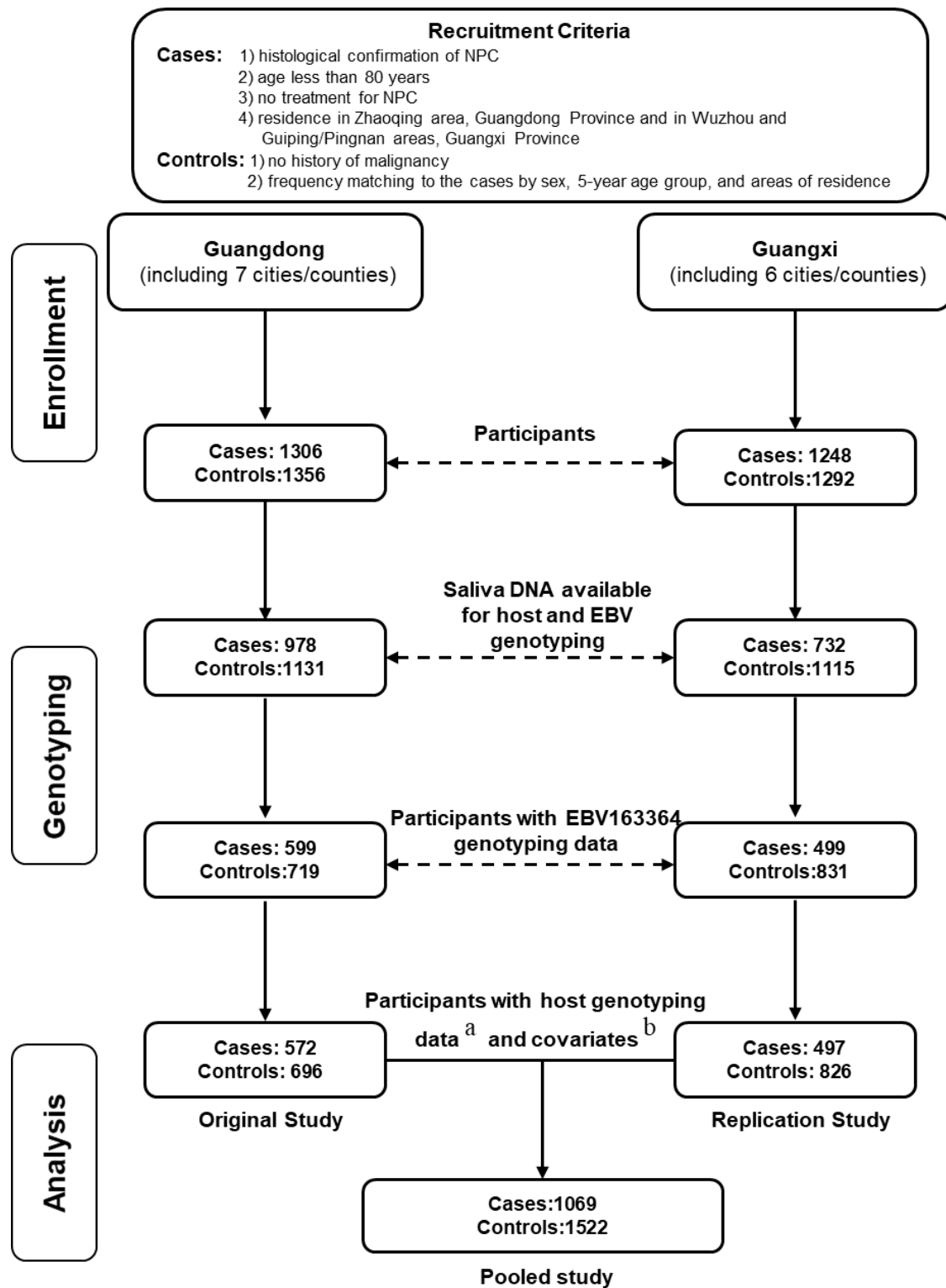
**Table S4. HLA-A\*0207 binding affinity with the peptides of NPC-low-risk and high-risk EBV subtypes, related to STAR Methods.**

Gene_Amino acid change	NPC-low-risk peptide		NPC-high-risk peptide		Reported	OR (95% CI)*
	Peptide	Binding rank %	Peptide*	Binding rank %		
<i>LMP2A_V254L/L255V</i>	<i>FLACVLVLI</i>	0.10	<i>FLACL<b>V</b>VLI</i>	0.14		2.7 (1.7, 4.2)
<i>LMP2A_C426S</i>	<i>CLGGLLTMV</i>	0.55	<i><b>S</b>LGGLLTMV</i>	0.30	Yes	2.2 (1.4, 3.5)
<i>EBNA3B_AA_36E*</i>	<i>GSDPISPEI</i>	1.88	<i><b>E</b>SEPISEI</i>	19.58		1.4 (1.0, 2.0)
<i>EBNA3A_AA_814G*</i>	<i>ALGYPLHAL</i>	0.91	<i>ALGY<b>A</b>LHGL</i>	1.04		1.8 (1.2, 2.8)
<i>BALF4_A743V</i>	<i>LVAGVVILV</i>	0.84	<i>L<b>V</b>VGVVILV</i>	2.07		2.8 (1.8, 4.4)
<i>BALF2_L700V</i>	<i>RLYGRRLPV</i>	0.69	<i><b>R</b>VYGRRLPV</i>	2.11		4.1 (2.4, 6.9)
<i>BNRF1_V1222I</i>	<i>FTNLGMPYV</i>	0.16	<i>FTNLGMPY<b>I</b></i>	0.53		2.3 (1.6, 3.4)
<i>BPLF1_L610I</i>	<i>QLPPSATTL</i>	0.36	<i><b>Q</b>IPPSATTL</i>	1.39		1.8 (1.2, 2.5)
<i>LMP1_L126F/M129I</i>	<i>YLLEMLWRL</i>	0.01	<i><b>Y</b>FLEILWRL</i>	0.34	Yes	2.2 (1.1, 4.3)

Abbreviation: OR, odds ratio; CI, confidence interval; NPC, nasopharyngeal carcinoma.

\*Amino acid changes in the high-risk peptides are highlighted in red.

\*Multiple peptide haplotypes are present. The high-risk peptide and a major low-risk peptide are shown. The OR indicates the NPC risk associated with the high-risk peptide compared to the other peptide variants.

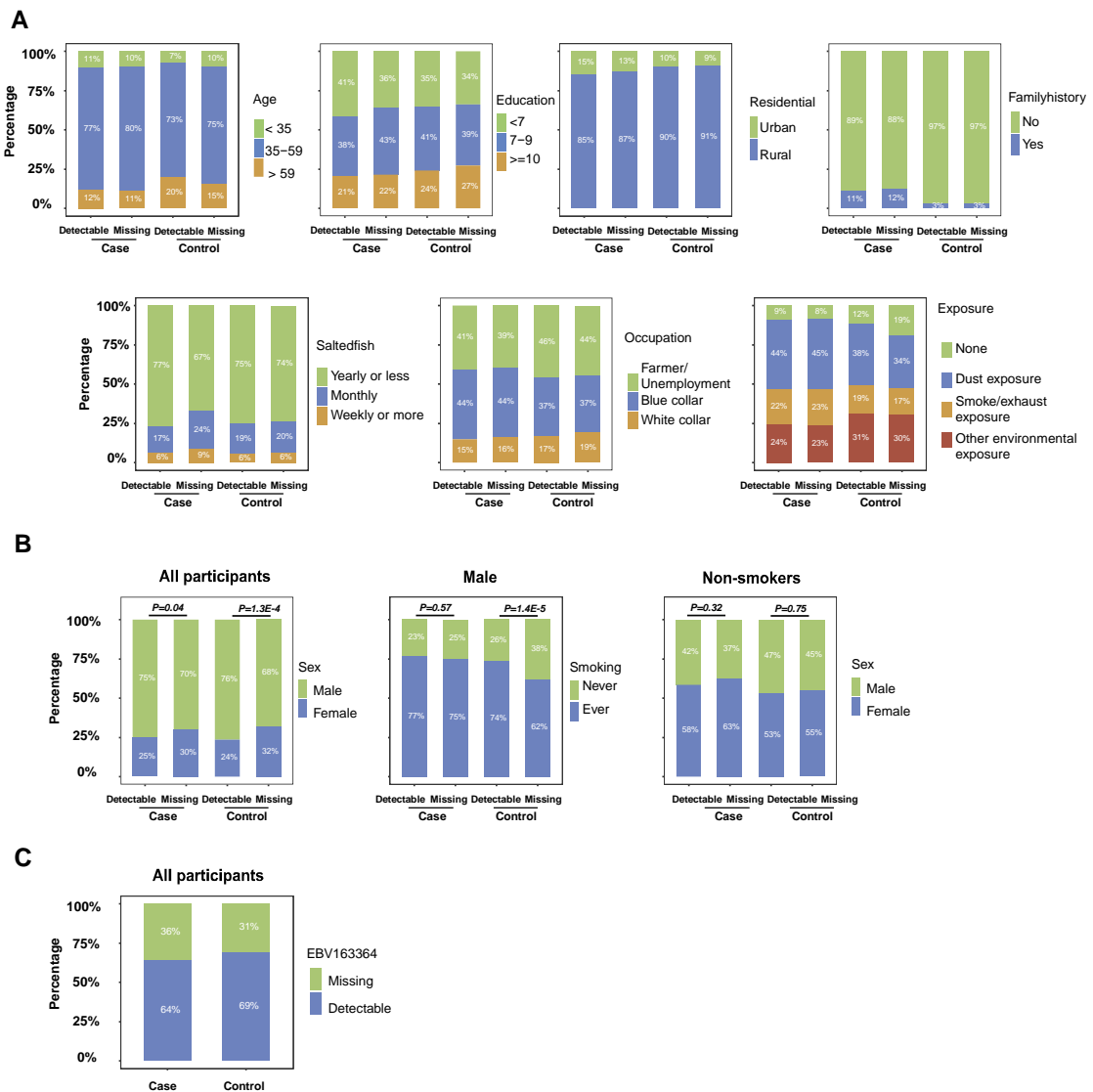


**Figure S1: Flowchart for the study design, related to STAR Methods.**

Abbreviation: NPC, nasopharyngeal carcinoma.

<sup>a</sup> Two host genetic SNPs: rs2860580 and rs2894207

<sup>b</sup> Covariates: sex, age, smoking, education level, salt-preserved fish consumption in 2000-2002, nasopharyngeal carcinoma history among first-degree relatives, rural or urban area of residence, current occupation, and environmental exposure.



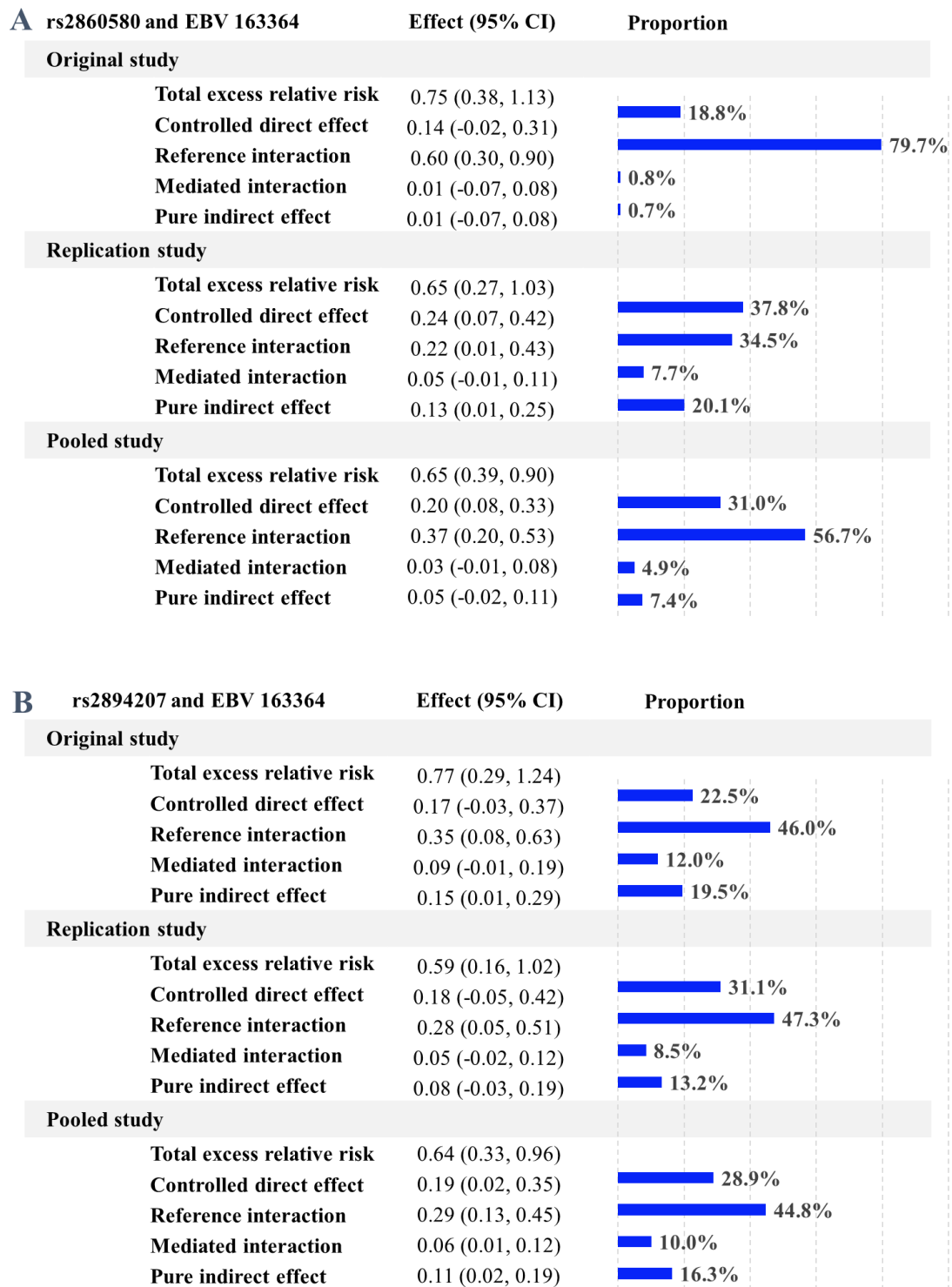
**Figure S2. Distribution of genotyping success or failure for EBV variant 163364 in the study participants available for saliva DNA and variable information, related to STAR Methods.** (A) Stacked bar plots of the distribution of genotyping success or failure for EBV SNP (163364). Variables: age at interview, education level, rural or urban area of residence, nasopharyngeal carcinoma history among first-degree relatives, salt-preserved fish consumption in 2000-2002, current occupation, and selected environmental exposures. (B) Stacked bar plots of the distribution of EBV genotyping success or failure by sex in all participants, by smoking status in men, and by sex among non-smokers. The values were calculated using  $\chi^2$  tests. (C) Stacked bar plots of EBV SNP genotyping failure among nasopharyngeal carcinoma cases and controls which were not associated with increased risk of nasopharyngeal carcinoma.

A	rs2860580	EBV 163364	Cases	Controls	OR (95% CI)*	Additive interaction	
	(0=AA; 1=AG; 2=GG)	(0=C; 1= T/CT)				RERI (95% CI)	P
<b>Original study</b>							
	0	0	9	50	0.88 (0.39, 1.99)		
	1	0	32	169	Reference		
	2	0	42	155	1.44 (0.85, 2.43)	3.34 (1.50, 5.19)	3.8E-4
	0	1	22	44	2.71 (1.41, 5.24)		
	1	1	179	139	6.76 (4.29, 10.64)		
	2	1	288	139	11.32 (7.26, 17.64)		
<b>Replication study</b>							
	0	0	6	68	0.43 (0.18, 1.05)		
	1	0	51	257	Reference		
	2	0	65	238	1.38 (0.92, 2.09)	2.48 (0.50, 4.46)	0.014
	0	1	21	20	5.11 (2.54, 10.28)		
	1	1	135	114	5.72 (3.84, 8.50)		
	2	1	219	129	8.77 (6.01, 12.79)		
<b>Pooled study</b>							
	0	0	15	118	0.62 (0.34, 1.12)		
	1	0	83	426	Reference		
	2	0	107	393	1.38 (1.00, 1.91)	3.08 (1.79, 4.37)	2.9E-6
	0	1	43	64	3.59 (2.26, 5.69)		
	1	1	314	253	6.23 (4.65, 8.35)		
	2	1	507	268	9.90 (7.46, 13.14)		
<b>B</b>							
B	rs2894207	EBV 163364	Cases	Controls	OR (95% CI)*	Additive interaction	
	(0=CC; 1=CT; 2=TT)	(0=C; 1= T/CT)				RERI (95% CI)	P
<b>Original study</b>							
	0	0	1	25	0.21 (0.03, 1.68)		
	1	0	27	130	Reference		
	2	0	55	219	1.23 (0.73, 2.06)	2.81 (0.63, 4.99)	0.012
	0	1	6	8	2.63 (0.82, 8.47)		
	1	1	114	103	5.54 (3.32, 9.23)		
	2	1	369	211	8.28 (5.22, 13.14)		
<b>Replication study</b>							
	0	0	2	26	0.40 (0.09, 1.79)		
	1	0	33	164	Reference		
	2	0	87	373	1.18 (0.75, 1.84)	2.39 (0.71, 4.07)	0.005
	0	1	7	5	7.54 (2.21, 25.72)		
	1	1	69	74	4.42 (2.66, 7.34)		
	2	1	299	184	8.24 (5.40, 12.58)		
<b>Pooled study</b>							
	0	0	3	51	0.29 (0.09, 0.95)		
	1	0	60	294	Reference		
	2	0	142	592	1.16 (0.83, 1.62)	2.57 (1.24, 3.90)	1.5E-4
	0	1	13	13	4.41 (1.91, 10.14)		
	1	1	183	177	5.02 (3.53, 7.15)		
	2	1	668	395	8.18 (6.00, 11.14)		

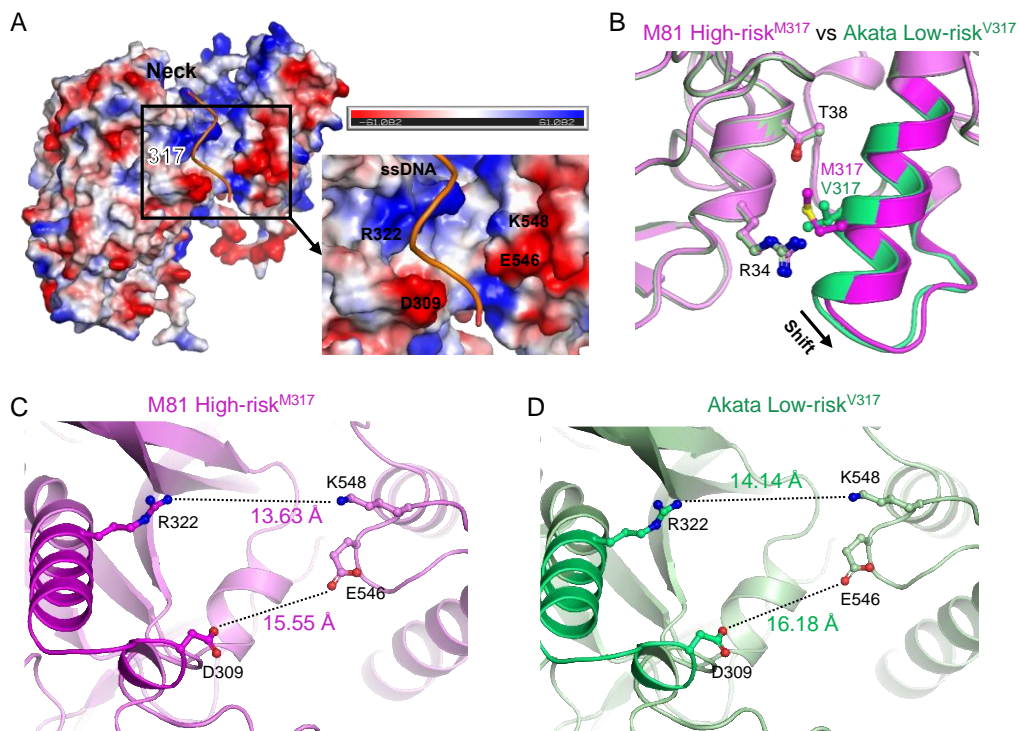
**Figure S3. Joint effect and additive interaction between EBV variant 163364 and per risk allele of the two host SNPs rs2860580 (A) and rs2894207 (B) on the risk of the nasopharyngeal carcinoma under the additive genetic model, related to Figure 1. The**



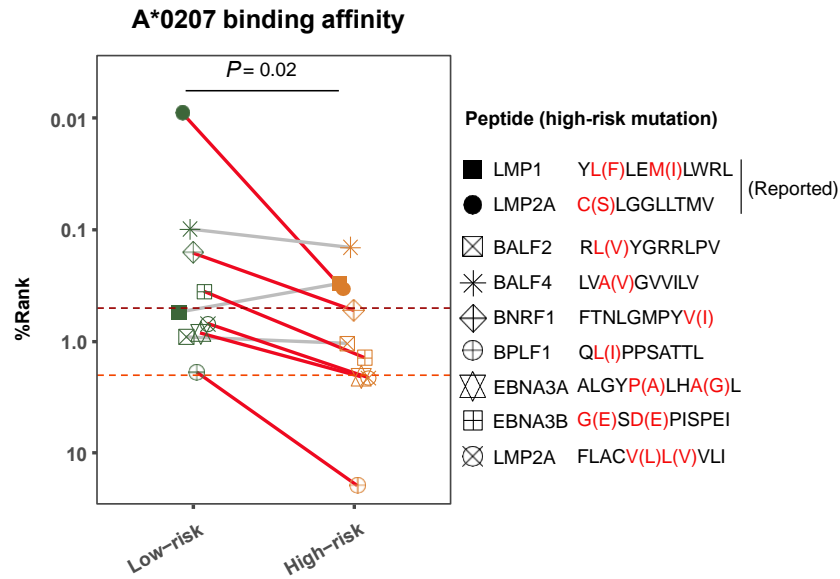
analyses were adjusted for age at interview, sex and smoking joint status, education level, salt-preserved fish consumption in 2000-2002, NPC history among first-degree relatives, rural or urban area of residence, current occupation, and environmental exposure. Abbreviation: OR, odds ratio; CI, confidence interval; RERI, relative excess risk due to interaction.



**Figure S4. Four-way decomposition of total excess relative risk for nasopharyngeal carcinoma associated with per risk allele of the two host SNPs rs2860580 (A) and rs2894207 (B) using additive genetic model, related to Figure 2.** The analyses were adjusted for age at interview, sex and smoking joint status, education level, salt-preserved fish consumption in 2000-2002, nasopharyngeal carcinoma history among first-degree relatives, rural or urban area of residence, current occupation, and environmental exposure.



**Figure S5. Predicted structures of BALF2 protein from high-risk M81 EBV and low-risk Akata EBV, related to STAR Methods.** (A) Predicted protein conformation of BALF2 in complex with single-stranded DNA (ssDNA, orange). The amino acid 317, encoded by the high-risk variant 163364, and the key amino acids interacting with ssDNA are indicated. (B) The V317M mutation induces an alpha-helix shift. Magenta and green indicate regional structures of BALF2 protein from high-risk M81 EBV and low-risk Akata EBV, respectively. Other two amino acids (R34 and T38) that retain their position, in contrast to V317M, are highlighted. (C-D) Spatial distances between amino acids interacting with ssDNA are indicated for high-risk M81 EBV (C) and low-risk Akata EBV (D) BALF2 proteins, respectively.



**Figure S6. HLA-A\*0207 binding affinity with the EBV peptides of nasopharyngeal carcinoma-low-risk and high-risk subtypes, related to STAR Methods.** The 9-mer peptides are indicated on the right, and mutations in the high-risk EBV subtype are highlighted in red. The LMP1 and LMP2A peptides have been verified with functional T cell response assays in previous studies, indicating that the mutant LMP2A peptide failed to elicit T cell responses in patients with nasopharyngeal carcinoma. The affinity is shown as the binding ranking percentile predicted with NetMHCpan-4.1. The dark red dashed line represents a ranking percentile of 0.5%, indicative of strong binding affinity. The red dashed line represents a ranking percentile of 2%, indicative of weak binding affinity.