Cell Genomics, Volume 4

# **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

	Allele frequency in cases N (%)	Allele frequency in controls N (%)	OR per risk allele* (95% Cl)	<i>P</i> value
Original study in Guar	ngdong			
rs2860580 (risk al	lele = G)			
А	273 (23.9)	496 (35.6)	Reference	6 00E 10
G	871 (76.1)	896 (64.4)	1.79 (1.49, 2.15)	0.002-10
rs2894207 (risk al	lele = T)			
С	155 (13.5)	299 (21.5)	Reference	
Т	989 (86.5)	1093 (78.5)	1.71 (1.37, 2.14)	2.74E-00
Replication study in G	Juangxi			
rs2860580 (risk al	lele = G)			
А	240 (24.1)	547 (33.1)	Reference	
G	754 (75.9)	1105 (66.9)	1.62 (1.34, 1.95)	3.02E-07
rs2894207 (risk al	lele = T)			
С	120 (12.1)	300 (18.2)	Reference	
Т	874 (87.9)	1352 (81.8)	1.63 (1.29, 2.06)	3.04E-05
Pooled study				
rs2860580 (risk al	lele = G)			
А	513 (24.0)	1043 (34.3)	Reference	
G	1625 (76.0)	2001 (65.7)	1.68 (1.47, 1.91)	3.76E-15
rs2894207 (risk al	lele = T)			
С	275 (12.9)	599 (19.7)	Reference	
Т	1863 (87.1)	2445 (80.3)	1.65 (1.40, 1.93)	7.48E-10

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

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.

	OR *	95% CI *	P value
rs2860580 and EBV 163364			
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
rs2894207 and EBV 163364			
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
Joint status of host SNPs and EBV 163364			
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

Table S2. Association between host genetic variants, rs2860580 and rs2894207, and their joint status and EBV variant 163364 <sup>+</sup>, related to Table 3.

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

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

 $\dagger$  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.

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

Gene Amino acid	NPC-low-risk peptide		NPC-high-risk peptide			
change	Peptide	Binding rank %	Peptide <sup>+</sup>	Binding rank %	Reported	OR (95% CI)*
LMP2A_V254L/L255V	FLACVLVLI	0.10	FLAC <mark>LV</mark> VLI	0.14		2.7 (1.7, 4.2)
LMP2A_C426S	CLGGLLTMV	0.55	<mark>S</mark> LGGLLTMV	0.30	Yes	2.2 (1.4, 3.5)
EBNA3B_AA_36E <sup>*</sup>	GSDPISPEI	1.88	<b>E</b> SEPISPEI	19.58		1.4 (1.0, 2.0)
EBNA3A_AA_814G <sup>*</sup>	ALGYPLHAL	0.91	ALGY <mark>A</mark> LH <mark>G</mark> L	1.04		1.8 (1.2, 2.8)
BALF4_A743V	LVAGVVILV	0.84	LV <mark>V</mark> GVVILV	2.07		2.8 (1.8, 4.4)
BALF2_L700V	RLYGRRLPV	0.69	<b>R<b>√</b>YGRRLPV</b>	2.11		4.1 (2.4, 6.9)
BNRF1_V12221	FTNLGMPYV	0.16	FTNLGMPY <mark>I</mark>	0.53		2.3 (1.6, 3.4)
BPLF1_L610I	QLPPSATTL	0.36	Q <mark>I</mark> PPSATTL	1.39		1.8 (1.2, 2.5)
LMP1_L126F/M129I	YLLEMLWRL	0.01	Y <mark>F</mark> LE <mark>I</mark> LWRL	0.34	Yes	2.2 (1.1, 4.3)

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

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.

	rs2860580	EBV 163364	~	~		Additive intera	ction
Α	(0=AA; 1=AG; 2=GG)	(0=C; 1= T/CT)	Cases	Controls	OR (95% CI)*	RERI (95% CI)	Р
Original s	tudy	)					
	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)		
	0	1	22	44	2.71 (1.41, 5.24)	3.34 (1.50, 5.19)	3.8E-4
	1	1	179	139	6.76 (4.29, 10.64)		
	2	1	288	139	11.32 (7.26, 17.64)		
Replicatio	on study						
	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)	2.48 (0.30, 4.40)	0.014
	1	1	135	114	5.72 (3.84, 8.50)		
	2	1	219	129	8.77 (6.01, 12.79)		
Pooled stu	ıdy						
	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)		
	0	1	43	64	3.59 (2.26, 5.69)	3.08 (1.79, 4.37)	2.9E-6
	1	1	314	253	6.23 (4.65, 8.35)		
	2	1	507	268	9.90 (7.46, 13.14)		
D	rs2894207	EBV 163364				Additive intera	ction
В	rs2894207 (0= CC; 1=CT; 2=TT)	EBV 163364 (0=C; 1=T/CT)	Cases	Controls	OR (95% CI)*	Additive intera RERI (95% CI)	ction P
<b>B</b> Original s	rs2894207 (0= CC; 1=CT; 2=TT)	EBV 163364 (0=C; 1= T/CT)	Cases	Controls	OR (95% CI)*	Additive intera	ction P
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) study 0	EBV 163364 (0=C; 1=T/CT)	Cases	Controls	<b>OR (95% CI)*</b>	Additive intera	ction P
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) tudy 0 1	EBV 163364 (0=C; 1=T/CT) 0 0	Cases	Controls	<b>OR (95% CI)*</b> 0.21 (0.03, 1.68) Reference	Additive intera	ction P
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) tudy 0 1 2	EBV 163364 (0=C; 1= T/CT) 0 0 0	<b>Cases</b> 1 27 55	<b>Controls</b> 25 130 219	<b>OR (95% CI)*</b> 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06)	Additive intera	ction P
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) tudy 0 1 2 0	EBV 163364 (0=C; 1=T/CT) 0 0 0	Cases 1 27 55 6	<b>Controls</b> 25 130 219 8	<b>OR (95% CI)*</b> 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47)	Additive intera RERI (95% CI) 2.81 (0.63, 4.99)	<b>ction</b> <i>P</i> 0.012
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) tudy 0 1 2 0 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1	<b>Cases</b> 1 27 55 6 114	<b>Controls</b> 25 130 219 8 103	<b>OR (95% CI)*</b> 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99)	<b>ction</b> <b>P</b> 0.012
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) tudy 0 1 2 0 1 2 0 1 2	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1	Cases 1 27 55 6 114 369	<b>Controls</b> 25 130 219 8 103 211	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99)	<b>ction</b> <b>P</b> 0.012
B Original s Replicatio	rs2894207 (0= CC; 1=CT; 2=TT) study 0 1 2 0 1 2 0 1 2 0 1 2	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1	Cases 1 27 55 6 114 369	Controls 25 130 219 8 103 211	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99)	ction <i>P</i> 0.012
B Original s Replication	rs2894207 (0= CC; 1=CT; 2=TT) ntudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0	Cases 1 27 55 6 114 369 2	<b>Controls</b> 25 130 219 8 103 211 26	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79)	Additive intera- RERI (95% CI)	ction <i>P</i> 0.012
B Original s Replication	rs2894207 (0= CC; 1=CT; 2=TT) tudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 1 1 2 1 1 1 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0 0 0 0	Cases 1 1 27 55 6 114 369 2 33	Controls 25 130 219 8 103 211 26 164	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference	Additive intera- RERI (95% CI)	ction <i>P</i> 0.012
B Original s Replication	rs2894207 (0=CC; 1=CT; 2=TT) tudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 2 0 1 2 2 0 1 2 2 0 1 2 2 0 1 2 2 1 1 0 1 2 1 0 1 2 1 1 1 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0	Cases 1 1 27 55 6 114 369 2 33 87	<b>Controls</b> 25 130 219 8 103 211 26 164 373	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84)	Additive intera- RERI (95% CI)	ction <i>P</i> 0.012
B Original s Replicatio	rs2894207 (0= CC; 1=CT; 2=TT) ntudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0 0 0 0 0 1 1	Cases 1 1 27 55 6 114 369 2 33 87 7	Controls 25 130 219 8 103 211 26 164 373 5	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction <i>P</i> 0.012 0.005
B Original s	rs2894207 (0=CC; 1=CT; 2=TT) ntudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 1 1 2 1 1 1 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1	Cases 1 1 27 55 6 114 369 2 33 87 7 69	Controls 25 130 219 8 103 211 26 164 373 5 74	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction <i>P</i> 0.012 0.005
B Original s	rs2894207 (0= CC; 1=CT; 2=TT) ntudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 2 0 1 2 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 33 87 7 69 299	Controls 25 130 219 8 103 211 26 164 373 5 74 184	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction <i>P</i> 0.012 0.005
B Original s Replication	rs2894207 (0= CC; 1=CT; 2=TT) ntudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 1 0 1 2 1 0 1 2 1 1 1 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 33 87 7 69 299	Controls 25 130 219 8 103 211 26 164 373 5 74 184	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction P 0.012 0.005
B Original s Replication	rs2894207 (0= CC; 1=CT; 2=TT) atudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 1 1 2 0 1 1 2 0 0 1 1 2 0 1 1 2 0 1 1 2 0 1 1 2 0 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 33 87 7 69 299 3	Controls 25 130 219 8 103 211 26 164 373 5 74 184 51	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58) 0.29 (0.09, 0.95)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction <i>P</i> 0.012 0.005
B Original s Replication	rs2894207 (0=CC; 1=CT; 2=TT) tudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 1 2	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 3 87 7 69 299 3 60	Controls 25 130 219 8 103 211 26 164 373 5 74 184 51 294	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58) 0.29 (0.09, 0.95) Reference	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction P 0.012 0.005
B Original s Replication	rs2894207 (0=CC; 1=CT; 2=TT) atudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 1 2	EBV 163364 (0=C; 1=T/CT) 0 0 1 1 1 1 1 2 0 0 0 0 1 1 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 3 87 7 69 299 3 60 142	Controls 25 130 219 8 103 211 26 164 373 5 74 184 51 294 592	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58) 0.29 (0.09, 0.95) Reference 1.16 (0.83, 1.62)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07)	ction <i>P</i> 0.012 0.005
B Original s Replication	rs2894207 (0=CC; 1=CT; 2=TT) atudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 0 1 1 2 0 0 1 2 0 0 1 1 2 0 0 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 2 0 0 1 2 0 0 1 2 0 1 2 0 1 2 0 1 1 2 0 1 2 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 2 0 1 1 2 0 0 1 1 2 0 0 1 2 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 2 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 2 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 0 1 1 2 0 1 1 2 0 1 1 2 0 1 1 2 0 1 1 2 0 1 1 2 1 1 1 2 1 1 1 1	EBV 163364 (0=C; 1=T/CT) 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 3 87 7 69 299 3 60 142 13	Controls 25 130 219 8 103 211 26 164 373 5 74 184 51 294 592 13	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58) 0.29 (0.09, 0.95) Reference 1.16 (0.83, 1.62) 4.41 (1.91, 10.14)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07) 2.57 (1.24, 3.90)	ction P 0.012 0.005 1.5E-4
B Original s Replication	rs2894207 (0= CC; 1=CT; 2=TT) atudy 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 0 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1	EBV 163364 (0=C; 1=T/CT) 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Cases 1 1 27 55 6 114 369 2 3 87 7 69 299 3 60 142 13 183	Controls 25 130 219 8 103 211 26 164 373 5 74 184 51 294 592 13 177	OR (95% CI)* 0.21 (0.03, 1.68) Reference 1.23 (0.73, 2.06) 2.63 (0.82, 8.47) 5.54 (3.32, 9.23) 8.28 (5.22, 13.14) 0.40 (0.09, 1.79) Reference 1.18 (0.75, 1.84) 7.54 (2.21, 25.72) 4.42 (2.66, 7.34) 8.24 (5.40, 12.58) 0.29 (0.09, 0.95) Reference 1.16 (0.83, 1.62) 4.41 (1.91, 10.14) 5.02 (3.53, 7.15)	Additive intera- RERI (95% CI) 2.81 (0.63, 4.99) 2.39 (0.71, 4.07) 2.57 (1.24, 3.90)	ction P 0.012 0.005 1.5E-4

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, saltpreserved 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.

A	rs2860580 and EBV 163364	Effect (95% CI)	Proportion
	Original study		
	Total excess relative risk	0.75 (0.38, 1.13)	
	<b>Controlled direct effect</b>	0.14 (-0.02, 0.31)	18.8%
	<b>Reference interaction</b>	0.60 (0.30, 0.90)	79.79
	<b>Mediated interaction</b>	0.01 (-0.07, 0.08)	0.8%
	Pure indirect effect	0.01 (-0.07, 0.08)	0.7%
	Replication study		
	Total excess relative risk	0.65 (0.27, 1.03)	
	<b>Controlled direct effect</b>	0.24 (0.07, 0.42)	37.8%
	<b>Reference interaction</b>	0.22 (0.01, 0.43)	34.5%
	Mediated interaction	0.05 (-0.01, 0.11)	<b>7.7%</b>
	Pure indirect effect	0.13 (0.01, 0.25)	20.1%
	Pooled study		
	Total excess relative risk	0.65 (0.39, 0.90)	
	<b>Controlled direct effect</b>	0.20 (0.08, 0.33)	31.0%
	<b>Reference interaction</b>	0.37 (0.20, 0.53)	56.7%
	Mediated interaction	0.03 (-0.01, 0.08)	<b>4.9%</b>
	Pure indirect effect	0.05 (-0.02, 0.11)	7.4%
B	rs2894207 and EBV 163364	Effect (95% CI)	Proportion
	Original study		
	Total excess relative risk	0.77 (0.29, 1.24)	22 5%
	Controlled direct effect	0.17 (-0.03, 0.37)	46 0%
	Reference interaction	0.35 (0.08, 0.63)	
	Mediated interaction	0.09 (-0.01, 0.19)	
	Pure indirect effect	0.15 (0.01, 0.29)	19.5%
	Replication study		
	Total excess relative risk	0.59 (0.16, 1.02)	21 10/
	Controlled direct effect	0.18 (-0.05, 0.42)	31.1%
	<b>Reference interaction</b>	0.28 (0.05, 0.51)	
	Mediated interaction	0.05 (-0.02, 0.12)	8.5%
	Pure indirect effect	0.08 (-0.03, 0.19)	13.2%
	Pooled study		
	Total excess relative risk	0.64 (0.33, 0.96)	
	<b>Controlled direct effect</b>	0.19 (0.02, 0.35)	28.9%
	<b>Reference interaction</b>	0.29 (0.13, 0.45)	44.8%
	Mediated interaction	0.06 (0.01, 0.12)	10.0%
	Pure indirect effect	0.11 (0.02, 0.19)	16.3%

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 highrisk 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.