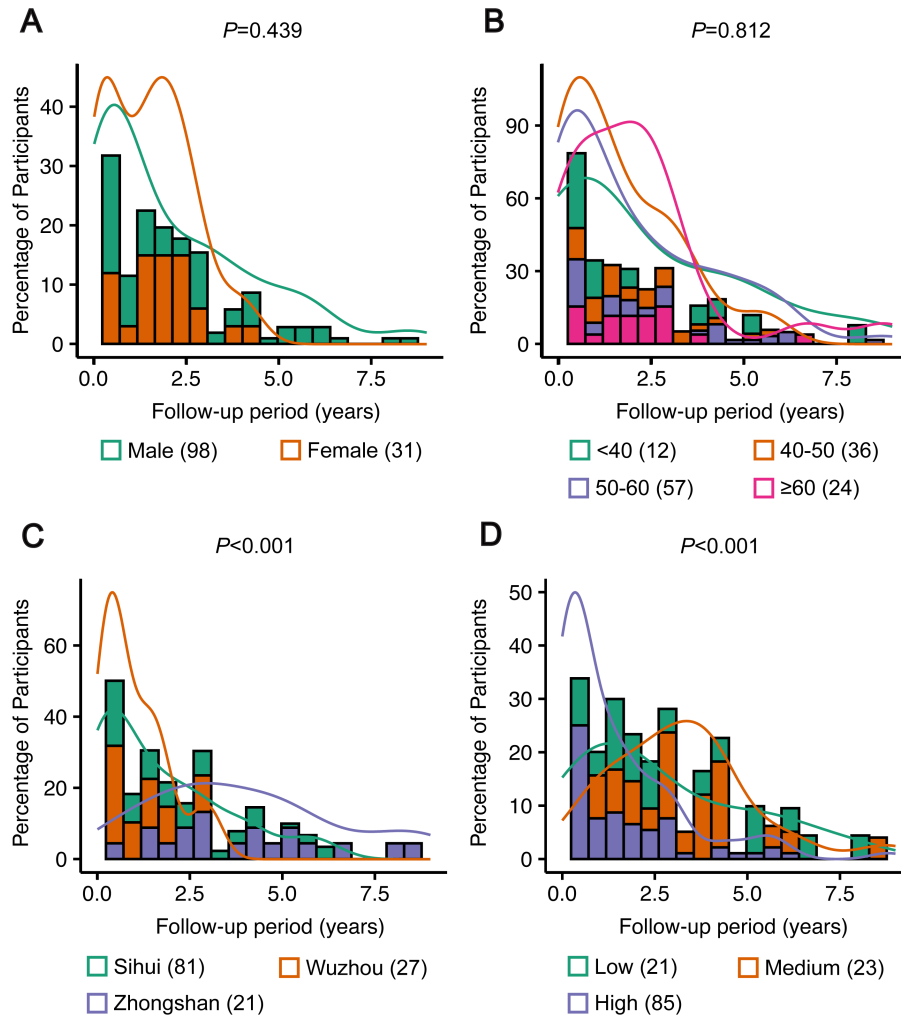
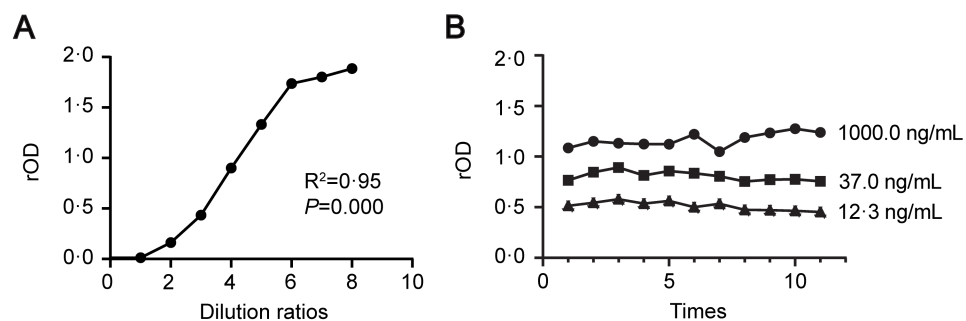
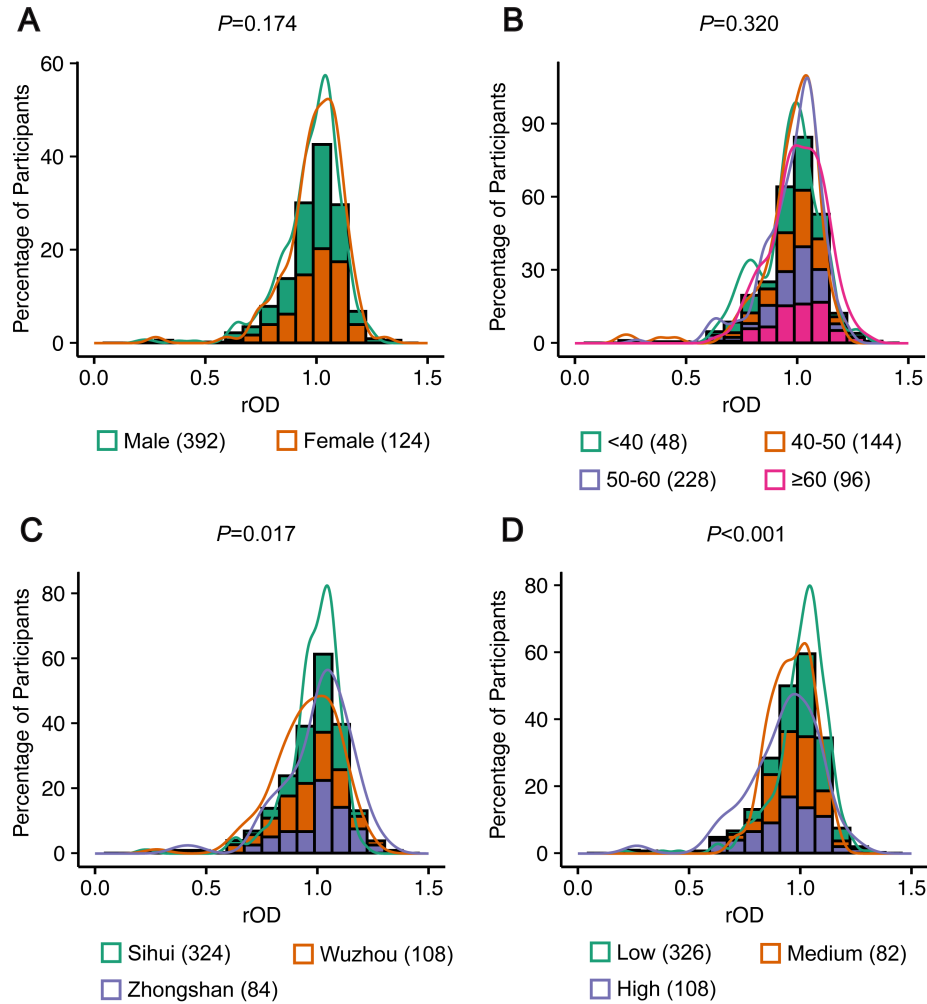




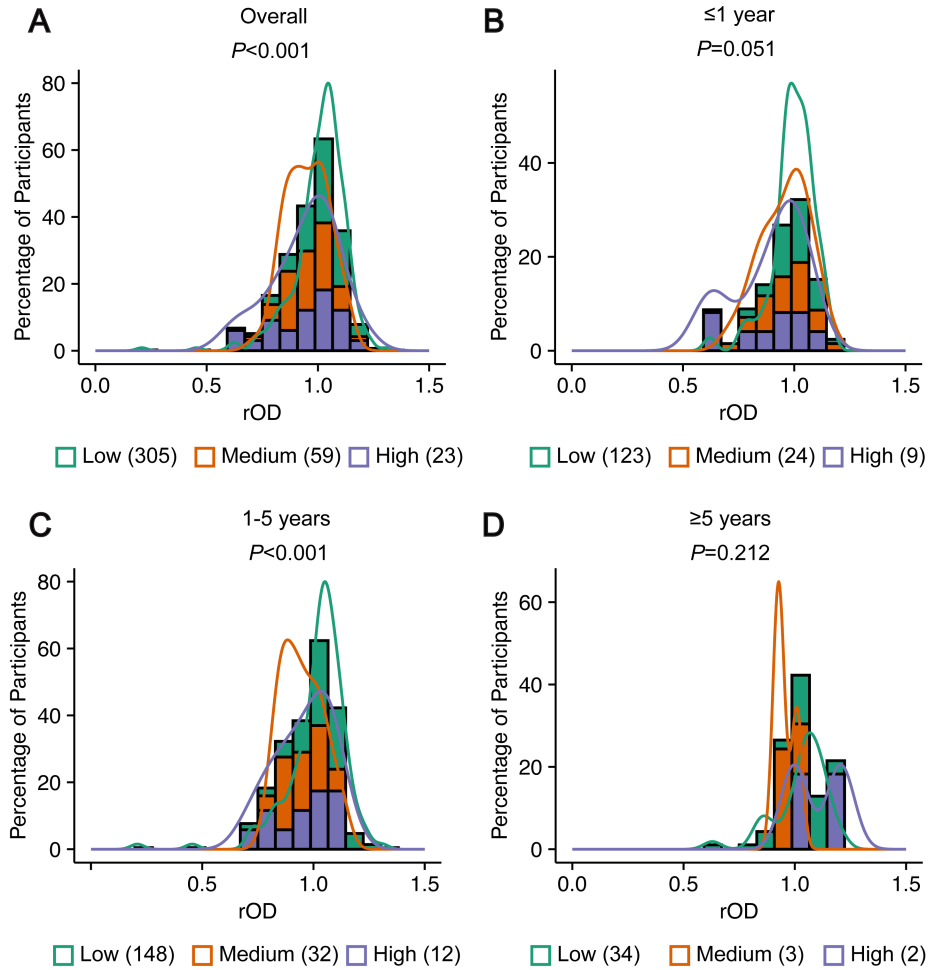
Supplemental Figure 1. Distribution of follow-up duration before nasopharyngeal carcinoma onset in the subgroups. The distribution of follow-up period before nasopharyngeal carcinoma onset stratified by sex (**A**), age (**B**), cohort (**C**), and NPC risk score (**D**). Continuous follow-up period was discretized into 20 consecutive bins. The lines at the top of each histogram represent kernel density estimations whose Y axes correspond to the density. Wilcoxon test (**A**) or Kruskal-Wallis test (**B-D**) were used to calculate P values. Numbers in the legends refer to the numbers of participants in each group.



Supplemental Figure 2. Quality control testing in ELISA experiments. (A) The linear range was assessed using 7 samples simulated by three-fold dilution of monoclonal human gp42 (1 μ g/mL) mixed with negative solvent. (B) The rOD values of monoclonal gp42-IgG standards at three different concentrations across 11 plates.

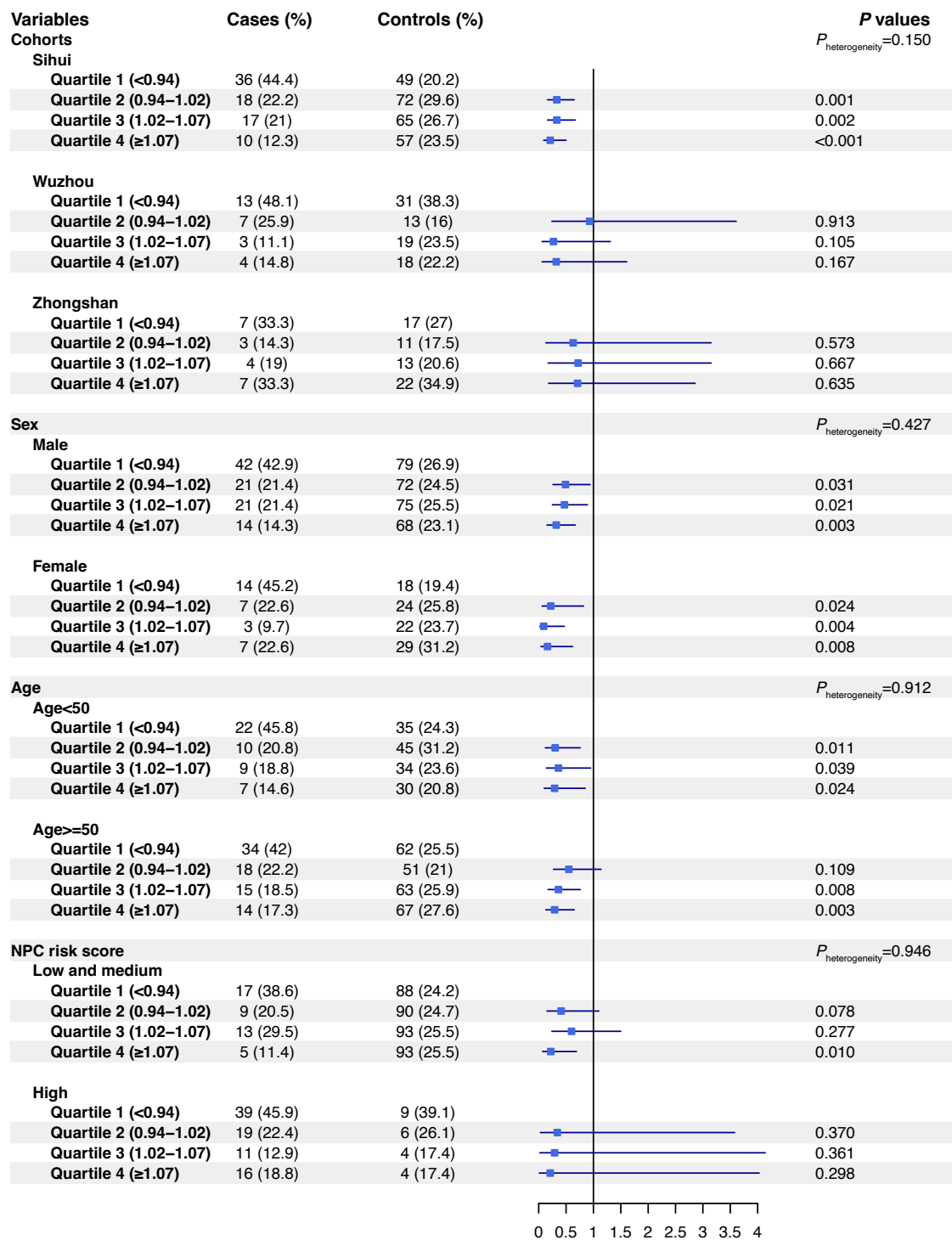


Supplemental Figure 3. Distribution of gp42-IgG levels in the subgroups. The distributions of gp42-IgG stratified by sex (**A**), age (**B**), cohort (**C**), and NPC risk score (**D**). Continuous rOD was discretized into 20 consecutive bins. The lines at the top of each histogram represent kernel density estimations whose Y axes correspond to the density. Wilcoxon test (**A**) or Kruskal-Wallis test (**B-D**) were used to calculate P values. Numbers in the legends refer to the numbers of participants in each group.

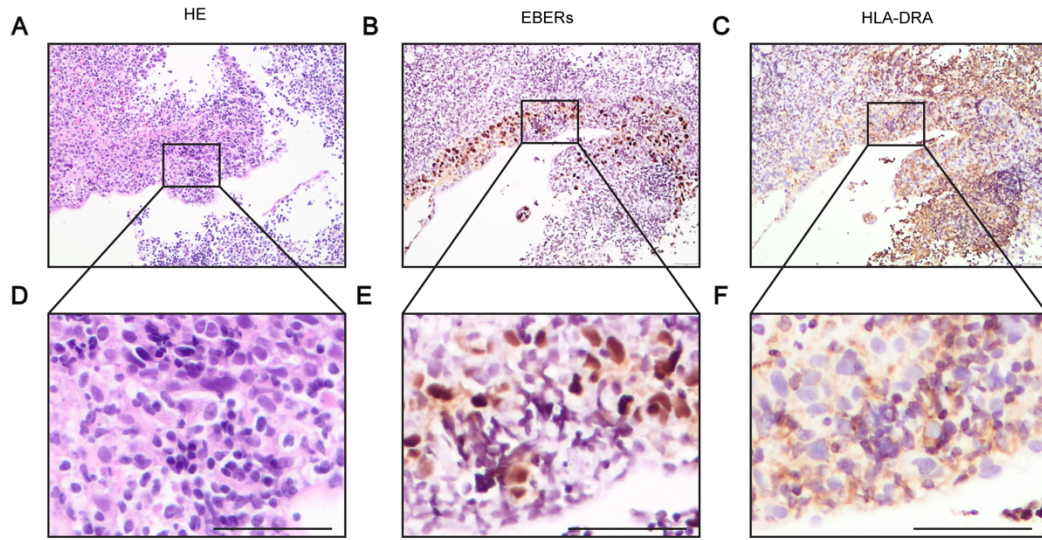


Supplemental Figure 4. Analysis of gp42-IgG titers at various EBV antibody NPC risk scores in the controls by different follow-up durations. Distributions of gp42-IgG titers in the overall cohort (**A**), the cohort with follow-up duration ≤ 1 year (**B**), 1-5 years (**C**), and ≥ 5 years (**D**). rOD was a standardized OD450-OD630 which is comparable across batches of tests. NPC risk score of each participant was determined by a risk prediction algorithm (Logit $P = -3.934 + 2.203 \times \text{VCA-IgA} + 4.797 \times \text{EBNA1-IgA}$) and the individuals were stratified by the scores (Low risk, $P < 0.65$; Medium risk, $0.65 \leq P < 0.98$; High risk, $P \geq 0.98$). Continuous rOD was discretized into 20 consecutive bins. The lines at the top of each histogram represent kernel

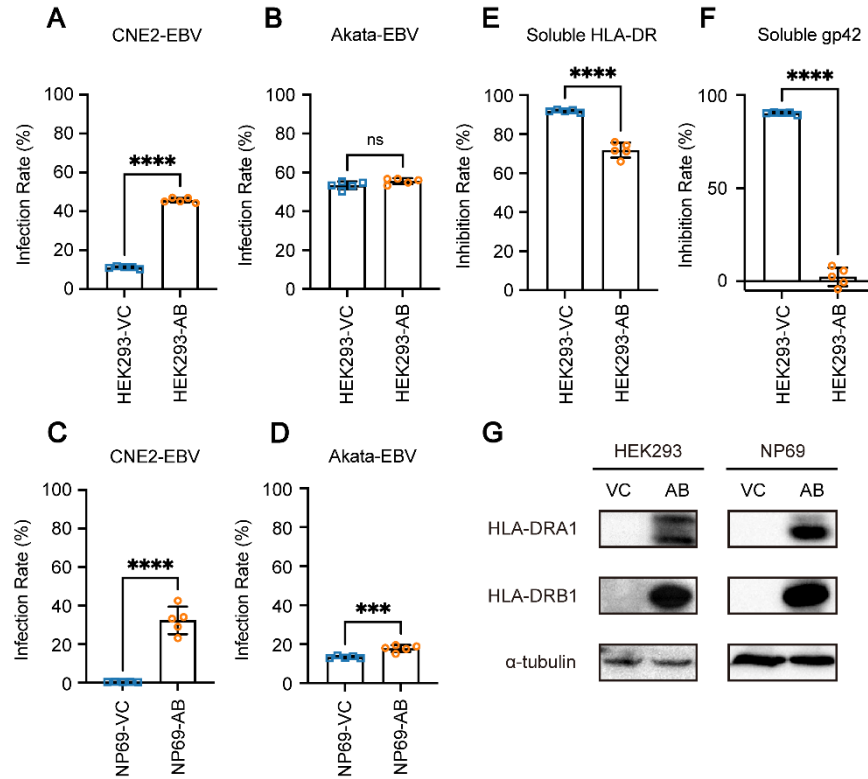
density estimations whose Y axes correspond to the density. Kruskal-Wallis test were used to calculate P values. Numbers in the legends refer to the numbers of participants in each group.



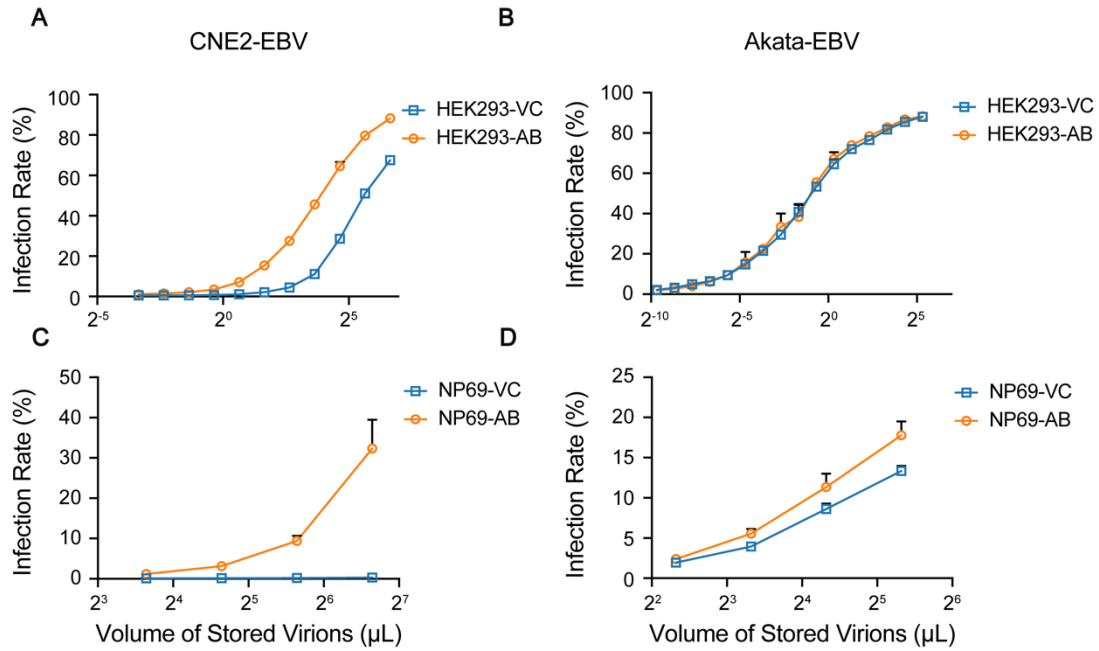
Supplemental Figure 5. Conditional logistic regression analysis reveals the risks of nasopharyngeal carcinoma in the stratified populations with various characteristics. Each square represents the point estimation of OR, the vertical bar represents the 95% CIs of ORs.



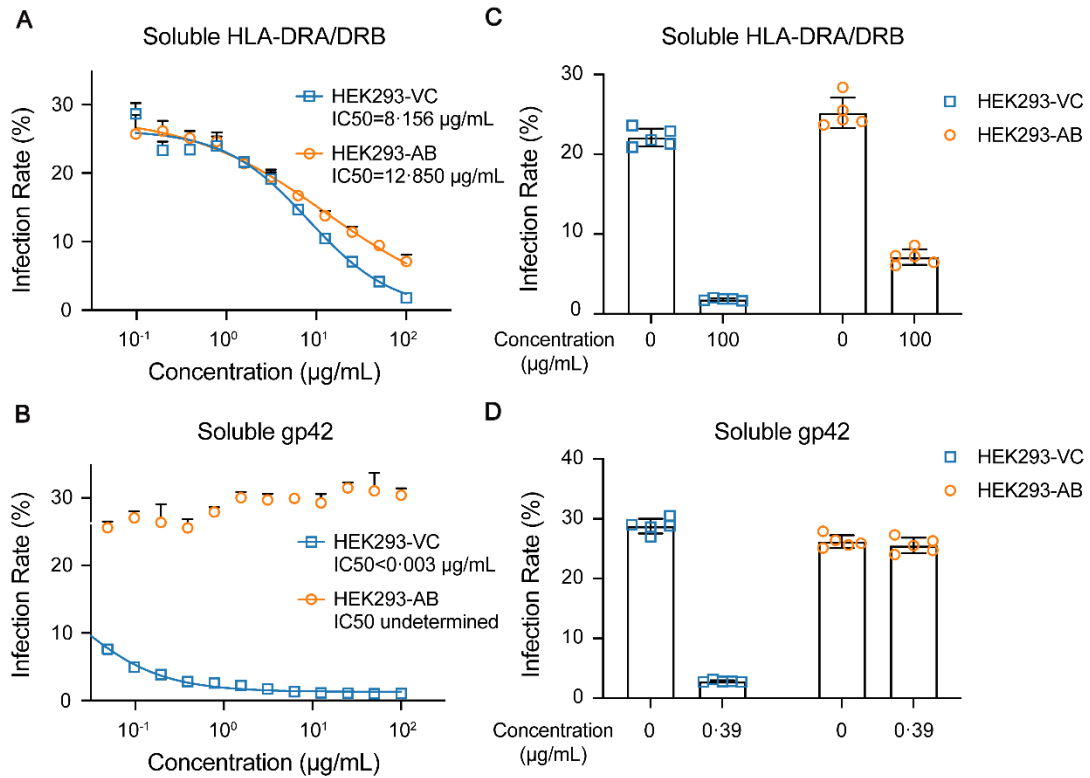
Supplemental Figure 6. HLA-II expression in nasopharyngeal atypical dysplasia. (A and D) HE staining. (B and E) In situ hybridization to detect EBERs. (C and F) Immunohistochemistry to detect HLA-DRA. Shown was a representative out of 27 samples. A-C were shot at 20×, D-F were zoomed in from the black square region, and scale bars at lower right represent 50 μm.



Supplemental Figure 7. Overexpression of HLA-II promotes epithelia origin EBV infection. We have established four cell lines called HEK293-VC, HEK293-AB, NP69-VC, and NP69-AB that stably transfected with either pLVX or pLVX-HLA-DRA1/DRB1. (A and B) CNE2-EBV (A) or Akata-EBV (B) infected HEK293-VC, HEK293-AB. (C and D) CNE2-EBV (C) or Akata-EBV (D) infected NP69-VC, NP69-AB. (E) Soluble HLA-DR at 100 $\mu\text{g/mL}$ inhibited CNE2-EBV infecting HEK293-VC and HEK293-AB. (F) Soluble gp42 at 0.39 $\mu\text{g/mL}$ inhibited CNE2-EBV infecting HEK293-VC and HEK293-AB. Inhibition rate was calculated by subtracting the relative infection rate from 100%. Each bar represents the mean of infection rate or inhibition rate. Error bars represent the SD deviation from the mean. Each group consisted of 5 samples. ns: not significant; ***: $P < 0.001$; ****: $P < 0.0001$. t-test was adopted. Residues complied normal distribution. (G) Verification of HLA-DRA and HLA-DRB expression by WB.



Supplemental Figure 8. Kinetics of EBV infection in various cell lines. HEK293-VC, HEK293-AB, NP69-VC, and NP69-AB that stably transfected with either pLVX or pLVX-HLA-DRA1/DRB1 were establishes. (A and B) Various volume of CNE2-EBV (A) or Akata-EBV (B) infects HEK293-VC or HEK293-AB. (C and D) Various volume of CNE2-EBV (C) or Akata-EBV (D) infects NP69-VC or NP69-AB. Each dot represents the mean infection rate of five replicates. Error bars represent SD. Each group consisted of 5 samples.



Supplemental Figure 9. Soluble HLA-DRA/DRB and gp42 significantly inhibit CNE2-EBV infecting HLA-DRA/DRB overexpressing epithelial cell lines. (A and B) CNE2-EBV Infection kinetics of two cell lines under various concentrations of soluble HLA-DRA/DRB (A) or soluble gp42 (B). Each dot represents the mean infection rate of five replicates. Error bars represent SD. (C and D) Bar plots of the infection rate of CNE2-EBV under indicated concentration of soluble HLA-DRA/DRB (C) or soluble gp42 (D). Each dot represents one test sample. The height of bars represents the mean infection rate. Error bars represent SD. Each group consisted of 5 samples.

Supplemental Table 1. Intra-plate and inter-plate variances of gp42-IgG standards

Sample ^a	Inter-plate variance (n=11)			Intra-plate variance(n=3)		
	Mean	SD	CV (%)	Mean	SD	CV (%)
1000.0 ng/mL	1.17	0.07	6.10	1.32	0.11	8.50
37.0 ng/mL	0.81	0.05	5.79	0.97	0.12	12.04
12.3 ng/mL	0.51	0.04	8.32	0.51	0.07	13.58
Overall			11.88			19.12

a. Three concentrations of the monoclonal gp42-IgG antibody standards were produced in house (ref 17)

Supplemental Table 2. Detection of EBERs in immunohistochemistry and HLA-DRA in situ hybridization of 27 atypical hyperplasia in nasopharynx.

HLA-DRA	EBERs		Total
	Positive	Negative	
Positive	12	1	13
Negative	8	6	14
Total	20	7	27