


Epstein Barr virus antibody and cancer risk in two prospective cohorts in Southern China

Received: 2 August 2024

Accepted: 6 June 2025

Published online: 01 July 2025

 Check for updates

Ming-Fang Ji^{1,14}, Yong-Qiao He^{2,14} , Min-Zhong Tang^{3,4,14}, Wen-Qiong Xue^{2,14}, Xia Yu^{1,14}, Hua Diao⁵, Da-Wei Yang⁵, Zhi-Ming Mai^{6,7}, Io Hong Cheong⁸, Zhi-Yang Zhao⁵, Biao-Hua Wu¹, Fu-Gui Li¹, Ji-Yun Zhan⁹, Chang-Ling Huang⁵, Hao-Lin Ma¹⁰, Jun Li^{3,4}, Yan-Cheng Li^{3,4}, Tong-Min Wang^{10,2}, Ying Liao^{10,2}, Xue-Yin Chen², Zhi-Heng Liang¹, Shi-Feng Lian¹, Yun Du¹¹, Xue-Jun Liang⁹, Zisis Kozlakidis¹², Jun Ma^{2,13} & Wei-Hua Jia^{10,2,5} 

Epstein-Barr virus (EBV) has been implicated in several human cancers, but its broader cancer risk remains unclear. We investigated the association between EBV VCA-IgA antibody levels and cancer risk in two large prospective cohorts from Southern China, comprising 73,939 adults. During around 8–10 years follow-up, 964 and 1026 incident cancer cases were identified in the Zhongshan and Wuzhou cohorts. VCA-IgA seropositivity was associated with higher age-standardized incidence rates for total cancer significantly. In pooled analyses, VCA-IgA seropositive individuals had higher risks of total cancer (HR 4.88, 95% CI: 2.84–8.37), lung cancer (1.76, 1.23–2.54), liver cancer (1.70, 1.10–2.63), nasopharyngeal carcinoma (26.05, 11.77–57.65), and lymphoma (3.20, 1.46–6.99) compared to seronegative individuals. The associations showed an increased dose-response pattern, and keep persistent even up to ten years prior to diagnosis. The population-attributable risk percentage for total cancer due to VCA-IgA seropositivity is estimated at 7.8%. These findings provide prospective evidence that EBV seropositivity is associated with increased risks of multiple cancers. This association results in a heightened attributed cancer burden in Southern China.

Epstein-Barr virus (EBV) is the most common and persistent human virus, affecting ~95% of the global population with asymptomatic life-long infection¹. While widespread in its host population, only a small proportion of EBV infections have been linked to human diseases, particularly those of lymphocytic and epithelial origin, with cancer being the most severe manifestation. Given its oncogenesis roles, EBV has been classified as a Group I carcinogen by the International Agency for Research on Cancer since 1997. Nevertheless, to date, only limited cancer types, such as certain histological subtypes of lymphomas, nasopharyngeal carcinoma (NPC), lymphoepithelioma-like carcinoma (LELC), and a minor proportion (~10%) of stomach cancer (although in controversial), have been directly linked to EBV infection^{2–4},

accounting for 239,700–357,900 new cases of cancer and 137,900–208,700 cancer deaths worldwide in 2020^{5,6}.

EBV primarily targets oropharyngeal epithelial cells and subsequently infects B cells, establishing a latent infection that allows the virus to evade the host's immune response. Under certain conditions, EBV can undergo spontaneous reactivation, switching from latent to lytic phase, wherein the virus replicates, produces new viral particles, and induces specific antibodies targeting various EBV proteins⁷. The detection of antibodies against EBV proteins in the blood has been widely observed in several EBV-associated malignancies, including NPC. Among the over 80 gene products of EBV, the Viral Capsid Antigen (VCA) emerges as a key structure protein expressed in the late

phase of lytic replication, playing a vital role in viral assembly and transmission⁸. While VCA-IgA is not a definitive marker of lytic reactivation, its elevated levels are strongly associated with increased EBV activity and cancer risk⁹. Numerous prospective cohort studies consistently demonstrate associations of an increased risk of NPC, with abnormal antibody elevation persisting several years before NPC diagnosis^{10,11}. At present, the detection of anti-VCA-IgA has been recommended by the Chinese Ministry of Health for NPC screening. Additionally, a series of nested case-control studies indicated associations between elevated EBV antibodies, including VCA-IgA, and the risks of Hodgkin lymphoma and Burkitt lymphoma^{12,13}.

Despite challenges in infecting most normal cells, recent comprehensive pan-cancer genome and whole-transcriptome analyses have revealed a notable prevalence of EBV in the host genome^{14,15}. A most recent study discovered a previously unappreciated link between EBV infection and host genomic instability across 2439 tumors from 38 cancer types, suggesting that EBV infection might predispose individuals to a broad spectrum of cancer types¹⁵. These recent discoveries, along with earlier individual cancer studies, provide potential evidence that certain uncontrolled EBV activity may contribute to the development of cancers widely beyond the established types. Despite these insights, a comprehensive understanding of the associations between EBV VCA-IgA antibodies and the risk of other cancer types remains limited.

In light of the hypothesis that uncontrolled EBV activity might broadly contribute to cancer development, we conducted a two-site, prospective cohort study. This systematic investigation assessed the association of EBV VCA-IgA antibodies with cancer risks in 73,939 adults with a median follow-up approaching a decade in Southern China, an endemic area for NPC. This study not only explores individual cancer risks but also assesses overall cancer risks, aiming to evaluate the EBV-attributed cancer burden. The serostatus of VCA-IgA served as a risk factor, measuring the host immune reaction against EBV.

Results

Characteristics of the cohorts at baseline

Among the 29,026 participants in the Zhongshan cohort (median age 46 years, 50.9% males), 1910 (6.6%) were VCA-IgA seropositive. In the univariable analysis, elder age, male, lower education level, current smoking and more frequent alcohol consumption were positively associated with risk of VCA-IgA seropositivity, while higher BMI was negatively associated with VCA-IgA seropositivity. In the multivariable analysis, age, education level, BMI and smoking history were independently associated with the risk of VCA-IgA. Among 44,913 participants in the Wuzhou cohort (mean age 47 years, 40.3% males), 2520 (5.6%) were VCA-IgA seropositive, with age consistently associated with VCA-IgA seropositive risk (Table 1, Supplemental Table 1, and Supplemental Data 1).

Age-standardized incidence rate of cancer according to VCA-IgA serostatus

In the Zhongshan cohort, 964 participants developed cancer during 298,313 person-years (PYs) of follow-up, resulting in a cumulative age-standardized incidence rate of 275.0 per 100,000 PYs. Similarly, in the Wuzhou cohort, 1026 subjects developed cancer during 376,790 PYs of follow-up, with an ASR of 285.17 per 100,000 PYs (Supplemental Table 2). The ASR of total cancer in the VCA-IgA seropositive group was significantly higher than that in the seronegative group in both cohorts (Zhongshan cohort: 592.73 vs. 253.77 per 100,000 PYs, $P=1.32 \times 10^{-9}$; Wuzhou cohort: 650.10 vs. 262.45, $P=2.81 \times 10^{-11}$). Consistent with previous studies, VCA-IgA seropositive individuals showed a significantly higher incidence of NPC compared to seronegative individuals. Furthermore, seropositive individuals had significantly higher ASRs for specific cancers in each cohort. In the Zhongshan cohort,

Table 1 | Baseline characteristics of participants in Zhongshan and Wuzhou cohorts by VCA-IgA serostatus

Variable	No. of VCA-IgA negative (%)	No. of VCA-IgA positive (%)	P *
Zhongshan cohort			
Age, years	46 ± 8	47 ± 8	2.67 × 10 ⁻¹³
Sex			0.03
Female	13,365 (49.3)	893 (46.8)	
Male	13,751 (50.7)	1017 (53.2)	
Education level			1.85 × 10 ⁻⁶
University or more	2350 (8.7)	131 (6.9)	
High school	15,513 (57.2)	1023 (53.6)	
Primary school or less	9253 (34.1)	756 (39.6)	
BMI			0.150
18.5–22.9	11,083 (40.9)	823 (43.1)	
<18.5	908 (3.3)	70 (3.7)	
23.0–27.5	11,805 (43.5)	783 (41.0)	
>27.5	3320 (12.2)	234 (12.3)	
Waist circumference			0.315
Normal (Male <85, female <80)	12,165 (44.9)	880 (46.1)	
Abnormal (Male ≥85, female ≥80)	14,951 (55.1)	1030 (53.9)	
Smoking history			1.80 × 10 ⁻⁹
Never smoker	22,002 (81.1)	1457 (76.3)	
Ex-smoker	1217 (4.5)	77 (4.0)	
Current smoker	3897 (14.4)	376 (19.7)	
Alcohol consumption			8.35 × 10 ⁻⁵
Less than monthly	22,312 (82.3)	1527 (79.9)	
Monthly or weekly	3512 (13.0)	250 (13.1)	
Everyday	1292 (4.8)	133 (7.0)	
Physical activity			0.040
Less than monthly	7096 (26.2)	497 (26.0)	
Monthly or weekly	9751 (36.0)	639 (33.5)	
Everyday	10,269 (37.9)	774 (40.5)	
Wuzhou cohort			
Age, years	47 ± 9	48 ± 9	1.44 × 10 ⁻⁹
Sex			0.744
Female	25,315 (59.7)	1496 (59.4)	
Male	17,078 (40.3)	1024 (40.6)	

BMI body mass index.
*The P values for comparisons between the VCA-IgA seropositive and seronegative groups were calculated using the Student's t-test for continuous variables (age) and the Chi-square test for categorical variables, including gender, education level, BMI, waist circumference, smoking history, alcohol consumption, and physical activity. All the tests were two-sided.

seropositive individuals showed significantly higher ASRs for lung cancer (62.75 vs. 37.16, $P=0.02$), liver cancer (60.65 vs. 31.30, $P=0.01$), nasopharyngeal carcinoma (233.67 vs. 13.23, $P=1.80 \times 10^{-99}$), and lymphoma (22.67 vs. 5.39, $P=1.59 \times 10^{-6}$). In contrast, in the Wuzhou cohort, significant differences in ASRs were observed for liver cancer (47.68 vs. 24.86, $P=0.04$) and nasopharyngeal carcinoma (345.47 vs. 8.36, $P<1.00 \times 10^{-100}$), while not for lung cancer or lymphoma (Fig. 1 and Supplemental Table 3).

Associations between VCA-IgA serostatus and the risk of total and site-specific cancers

In the combined analysis of two cohorts, the adjusted hazard ratios of VCA-IgA seropositive, compared to VCA-IgA seronegative counterparts, were 4.88 (95% confidence interval [CI]:

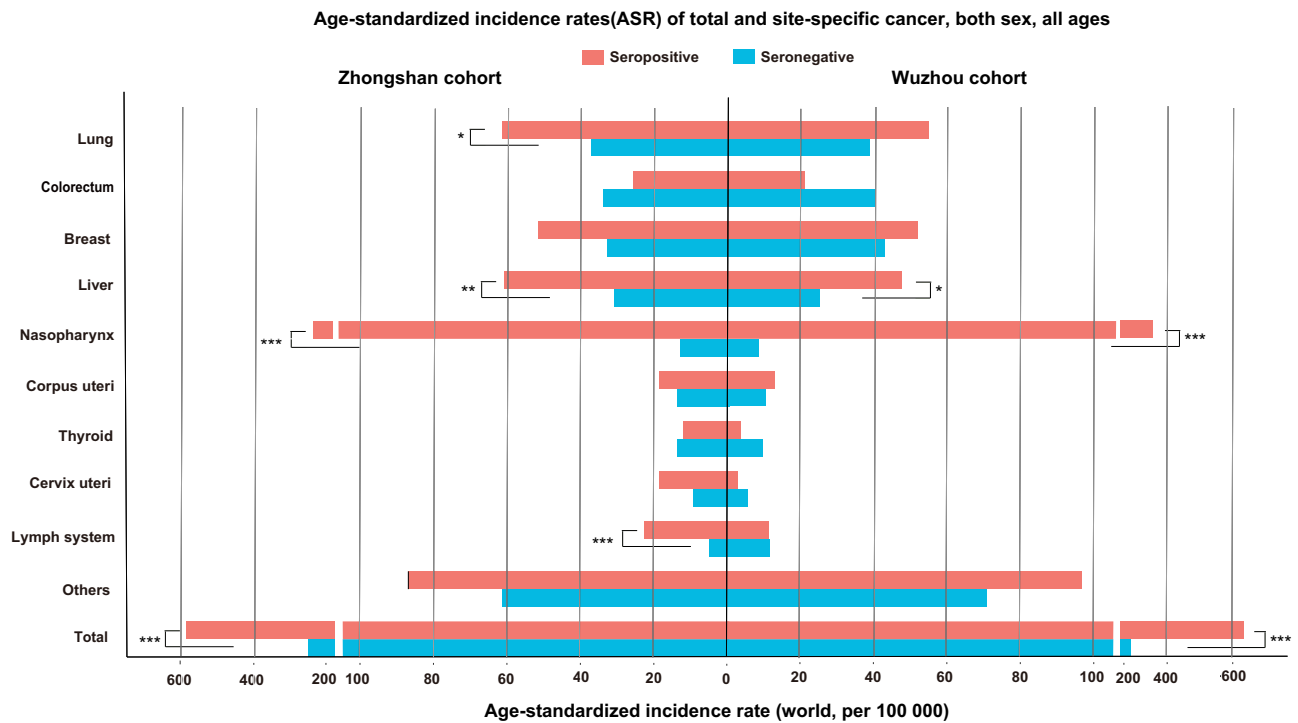


Fig. 1 | Age-standardized cancer incidence in Zhongshan and Wuzhou cohorts by VCA-IgA serostatus. The age-standardized incident rates were generated based on the world standard population. The comparison between the two groups were conducted using standardized rate ratios. All the tests were two-sided. The individuals with seropositive and seronegative VCA-IgA antibodies were indicated using

red and blue colors, respectively. The comparison between seropositive and seronegative individuals was marked with stars if the incidences changed statistically significantly. * represents P values less than 0.05 and more than 0.01; ** represents P values less than 0.01 and more than 0.001; *** represents P values less than 0.001. Source data are provided as a Source Data file.

2.84–8.37) for total cancer, 1.76 (95%CI: 1.23–2.54) for lung cancer, 1.70 (1.10–2.63) for liver cancer, 26.05 (11.77–57.65) for nasopharyngeal carcinoma, and 3.20 (1.46–6.99) for lymphoma (see Fig. 2 and Supplemental Table 4). In Zhongshan cohort, the adjusted hazard ratios for VCA-IgA seropositive participants were 3.70 (2.67–5.14) for total cancer, 1.67 (1.02–2.74) for lung cancer, 1.79 (1.02–3.14) for liver cancer, 17.34 (11.21–26.82) for nasopharyngeal carcinoma, 5.00 (1.97–12.72) for lymphoma. In Wuzhou cohort, the hazard ratios were 6.42 (4.65–8.85) for total cancer, 1.88 (1.10–3.21) for lung cancer, 1.58 (0.80–3.14) for liver cancer, 39.00 (25.55–59.52) for nasopharyngeal carcinoma, 1.10 (0.26–4.63) for lymphoma (Supplemental Data 2).

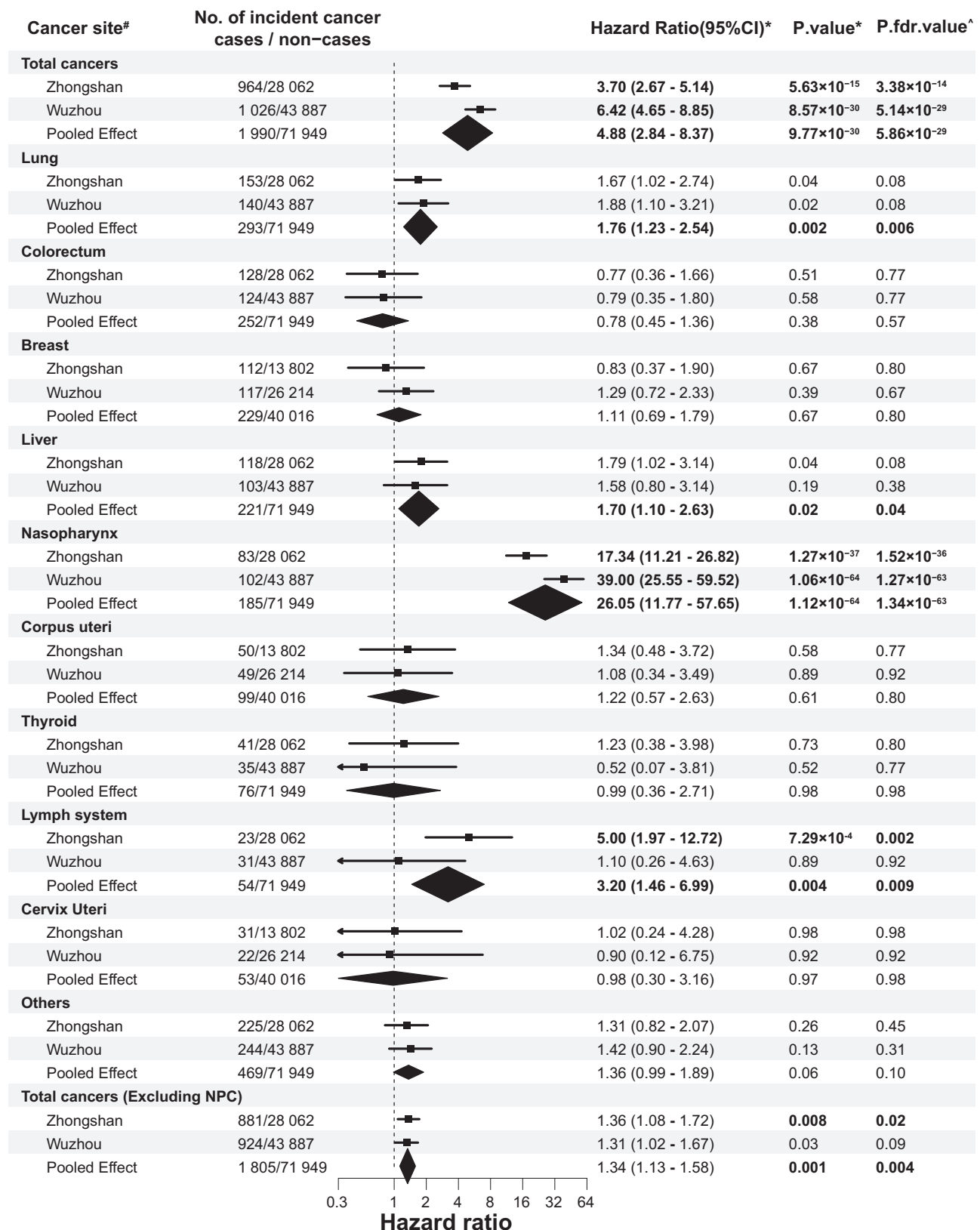
To evaluate the robustness of the results, we conducted sensitivity analyses by regarding cancers other than the outcome as competing events, as well as excluding patients diagnosed within the initial one to three years of follow-up. The associations between VCA-IgA antibody and cancer risks (including total cancers, lung, liver, nasopharynx, and lymphoma) were consistent with our primary findings (Supplemental Data 3, 4). Stratified analyses by gender and age also demonstrated similar associations across various subgroups (Supplemental Tables 5, 6). Subgroup analyses by histological type showed significant associations between VCA-IgA antibody and increased risks of small-cell lung cancer and hepatocellular carcinoma (Supplemental Table 7). Matching the age distribution in the Zhongshan cohort, we analyzed participants aged 30–59 in the Wuzhou cohort. The outcomes paralleled those observed for ages 30–69 (Supplemental Table 8). After adjusting for HBV infection status, the association between elevated VCA-IgA antibody levels and liver cancer risk remained statistically significant (adjusted HR = 2.02, 95% CI: 1.06–3.87, P = 0.034, Supplemental Table 9). This

finding demonstrates that the observed relationship between VCA-IgA and liver cancer risk may not be confounded by HBV infection. Additionally, we observed a significant synergistic interaction between VCA-IgA and HBsAg in relation to liver cancer risk. Compared to individuals negative for both VCA-IgA and HBsAg, the HRs were 1.51 for those with VCA-IgA positivity alone (VCA-IgA (+) & HBsAg (–)), 34.93 for those with HBsAg positivity alone (VCA-IgA (–) & HBsAg (+)), and 73.15 for individuals positive for both markers (VCA-IgA (+) & HBsAg (+)). The attributable proportion (AP) for interaction was 0.52 (95% CI: 0.18–0.85, P for interaction = 0.001, Supplemental Table 10). These results suggest a synergistic effect between elevated VCA-IgA levels and HBV infection, significantly amplifying liver cancer risk. To further evaluate the robustness of our findings, we conducted sensitivity analyses using multiple adjustment models in the Zhongshan cohort, ranging from unadjusted to partially and fully adjusted models for all available covariates. Across all eight models, the association between VCA-IgA antibody levels and cancer risk remained stable and statistically significant (Supplemental Table 11).

Further adjustments to different covariables produced similar associations between VCA-IgA and cancer risks in the Zhongshan cohort (Supplemental Table 11).

Dose-response and time-dependent associations between VCA-IgA antibody level and cancer risks

In the Zhongshan cohort, the restricted cubic spline model, using VCA-IgA antibody level as a continuous exposure, revealed a positive association between VCA-IgA antibody level and risk of total cancer. This indicates that the risk of total cancer is enhanced with a dose-response relationship corresponding to the



elevation of VCA-IgA antibody ($P = 6.95 \times 10^{-5}$, Fig. 3A). Similar trends were also observed for lung cancer, liver cancer, NPC and lymphoma (Supplemental Tables 12, 13). Additionally, we categorized EBV antibody levels into four groups based on the quartiles and observed a consistent trend that the adjusted HR for

total cancer increased as the VCA-IgA antibody level increased ($P_{\text{trend}} = 2.90 \times 10^{-8}$, Supplemental Table 14). Furthermore, elevated levels of VCA-IgA were consistently and significantly associated with an increased risk of cancer, even up to 10 years prior to diagnosis (Fig. 3B).

Fig. 2 | Associations of VCA-IgA serostatus with cancer risk in Zhongshan and Wuzhou cohorts. # According to the results of the proportional hazards (PH) assumption test, hazard ratios (HRs), and *P* values for total cancer were estimated using a time-dependent Cox regression model, where VCA-IgA was modeled as a time-varying covariate. For site-specific cancers, the PH assumption was satisfied, and HRs were estimated using standard Cox proportional hazards models with fixed covariates. *In the Zhongshan cohort, the HR and *P* values in the multivariable analysis were adjusted by age, sex (male and female), education level (classified as primary school or less, high school, and university or more), body mass index (normal weight: 18.5–22.9, underweight: <18.5, overweight: 23.0–27.5 and obese: >27.5 kg/m²), waist circumference (normal: <85 cm for male or <80 cm for female; abnormal: ≥85 cm for male or ≥80 cm for female), smoking history (categorized as

never smoker, ex-smoker, current smoker), alcohol consumption (grouped into less than monthly, monthly or weekly, once or more a day) and physical activity (grouped into less than monthly, monthly or weekly, once or more a day). In the Wuzhou cohort, the HR and *P* values were calculated using Cox regression models adjusting for age and sex (male and female). HR and *P* values for total cancers were adjusted for the aforementioned variables and for the time-varying variable of VCA-IgA. All the tests were two-sided. The solid squares in the center of the error bars are the HR values, and the error bars are the corresponding 95% confidence intervals of the HRs. The solid diamonds represent the pooled HR and the corresponding 95% confidence intervals of Zhongshan and Wuzhou cohorts. All the tests were two-sided. **P*. FDR value represents that the adjusted *P* value was calculated by the false discovery rate (FDR) method.

Cancer burden attributed to VCA-IgA seropositivity in Southern China

Based on the identified EBV-cancer associations, we estimated the cancer burden attributed to VCA-IgA seropositivity. The attributable risk percentage (ARP) of total cancers among VCA-IgA seropositive individuals was 58.5% (95% CI: 53.3–63.7%), indicating that ~58.5% of all VCA-IgA seropositive cancer cases can be attributed to EBV. The PARP for total cancer was 7.8% (6.3–9.4%), indicating that about 7.8% of cancers can be attributed to VCA-IgA seropositivity at the population level in Southern China. The calculated exposure impact number (EIN) of 28.6 suggests that, among every 28.6 VCA-IgA seropositive individuals, one additional cancer patient can be attributed to VCA-IgA seropositivity. Similarly, the case impact number (CIN) of 12.8 indicates that for every 12.8 cancer cases, one case would be attributable to VCA-IgA seropositivity. These figures highlight the notable etiological impact of VCA-IgA seropositivity, thereby reinforcing the significant burden of EBV-associated cancers within the region. Furthermore, when excluding NPC from the total cancer burden calculation, the PARP was ~2.6% (1.3–4.0%), meaning that about 2.6% of total cancers, excluding NPC in these areas, can be attributed to VCA-IgA seropositivity (Table 2 and Supplemental Table 15).

Epstein-Barr Virus-encoded RNA (EBER) detection in lung and liver cancer tissues

We performed EBER-ISH analysis on tumor tissues from 129 lung cancer and 348 liver cancer cases. EBER positivity was observed in 1.6% (2/129) of lung cancer samples and 1.1% (4/348) of liver cancer samples, while none of the adjacent normal tissues showed positivity. Among the EBER-positive cases, the lung cancer samples were squamous cell carcinoma, and the liver cancer samples were intrahepatic cholangiocarcinoma. In both lung and liver cancer, EBER signals were primarily detected in tumor cells (Fig. 4).

Discussion

In this population-based prospective cohort study across two sites involving over 70,000 individuals, with nearly a decade of follow-up, we provided robust epidemiological evidence supporting the associations between EBV VCA-IgA antibodies and a wide spectrum of cancer risks. This comprehensive assessment expands beyond traditional EBV-associated cancers, such as NPC and lymphoma, to include prevalent malignancies such as lung and liver cancers.

During the past 60 years since the discovery of EBV, epidemiologists worldwide have endeavored to delineate the intricate relationships between EBV activity status, host immune response, and cancer risk. These efforts, crucial in guiding mechanistic research and influencing clinical diagnosis and treatment, have predominantly linked specific EBV-related malignancies—such as endemic Burkitt lymphoma in Africa, NPC in Asia, and rare lymphoepithelioma-like carcinoma (LELC)—to EBV carcinogenesis¹. However, common cancer types, notably lung and liver cancers, have seldom been investigated in large-scale epidemiological studies,

leading to an insufficient exploration of the overall cancer burden attributed to EBV.

Although previous epidemiological studies have individually investigated the associations between anti-EBV antibodies and risks of site-specific cancer^{12,13,16–22}, most of which were retrospective case-control designs, few utilized nested case-control designs, often limited by modest or small sample sizes. Notable exceptions include large-scale screening cohorts for NPC established in high-incidence areas of East Asia, such as Guangdong¹¹, Guangxi²³, and Taiwan¹⁰. Our study builds upon this foundation, reinforcing the association of EBV seropositivity with NPC and revealing a 20–30-fold higher risk in seropositive individuals. Regarding lymphoma, earlier studies have observed the associations between EBV antibodies and the risk of Burkitt lymphoma in children and Hodgkin lymphoma in adults. A prospective cohort study conducted in Uganda in 1978 revealed that children with high antibody titers to the VCA face an elevated risk of developing Burkitt Lymphoma, utilizing a nested case-control design with 14 incident cases and 69 controls¹². Similarly, another prospective cohort carried out in the United States in 1989 demonstrated that participants with elevated VCA-IgA antibody had a significantly higher risk for Hodgkin's disease compared to controls. This study employed a nested case-control design and included 43 incident cases and 96 controls¹³. Our study provided large-scale prospective evidence between elevated EBV antibodies and lymphoma risk in Southern China. Seropositive individuals exhibited an approximately threefold increased risk of lymphoma compared to the seronegative ones in the combined analysis. However, the significant association between VCA-IgA and lymphoma was observed in the Zhongshan cohort, while no such association was found in the Wuzhou cohort. This discrepancy may be due to regional differences in lymphoma incidence. According to the China Cancer Registry Annual Report (2014), the incidence of lymphoma is higher in Zhongshan compared to Wuzhou (ASR: 7.19 vs. 4.06 per 100,000 PYs for males and 4.53 vs. 2.42 per 100,000 PYs for females). Additionally, previous studies have indicated that different antibodies against EBV, such as VCA-IgG, not VCA-IgA, are more commonly associated with the risk of lymphomas, especially for Burkitt lymphoma and non-Hodgkin lymphoma²⁴. These differences highlight the complex relationship between EBV serology and lymphoma risk, which may depend on the specific antibody type, lymphoma subtype.

Beyond NPC and lymphomas, studies linking EBV to other cancer types were limited, especially in prospective designs. Our study establishes prospective links between VCA-IgA seropositivity and the risks of common cancers such as lung and liver cancers. EBV is well-known for its association with lymphoepithelioma-like lung cancer and lymphoepithelioma-like intrahepatic cholangiocarcinoma, rare subtypes characterized by undifferentiated carcinoma cells surrounded by dense lymphoid stroma. In these cancers, EBV plays a direct and central oncogenic role, as evidenced by its frequent detection within tumor cells^{25–27}. However, the cancer types analyzed in our study,

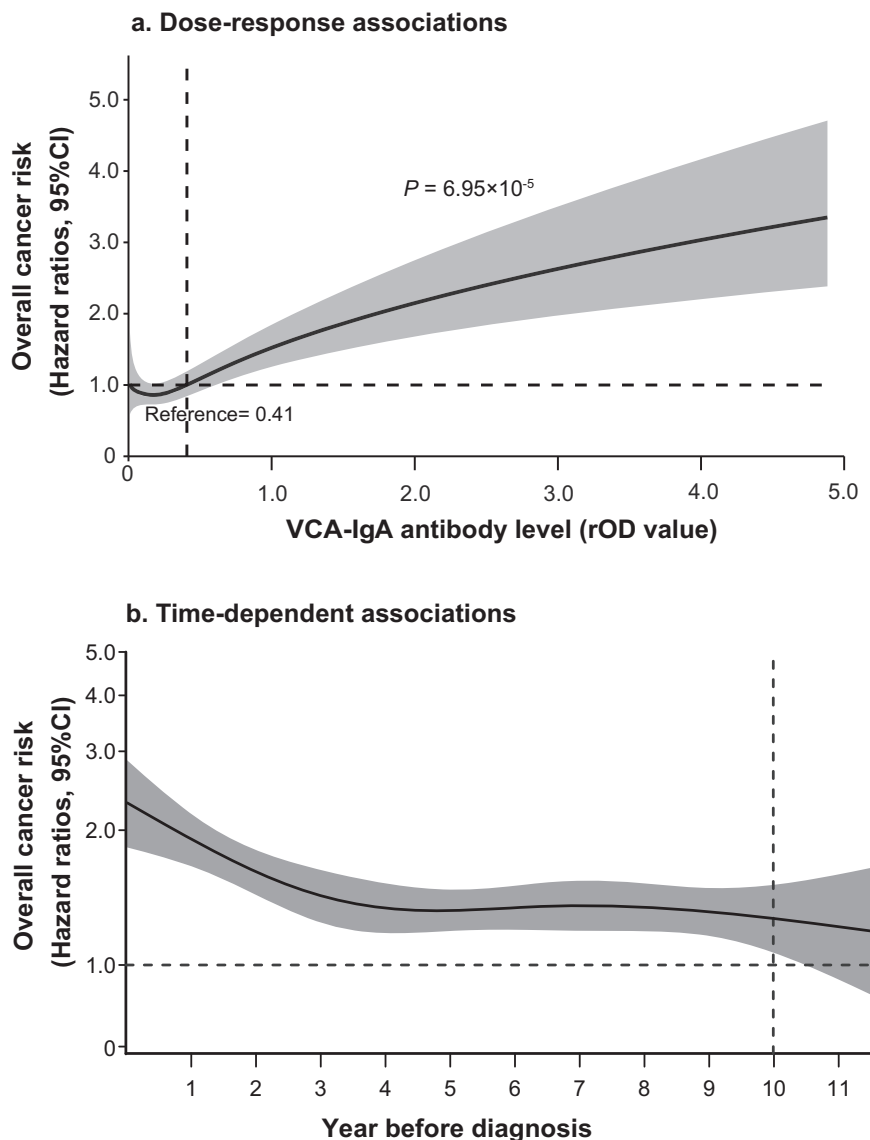


Fig. 3 | Dose-response and time-dependent associations between VCA-IgA levels and cancer risk in Zhongshan cohort. a Dose-response association of VCA-IgA antibody level with overall cancer risk. The horizontal dashed line represents the referent hazard ratio (HR) of 1. The vertical line indicates that the HR of 1 corresponds to a relative optical density (rOD) of VCA-IgA of 0.41; The shaded area represents the 95% confidence interval for the HRs. **b** Time-dependent associations of VCA-IgA antibody with overall cancer risk. The horizontal dashed line indicates a referent HR of 1. The vertical line signifies that elevated VCA-IgA antibody was significantly associated with overall cancer risk up to ten years before cancer diagnosis. The solid line depicts the variation in overall cancer risk with VCA-IgA antibodies over time prior to diagnosis. The shaded area represents the 95%

confidence interval for the HRs. The HR and P values were calculated using cox regression models adjusted by age, sex (male and female), education level (classified as primary school or less, high school, and university or more), body mass index (normal weight: 18.5–22.9, underweight: <18.5, overweight: 23.0–27.5 and obese: >27.5 kg/m²), waist circumference (normal: <85 cm for male or <80 cm for female; abnormal: ≥85 cm for male or ≥80 cm for female), smoking history (categorized as never smoker, ex-smoker, current smoker), alcohol consumption (grouped into less than monthly, monthly or weekly, once or more a day) and physical activity (grouped into less than monthly, monthly or weekly, once or more a day). All the tests were two-sided.

including small-cell lung cancer and hepatocellular carcinoma, are more prevalent and exhibit complex, multifactorial etiologies. The significant associations observed between VCA-IgA antibody levels and the risks of these cancers, along with the limited presence of EBV in tumor tissues, suggest that EBV may contribute indirectly to tumorigenesis in these cancers. Potential mechanisms include immune modulation, chronic inflammation, host genomic instability, or synergistic interactions with established risk factors, such as smoking for lung cancer and hepatitis virus infections for liver cancer. These findings suggest that while EBV's oncogenic role may vary across cancer subtypes, its indirect contributions to tumorigenesis warrant further investigation.

We hypothesize that environmental risk factors, such as smoking and alcohol consumption, may influence EBV activity, thereby modulating cancer risk. Smoking-induced chronic inflammation, oxidative stress, and immune suppression could enhance EBV activity, as supported by prior studies showing that cigarette smoke extract upregulates EBV lytic gene expression⁹. Additionally, our unpublished data identified a significant interaction between smoking and VCA-IgA antibody levels, suggesting a synergistic effect that elevates NPC risk. These findings support the hypothesis that smoking-EBV interactions may contribute to lung carcinogenesis, warranting further investigation to clarify their role in lung cancer development. For liver cancer, chronic alcohol consumption, known to induce liver inflammation and

Table 2 | Attributable fractions and impact numbers for cancer risk due to VCA-IgA seropositivity

Cancer types	Attributable fractions and impact numbers	Zhongshan cohort	Wuzhou cohort	Combined cohort
Total cancers [*]				
Attributable risk percentage, % (95% CI)		52.7 (44.1–61.3)	62.4 (55.8–68.9)	58.5 (53.3–63.7)
Population attributable risk percentage, % (95% CI) [*]		6.3 (4.0–8.5)	9.1 (6.87–11.2)	7.8 (6.3–9.4)
Exposure impact number		29.0 (21.8–43.1)	28.9 (22.9–39.0)	28.6 (23.8–35.8)
Case impact number		14.6 (11.0–21.6)	11.7 (9.35–15.8)	12.8 (10.7–16.0)
Total cancers excluding nasopharyngeal carcinoma [*]				
Attributable risk percentage, % (95%CI)		31.0 (15.5–46.6)	28.4 (11.0–45.7)	31.0 (19.6–42.3)
Population attributable risk percentage, % (95%CI) [*]		2.6 (0.6–4.6)	2.3 (0.6–4.1)	2.6 (1.3–4.0)
Exposure impact number		75.1 (44.0–256.0)	125.0 (68.1–768.0)	93.5 (61.5–195.0)
Case impact number		35.5 (20.8–121.0)	47.2 (25.7–287.0)	39.1 (25.8–81.5)

^{*} “Total cancers” refers to all incident cancer cases.
^{*} The prevalence of VCA-IgA seropositivity is 0.06.
^{*} “Total cancers excluding nasopharyngeal carcinoma” refers to all incident cancer cases except for nasopharyngeal carcinoma.

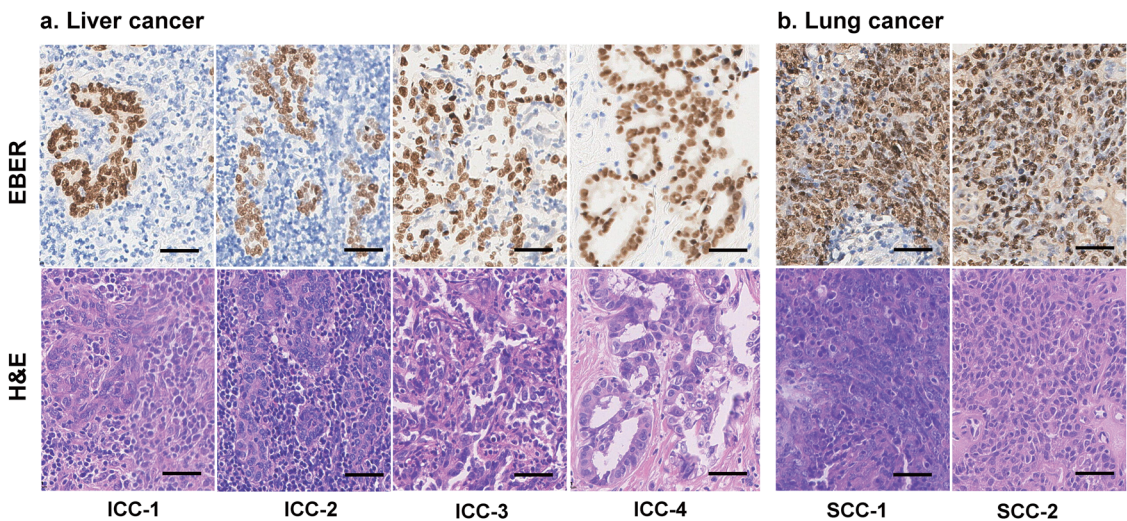


Fig. 4 | Histopathological images of Epstein-Barr virus-encoded small RNA (EBER)-positive tumor tissues. In situ hybridization (ISH) for Epstein-Barr virus-encoded RNA (EBER) and corresponding hematoxylin and eosin (H&E) staining were performed in EBER-positive liver (a) and lung (b) tumor tissues. Images are

shown at 200×magnification; scale bar = 50 μm. ICC intrahepatic cholangiocarcinoma, SCC squamous cell carcinoma. Each staining was independently performed twice per sample with consistent results.

immune dysfunction²⁸, may similarly interact with EBV to facilitate carcinogenesis. Notably, our findings demonstrated a significant interaction between VCA-IgA and HBV infection, underscoring the intricate interplay between EBV and environmental factors in liver cancer. These observations highlight the urgent need for longitudinal studies and experimental models to confirm these associations and unravel the mechanisms underlying these complex interactions.

Furthermore, the detection of EBER positivity in a small subset of lung (1.6%) and liver (1.1%) cancer tissues suggests a potential role for EBV in these malignancies. EBER signals were primarily localized to tumor cells, providing direct evidence of EBV presence in these cancers. For lung cancer, our findings are consistent with previous reports of EBV detection in histological subtypes such as squamous cell carcinoma²⁹. Similarly, for liver cancer, the detection of EBV in intrahepatic cholangiocarcinoma aligns with molecular studies identifying EBV transcripts in liver cancer cases among Chinese and Japanese populations^{30,31}. Recent genomic and transcriptomic studies have highlighted the broad presence of EBV across various cancers^{14,15}. Notably, EBV has been implicated in promoting host genomic instability, with the Epstein-Barr nuclear antigen-1 (EBNA1) protein known to induce chromosomal breaks at 11q23, thereby contributing to tumorigenesis¹⁵. Additionally, EBV-related proteins and transcripts have been shown to modulate the immune response, drive epigenetic

alterations, and facilitate tumor progression³². The sporadic detection of EBV in these cancer types may be explained by the “hit and run” hypothesis³³. This theory posits that EBV may act transiently during the early stages of tumorigenesis, inducing genetic or epigenetic alterations that initiate cancer development, but its presence diminishes in later stages. Such transient viral activity could lead to genomic instability, chronic inflammation, or immune modulation, contributing to a tumor-promoting environment. It is important to note that the EBV positivity rate in these tissues may be underestimated due to the limited sensitivity of ISH in detecting low viral loads, latent viral states, or RNA degradation^{34–36}.

Taken together, this study expands the spectrum of cancers associated with elevated VCA-IgA levels, highlighting a broader role for abnormal EBV activity in cancer development. Our findings align with emerging evidence indicating that uncontrolled EBV activity plays a role in oncogenesis by facilitating viral spread and regulating cellular oncogenic pathways mediated by lytic proteins and miRNAs^{25,32,37,38}, accompanied by compromised host immune defense. Uncontrolled EBV activity, characterized by the transition from latency to lytic replication due to impaired host immune regulation, plays a pivotal role in the development of EBV-associated cancers. Typically, EBV establishes a lifelong asymptomatic latent infection after primary infection, maintained under control by the host immune system.

However, disruptions to this balance, triggered by immune suppression, chronic inflammation, or other stressors-can lead to lytic reactivation, resulting in increased viral load and expression of oncogenic viral proteins. Both latent proteins, such as EBNA1 and LMP1²⁵, and lytic proteins like BGLF5³⁹, contribute to cancer progression through mechanisms such as immune evasion, genomic instability, and the modulation of cellular processes. These mechanisms collectively underscore the multifaceted role of uncontrolled EBV activity in cancer development, highlighting its broader implications and emphasizing the need for further research to clarify its contributions to tumorigenesis across diverse malignancies.

This study evaluated attributable fractions and impact numbers for overall cancer risk associated with VCA-IgA seropositivity in Southern China. With most cancers exhibiting positive associations, we identified an average population attributable risk percentage of overall cancer at 7.8%, which was over fourfold higher than the current global estimates of 1.9%⁶. This evidence emphasized that EBV activity, particularly under compromised immune response, could result in a substantial impact on cancer burden than previously anticipated. While our study focused on specific regions in Guangdong and Guangxi provinces, representing a potential risk for 180 million individuals, it may serve as a representative population for similar areas. Notably, excluding NPC reduces the PARP to 2.6%, indicating NPC's significant contribution to the overall cancer risk observed in Southern China. Additionally, removing lymphoma results in a slight decrease of the PAR to 2.41%. This suggests that EBV's oncogenic influence is more pronounced in specific cancers like NPC and lymphoma but less pronounced in others.

Our study has several limitations. Firstly, it was conducted in Southern China, where the local populations exhibit a distinct genetic background, and specific EBV subtypes are associated with NPC risk⁴⁰. Specifically, while heterogeneity analyses (I^2 and Cochrane Q tests) were performed to evaluate consistency between the Zhongshan and Wuzhou cohorts, finding minimal statistical differences for most cancer types, inherent differences in cohort characteristics, such as recruitment periods, sex distribution, and covariate availability, remain a limitation. These variations may introduce confounding that cannot be fully addressed through meta-analysis. Future studies with harmonized data collection across diverse populations are needed to confirm these findings.

Secondly, residual confounding should be carefully assessed. The Wuzhou cohort lacked detailed covariate information, potentially influencing hazard ratios, particularly regarding smoking. However, given that less than 1.0% of females are smokers in China⁴¹, subgroup analysis for females showed consistent results with total and male data, minimizing this concern. Sensitivity analyses in the Zhongshan cohort supported stable associations after adjusting for different covariables, reinforcing the reliability of our findings. We acknowledged that the ASRs for lung cancer and lymphoma in VCA-IgA seropositive individuals were not significantly higher than the seronegative ones in the Wuzhou cohort. This discrepancy may reflect variability in cohort characteristics or other region-specific factors. As such, replication in other studies is essential to confirm these findings. We also emphasize that our two-stage study design provides observational associations that should be interpreted with caution to avoid overstating conclusions. Thirdly, the potential for reverse causation was addressed. Sustained elevation of EBV antibody levels up to ten years prior to a cancer diagnosis and sensitivity analysis excluding incident cases during the first, second and third years of follow-up support the temporal validity of our results. Fourthly, different VCA-IgA detection methods were used in the two methods. Since both methods are widely used for NPC screening with comparable seropositive rates (6.6% and 5.6%), it is unlikely that the different methods influenced the main results. While our study focused on VCA-IgA as a biomarker for NPC screening, we recognize the potential value of examining

additional EBV-related antibodies to provide a more comprehensive understanding of EBV's role in carcinogenesis. For instance, VCA-IgG, which reflects past EBV infection, has been associated with other EBV-related malignancies, such as Hodgkin lymphoma²⁴. Antibodies against early antigen (EA) and lytic or latent-phase proteins like Zta, Rta, EBNA1, and BNLF2b⁴² could offer valuable insights into different stages of EBV activity. Lastly, the use of a single time-point measurement of VCA-IgA restricts our ability to assess its longitudinal fluctuations and their potential implications for cancer risk. Future research leveraging longitudinal designs and serial measurements of EBV-related biomarkers would provide deeper insights into the temporal dynamics of VCA-IgA antibodies and their association with cancer risks.

In conclusion, this study broadened the spectrum of cancers associated with EBV, revealing varying degrees of strength in these associations and underscoring a greater cancer burden than previously estimated. These findings emphasize the need for further research to clarify EBV's role in diverse malignancies and the mechanisms underlying its oncogenic activity. Establishing causal links between EBV activity and cancer development requires robust mechanistic studies and interventional trials. Additionally, longitudinal and geographically diverse studies are essential to better evaluate the contribution of EBV abnormal activity, such as the event of reactivation, to cancer risks while accounting for potential confounding factors. Longitudinal studies are warranted to explore these associations across wide geographical regions, providing a more comprehensive evaluation of the contribution of EBV activity to the global cancer burden.

Methods

The study was approved by the ethical review committees of Sun Yat-sen University Cancer Center (B-2023-714-01), adhering to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. All participants signed informed consent.

Study design and population

As a part of China's National Cancer Screening Project, we initiated large-scale, cluster-randomized controlled trials for screening of NPC and other cancers in endemic areas of Southern China since 2009 (ClinicalTrials.gov identifiers: NCT04085900; NCT00941538). The primary objective of this trial was to compare the early diagnosis rate and mortality rate of NPC between the screening and control groups. In the screening group, participants underwent EBV antibody tests to identify high-risk individuals for further clinical evaluations, while the control group received no interventions. The eligibility criteria have been outlined previously (Supplementary Note S1)^{43–46}.

Zhongshan city in Guangdong province and Wuzhou city in Guangxi province were among the pioneers to participate in the National NPC screening project in Southern China. Based on this screening project, we executed a population-based prospective cohort study, encompassing all participants from the screening groups in Zhongshan and Wuzhou cities who already had EBV antibody tests, to investigate the association between EBV antibodies and cancer risks (ClinicalTrials.gov identifiers: NCT06203314). Briefly, the Zhongshan cohort, consisted of 29,026 individuals aged 30–59, recruited from Zhongshan city between 2009 and 2014. The Wuzhou cohort, comprised 44,913 participants aged 30–69, enrolled in 2014.

Follow-up and outcomes

The participants were followed up annually for cancer incidence, vital status and immigration status through the Cancer Registry, Death Registry, and Population Registry of Zhongshan City and Wuzhou City. Incident cancer cases were ascertained from the Cancer Registry of both cities, both of which are included in the WHO-IARC Cancer Incidence in Five Continents. Detailed diagnosis and clinical data for all incident cancer cases were reviewed by an expert panel of two

pathologists who were blinded to any unique information in this study. The primary outcome of the study was the diagnosis of any type of cancer. The cancer type for each participant was defined as their initial and primary cancer diagnosis. The cancer type was categorized based on ICD-10 and ICD Oncology codes (ICD-O-1–3). To ensure sufficient power for subsequent analyses, we showed cancer types with more than fifty patients in the cohorts. Based on the ICD-10 and ICD-O-1–3, we categorized cancers into nine types: lung (C33–34), colon and rectum (C18–20), breast (C50), liver (C22), nasopharynx (C11), corpus uteri (C54), thyroid (C73), cervix uterus (C53), and lymphomas (C81–85, C88, C90, C96). Types with fewer than fifty patients were combined and assigned to the “others” group. The follow-up time was defined as the duration from the baseline visit to the diagnosis for incident cancer cases, or until the last follow-up date (Zhongshan cohort: June 30, 2021; Wuzhou cohort: December 31, 2022) for non-cases.

EBV antibody test

The methodology for the VCA-IgA antibody test in both the Zhongshan and Wuzhou cohorts adhered to standard laboratory procedures and has been widely applied in Southern China^{11,23,43,46,47}. In the Zhongshan cohort, the serum VCA-IgA antibody level was determined using an ELISA kit (EUROIMMUN AG, Lübeck, Germany) following the manufacturer's instructions (Supplementary Note S2)^{11,43}. For each tested sample, the results were expressed as relative optical density (rOD), which was calculated as the ratio of the sample's OD value to that of the reference calibrator. A sample with a rOD value ≥ 1 was considered positive for VCA-IgA, while a sample with a rOD value < 1 was considered negative for VCA-IgA. In the Wuzhou cohort, serum VCA-IgA antibody levels were determined by titration using an Immunoenzymatic assay (Supplementary Note S2)^{23,46,47}. In brief, cell smears from B95-8 cultures were prepared, fixed in acetone and used in the indirect Immunoenzymatic method with peroxidase-conjugated anti-human IgA antibody. A serum was considered positive if the cells in the well containing the 1:5 dilution showed brown color; otherwise, the sample was considered negative for VCA-IgA.

HBV infection test

To assess the potential confounding effect of HBV infection, we utilized HBV infection data from a mass liver cancer screening program conducted in Zhongshan City in 2012⁴⁸. A total of 16,052 participants from the Zhongshan cohort were included in this program. Serum samples were collected at baseline and tested for hepatitis B surface antigen (HBsAg) using enzyme-linked immunosorbent assay (ELISA) kits (Autobio Diagnostics Co., China) according to the manufacturer's instructions. During the follow-up period, 70 incident liver cancer cases were diagnosed. For our analysis, we included these 70 primary liver cancer cases and 15982 controls with available HBsAg data.

Detection of EBV-encoded RNA in tumor tissues

To investigate EBV presence in lung and liver cancer tissues, we performed *in situ* hybridization (ISH) targeting Epstein-Barr encoded RNA (EBER). Tissue samples were obtained from the Sun Yat-sen University Cancer Center biobank. For lung cancer, the samples included 129 tumor tissues and 127 matched adjacent normal tissues from 129 cases. For liver cancer, the analysis included 348 tumor tissues and 142 matched adjacent normal tissues from 348 cases. Formalin-fixed, paraffin-embedded (FFPE) tissue sections were deparaffinized, rehydrated, and treated with proteinase K to facilitate probe penetration. EBER hybridization was conducted using a commercial kit (Zhongshan Jinqiao Biotechnology Co., Ltd.) following the manufacturer's protocol. EBER staining results were independently evaluated by two pathologists blinded to clinical and pathological data. To confirm the localization of EBER-positive signals, the corresponding matched slide underwent hematoxylin and eosin (HE) staining.

Covariates measurement

In the Zhongshan cohort, covariates included age, sex (male and female), education level (classified as primary school or less, high school, and university or more), body mass index (BMI), waist circumference, smoking history (categorized as never smoker, ex-smoker, current smoker), alcohol consumption (grouped into less than monthly, monthly or weekly, once or more a day) and physical activity (grouped into less than monthly, monthly or weekly, once or more a day). BMI was categorized into four groups (normal weight: 18.5–22.9, underweight: < 18.5 , overweight: 23.0–27.5 and obese: > 27.5 kg/m²) based on Asian-based criteria⁴⁹. The waist circumference was categorized into two groups (normal: < 85 cm for males or < 80 cm for females; abnormal: ≥ 85 cm for males or ≥ 80 cm for females⁵⁰). Participants were categorized as smokers if they had smoked at least 100 cigarettes during their lifetime, and Ex-smokers were those who had quit smoking for at least one year prior to the interview⁹. However, there was mild missing data for certain variables: education level (17.2% missing), BMI (14.4% missing), waist circumference (18.5% missing), smoking history (15.4% missing), alcohol consumption (15.2% missing), and physical activity (15.7% missing). To mitigate potential biases arising from missing covariables, a multiple imputation analysis was applied⁵¹, assuming that missing values were missing at random. We employed the multivariate imputation by chained equations (MICE) using the R package MICE (version 3.12.0). In the Wuzhou cohort, only age and sex were available as covariates. In both cohorts, there were no missing data for VCA-IgA antibody level and cancer risk outcomes.

Statistical analysis

We calculated the age-standardized incidence rate (ASR) of total cancers and specific cancer types for the cohort population. The ASR and its 95% confidence interval were generated based on the world standard population⁵². We generated a relevant confidence interval (CI) for the seropositive and seronegative individuals of EBV VCA-IgA antibody based on the gamma distribution⁵³. We compared these two groups using standardized rate ratios implemented in the R package “dsr”(v0.2.2)^{53,54}. To assess the associations between VCA-IgA antibodies and cancer risks, we used Cox proportional hazards regression models to calculate crude and adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) implemented in the R package “survival”(v3.7.0). The proportionality of hazard assumptions was checked based on Schoenfeld's residuals⁵⁵. In the Zhongshan cohort, the HR and *P* values in the multivariable analysis were adjusted for the following covariates: age, sex (male and female), education level (classified as primary school or less, high school, and university or more), body mass index (normal weight: 18.5–22.9, underweight: < 18.5 , overweight: 23.0–27.5 and obese: > 27.5 kg/m²), waist circumference (normal: < 85 cm for male or < 80 cm for female; abnormal: ≥ 85 cm for male or ≥ 80 cm for female), smoking history (categorized as never smoker, ex-smoker, and current smoker), alcohol consumption (grouped into less than monthly, monthly or weekly, once or more a day) and physical activity (grouped into less than monthly, monthly or weekly, once or more a day). In the Wuzhou cohort, the HR and *P* values in the multivariable analysis were adjusted by age and sex (male and female). To assess the dose-response association between continuous VCA-IgA levels and cancer risk, we used a restricted cubic spline Cox proportional hazards model implemented via the *cph* function in the “rms” package (v6.8.2)⁵⁶. HRs and 95% CIs were derived using the *Predict* function based on the fitted model. To evaluate time-dependent associations, we applied a Cox regression model with time-dependent covariates using the *coxph* function in R. HRs and 95% CIs were estimated dynamically at each time point to visualize changing risk patterns over time. The model is defined as follows:

$$h(t) = h_0(t)e^{\beta X(t)} \quad (1)$$

where $h(t)$ represents the hazard function at time t , $h_0(t)$ is the baseline hazard, $X(t)$ denotes the time-dependent covariate, and β is the regression coefficient. This method enables a dynamic assessment of the association between VCA-IgA antibody levels and cancer risk over time. The models controlled for the same covariates as in the Cox proportional hazards regression models, including age, sex, and other relevant variables. To evaluate the overall effect size of both the Zhongshan and Wuzhou cohorts, we conducted a meta-analysis to estimate the combined effect size of these two cohorts. Effect sizes and 95% CIs were extracted from each study, and heterogeneity was assessed using the I^2 statistic calculated by the Cochrane Q test. If I^2 was less than 75%, a fixed-effects model was used, assuming a common effect size across studies. For I^2 values above 75%, a random-effects model was employed to account for both within-study and between-study variance⁵⁷.

We also conducted several sensitivity analyses to enhance the robustness of our findings. First, using Fine and Gray competing risk proportional hazards models implemented in the R package “cmprsk” (v2.2.11)⁵⁸, we conducted competing risk analysis by treating other cancers as competing events. Second, we excluded events that occurred within the initial 1 to 3 years of follow-up to reduce potential reverse causation. Third, we stratified our analyses by sex (males and females), age (≤ 46 and >46 years, defined as the median age in the cohort population), and the primary histological types of specific cancer, particularly for lung cancer and liver cancer. Fourth, in the Zhongshan cohort, the original VCA-IgA antibody level was expressed as a continuous rOD value. We repeated our analyses by replacing the binary VCA-IgA antibody serostatus with this continuous rOD value. We then classified the EBV antibody levels into four quartiles as follows: “P0-P25” ($rOD \leq 0.17$), “P25-P50” ($0.17 < rOD \leq 0.28$), “P50-P75” ($0.28 < rOD \leq 0.48$), and “P75-P100” ($rOD > 0.48$). We analyzed the relationship between these defined EBV antibody categories and the associated risks of cancer. Fifth, we confined our analysis in the Wuzhou cohort to those participants aged 30 to 59 years to coincide with the age distribution in the Zhongshan cohort. Sixth, to further evaluate the association between VCA-IgA antibody levels and liver cancer risk, we incorporated HBV infection data into our analysis. Specifically, the association was assessed after adjusting for HBV infection status using HBsAg seropositivity. Additionally, we adjusted various covariables in the Cox model in the Zhongshan cohort analysis.

To assess the cancer burden associated with EBV seropositivity, we calculated the attributable risk percentage (ARP) via the following formula:

$$ARP = \frac{CI_e - CI_u}{CI_e} \quad (2)$$

Where CI_e represents cumulative incidence in the EBV seropositive group, while CI_u represents cumulative incidence in the EBV seronegative group. The population attributable risk percentage (PARP) was calculated using ARP multiplied by the prevalence of EBV seropositivity in the population. In addition, exposure impact number (EIN) was calculated to represent the number of EBV seropositive individuals among whom one excess cancer patient was attributed to EBV seropositivity using the following formula:

$$EIN = \frac{1}{CI_e - CI_u} \quad (3)$$

Furthermore, the case impact number (CIN) was calculated as the reciprocal of the PARP, which represents the number of cancer cases among whom one case would be attributable to the EBV seropositivity⁵⁹. Considering the uneven global prevalence of NPC, we also calculated the cancer burden attributed to EBV seropositivity by excluding NPC to prevent potential overestimation in regions

where NPC is not endemic. These risk fractions and impact numbers related to cancer burden were calculated using the R package “twotxtwo” (v0.1.0).

Statistical significance was indicated by a two-tailed P value of less than 0.05. When analyzing the associations across multiple cancer types, false discovery rate (FDR) adjustment was applied to account for multiple testing, with an FDR-adjusted p value < 0.05 considered statistically significant. All statistical tests were performed using R 4.1.1.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The Individual-level epidemiological data generated in this study have been deposited in the Research Data Deposit public platform (www.researchdata.org.cn) under accession code RDDA2025560963. These data were available under restricted access due to privacy and ethical considerations. Researchers may request access for academic use by submitting a formal application through the repository or by contacting the corresponding author and obtaining approval from the institutional ethics committee. The raw data are protected and are not available due to data privacy laws. The aggregate baseline information and summary statistics used in this study are provided in the Supplementary Information. Source data are provided with this paper.

References

- Young, L. S., Yap, L. F. & Murray, P. G. Epstein-Barr virus: more than 50 years old and still providing surprises. *Nat. Rev. Cancer* **16**, 789–802 (2016).
- Murray, P. G. & Young, L. S. An etiological role for the Epstein-Barr virus in the pathogenesis of classical Hodgkin lymphoma. *Blood* **134**, 591–596 (2019).
- Chen, Y. P. et al. Nasopharyngeal carcinoma. *Lancet* **394**, 64–80 (2019).
- Young, L. S. & Rickinson, A. B. Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer* **4**, 757–768 (2004).
- de Martel, C. et al. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob. Health* **8**, e180–e190 (2020).
- Wong, Y. et al. Estimating the global burden of Epstein-Barr virus-related cancers. *J. Cancer Res. Clin. Oncol.* **148**, 31–46 (2022).
- Damania, B., Kenney, S. C. & Raab-Traub, N. Epstein-Barr virus: biology and clinical disease. *Cell* **185**, 3652–3670 (2022).
- Longnecker, R. M., Kieff, E. & Cohen, J. I. in *Fields Virology* (Ch. 61) (Wolters Kluwer Health Adis, 2013).
- Xu, F. H. et al. An epidemiological and molecular study of the relationship between smoking, risk of nasopharyngeal carcinoma, and Epstein-Barr virus activation. *J. Natl Cancer Inst.* **104**, 1396–1410 (2012).
- Chien, Y. C. et al. Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *N. Engl. J. Med.* **345**, 1877–1882 (2001).
- Ji, M. F. et al. Incidence and mortality of nasopharyngeal carcinoma: interim analysis of a cluster randomized controlled screening trial (PRO-NPC-001) in southern China. *Ann. Oncol.* **30**, 1630–1637 (2019).
- de-The, G. et al. Epidemiological evidence for causal relationship between Epstein-Barr virus and Burkitt’s lymphoma from Ugandan prospective study. *Nature* **274**, 756–761 (1978).
- Mueller, N. et al. Hodgkin’s disease and Epstein-Barr virus. Altered antibody pattern before diagnosis. *N. Engl. J. Med.* **320**, 689–695 (1989).
- Zapatka, M. et al. The landscape of viral associations in human cancers. *Nat. Genet.* **52**, 320–330 (2020).

15. Li, J. S. Z. et al. Chromosomal fragile site breakage by EBV-encoded EBNA1 at clustered repeats. *Nature* **616**, 504–509 (2023).
16. He, J. R. et al. Joint effects of Epstein-Barr virus and polymorphisms in interleukin-10 and interferon-gamma on breast cancer risk. *J. Infect. Dis.* **205**, 64–71 (2012).
17. Lehtinen, T. et al. Increased risk of malignant lymphoma indicated by elevated Epstein-Barr virus antibodies—a prospective study. *Cancer Causes Control* **4**, 187–193 (1993).
18. Levine, P. H. et al. Elevated antibody titers to Epstein-Barr virus prior to the diagnosis of Epstein-Barr-virus-associated gastric adenocarcinoma. *Int. J. Cancer* **60**, 642–644 (1995).
19. Koshiol, J. et al. Epstein-Barr virus serology and gastric cancer incidence and survival. *Br. J. Cancer* **97**, 1567–1569 (2007).
20. Kim, Y. et al. Epstein-Barr virus antibody level and gastric cancer risk in Korea: a nested case-control study. *Br. J. Cancer* **101**, 526–529 (2009).
21. Bertrand, K. A. et al. A prospective study of Epstein-Barr virus antibodies and risk of non-Hodgkin lymphoma. *Blood* **116**, 3547–3553 (2010).
22. Coghill, A. E. et al. The association between the comprehensive Epstein-Barr virus serologic profile and endemic Burkitt lymphoma. *Cancer Epidemiol. Biomark. Prev.* **29**, 57–62 (2020).
23. Zeng, Y. et al. Prospective studies on nasopharyngeal carcinoma in Epstein-Barr virus IgA/VCA antibody-positive persons in Wuzhou City, China. *Int. J. Cancer* **36**, 545–547 (1985).
24. Coghill, A. E. & Hildesheim, A. Epstein-Barr virus antibodies and the risk of associated malignancies: review of the literature. *Am. J. Epidemiol.* **180**, 687–695 (2014).
25. Li, H. et al. Epstein-Barr virus lytic reactivation regulation and its pathogenic role in carcinogenesis. *Int. J. Biol. Sci.* **12**, 1309–1318 (2016).
26. Hong, S. et al. The genomic landscape of Epstein-Barr virus-associated pulmonary lymphoepithelioma-like carcinoma. *Nat. Commun.* **10**, 3108 (2019).
27. Huang, Y. H. et al. Clinicopathologic features, tumor immune microenvironment and genomic landscape of Epstein-Barr virus-associated intrahepatic cholangiocarcinoma. *J. Hepatol.* **74**, 838–849 (2021).
28. Hu, J. et al. Human cytomegalovirus and Epstein-Barr virus infections, risk factors, and their influence on the liver function of patients with acute-on-chronic liver failure. *BMC Infect. Dis.* **18**, 577 (2018).
29. Osorio J. C. et al. Epstein-Barr virus infection in lung cancer: insights and perspectives. *Pathogens* **11**, 132 (2022).
30. Sugawara, Y. et al. Detection of Epstein-Barr virus (EBV) in hepatocellular carcinoma tissue: a novel EBV latency characterized by the absence of EBV-encoded small RNA expression. *Virology* **256**, 196–202 (1999).
31. Li, W. et al. Epstein-Barr virus in hepatocellular carcinogenesis. *World J. Gastroenterol.* **10**, 3409–3413 (2004).
32. Munz, C. Latency and lytic replication in Epstein-Barr virus-associated oncogenesis. *Nat. Rev. Microbiol.* **17**, 691–700 (2019).
33. Ambinder, R. F. Gammaherpesviruses and “hit-and-run” oncogenesis. *Am. J. Pathol.* **156**, 1–3 (2000).
34. Chen, K., Wang, M., Zhang, R. & Li, J. Detection of Epstein-Barr virus encoded RNA in fixed cells and tissues using CRISPR/Cas-mediated RCasFISH. *Anal. Biochem.* **625**, 114211 (2021).
35. Kimura, H. et al. Identification of Epstein-Barr virus (EBV)-infected lymphocyte subtypes by flow cytometric in situ hybridization in EBV-associated lymphoproliferative diseases. *J. Infect. Dis.* **200**, 1078–1087 (2009).
36. AbuSalah, M. A. H. et al. Recent advances in diagnostic approaches for Epstein-Barr virus. *Pathogens* **9**, 226 (2020).
37. Rosemarie, Q. & Sugden, B. Epstein-Barr virus: how its lytic phase contributes to oncogenesis. *Microorganisms* **8**, 1824 (2020).
38. Okuno, Y. et al. Defective Epstein-Barr virus in chronic active infection and haematological malignancy. *Nat. Microbiol.* **4**, 404–413 (2019).
39. Wu, C. C. et al. Epstein-Barr virus DNase (BGLF5) induces genomic instability in human epithelial cells. *Nucleic Acids Res.* **38**, 1932–1949 (2010).
40. Xue, W. Q. et al. A comprehensive analysis of genetic diversity of EBV reveals potential high-risk subtypes associated with nasopharyngeal carcinoma in China. *Virus Evol.* **7**, veab010 (2021).
41. Chen, Z. et al. Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. *Lancet* **386**, 1447–1456 (2015).
42. Li, T. et al. Anti-Epstein-Barr virus BNLf2b for mass screening for nasopharyngeal cancer. *N. Engl. J. Med.* **389**, 808–819 (2023).
43. Chen, G. H. et al. Prospective assessment of a nasopharyngeal carcinoma risk score in a population undergoing screening. *Int. J. Cancer* <https://doi.org/10.1002/ijc.33424> (2020).
44. Liu, Y. et al. Establishment of VCA and EBNA1 IgA-based combination by enzyme-linked immunosorbent assay as preferred screening method for nasopharyngeal carcinoma: a two-stage design with a preliminary performance study and a mass screening in southern China. *Int. J. Cancer* **131**, 406–416 (2012).
45. Liu, Z. 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.* **177**, 242–250 (2013).
46. Chen, Y. et al. Nasopharyngeal Epstein-Barr virus load: an efficient supplementary method for population-based nasopharyngeal carcinoma screening. *PLoS ONE* **10**, e0132669 (2015).
47. Yi, Z. et al. Application of an immunoenzymatic method and an immunoradiographic method for a mass survey of nasopharyngeal carcinoma. *Intervirology* **13**, 162–168 (1980).
48. Ji, M. et al. Mass screening for liver cancer: results from a demonstration screening project in Zhongshan City, China. *Sci. Rep.* **8**, 12787 (2018).
49. Consultation WHOE Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* **363**, 157–163 (2004).
50. Lear, S. A., James, P. T., Ko, G. T. & Kumanyika, S. Appropriateness of waist circumference and waist-to-hip ratio cutoffs for different ethnic groups. *Eur. J. Clin. Nutr.* **64**, 42–61 (2010).
51. Sterne, J. A. et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338**, b2393 (2009).
52. Ahmad, O. B. et al. Age Standardization of Rates: A New WHO Standard. *GPE Discussion Paper Series: No 31* (WHO, 2001).
53. Fay, M. P. & Feuer, E. J. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat. Med.* **16**, 791–801 (1997).
54. Elandt-Johnson, R. C. & Johnson, N. L. *Survival Models and Data Analysis* (John Wiley and Sons, 1980).
55. Abeysekera, W. W. M. & Sooriyachchi, R. Use of Schoenfeld’s global test to test the proportional hazards assumption in the Cox proportional hazards model: an application to a clinical study. *J. Natn. Sci. Found. Sri Lanka* **37**, 41–45 (2009).
56. Gauthier, J., Wu, Q. V. & Gooley, T. A. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transpl.* **55**, 675–680 (2020).
57. Haidich, A. B. Meta-analysis in medical research. *Hippokratia* **14**, 29–37 (2010).
58. Fine, J. P. G. R. A Proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* **94**, 496–509 (1999).
59. Heller, R. F., Dobson, A. J., Attia, J. & Page, J. Impact numbers: measures of risk factor impact on the whole population from case-control and cohort studies. *J. Epidemiol. Community Health* **56**, 606–610 (2002).

Acknowledgements

This study was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501000), the National Natural Science Foundation of China (82473703, 82273705, 82373656 and 82404339), the Science and Technology Planning Project of Guangzhou, China (2024A04J00693 and 2024A04J4560), Young Science and Technology Talent Support Program of Guangdong Precision Medicine Application Association (YSTTGDPMAA202502), the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (24qnpy292 and 24ykqb002), the Young Talents Program of Sun Yat-sen University Cancer Center (YTP-SYSUCC-0081 and YTP-SYSUCC-0076), Cancer Innovative Research Program of Sun Yat-sen University Cancer Center (CIRP-SYSUCC-0017), the Young Talent Support Project of Guangzhou Association for Science and Technology (QT2024-030). Where authors are identified as personnel of the International Agency for Research on Cancer/WHO, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/WHO.

Author contributions

W.-H.J., M.-F.J. and Y.-Q.H. supervised the project. M.-F.J., Y.-Q.H., M.-Z.T., W.-Q.X., X.Y., H.D., D.-W.Y., B.-H.W., F.-G.L., J.-Y.Z., H.-L.M., J.L., Y.-C.L., T.-M.W., Y.L., X.-Y.C., Z.-H.L., S.-F.L., Y.D. and X.-J.L. conducted the cohort enrollment, investigation and follow-up. Y.-Q.H., M.-Z.T., W.-Q.X., and X.Y. cleaned the data and Y.-Q.H., W.-Q.X., Z.-Y.Z., and C.-L.H. performed the analysis under the supervision of W.-H.J. Y.-Q.H. and W.-Q.X. drafted the manuscript with editing from W.-H.J. All authors reviewed and approved the final manuscript. Z.-M.M., I.H.C., Z.K., J.M., and W.-H.J. reviewed and revised the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing interests

Authors declare no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60999-5>.

Correspondence and requests for materials should be addressed to Yong-Qiao He or Wei-Hua Jia.

Peer review information *Nature Communications* thanks Paola Chabay, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Cancer Research Institute of Zhongshan City, Zhongshan City People's Hospital, Zhongshan, Guangdong, P. R. China. ²State 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, P. R. China. ³Wuzhou Red Cross Hospital, Wuzhou, Guangxi, P.R. China. ⁴Wuzhou Cancer Center, Wuzhou, Guangxi, P.R. China. ⁵School of Public Health, Sun Yat-Sen University, Guangzhou, Guangdong, P. R. China. ⁶Radiation Epidemiological Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA. ⁷School of Public Health, The University of Hong Kong, Hong Kong S.A.R., P. R. China. ⁸State Key Laboratory of Systems Medicine for Cancer, Centre for Single-Cell Omics, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, P. R. China. ⁹Public Health Service Center of Xiaolan Town, Zhongshan City, Guangdong, P. R. China. ¹⁰Guilin Medical University, Guilin, Guangxi, P.R. China. ¹¹Southwest Hospital, Chongqing, P. R. China. ¹²International Agency for Research on Cancer, World Health Organization, Lyon, France. ¹³Department of Radiation Oncology, State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, P.R. China. ¹⁴These authors contributed equally: Ming-Fang Ji, Yong-Qiao He, Min-Zhong Tang, Wen-Qiong Xue, Xia Yu.

✉ e-mail: heyq@sysucc.org.cn; jiawh@sysucc.org.cn