Trials

RESEARCH Open Access

No genetic causal association between human papillomavirus and lung cancer risk: a bidirectional two-sample Mendelian randomization analysis

Yizhuo Chen[1], Ziqing Xu[1], Zhouqi Zhang[1], Xin Wang[2] and Ming Dong[1*]

Abstract

Introduction Several observational or retrospective studies have previously been conducted to explore the possible association between lung cancer and human papillomavirus (HPV) infection. However, there may be inconsistencies in the data and conclusions due to differences in study design and HPV testing methods. There are currently no studies that provide conclusive evidence to support the involvement of HPV in the occurrence and development of lung cancer. Therefore, the relationship between HPV and lung cancer remains controversial and uncertain. This study aimed to explore whether HPV infection is causally related to lung cancer risk by systematically performing a two-way Two-Sample Mendelian Randomization (TSMR) analysis.

Methods In the International Lung Cancer Consortium (ILCCO) genome-wide association study dataset, we included 11,348 lung cancer (LUCA) cases, including 3275 squamous cell carcinoma (LUSC) cases, 3442 adenocarcinoma (LUAD) cases, and 15,861 cases of control. Using genetic variants associated with the HPV E7 protein as instrumental variables, we summarized statistics associated with HPV infection in the MRC IEU OpenGWAS database, which included the HPV-16 E7 protein and the HPV-18 E7 protein. Two-sample Mendelian randomization (MR) results are expressed as odds ratios (OR) and 95% confidence intervals (CI).

Results Based on a comprehensive analysis of genome-wide association study (GWAS) data from public databases, we mainly used inverse-variance weighted (IVW) to estimate causal relationships, while using MR-Egger, weighted median, simple mode, and weighted mode, and other four methods as supplements. Two-sample MR Analysis revealed no causal relationship between exposure factors (HPV-16 E7 protein and HPV-18 E7 protein) and outcome factors (lung cancer (LUCA) and its subtypes squamous cell carcinoma (LUSC) and adenocarcinoma (LUAD)) in forward MR Analysis using the IVW approach.HPV-16 E7 protein and LUCA and its subtypes LUSC and LUAD by IVW method results: [OR] = 1.002; 95% [CI]: 0.961 – 1.045; p = 0.920; [OR] = 1.023; 95% [CI]: 0.966 – 1.084; p = 0.438; [OR] = 0.994; 95% [CI]: 0.927 – 1.066; p = 0.872); HPV-18 E7 protein and LUCA and its subtypes LUSC and LUAD by IVW method results: [OR] = 0.965; 95% [CI]: 0.914 – 1.019; p = 0.197; [OR] = 0.933; 95% [CI]: 0.834 – 1.043; p = 0.222; [OR] = 1.028; 95% [CI]: 0.945 – 1.118; p = 0.524. It was observed through reverse MR that LUCA and its subtypes LUSC and LUAD were used as exposure factors, and HPV infection (HPV-16 E7 protein and HPV-18 E7 protein) was used as the outcome factors, the results of the IVW method are also invalid.LUCA and HPV-16 E7 protein and HPV-18 E7 protein by IVW method

*Correspondence:
Ming Dong
dongming@tmu.edu.cn
Full list of author information is available at the end of the article

results: [OR] = 1.036; 95% [CI]: 0.761 – 1.411; p = 0.82; [OR] = 1.318; 95% [CI]: 0.949 – 1.830; p = 0.099; LUSC and HPV-16 E7 protein and HPV-18 E7 protein by IVW method results: [OR] = 1.123; 95% [CI]0.847 – 1.489; p = 0.421; [OR] = 0.931; 95% [CI]: 0.660 – 1.313; p = 0.682; LUAD and HPV-16 E7 protein and HPV-18 E7 protein by IVW method results: [OR] = 1.182; 95% [CI] 0.983 – 1.421; p = 0.075; [OR] = 1.017; 95% [CI]: 0.817 – 1.267; p = 0.877.Our results indicate that there is no causal relationship between genetically predicted HPV infection and LUCA and its subtypes LUSC and LUAD. In addition, in the reverse MR analysis, we did not observe a significant causal relationship between LUCA and its subtypes LUSC and LUAD on HPV infection.

Conclusions Our findings do not support a genetic association between HPV infection and lung cancer.

Keywords Mendelian randomization, HPV infection, Human papillomavirus, Lung cancer, Lung adenocarcinoma, Lung squamous cell carcinoma, Causal relationship

Introduction

According to the latest global cancer statistics analysis from the International Agency for Research on Cancer (IARC) in 2022, compared to the data from 2020, the number of new lung cancer cases has increased from 2.2 million to nearly 2.5 million, raising its proportion of total cancer cases from 11.4% to 12.4%; although the number of deaths caused by lung cancer remained unchanged in absolute terms (about 1.8 million), its share of total cancer deaths worldwide increased slightly from 18.0% to 18.7%. These data suggest that lung cancer is further increasing in importance in the global cancer burden, not only topping the list in terms of incidence but also remaining the leading cause of cancer death [1, 2]. Smoking as a risk factor for causing lung cancer is widely acknowledged in the medical field. The association between the two is considered one of the strongest and longest-known risk associations among modifiable lifestyle factors and specific types of cancer [3–5]. In many countries, men's smoking prevalence and cumulative smoking exposure are generally higher than women's, and men's smoking cessation rate is lower, which is also an important reason why men have a higher incidence of lung cancer than women [5]. However, the incidence of lung cancer in non-smokers still exists and may be increasing in certain regions and populations [6–8]. The research by REVEL M and colleagues indicates that in most European countries, it is anticipated that the mortality rate from lung cancer in females will surpass that of breast cancer [9]. Therefore, given these facts, there is a growing focus on cancer risk factors other than smoking, such as viral infections, chronic inflammation, genetic variants, and environmental exposures.

The HPV belongs to the Papillomaviridae family and is a DNA virus that infects the epithelial cells of the skin or mucous membranes [10]. Multiple studies indicate that HPV is one of the most prevalent sexually transmitted infections globally [11–14] 0.4.5% of global cancer cases (630,000 new cancer cases annually) are attributed to HPV infection, with 8.6% in females and 0.8% in males [15, 16]. Nearly all cases of cervical cancer are caused by HPV infection [17]. HPV is also a significant driving factor for head and neck cancers, and anogenital cancers, and its role in the etiology of oropharyngeal cancer is increasingly prominent [18]. For example, in the United States, oropharyngeal cancer has become the most common malignancy associated with HPV [19] Currently, there are 12 HPV genotypes classified as carcinogenic, including types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59. Among these, HPV16 and HPV18 are the most common carcinogenic types. These types are closely associated with cervical cancer, related lesions, and precancerous dysplasia, playing a crucial role in the formation of malignant cervical tumors [17, 20]. Among them, the clearance rate of HPV type 16 is the lowest among all HPV types [21]. The early region (E) oncoproteins of HPV, including E1, E2, E4, E5, E6, and E7, are associated with pathogenic mechanisms and play important roles in cancer progression. Among the six E proteins, E6 and E7 are the main regulators of viral pathogenicity, and they are also significantly involved in the development of cervical cancer [22]. Almost all sexually active individuals may become infected with HPV at some point in their lives. Most people do not exhibit symptoms, and infections are generally undetectable within the first 2 years after exposure [14]. Approximately 10% of HPV infections may persist for an extended period, potentially leading to the development of precancerous lesions, with only a small proportion of these lesions progressing to HPV-related tumor diseases [12].

Many previous research reports indicate that HPV infection may be a potential risk factor for the occurrence and development of lung cancer, showing a positive correlation with the risk of lung cancer [23–28]. Some studies suggest that the risk of lung cancer is most closely associated with pulmonary infections caused by HPV types 16 and 18. Moreover, the prevalence of HPV infection is higher in squamous cell carcinoma compared to adenocarcinoma [25, 28]. An association between HPV and lung cancer was found in a study by Rezaei et al. [29]. Specifically, they found a significant increase in the

expression of inflammatory cytokines in HPV-positive lung cancer samples and control tissues compared to HPV-negative lung cancer and HPV-negative control tissues. Therefore, the authors suggest that HPV infection may trigger inflammation and epithelial-mesenchymal transition (EMT), which may contribute to the development of lung cancer. In a meta-analysis, Drokow et al. noted that patients with HPV type 16 had a higher risk of developing non-small cell lung cancer (NSCLC) compared with those with HPV type 18 infection (OR = 1.95, 95% CI: 1.00–3.79) [28].

Contradicting the aforementioned findings, the study by Jing-Yang Huang et al. suggests that HPV infection is associated with the occurrence of lung adenocarcinoma but not with lung squamous cell carcinoma [30]. Simultaneously, there are other studies suggesting that there is no correlation between HPV infection and an increased risk of lung cancer [31–33]. The conflicting results of different studies on the association between HPV infection and lung cancer may stem from variations in research methods, detection methods, geographical and population differences, tumor types, control of confounding factors, and data analysis methods. At the same time, previous research has triggered the question we have to discuss in this article: Is there a causal relationship between HPV infection and lung cancer? If it exists, which factor causes another?

Mendelian Randomization (MR) is a method for assessing the causal impact between modifiable risk factors and diseases, using genetic variations as instrumental variables for exposure. Therefore, this approach is more adept at avoiding common pitfalls in traditional clinical research, such as measurement errors, confounding, and reverse causation [34]. Sample Mendelian Randomization (MR) analysis, by utilizing single nucleotide polymorphisms (SNPs) from Genome-Wide Association Study (GWAS) data, offers a more rigorous method for causal inference to assess the potential relationship between modifiable risk factors and diseases [35]. Therefore, we conducted a two-sample bidirectional Mendelian Randomization analysis to enhance our understanding of the causal relationship between HPV infection and lung cancer while improving the credibility of this study.

Materials and methods

Study design

We followed the latest STROBE-MR (Strengthening the Reporting of Observational Studies in Epidemiology for Mendelian Randomization) guidelines, conducting the study using a bidirectional Two-Sample Mendelian Randomization (TSMR) approach to explore the reciprocal associations between HPV infection and lung cancer [36]. In the forward MR analysis, HPV infection was considered as the exposure. Lung cancer, including its subtypes squamous cell carcinoma and adenocarcinoma, was analyzed as the outcomes. The reverse MR analysis, on the other hand, treated lung cancer and its subtypes, squamous cell carcinoma and adenocarcinoma, as exposures and HPV infection as the outcome.In our Mendelian randomization (MR) analysis, genetic variants were utilized as instrumental variables (IVs) to estimate causal effects.Each genetic variant was treated as an instrumental variable, and it had to satisfy the following three core assumptions: ① The chosen IV is strongly associated with the exposure; ② The IV is unrelated to confounding factors associated with both the exposure and the outcome; ③ The IV influences the outcome solely through the exposure (Fig. 1). We conducted the analysis using publicly available aggregated statistical data, hence ethical approval was not required.

Data sources

The data sources utilized in the study were sourced from the MRC IEU OpenGWAS developed by the MRC Integrative Epidemiology Unit at the University of Bristol (https://gwas.mrcieu.ac.uk/, Version: v6.5.2–2022-04–11) [36]. The HPV data included in this study have the following GWAS IDs: prot-c-2623_54_4 (HPV E7 Type 16) and prot-c-2624_31_2 (HPV E7 Type 18). The summary data for the genome-wide association study (GWAS) on lung cancer were retrieved through the IEU-OpenGWAS online platform. These GWAS data sources originate from the International Lung Cancer Consortium (ILCCO) and encompass 11,348 cases of lung cancer (LUCA) and 15,861 control subjects. Within the LUCA cases, there is further stratification based on histological subtypes, including 3,275 cases of squamous cell carcinoma (LUSC) and 3,442 cases of adenocarcinoma (LUAD). The GWAS IDs for LUCA and its subtypes LUSC and LUAD are "ieu-a-966," "ieu-a-967," and "ieu-a-965," respectively [37]. These samples are all restricted to individuals of European ancestry, to some extent mitigating biases introduced by confounding factors related to race. Detailed information on the data resources is listed in Table 1.

Selection of instrumental variables

In order to identify suitable genetic instrumental variables (IVs), we implemented a series of quality control steps to ensure the robustness and confidence of Mendelian randomization (MR) analyses. Initially, we used a filtration criterion of $P < 5 \times 10^{(-8)}$ to extract independent SNPs associated with HPV E7 protein and lung cancer causality. However, because there were few or no SNPS that met this criterion, we adjusted the significance threshold to a more lenient criterion of $P < 5 \times 10^{(-5)}$.

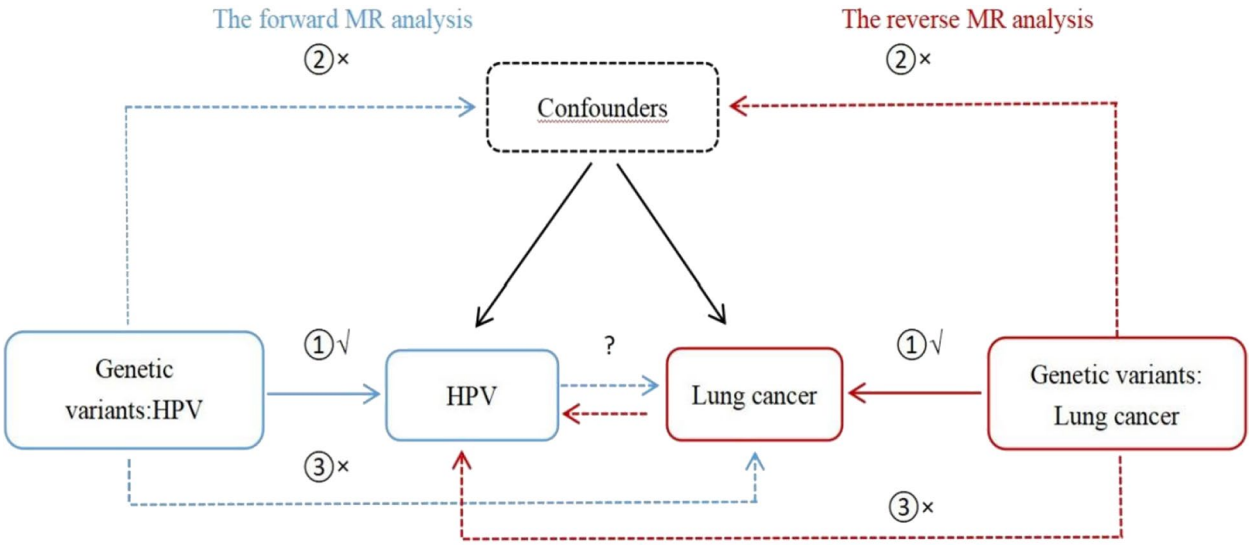


Fig. 1 Schematic representation of the bidirectional bidirectional Mendelian randomization (MR) study design. In the figure, blue indicates forward MR analysis with HPV as the exposure and lung cancer as the outcome, while red represents reverse MR analysis with lung cancer as the exposure and HPV as the outcome. Dashed lines indicate unrelated relationships, while solid lines denote associations between variables and confounding factors or outcomes that cannot be overcome

Table 1 Brief description of data sources involved in mendelian randomization studies

GWAS ID	Exposure and outcome	Year	Consortium	Sample size	Ncases	Ncontrols	Nsnp	Ancestry	Access link
prot-c-2623_54_4	HPV E7 Type 16	2019	NA	-	-	-	501428	European	https://gwas.mrcieu.ac.uk/datasets/prot-c-2623_54_4/
prot-c-2624_31_2	HPV E7 Type18	2019	NA	-	-	-	501428	European	https://gwas.mrcieu.ac.uk/datasets/prot-c-2624_31_2/
ieu-a-966	LUCA	2014	ILCCO	27,209	11,348	15,861	8,945,893	European	https://gwas.mrcieu.ac.uk/datasets/ieu-a-966/
ieu-a-967	LUSC	2014	ILCCO	18,313	3,275	15,038	8,893,750	European	https://gwas.mrcieu.ac.uk/datasets/ieu-a-967/
ieu-a-965	LUAD	2014	ILCCO	18,336	3,442	14,894	8,881,354	European	https://gwas.mrcieu.ac.uk/datasets/ieu-a-965/

The reasons for this change are as follows: First, to improve the feasibility of the analysis. A strict $P<5\times10^{-8}$ threshold can make it difficult to find enough SNPs for MR Analysis, whereas adopting a $P<5\times10^{-5}$ threshold can increase the number of SNPS available to make analysis possible. Second, preserve the significance of instrumental variables. Although the $P<5\times10^{-5}$ threshold is more lenient than $P<5\times10^{-8}$, it still has strong statistical significance, ensuring that the selected SNPs is significantly associated with the exposure variable. Third, improve the statistical power of the analysis. The looser significance threshold allows for the inclusion of more SNPs, increasing statistical power and allowing us to better detect potential causation. However, the use of looser significance thresholds may increase the risk of false positives, weaken the strength of instrumental variables, and increase the complexity of interpretation of results. To mitigate these risks, we excluded SNPs in strongly linked disequilibrium (LD) ($r^2<0.001$, window size = 10,000 kb), using LD estimates from the 1000 Genome Project European population [38]. In addition, we calculated an F statistic to assess the degree of association between IVs and exposure risk and tool strength, with an F statistic greater than 10 considered strong enough. The formula for calculating the F statistic is $F=R^2\times(n-2)/(1-R^2)$, where R^2 represents the

variation in the exposure variable for each IV interpretation and N represents the sample size of the exposed GWAS. We also use PhenoScanner to search for SNPs that may be pleiotropic to assess the association of these SNPs with multiple phenotypes to determine whether they might influence the study results. Finally, we ensure consistency of allelic effects in the exposure and outcome datasets by excluding fuzzy SNPs with inconsistent alleles and palindromic SNPs with intermediate allelic frequencies.

Mendelian randomization analysis

This study conducted data analysis using R software (version 4.3.2, www.r-project.org/) and employed the TwoSampleMR package (version 0.5.7). The primary analytical approach utilized was the Inverse Variance Weighted (IVW, random effects) method [39]. Additionally, various complementary MR detection methods, including MR-Egger, Weighted Median, Simple Mode, and Weighted Mode, were employed to precisely test causal effects and correct for pleiotropy effects [40, 41]. Inverse Variance Weighted (IVW) is an effective analytical method that assumes all genetic variations are valid instrumental variables and possesses robust capabilities in detecting causal relationships. It achieves this by calculating the weighted average of the estimates of the Wald ratio [42]. The MR-Egger regression is capable of detecting and correcting for pleiotropy but is susceptible to the influence of outlying genetic variants, potentially reducing statistical power [43]. The Weighted Median method can mitigate the impact of invalid instruments and still provide consistent estimates of causal effects when analyzing information from 50% of genetic variations of invalid instruments [44]. While the Simple Mode may not be as powerful as IVW, it demonstrates stability in the presence of pleiotropy [45]. Lastly, for mode assessment, Weighted Mode is highly sensitive to the inclusion of hard-thresholded instruments [46].

We conducted various sensitivity analyses to validate the robustness of the MR results, including Cochran's Q test, MR-Egger intercept test, MR-PRESSO, and leave-one-out analysis. Cochran's Q is a heterogeneity test, that mainly uses the IVW analysis method and MR-Egger regression. The test result $P > 0.05$ indicates that there is no heterogeneity among IVs [47]. The intercept value in MR-Egger is used to evaluate pleiotropy, and $P > 0.05$ indicates the absence of horizontal pleiotropy [48]. In the presence of heterogeneity or pleiotropy in MR results, we employed MR-PRESSO to detect potential pleiotropic distortion outliers and mitigated horizontal pleiotropy by excluding significant outliers [49].To identify potential heterogeneous SNPs, we performed a leave-one-out analysis by systematically excluding each SNP to assess the robustness and consistency of the results. Additionally, we generated forest plots, scatter plots, funnel plots, and leave-one-out analysis plots to visually present the results in a highly illustrative manner. Specifically, the forest plot vividly illustrates the impact of each SNP on the results; the leave-one-out analysis plot assesses the visual reliability of the results; scatter plots display the fitting results of different MR analyses; and the funnel plot provides an intuitive assessment of the heterogeneity of instrumental variables.

Results

Instrumental variable (IV) selection

By filtering SNPs associated with the exposure, removing those in linkage disequilibrium (LD), and excluding weak instrumental variables with $F < 10$, we obtained 23 SNPs associated with HPV 16 E7 protein (F-statistic > 10) and 13 SNPs associated with HPV 18 E7 protein (F-statistic > 10). Simultaneously, we identified 105 SNPs associated with lung cancer (F-statistic > 10), 86 SNPs associated with squamous cell carcinoma (F-statistic > 10), and 88 SNPs associated with adenocarcinoma (F-statistic > 10). Numerous studies have confirmed that smoking is one of the primary risk factors for lung cancer. Simultaneously, there exists a complex interplay between smoking and HPV infection, with smoking being considered a potential risk factor for HPV infection. Therefore, we excluded SNPs associated with smoking to ensure the accuracy of the study. In the forward analysis, no pleiotropic instrumental variables related to HPV were identified. In the reverse analysis, for the 105 SNPs associated with lung cancer (F-statistic > 10), four pleiotropic instrumental variables related to smoking were removed. Similarly, for the 86 SNPs associated with squamous cell carcinoma (F-statistic > 10), three pleiotropic instrumental variables related to smoking were removed. Additionally, for the 88 SNPs associated with adenocarcinoma (F-statistic > 10), three pleiotropic instrumental variables related to smoking were removed. In the forward analysis with HPV as the exposure, we identified 23 SNPs associated with HPV-16 E7 protein and lung cancer, including its subtypes (squamous cell carcinoma and adenocarcinoma). One palindromic SNP (rs2864426) was excluded. For HPV-18 E7 protein, there were 12 SNPs associated with lung cancer and its subtypes (squamous cell carcinoma and adenocarcinoma), with no palindromic SNPs. In the reverse analysis with lung cancer and its subtypes (squamous cell carcinoma and adenocarcinoma) as the exposure, no palindromic SNPs were found. There were a total of 11 SNPs for lung cancer and HPV (HPV-16 E7 protein、HPV-18 E7 protein) combined, 6 SNPs for squamous cell carcinoma and HPV, and 14 SNPs for adenocarcinoma and HPV. (Supplementary File S1).

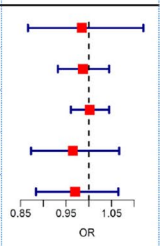
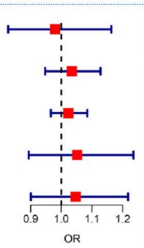
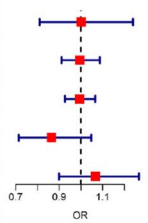
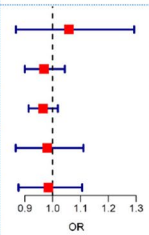
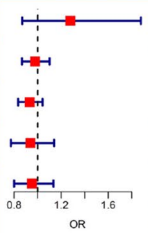
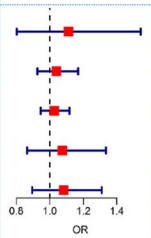
Exposure	Outcome	MR method	No. of SNPs	OR	(95% CI)	P value	Forest plot
HPV E7 Type 16	LUCA	MR Egger	22	0.985	0.866–1.120	0.821	
		Weighted median		0.987	0.932–1.045	0.647	
		IVW		1.002	0.961–1.045	0.92	
		Simple mode		0.965	0.873–1.067	0.498	
		Weighted mode		0.970	0.884–1.065	0.534	
	LUSC	MR Egger		0.980	0.826–1.163	0.821	
		Weighted median		1.034	0.947–1.128	0.46	
		IVW		1.023	0.966–1.084	0.438	
		Simple mode		1.051	0.894–1.235	0.554	
		Weighted mode		1.047	0.901–1.218	0.555	
	LUAD	MR Egger		1.002	0.810–1.239	0.987	
		Weighted median		0.995	0.911–1.087	0.92	
		IVW		0.994	0.927–1.066	0.872	
		Simple mode		0.865	0.715–1.047	0.152	
		Weighted mode		1.067	0.900–1.266	0.464	
HPV E7 Type 18	LUCA	MR Egger	12	1.059	0.868–1.293	0.584	
		Weighted median		0.969	0.900–1.044	0.407	
		IVW		0.965	0.914–1.019	0.197	
		Simple mode		0.981	0.867–1.110	0.768	
		Weighted mode		0.985	0.877–1.106	0.806	
	LUSC	MR Egger		1.278	0.870–1.877	0.24	
		Weighted median		0.978	0.867–1.102	0.711	
		IVW		0.933	0.834–1.043	0.222	
		Simple mode		0.939	0.774–1.140	0.539	
		Weighted mode		0.952	0.799–1.135	0.597	
	LUAD	MR Egger		1.113	0.802–1.544	0.536	
		Weighted median		1.04	0.926–1.169	0.507	
		IVW		1.028	0.945–1.118	0.524	
		Simple mode		1.075	0.865–1.336	0.528	
		Weighted mode		1.083	0.895–1.311	0.428	

Fig. 2 Causal estimates of the relationship between HPV (HPV-16 E7 protein, HPV-18 E7 protein) and lung cancer (LUCA), including its subtypes squamous cell carcinoma (LUSC) and adenocarcinoma (LUAD), represented by odds ratios (OR) and 95% confidence intervals (CI). MR: Mendelian randomization; IVW: Inverse variance weighting; LUCA: Lung cancer; LUSC: Squamous cell carcinoma; LUAD: Adenocarcinoma

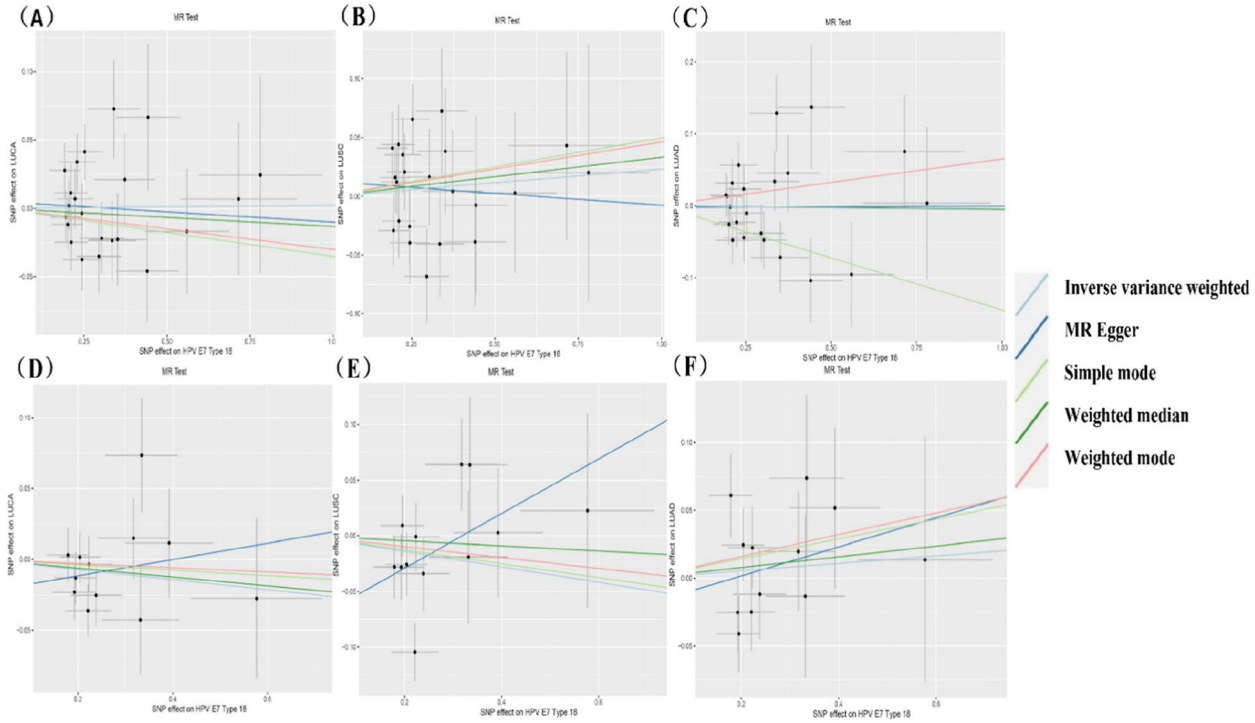


Fig. 3 Scatter plots assessing the causal relationship between HPV (HPV-16 E7 protein, HPV-18 E7 protein) and lung cancer, as well as its subtypes. Specifically, (**A**) causal estimates of HPV-16 E7 protein on lung cancer, (**B**) causal estimates of HPV-16 E7 protein on squamous cell carcinoma, (**C**) causal estimates of HPV-16 E7 protein on adenocarcinoma, (**D**) causal estimates of HPV-18 E7 protein on lung cancer, (**E**) causal estimates of HPV-18 E7 protein on squamous cell carcinoma, and (**F**) causal estimates of HPV-18 E7 protein on adenocarcinoma. The slope of each line corresponds to the causal estimate of each method. Individual SNP effects on both the outcome and exposure are depicted by vertical and horizontal lines, respectively. LUCA: Lung cancer; LUSC: Squamous cell carcinoma; LUAD: Adenocarcinoma

The causal impact of HPV on lung cancer

In forward Mendelian randomization (MR) analysis, with the exposure being HPV E7 proteins (HPV-16 E7 protein、HPV-18 E7 protein), and the outcomes being lung cancer (LUCA) and its subtypes, squamous cell carcinoma (LUSC) and adenocarcinoma (LUAD), the results indicate that genetic variations associated with HPV infection are not causally linked to the risk of lung cancer. The IVW method indicates no significant evidence of a causal relationship between HPV infection and lung cancer. Estimates from MR-Egger, weighted median, simple mode, and weighted mode all confirm this null result (Fig. 2). Scatter plots depicting the effect sizes of SNPs for HPV-16 E7 protein and HPV-18 E7 protein about lung cancer (LUCA) and its subtypes, squamous cell carcinoma (LUSC) and adenocarcinoma (LUAD), are illustrated in Fig. 3.

In the forward analysis, heterogeneity in individual SNP estimates was detected only in the MR analysis of HPV-18 E7 protein and squamous cell carcinoma using Cochran's Q statistic and the MR-IVW method ($Q=20.525$, $P=0.039$). MR-Egger regression did not reveal horizontal pleiotropy, and the MR-Egger intercept did not show significant evidence of directional pleiotropy ($P>0.05$). However, the MR-PRESSO test indicated significant horizontal pleiotropy in the MR analysis of HPV-18 E7 protein and squamous cell carcinoma ($p<0.012$) and identified rs4702371 as an outlier (Table 2). The funnel plot illustrates positions where directional pleiotropy might be present in each outcome, but assessing funnel plot symmetry is challenging due to the limited number of genetic instruments (Fig. 4). Leave-one-out analysis results indicate that SNPs with potential influence may impact the analysis, cautioning against drawing definitive conclusions (Fig. 5). The forest plot displays effect estimates and 95% confidence intervals using the TSMR method (Fig. 6). After removing outliers, a re-analysis of Mendelian randomization indicates that there is still no genetic causal relationship between HPV-18 E7 protein and squamous cell carcinoma (IVW, [OR] = 0.987, [CI] = 0.904 − 1.077, $p=0.76$, see Supplementary File S2.Supplementary Figure S1). Further sensitivity analysis indicates no heterogeneity between SNPs and no evidence of horizontal pleiotropy. The MR-PRESSO test also did not identify any outliers (Table 2).

Table 2 Sensitivity analysis of MR analysis results for exposure and outcome

Sensitivity analysis of MR										
	Exposure	Exposure	MR egger pleiotropy test		Heterogeneity test				MR PRESSO	
					Method				Global test p	Outliers
					MR Egger		IVW			
			Intercept	*P* value	Q	*P* value	Q	*P* value		
Forward MR analysis	HPV E7 Type 16	LUCA	0.005	0.784	25.290	0.191	25.387	0.231	None	
		LUSC	0.013	0.607	19.806	0.470	20.079	0.516	None	
		LUAD	-0.002	0.942	30.490	0.062	30.498	0.082	None	
	HPV E7 Type 18	LUCA	-0.023	0.364	10.006	0.440	10.911	0.451	None	
		LUSC	-0.078	0.127	16.064	0.098	20.525	0.039	<0.012	rs4702371
		LUSC[a]	-0.058	0.158	5.110	0.825	7.482	0.679	None	
		LUAD	-0.020	0.632	10.905	0.365	11.172	0.429	None	
Reverse MR analysis	LUCA	HPV E7 Type 16	0.003	0.953	11.901	0.219	11.906	0.291	None	
		HPV E7 Type 18	-0.079	0.143	10.527	0.310	13.537	0.195	None	
	LUSC	HPV E7 Type 16	-0.019	0.862	3.089	0.543	3.123	0.681	None	
		HPV E7 Type 18	-0.184	0.166	4.354	0.360	7.474	0.188	None	
	LUAD	HPV E7 Type 16	0.036	0.408	12.653	0.317	13.505	0.333	None	
		HPV E7 Type 18	0.059	0.255	16.991	0.108	19.216	0.083	None	

LUCA Lung cancer, *LUSC* Squamous cell carcinoma, *LUAD* Adenocarcinoma, *IVW* Inverse Variance Weighting

LUSC[a] MR analysis of HPV-18 E7 protein and squamous cell carcinoma after excluding outlier SNP (rs44702371)

The causal impact of lung cancer on HPV

In the reverse study, there is no evidence indicating a causal relationship between lung cancer (LUCA) and its subtypes (squamous cell carcinoma (LUSC), adenocarcinoma (LUAD)) with HPV (HPV-16 E7 protein、 HPV-18 E7 protein) (Supplementary File S2, Supplementary Figure S2, S3). The Cochran's Q test report does not indicate the presence of heterogeneity ($P > 0.05$). MR-Egger regression results show that genetic pleiotropy does not impact the outcomes ($P > 0.05$). The distortion test in MR-PRESSO analysis did not detect any outliers, further affirming the absence of evidence supporting the existence of horizontal pleiotropy ($P > 0.05$) (Table 2). The funnel plots and leave-one-out analysis indicate minimal individual SNP bias in the results, suggesting the robustness of the MR analysis (Supplementary Figure S4, S5). Forest plots illustrating the causal effects of individual SNPs between lung cancer (LUCA) and its subtypes (squamous cell carcinoma (LUSC), adenocarcinoma (LUAD)) with HPV (HPV-16 E7 protein、HPV-18 E7 protein) are presented in Supplementary File S2 and Supplementary Figure S6.

The lack of a causal relationship between HPV infection and lung cancer may be explained by several biological mechanisms: firstly, the viral load of HPV infection in the lungs might be insufficient to cause cellular transformation and carcinogenesis. Studies have shown that the persistence of viral infection and a high viral load are critical for its carcinogenic potential. Secondly, the host's immune response might play a protective role in HPV infection in the lungs, effectively controlling the spread and replication of the virus, thereby preventing the progression of infection and the development of cancer. Thirdly, the development of lung cancer involves multiple cellular signaling pathways and genomic alterations, such as EGFR mutations, KRAS mutations, and TP53 mutations. A single HPV infection may be insufficient to play a dominant role in these complex mechanisms. Lastly, differences in research methodologies, including sample selection and biomarker detection techniques, may lead to inconsistencies in study results. For instance, some studies might have used more sensitive detection techniques or stricter sample selection criteria, affecting the detection of HPV DNA. Additionally, variations in the geographical, racial, and clinical characteristics of the samples could also have a significant impact on the results. Further exploration of these potential biological mechanisms can provide a better understanding of the complex relationship between HPV infection and lung cancer and offer new directions and insights for future research.

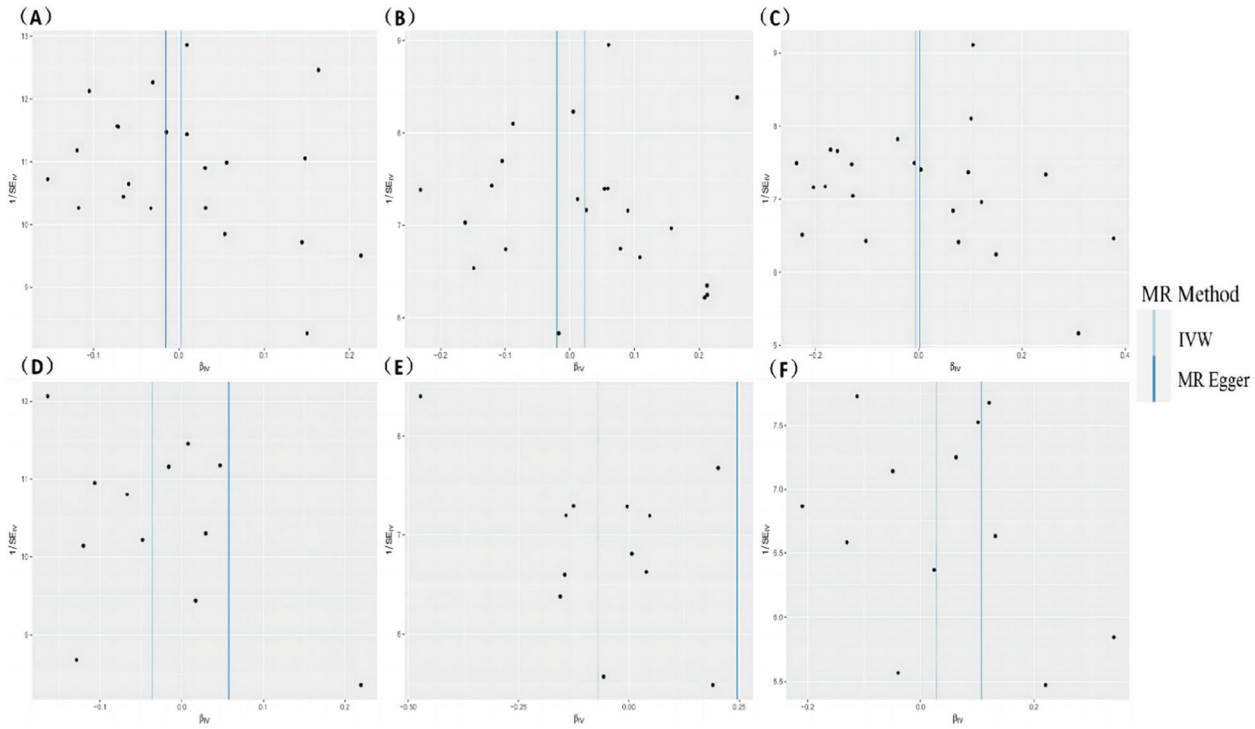


Fig. 4 Funnel plots depicting overall heterogeneity in MR estimates of the impact of HPV (HPV-16 E7 protein, HPV-18 E7 protein) on lung cancer and its subtypes. **A** Funnel plot for the causal effect of HPV-16 E7 protein on lung cancer. **B** Funnel plot for the causal effect of HPV-16 E7 protein on squamous cell carcinoma. **C** Funnel plot for the causal effect of HPV-16 E7 protein on adenocarcinoma. **D** Funnel plot for the causal effect of HPV-18 E7 protein on lung cancer. **E** Funnel plot for the causal effect of HPV-18 E7 protein on squamous cell carcinoma. **F** Funnel plot for the causal effect of HPV-18 E7 protein on adenocarcinoma. IVW, Inverse Variance Weighting

Discussion

This study is the first to comprehensively investigate the bidirectional causal relationship between HPV infection and lung cancer using multiple complementary Mendelian Randomization (MR) methods. Our MR Analysis using large-scale GWAS data consistently showed no evidence supporting a causal relationship between HPV infection and increased lung cancer risk. Similarly, reverse MR Analysis did not find a causal relationship between genetic susceptibility to lung cancer and HPV infection.

The results of this study contradict some previous reports on the association between HPV infection and lung cancer. A study by NIE Z et al. suggests that HPV16 infection may influence the development of lung cancer, particularly by regulating the SNHG1 gene and promoting angiogenesis, which is crucial for tumor growth and spread [50]. A study involving 152 cases of primary lung cancer patients as the lung cancer group and 87 individuals with benign lung lesions as the control group revealed that the incidence of HPV infection in primary lung cancer patients was higher than in those with benign lung lesions. Furthermore, the study found a close association between HPV infection and patients' TNM staging, differentiation degree, and lymph node metastasis. From these findings, it is inferred that HPV infection not only increases the risk of primary lung cancer but is also closely related to its clinical and pathological characteristics [21]. Researchers, including HARABAJSA.S, concluded from the analysis of 67 lung adenocarcinoma samples that non-small cell lung cancer patients with EGFR mutations are more likely to be infected with HPV. Additionally, high-risk HPV infection is more prevalent in lung adenocarcinomas with EGFR mutations [51]. An epidemiological study on the global role and mechanisms of high-risk human papillomavirus (HR-HPV) in lung cancer found that HR-HPV is involved in the occurrence of different subtypes of lung cancer in both smokers and non-smokers. The study proposed several potential mechanisms [52]. In summary, the evidence supports HPV infection as a cause of the occurrence and development of lung cancer, but there is no evidence indicating whether lung cancer increases the risk of HPV infection.

However, not every study has arrived at the same conclusion regarding the association between HPV and the risk of lung cancer. A recent meta-analysis on global lung cancer HPV DNA infection, stratified by pathological type and geographic region, indicates that despite

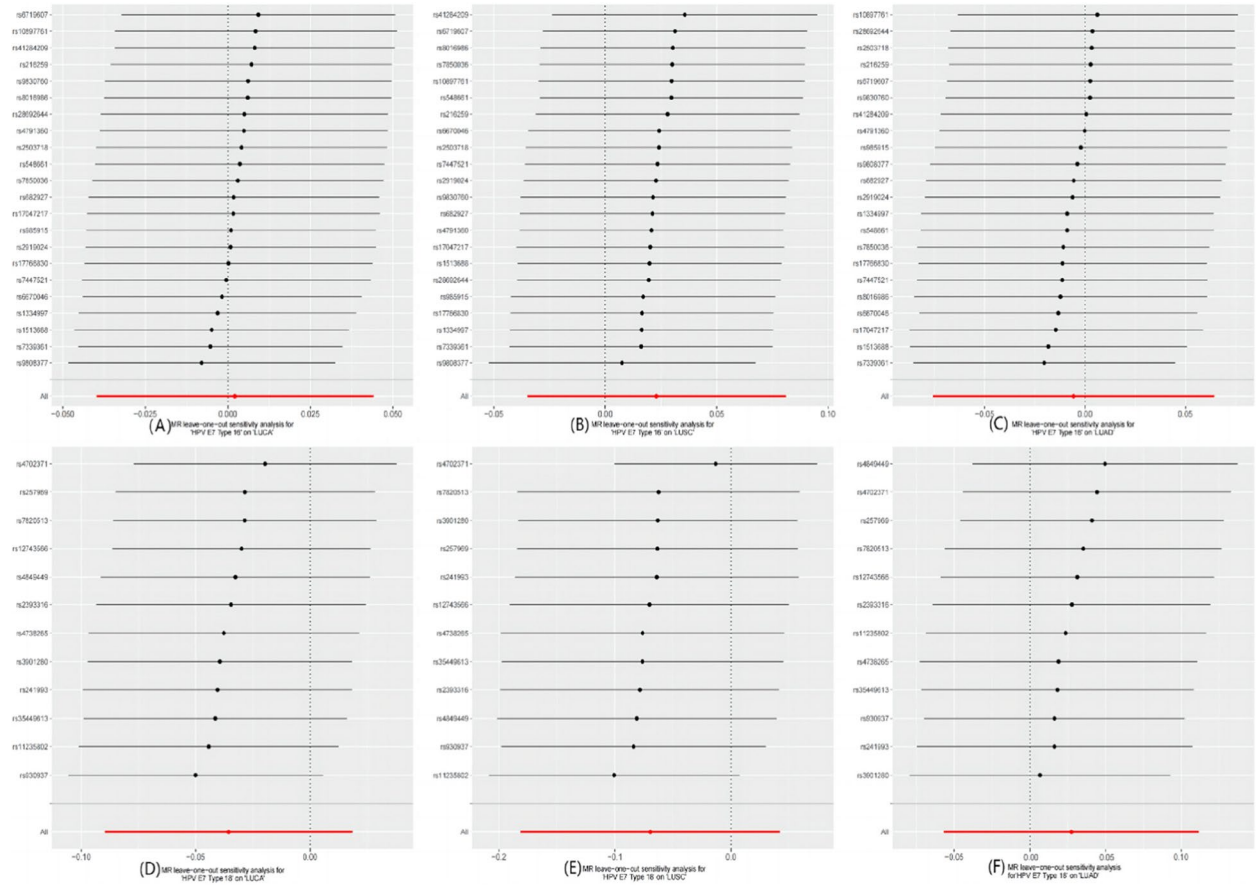


Fig. 5 Leave-one-out analysis of the causal relationship between HPV (HPV-16 E7 protein, HPV-18 E7 protein) and lung cancer and its subtypes. The red line represents the result of the random-effects IVW analysis. **A** HPV-16 E7 protein and lung cancer. **B** HPV-16 E7 protein and squamous cell carcinoma. **C** HPV-16 E7 protein and adenocarcinoma. **D** HPV-18 E7 protein and lung cancer. **E** HPV-18 E7 protein and squamous cell carcinoma. **F** HPV-18 E7 protein and adenocarcinoma. LUCA: Lung cancer; LUSC: Squamous cell carcinoma; LUAD: Adenocarcinoma. IVW: Inverse Variance Weighting

the presence of global HPV DNA positivity in lung cancer, there is a lack of conclusive evidence confirming the presence of HPV DNA in tumors. This makes it challenging to determine its carcinogenic role in the development of lung cancer, as there is a lack of robust evidence demonstrating HPV's potential involvement in the occurrence of lung cancer [29]. Data from the study conducted by Estela Maria Silva and colleagues indicate the absence of HPV DNA in a series of non-small cell lung cancers (NSCLC), further questioning the association between HPV and this specific subtype of lung cancer [28].

The controversial findings mentioned above complicate the interpretation of the causal relationship between HPV and lung cancer. Furthermore, due to the expensive human and material costs associated with randomized controlled trials (RCTs) and the involvement of numerous ethical issues, using RCTs to explore this association becomes exceedingly challenging. Therefore, we conducted this MR study. Compared to previous observational studies, studies using a bidirectional MR design are less susceptible to confounding factors and reverse causation. Simultaneously, we implemented a series of measures to fulfill the core assumptions of MR. By applying various MR methods, utilizing the PhenoScanner database, and excluding SNPs associated with confounding factors, we mitigated the potential impact of pleiotropy on the results, ensuring the robustness of our findings. Another notable feature of this study is the utilization of large sample size and SNPs from GWAS, which not only provides the study with sufficient statistical power to accurately estimate causal relationships but also enhances the credibility of the study results.

While our study has achieved significant results, it is important to note several limitations when evaluating our research. Firstly, the dataset we utilized is entirely based on individuals of European ancestry. Given the heterogeneity among racial groups, the generalizability of our study findings may be somewhat limited. Caution

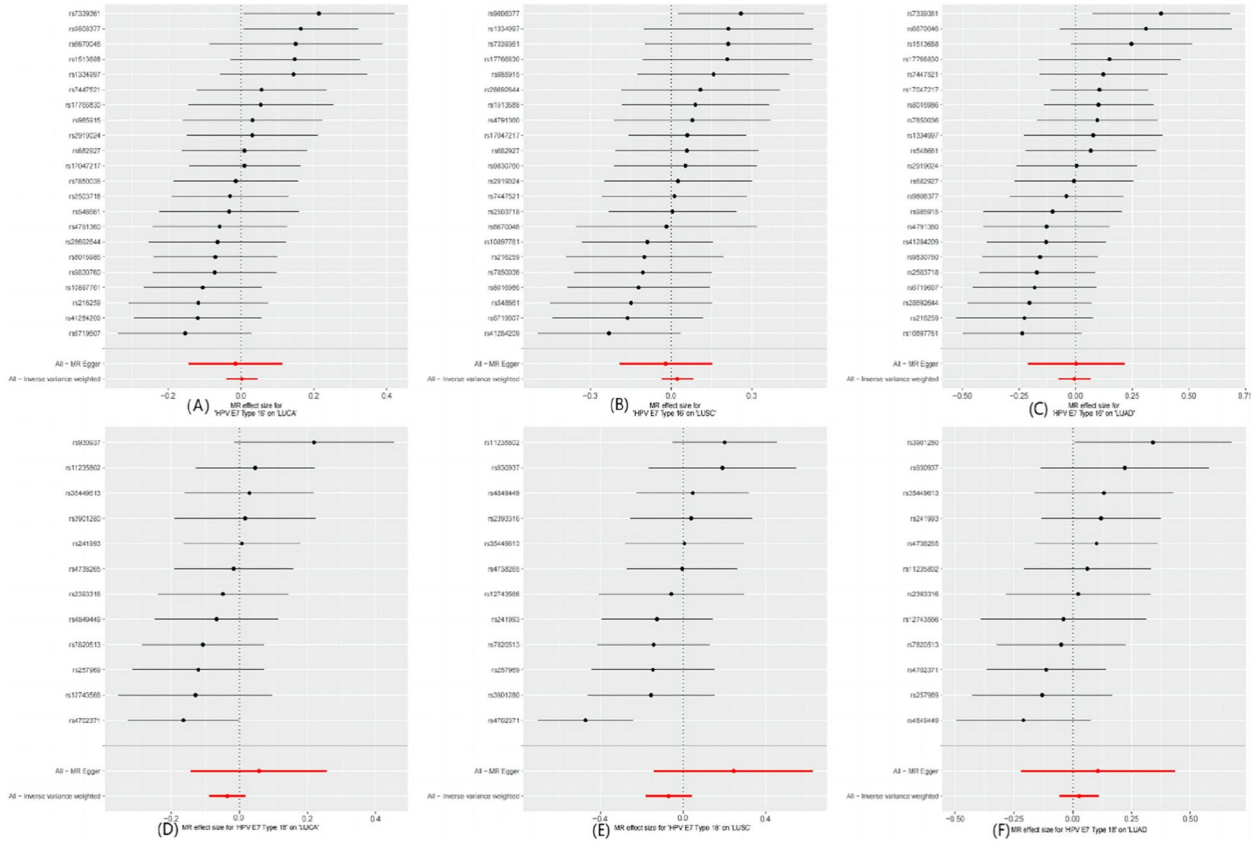


Fig. 6 Forest plots of the causal effects of single nucleotide polymorphisms (SNPs) associated with HPV (HPV-16 E7 protein, HPV-18 E7 protein) on lung cancer and its subtypes. **A** Forest plot for the causal effect of HPV-16 E7 protein on lung cancer. **B** Forest plot for the causal effect of HPV-16 E7 protein on squamous cell carcinoma. **C** Forest plot for the causal effect of HPV-16 E7 protein on adenocarcinoma. **D** Forest plot for the causal effect of HPV-18 E7 protein on lung cancer. **E** Forest plot for the causal effect of HPV-18 E7 protein on squamous cell carcinoma. **F** Forest plot for the causal effect of HPV-18 E7 protein on adenocarcinoma. LUCA: Lung cancer; LUSC: Squamous cell carcinoma; LUAD: Adenocarcinoma

is needed when extrapolating the research results to other ethnic populations and requires careful validation. Secondly, although we used $F > 10$ as the criterion for selecting strong instrumental variables in this study, in bidirectional MR, we chose instrumental variables based on a relatively lenient significance threshold of $P < 5 \times 10$^(-5), rather than the traditional $P < 5 \times 10$^(-8). Thirdly, we only obtained datasets involving the level of HPV E7 protein, and despite extensive searching of the GWAS database, we did not identify other potential instrumental variables related to different aspects of HPV infection, such as the presence of HPV DNA or other HPV proteins. This highlights the dependence of Mendelian randomization analysis on effective instrumental variables and its limitations in this regard, potentially leading to incomplete or biased interpretations of study results.

To improve the study, it is crucial to acknowledge these limitations, particularly the reliance on European ancestry data and constraints in instrumental variable selection. Future research directions should focus on validating findings in more diverse populations to enhance generalizability, or integrate interdisciplinary approaches using various types of HPV-related genetic data to enhance the interpretability and applicability of study results.

Conclusion

Overall, our bidirectional TwoSampleMR study results indicate that there is no causal relationship between HPV infection and lung cancer at the genetic level. Similarly, genetic susceptibility to lung cancer does not causally affect HPV infection. While the HPV vaccine remains crucial in preventing HPV-related cancers such as cervical cancer [14, 53], our study suggests that its role in preventing lung cancer may not be significant. This underscores the importance of continuing large-scale genetic studies and longitudinal research to gain a deeper understanding of the complex interplay between HPV infection and lung cancer risk.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13063-024-08366-5.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
We thank the MRC Integrative Epidemiology Unit OpenGWAS and the International Lung Cancer Consortium (ILCCO) for providing data support. Special thanks to the authors for their valuable advice and assistance during data analysis and manuscript revision. We appreciate the funding and support from all funding agencies. All authors express sincere gratitude for their contributions.

Authors' contributions
CYZ (Yizhuo Chen) and DM (Ming Dong) contributed to the concept and design of the study. CYZ organized the database. CYZ and ZZQ (Zhouqi Zhang) performed statistical analysis. CYZ, ZZQ, and XZQ (Ziqing Xu) wrote the first draft of the manuscript. WX (Xin Wang) and XZQ wrote parts of the manuscript. All authors participated in manuscript revision, read, and approved the submitted version.

Funding
1: Beijing science and technology innovation medical development foundation, KC2021-JX-0186-53 and KC2023-JX-0288-PQ87 to Ming Dong.
2: Tianjin key Medical Discipling (Specialty) Construction Project: TJYXZDXK-061B to Ming Dong.
3: Diversified Input Project of Tianjin Natural Science Foundation, 21JCYBJC01310 to Xin Wang.
4: Tianjin Advanced Medical Professionals Training Program: TJSQNYXXR-D2-069 to Xin Wang.

Availability of data and materials
The data sources utilized in this study were obtained from the MRC IEU OpenGWAS, developed by the MRC Integrative Epidemiology Unit at the University of Bristol (Version: v6.5.2-2022-04-11), accessible at https://gwas.mrcieu.ac.uk/. Specifically, the HPV data utilized in this study include the following GWAS IDs: prot-c-2623_54_4 (HPV E7 Type 16) and prot-c-2624_31_2 (HPV E7 Type 18).
The summary data for the genome-wide association study (GWAS) on lung cancer were retrieved through the IEU-OpenGWAS online platform, sourced from the International Lung Cancer Consortium (ILCCO). This dataset comprises 11,348 cases of lung cancer (LUCA) and 15,861 control subjects. Within the LUCA cases, there is further stratification based on histological subtypes, including 3,275 cases of squamous cell carcinoma (LUSC) and 3,442 cases of adenocarcinoma (LUAD). The GWAS IDs for LUCA and its subtypes LUSC and LUAD are "ieu-a-966," "ieu-a-967," and "ieu-a-965," respectively.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that there are no competing interests that could influence the results and conclusions of this study.

Author details
[1]Department of Lung Cancer Surgery, Tianjin Medical University General Hospital, Anshan Road No.154, Heping District, Tianjin 300052, China. [2]Department of Pediatric Surgery, Tianjin Children's Hospital (Tianjin University Children's Hospital), 238 LongYan Road, Tianjin 300134, China.

Received: 4 March 2024 Accepted: 1 August 2024
Published online: 03 September 2024

References

1. Li C, Lei S, Ding L, et al. Global burden and trends of lung cancer incidence and mortality. Chin Med J. 2023;136(13):1583–90. https://doi.org/10.1097/CM9.0000000000002529.
2. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63. https://doi.org/10.3322/caac.21834.
3. Hawrysz I, Wadolowska L, Slowinska MA, et al. Lung cancer risk in men and compliance with the 2018 WCRF/AICR cancer prevention recommendations. Nutrients. 2022;14(20):4295. https://doi.org/10.3390/nu14204295.
4. Shams-White MM, Brockton NT, Mitrou P, et al. Operationalizing the 2018 World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) Cancer Prevention Recommendations: a standardized scoring system. Nutrients. 2019;11(7):1572. https://doi.org/10.3390/nu11071572.
5. O'Keeffe LM, Taylor G, Huxley RR, et al. Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis. BMJ Open. 2018;8(10):e021611. https://doi.org/10.1136/bmjopen-2018-021611.
6. Siegel DA, Fedewa SA, Henley SJ, et al. Proportion of never smokers among men and women with lung cancer in 7 US States. JAMA Oncol. 2021;7(2):302–4. https://doi.org/10.1001/jamaoncol.2020.6362.
7. Rissanen E, Heikkinen S, Seppä K, et al. Incidence trends and risk factors of lung cancer in never smokers: pooled analyses of seven cohorts. Int J Cancer. 2021;149(12):2010–9. https://doi.org/10.1002/ijc.33765.
8. Wang P, Sun S, Lam S, et al. New insights into the biology and development of lung cancer in never smokers-implications for early detection and treatment. J Transl Med. 2023;21(1):585. https://doi.org/10.1186/s12967-023-04430-x.
9. Revel MP, Chassagnon G. Ten reasons to screen women at risk of lung cancer. Insights Imaging. 2023;14:176. https://doi.org/10.1186/s13244-023-01512-8.
10. Sun J, Xu J, Liu C, et al. The association between human papillomavirus and bladder cancer: Evidence from meta-analysis and two-sample mendelian randomization. J Med Virol. 2023;95(1):e28208. https://doi.org/10.1002/jmv.28208.
11. You EL, Henry M, Zeitouni AG. Human papillomavirus–associated oropharyngeal cancer: review of current evidence and management. Curr Oncol. 2019;26(2):119–23. https://doi.org/10.3747/co.26.4819.
12. Osmani V, Klug SJ. HPV-Impfung zur Prävention von Genitalwarzen und Krebsvorstufen – Evidenzlage und Bewertung. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2021;64(5):590–9. https://doi.org/10.1007/s00103-021-03316-x.
13. Scott-Wittenborn N, Fakhry C. Epidemiology of HPV Related Malignancies. Seminars in radiation oncology. 2021;31(4):286–96. https://doi.org/10.1016/j.semradonc.2021.04.001.
14. Shapiro GK. HPV Vaccination: An Underused Strategy for the Prevention of Cancer. Curr Oncol. 2022;29(5):3780–92. https://doi.org/10.3390/curroncol29050303.
15. de Martel C, Plummer M, Vignat J, et al. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer. 2017;141(4):664–70. https://doi.org/10.1002/ijc.30716.
16. Roman BR, Aragones A. Epidemiology and incidence of HPV-related cancers of the head and neck. J Surg Oncol. 2021;124(6):920–2. https://doi.org/10.1002/jso.26687.
17. Nelson CW, Mirabello L. Human papillomavirus genomics: understanding carcinogenicity. Tumour Virus Res. 2023;15:200258. https://doi.org/10.1016/j.tvr.2023.200258.
18. Szymonowicz KA, Chen J. Biological and clinical aspects of HPV-related cancers. Cancer Biol Med. 2020;17(4):864–78. https://doi.org/10.20892/j.issn.2095-3941.2020.0370.

19. Timbang MR, Sim MW, Bewley AF, et al. HPV-related oropharyngeal cancer: a review on burden of the disease and opportunities for prevention and early detection. Hum Vaccin Immunother. 2019;15(7–8):1920–8. https://doi.org/10.1080/21645515.2019.1600985.
20. Oyouni AAA. Human papillomavirus in cancer: Infection, disease transmission, and progress in vaccines. J Infect Public Health. 2023;16(4):626–31. https://doi.org/10.1016/j.jiph.2023.02.014.
21. Wood ZC, Bain CJ, Smith DD, et al. Oral human papillomavirus infection incidence and clearance: a systematic review of the literature. J Gen Virol. 2017;98(4):519–26. https://doi.org/10.1099/jgv.0.000727.
22. Bhattacharjee R, Das SS, Biswal SS, el. Mechanistic role of HPV-associated early proteins in cervical cancer: Molecular pathways and targeted therapeutic strategies. Crit Rev Oncol Hematol. 2022;174:103675. https://doi.org/10.1016/j.critrevonc.2022.103675.
23. Ping L, Yi Li, Bo W, et al. TLR4 gene polymorphism and susceptibility to primary lung cancer in association with chlamydia infection and HPV infection. Chin J Nosocomial Infect. 2021;31(17):2618–22.
24. Guangping Li, Hongxin Z, Lei Z, et al. Relationship Between Survivalin and VEGF Expression in Non-Small Cell Lung Cancer and HPV Infection. Chinese Journal of Nosocomial Infection. 2021;31(4):544–8.
25. Karnosky J, Dietmaier W, Knuettel H, et al. HPV and lung cancer: a systematic review and meta-analysis. Cancer Rep. 2021;4(4):e1350. https://doi.org/10.1002/cnr2.1350.
26. Huang JY, Lin C, Tsai SCS, et al. Human Papillomavirus Is Associated With Adenocarcinoma of Lung: A Population-Based Cohort Study. Front Med. 2022;9: 932196. https://doi.org/10.3389/fmed.2022.932196.
27. Yang Xuejiao Wu, Yanfei. Study on the correlation between HPV infection, IL-1β, EGFR polymorphisms, and non-smoking female lung cancer. Public Health Prev Med. 2022;33(6):115–8.
28. Drokow EK, Effah CY, Agboyibor C, et al. Microbial infections as potential risk factors for lung cancer: Investigating the role of human papillomavirus and chlamydia pneumoniae. AIMS Public Health. 2023;10(3):627–46. https://doi.org/10.3934/publichealth.2023044.
29. Rezaei M, Mostafaei S, Aghaei A, et al. The association between HPV gene expression, inflammatory agents and cellular genes involved in EMT in lung cancer tissue. BMC Cancer. 2020;20:916. https://doi.org/10.1186/s12885-020-07428-6.
30. Huang JY, Lin C, Tsai SCS, et al. Human Papillomavirus Is Associated With Adenocarcinoma of Lung: A Population-Based Cohort Study. Frontiers in medicine, 2022, 9((Huang J.-Y.) Laboratory of Statistics, Department of Medical Research, Chung Shan Medical University Hospital, Taichung, Taiwan): 932196. https://doi.org/10.3389/fmed.2022.932196.
31. EM S, VS M, PRA P, et al. Human papillomavirus is not associated to non-small cell lung cancer: data from a prospective cross-sectional study. Infectious agents and cancer. 2019;14[2023–11–13]. https://pubmed.ncbi.nlm.nih.gov/31388352/. https://doi.org/10.1186/s13027-019-0235-8.
32. Li X, Ling Y, Hu L, et al. Detection of Human Papillomavirus DNA, E6/E7 Messenger RNA, and p16INK4a in Lung Cancer: A Systematic Review and Meta-analysis. J Infect Dis. 2023;228(8):1137–45. https://doi.org/10.1093/infdis/jiad295.
33. He F, Xiong W, Yu F, et al. Human papillomavirus infection maybe not associated with primary lung cancer in the Fujian population of China. Thoracic Cancer. 2020;11(3):561–9. https://doi.org/10.1111/1759-7714.13282.
34. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ (Clinical research ed). 2018;362. https://doi.org/10.1136/bmj.k601.
35. Kurilshikov A, Medina-Gomez C, Bacigalupe R, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53(2):156–65. https://doi.org/10.1038/s41588-020-00763-1.
36. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA. 2021;326(16):1614–21. https://doi.org/10.1001/jama.2021.18236.
37. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet. 2014;46(7):736–41. https://doi.org/10.1038/ng.3002.
38. 1000 GENOMES PROJECT CONSORTIUM, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061–73. https://doi.org/10.1038/nature09534.
39. Z L, Y D, W P. Combining the strengths of inverse-variance weighting and Egger regression in Mendelian randomization using a mixture of regressions model. PLoS Genet. 2021;17(11)[2023–11–21]. https://pubmed.ncbi.nlm.nih.gov/34793444/. https://doi.org/10.1371/journal.pgen.1009922.
40. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377–89. https://doi.org/10.1007/s10654-017-0255-x.
41. Verbanck M, Chen CY, Neale B, et al. Publisher Correction: Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(8):1196. https://doi.org/10.1038/s41588-018-0164-2.
42. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46(6):1734–9. https://doi.org/10.1093/ije/dyx034.
43. Causal Relationships Between Social Isolation and Osteoarthritis: A Mendelian Randomization Study in European Population - PMC[EB/OL]. [2023–11–21]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8523904/.
44. Bowden J, Davey Smith G, Haycock PC, et al. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40(4):304–14. https://doi.org/10.1002/gepi.21965.
45. Li C, Niu M, Guo Z, et al. A mild causal relationship between tea consumption and obesity in general population: a two-sample mendelian randomization study. Front Genet. 2022;13:795049. https://doi.org/10.3389/fgene.2022.795049.
46. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. Oxford Academic[EB/OL]. 2023. https://academic.oup.com/ije/article/46/6/1985/3957932.
47. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife;7:e34408. https://doi.org/10.7554/eLife.34408.
48. Shu MJ, Li JR, Zhu YC, et al. Migraine and Ischemic Stroke: A Mendelian Randomization Study[J/OL]. Neurology and Therapy. 2022;11(1):237–46. https://doi.org/10.1007/s40120-021-00310-y.
49. Chang MJ, Liu MT, Chen MR, et al. Mendelian randomization analysis suggests no associations of herpes simplex virus infections with systemic lupus erythematosu. J Med Virol. 2023;95(3):e28649. https://doi.org/10.1002/jmv.28649.
50. Nie Z, Zhang K, Li Z, el. Human papillomavirus 16 E6 promotes angiogenesis of lung cancer via SNHG1[J/OL]. Cell Biochem Biophys. 2023;81(2):325–36. https://doi.org/10.1007/s12013-022-01121-0.
51. Harabajsa S, Šefčić H, Klasić M, et al. Infection with human cytomegalovirus, Epstein-Barr virus, and high-risk types 16 and 18 of human papillomavirus in EGFR-mutated lung adenocarcinoma. Croatian Med J. 2023;64(2):84–92. https://doi.org/10.3325/cmj.2023.64.84.
52. Osorio JC, Candia-Escobar F, Corvalán AH, et al. High-risk human papillomavirus Infection in lung cancer: mechanisms and perspectives. Biol Basel. 2022;11(12):1691. https://doi.org/10.3390/biology11121691.
53. Dykens JA, Peterson CE, Holt HK, et al. Gender neutral HPV vaccination programs: Reconsidering policies to expand cancer prevention globally. Front Public Health. 2023;11:1067299. https://doi.org/10.3389/fpubh.2023.1067299.

Publisher's Note