The influence of human papillomavirus infection on risk of colorectal cancer: a mendelian randomization study

Bo Pei¹, Peijun Liu², Shixuan Peng³, and Fuxiang Zhou¹

¹Zhongnan Hospital of Wuhan University Department of Oncology
²Central Hospital of Enshi Tujia and Miao Autonomous Prefecture
³University of South China

September 1, 2023

Abstract

Despite the well-established link between human papillomavirus (HPV) and cervical and anogenital cancers, there is ongoing debate regarding the relationship between HPV and colorectal cancer (CRC). This study aimed to evaluate the causal connection between HPV infection and CRC. To achieve this, we conducted a Mendelian randomization analysis utilizing data from genomewide association studies (GWAS) to explore the association between HPV and CRC. Our analysis revealed a significant association between genetically predicted HPV-16 infection and the risk of paternal colorectal adenocarcinoma (HPV-16: OR 1.058, 95% CI 1.013 to 1.102; p = 0.011), as well as CRC (HPV-16: OR 1.045 95% CI 1.005 to 1.085; p = 0.025). These findings provide compelling evidence for a causal effect of HPV-16 on the development of CRC. Further investigations into the underlying mechanisms and elucidation of this association are necessary to identify viable interventions for the prevention and treatment of HPV-16-associated CRC.

The influence of human papillomavirus infection on risk of colorectal cancer: a mendelian randomization study

Bo Pei¹,²*, Peijun Liu³*, Shixuan Peng⁴*, Fuxiang Zhou¹#

¹Department of Radiation Oncology and Medical Oncology, Zhongnan Hospital of Wuhan University, Hubei Cancer Clinical Study Center, Hubei Key Laboratory of Tumor Biological Behaviors, Wuhan, China
²Department of Oncology, The Central Hospital of Enshi Tujia and Miao Autonomous Prefecture, Enshi Clinical College of Wuhan University, Enshi, China
³Department of Respiratory and Critical Care Medicine, The Central Hospital of Enshi Tujia and Miao Autonomous Prefecture, Enshi Clinical College of Wuhan University, Enshi, China
⁴Department of Oncology, Graduate Collaborative Training Base of The First People’s Hospital of Xiangtan City, Hengyang Medical School, University of South China, Hengyang, China

Bo Pei, Peijun Liu, Shixuan Peng contributed equally to this study and share first authorship.

* these authors contributed equally to this work.

# Corresponding Author: Fuxiang Zhou, happyzfx@sina.com

Correspondence

Fuxiang Zhou, Department of Radiation Oncology and Medical Oncology, Zhongnan Hospital of Wuhan University, Hubei Cancer Clinical Study Center, Hubei Key Laboratory of Tumor Biological Behaviors, Wuhan, China.
Email: happyzfx@sina.com

Funding information
Natural Science Foundation of Enshi Tujia and Miao Autonomous Prefecture Government

Abstract
Despite the well-established link between human papillomavirus (HPV) and cervical and anogenital cancers, there is ongoing debate regarding the relationship between HPV and colorectal cancer (CRC). This study aimed to evaluate the causal connection between HPV infection and CRC. To achieve this, we conducted a Mendelian randomization analysis utilizing data from genomewide association studies (GWAS) to explore the association between HPV and CRC. Our analysis revealed a significant association between genetically predicted HPV-16 infection and the risk of paternal colorectal adenocarcinoma (HPV-16: OR 1.058, 95% CI 1.013 to 1.102; p = 0.011), as well as CRC (HPV-16: OR 1.045 95% CI 1.005 to 1.085; p = 0.025). These findings provide compelling evidence for a causal effect of HPV-16 on the development of CRC. Further investigations into the underlying mechanisms and elucidation of this association are necessary to identify viable interventions for the prevention and treatment of HPV-16-associated CRC.

KEYWORDS
Human papillomavirus, Colorectal cancer, Mendelian randomization

1 Introduction
Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer death around the world, with an estimated 1.9 million new cases and 935,000 deaths in 2020[1]. In the United States, CRC ranks second in cancer-related deaths overall and is the leading cause in men younger than 50 years[2]. Although lifestyle, environmental factors, genetic and obesity might have some association, the exact causes of this worrying rise are not completely understood[3]. Numerous factors contributed to the etiology of CRC, such as positive family history, male sex, age, increased bodyweight, red and processed meat intake, smoking and excessive alcohol intake[4-10]. Among the established risk factors for the formation and development of CRC are infections due to specific bacterial species or viruses. For example, microbiota research demonstrates that Bacteroides fragilis and Fusobacterium nucleatum may increase the risk for CRC[11, 12].

Among the viruses, the human papillomavirus (HPV) is a DNA virus that has a generally accepted aetiological role in cervical and anogenital cancer[13, 14]. However, although greater attention has been focused on the issue of HPV infection have association with CRC Incidence since Kirgan et al. first reported a potential link between HPV infection and colon cancer in 1990, the association of HPV and the risk of CRC still remains inconclusive[15, 16]. One subset of studies support the notion that HPV is closely associated with occurrence of CRC[16-29]. On the contrary, the other subset of studies are in a seemingly opposite direction, HPV infection has not been implicated as a risk for CRC[30-37]. Given the lack of consensus on the association between HPV infection and colorectal carcinogenesis, there is a need to conduct a Mendelian randomization (MR) analysis.

MR is a method which attempts to minimise measurement error, reverse causation and confounding by utilizing genetic variants which are known to be reliably related to modifiable risk factors of interest, to evaluate the causal effects for these risk factors on clinically relevant outcomes[38]. It presents a valuable tool and has become a widely used approach to explore the potential causal association between risk factors or modifiable exposures and disease outcomes[39]. The role of human papillomavirus in CRC remains unclear and evidence from observational studies may be subject to confounding and selection bias. To thoroughly evaluate the causal effects of HPV infection on the risk of CRC, MR analysis was conducted in this study.

Methods and Materials
2.1 Sources of Data and Selection of Ivs
The summary statistics for HPV E7 phenotypes were extracted from the IEU GWAS database. The GWAS ID of summary-level data included in this study were: prot-c-2623_54_4 (E7 protein of HPV16), prot-c-2624_31_2 (E7 protein of HPV18). The candidate genetic instruments for colorectal adenocarcinoma and colorectal cancer were sourced from the FinnGen database. It is important to note that all individuals included in the FinnGen data were of European ancestry.

In order to select IVs, SNPs need to satisfy several criteria. First and foremost, SNPs should exhibit a significant association with exposure at the genome-wide level. These SNPs were considered as instrumental variables to evaluate the causality of HPV E7 protein ($p < 5 \times 10^{-5}$, $r^2 < 0.001$ and clump distance >10000 kb). Five MR methods were used: Weighed median regression, Inverse variance weighting (IVW), Mendelian randomization-Egger (MR Egger), Simple mode, and Weighed mode.

2.2 Statistical analysis and sensitivity analyses

The heterogeneity test, specifically Cochran’s Q test in the IVW method, is utilized to assess variations between each instrumental variable (IV). If the p-value is less than 0.05, it suggests significant heterogeneity among the groups being examined. Pleiotropy refers to the phenomenon where a single gene impacts two or more traits. The MR Pleiotropy RESidual Sum and Outlier (PRESSO) method is employed to detect horizontal pleiotropy and identify outlier variants. Upon the removal of these outlier SNPs, an unbiased causal estimation can be obtained through outlier-corrected Mendelian randomization analysis. The MR-Egger intercept test is used to assess directional pleiotropy in the set of IVs. If the intercept test yields a non-zero value, it indicates the presence of directional pleiotropy within the IVs. The leave-one-out analysis is a testing method in which each SNP is sequentially removed, and the analysis is rerun to observe if the removal leads to significant changes in the results. The leave-one-out analysis is performed to assess whether the results of MR are strongly driven by a specific SNP (The flow chart of the study is shown in Figure 1).

Results

In this MR, HPV-16 increased the risk of paternal colorectal adenocarcinoma (HPV-16: OR 1.058, 95% CI 1.013 to 1.102; $p = 0.011$) and CRC (HPV-16: OR 1.045 95% CI 1.005 to 1.085; $p = 0.025$) (Fig 2A, 2B). But similar result was not shown in the subtype of HPV-18, colorectal adenocarcinoma (HPV-18: OR 0.955, 95% CI 0.893 to 1.022; $p = 0.184$), CRC (HPV-18: OR 0.949, 95% CI 0.889 to 1.011; $p = 0.106$) (Fig 2C, 2D).

We conducted a Mendelian randomization analysis to investigate the relationship between HPV-16 and colorectal adenocarcinoma. From our investigation, 23 SNPs were identified and subsequently chosen for further Mendelian randomization analysis in the colorectal adenocarcinoma group (Table S1). The IVW method has the advantage of simultaneously handling multiple SNPs and remains effective even when the correlation between SNPs is weak. Heterogeneity was observed by Cochran’s Q test ($Q = 21.54, p = 0.032$). The MR-Egger intercept analysis indicated the absence of directional pleiotropy ($p = 0.089$). Figure 3A presents scatterplots illustrating the results of these analyses. In the leave-one-out analysis, no individual SNP exerted a significant influence on the overall effect of HPV-16 (exposure) on colorectal adenocarcinoma (outcome) (Figure 3B).

In the Mendelian randomization analysis examining the relationship between HPV-16 and CRC, Figure 4A displays scatterplots that depict the findings. The leave-one-out analysis revealed that no single SNP significantly affected the comprehensive impact of HPV-16 (exposure) on CRC (outcome), as demonstrated in Figure 4B.

Table S1 SNPs were extracted from prot-c-2623_54_4

<table>
<thead>
<tr>
<th>chr.exposure</th>
<th>sample.size.exposure</th>
<th>pval.exposure</th>
<th>se.exposure</th>
<th>pos.exposure</th>
<th>id.exposure</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>994</td>
<td>3.39E-05</td>
<td>0.0461465</td>
<td>71154117</td>
<td>prot-c-2623_54_4</td>
<td>rs1334997</td>
</tr>
<tr>
<td>1</td>
<td>995</td>
<td>1.62E-05</td>
<td>0.045165</td>
<td>37479392</td>
<td>prot-c-2623_54_4</td>
<td>rs548661</td>
</tr>
<tr>
<td>1</td>
<td>996</td>
<td>5.15E-06</td>
<td>0.0966405</td>
<td>194104179</td>
<td>prot-c-2623_54_4</td>
<td>rs6670046</td>
</tr>
<tr>
<td>Rank</td>
<td>ID</td>
<td>p-value</td>
<td>log10(p-value)</td>
<td>Chromosome</td>
<td>Position</td>
<td>Tag</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>----------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>--------------</td>
</tr>
<tr>
<td>1</td>
<td>995</td>
<td>2.77E-05</td>
<td>0.185467</td>
<td>242522042</td>
<td>prot-c-2623_54_4</td>
<td>rs2919024</td>
</tr>
<tr>
<td>2</td>
<td>997</td>
<td>1.96E-06</td>
<td>0.0527163</td>
<td>179421694</td>
<td>prot-c-2623_54_4</td>
<td>rs9808377</td>
</tr>
<tr>
<td>2</td>
<td>996</td>
<td>3.55E-05</td>
<td>0.0589071</td>
<td>60866057</td>
<td>prot-c-2623_54_4</td>
<td>rs6719607</td>
</tr>
<tr>
<td>2</td>
<td>994</td>
<td>4.72E-05</td>
<td>0.175043</td>
<td>55970893</td>
<td>prot-c-2623_54_4</td>
<td>rs17047217</td>
</tr>
<tr>
<td>3</td>
<td>996</td>
<td>3.69E-05</td>
<td>0.0571532</td>
<td>145529197</td>
<td>prot-c-2623_54_4</td>
<td>rs2864426</td>
</tr>
<tr>
<td>5</td>
<td>997</td>
<td>4.18E-05</td>
<td>0.0907434</td>
<td>27610789</td>
<td>prot-c-2623_54_4</td>
<td>rs7447521</td>
</tr>
<tr>
<td>6</td>
<td>997</td>
<td>9.99E-06</td>
<td>0.126081</td>
<td>81380831</td>
<td>prot-c-2623_54_4</td>
<td>rs2503718</td>
</tr>
<tr>
<td>6</td>
<td>997</td>
<td>3.59E-05</td>
<td>0.0538665</td>
<td>65433784</td>
<td>prot-c-2623_54_4</td>
<td>rs985915</td>
</tr>
<tr>
<td>6</td>
<td>997</td>
<td>2.14E-05</td>
<td>0.0496137</td>
<td>24931083</td>
<td>prot-c-2623_54_4</td>
<td>rs216259</td>
</tr>
<tr>
<td>9</td>
<td>993</td>
<td>8.34E-06</td>
<td>0.0458036</td>
<td>1358244</td>
<td>prot-c-2623_54_4</td>
<td>rs682927</td>
</tr>
<tr>
<td>9</td>
<td>997</td>
<td>4.49E-06</td>
<td>0.052959</td>
<td>95589160</td>
<td>prot-c-2623_54_4</td>
<td>rs7850036</td>
</tr>
<tr>
<td>11</td>
<td>996</td>
<td>3.55E-06</td>
<td>0.0944671</td>
<td>80884316</td>
<td>prot-c-2623_54_4</td>
<td>rs10897761</td>
</tr>
<tr>
<td>13</td>
<td>996</td>
<td>1.34E-05</td>
<td>0.0778514</td>
<td>109504132</td>
<td>prot-c-2623_54_4</td>
<td>rs7339361</td>
</tr>
<tr>
<td>13</td>
<td>997</td>
<td>3.60E-05</td>
<td>0.0846543</td>
<td>50005050</td>
<td>prot-c-2623_54_4</td>
<td>rs28692644</td>
</tr>
<tr>
<td>13</td>
<td>990</td>
<td>9.84E-06</td>
<td>0.0664491</td>
<td>48650492</td>
<td>prot-c-2623_54_4</td>
<td>rs41284209</td>
</tr>
<tr>
<td>14</td>
<td>997</td>
<td>3.90E-06</td>
<td>0.0721241</td>
<td>87365603</td>
<td>prot-c-2623_54_4</td>
<td>rs8016986</td>
</tr>
<tr>
<td>17</td>
<td>997</td>
<td>1.56E-05</td>
<td>0.0463611</td>
<td>9713505</td>
<td>prot-c-2623_54_4</td>
<td>rs4791360</td>
</tr>
<tr>
<td>18</td>
<td>994</td>
<td>2.46E-05</td>
<td>0.0540585</td>
<td>37304602</td>
<td>prot-c-2623_54_4</td>
<td>rs1513688</td>
</tr>
<tr>
<td>18</td>
<td>995</td>
<td>2.44E-05</td>
<td>0.0496934</td>
<td>44040660</td>
<td>prot-c-2623_54_4</td>
<td>rs17766830</td>
</tr>
</tbody>
</table>

**Fig. 1** Research Flowchart
Fig. 2 Univariable mendelian randomization analysis of HPV Infection on colorectal cancer risk.
Table 1: Comparison of different methods for analyzing the relationship between HPV infection and colorectal adenocarcinoma.

<table>
<thead>
<tr>
<th>Method</th>
<th>p-value</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR bigger</td>
<td>0.014</td>
<td>0.940(0.833-1.057)</td>
</tr>
<tr>
<td>Weighted median</td>
<td>0.011</td>
<td>1.001(0.990-1.012)</td>
</tr>
<tr>
<td>Inverse variance weighted</td>
<td>0.011</td>
<td>1.043(1.000-1.143)</td>
</tr>
<tr>
<td>Simple mode</td>
<td>0.185</td>
<td>1.000(0.999-1.001)</td>
</tr>
<tr>
<td>Weighted mode</td>
<td>0.118</td>
<td>0.990(0.973-1.007)</td>
</tr>
</tbody>
</table>

Fig. 3 Scatter and leave one out plots demonstrating influential outliers in univariable MR of HPV infection and colorectal adenocarcinoma.
Discussion

Oncogenic papillomaviruses have been linked to both benign and malignant cervix and other anogenital diseases[14]. Given HPV’s well-known function in cervical and anal carcinogenesis, several studies have shown that HPV is also linked to CRC[13, 16–28]. DNA viruses have been shown to activate proto-oncogenes (including p53, pRB, and c-myc)[40, 41], and some studies have hypothesized that the ras oncogene collaborates with the HPV E6/E7 genes to cause full transformation of normal cells[42, 43]. The existence of HPV DNA in colon cancers is a source of contention. While previous investigations failed to find HPV DNA in colon biopsy samples[25], a growing corpus of current data shows that HPV 16 and 18 infections may play a role in the development of colorectal cancer[16].

This is the first study to employ Mendelian randomization to investigate whether high-risk HPV infection is genetically linked to colorectal carcinogenesis as well as independent effects. The IEU GWAS database was used to provide summary statistics for HPV E7 phenotypes. We discovered the influence of human papillomavirus infection on colorectal cancer risk using this extensive statistical data[16]. The role of HPV in the carcinogenesis of colorectal cancer is a contentious issue, with HPV-16 being the most prevalent, high-risk subtype of the causative type; HPV-18, while a high-risk causative subtype in cervical cancer, is infected in a lower percentage of cases than HPV-16, and we discovered that the database samples contained many samples that had not been tested for HPV-18. We discovered that while HPV-18 infection was not linked to an elevated risk of colon cancer, HPV-16 infection was[33]. In this work, we used Mendelian randomization techniques to identify the causal, independent relationship between the risk of colorectal cancer and the subtypes of the HPV virus. In order to determine the relationship between human papillomavirus infection and colorectal cancer, Marina K. Ibragimova et al. examined the independent impact of HPV on the risk of developing colorectal cancer[24]. The amount of HPV infection in colorectal tumor tissue was found to be statistically significant, and the relative risk of developing colorectal cancer as a result of HPV infection was found to be RR (95% CI)=2.97 (1.42-6.22), p=0.0039, based on the data obtained[24]. Xianhui Zhang et al. identified a pooled prevalence and OR of 0.31 (95% CI: 0.24-0.37) and 2.03 (95% CI: 0.79-5.26), respectively, among 193 patients with colorectal adenomas in another meta-analysis based on a Chinese sample. The prevalence of HPV16 and HPV18 in HPV-positive malignancies varied from 57.9% to 100% and 0% to 39.7%, respectively[16].

The correlation between HPV and the risk of colorectal cancer has been linked to a number of processes. The relationship between HPV infection, constitutive Stat3 activity, and IL-17 levels was first discovered by Yi Xin Li et al[44]. This relationship may orchestrate a pro-inflammatory microenvironment in the colorectum, which in turn may promote carcinogenesis and may aid in the progression of colorectal cancer CRC[44]. 14.2% of the CRC samples included HPV DNA, with HPV16 being the most common kind. No discernible
variations between clinical and pathological Except for age, there were no discernible variations in clinical
and pathological characteristics between CRCs with and without HPV. Patients who tested positive for HPV
were noticeably younger (p = 0.05). There was no significant correlation between the presence of HPV and
overexpression of p16INK4A (p = 0.325).\[45]\]

Mendelian randomization was employed in this investigation to ascertain the independent and causative
effects of high-risk HPV infection on colorectal carcinogenesis. Sensitivity analysis was used to ensure the
reliability of the findings and further supported the effect of HPV-16 on the risk of colorectal cancer\[28].
Additionally, we conducted a combined analysis using information from other research, which improved the
validity and applicability of our findings. Additionally, our study has several drawbacks. We were unable
to identify how much HPV infection changed this risk. It is challenging to directly compare outcomes be-
tween research because of variations in methodologies and assumptions. Although there are differences in
methodologies, estimates, and levels of interpretation, it is impossible to directly compare the independent
effects of HPV infection in our investigation with those in earlier observational studies. When compared to
the short-term impacts documented in observational research, estimates of MR may reflect long-term HPV
infection effects. To ascertain this, additional examination of the individual data is required. We still need
more research to examine more intricate molecular pathways and biological processes in order to completely
comprehend the association between HPV and colorectal cancer, even though MR approaches can offer inde-
pendent estimations of the effects of HPV infection. Our research adds to the body of knowledge supporting
the link between the development of colorectal cancer and clusters of high-risk human papillomavirus. This is
crucial for comprehending how colorectal cancer develops and for creating public health initiatives. To fully
understand the role of HPV infection in the emergence of colorectal cancer, additional research is required,
including more in-depth analyses of individual data as well as studies of molecular mechanisms.

Conclusion
Our findings suggest that HPV-16 infection is associated with an increased risk of developing
colorectal cancer. However, no significant association was observed between HPV-18 infection and colorectal
cancer risk. Our MR analysis provides strong evidence for a causal effect of HPV infection on CRC risk
and contributes to a better understanding of its etiology. Further elucidation of this association and the
underlying mechanisms is needed to identify feasible interventions to promote the prevention and treatment
of HPV-16-associated CRC.

AUTHOR CONTRIBUTIONS
Bo Pei, Peijun Liu, Shixuan Peng conceived and initiated the project, conducted this MR analysis, and
wrote the manuscript. Fuxiang Zhou: reviewed and edited the manuscript. All authors contributed to the
article and approved the submitted version.

ACKNOWLEDGMENTS
We are grateful to the IEU OpenGWAS participants and genome-wide association study
consortia which made their summary statistics publicly available for this study, without which this en-
deavor would not have been possible. This work was supported by the Natural Science Foundation of Enshi
Tujia and Miao Autonomous Prefecture Government (D20220060).

CONFLICTS OF INTEREST
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
All data relevant to the study are included in the article are available from the corresponding author.

References


