Health Science Reports

WILEY

Health Science Reports
Open Access

ORIGINAL RESEARCH OPEN ACCESS

Genetic Association Between High-Risk HPV (HPV16 and HPV18) Infection and Tumor Development: A Mendelian Randomization Analysis

You Wu | Haiyang Xia | Feng Du | Jian Zhang | Xulin Jiang | Jun Tang | Zhihong Li

Department of Gastrointestinal Surgery, Ezhou Central Hospital, Affiliated to Hubei University of Science and Technology, Ezhou, Hubei, People Republic of China

Correspondence: Zhihong Li (lizhihongdoctor@163.com)

Received: 22 October 2024 | **Revised:** 26 March 2025 | **Accepted:** 2 April 2025

Funding: This study was supported by Ezhou Science and Technology Bureau Basic and Talent Innovation Project (Grants EZ01-007-20220006 and EZ01-007-20230125).

Keywords: cancers | HPV | human papillomavirus | Mendelian randomization | neoplasms

ABSTRACT

Background and Aims: Observational and experimental studies have provided substantial evidence supporting a link between cervical cancer and human papillomavirus (HPV) infection. Nevertheless, there is uncertainty regarding the association of benign and malignant cancers with HPV infection.

Methods: The study was divided into two approaches, Mendelian randomization (MR) and multivariate Mendelian randomization (MVMR), to investigate the link between HPV and both benign and malignant cancers. This study employed genetic variants as instrumental variables to mitigate potential biases arising from confounding factors and reverse causality. In 338 cases and 1000 controls in the European ancestry of Germany, independent genetic variations were identified as having a substantial correlation ($p < 5 \times 10^{-5}$) with exposure and HPV infection. The outcome variables data of various carcinomas were acquired from the Genome-wide association summary data. Meanwhile, benign tumors from the FinnGen and UK Biobank (UKB) consortium were acquired as well.

Results: Following correction for multiple testing, the MVMR method was employed and the causal association was investigated between genetic liability to HPV infection and various malignancies, including bone and articular cartilage, bladder cancer, secondary malignant neoplasm of the liver, prostate cancer, as well as benign tumors including melanocytic naevi of the lip, brain, bronchus and lung, lip, mouth and pharynx, pancreas, and haemangioma and lymphangioma, and female genitalia.

Conclusions: From a genetic standpoint, HPV may contributes to the formation of benign and malignant tumors in female genital cancer as well as malignancies in other regions of the body, which should inform public health policy.

Abbreviations: CI, confdence interval; E2F, transcription factor; GWAS, Genome-Wide Association Study; IVW, inverse variance weighting; LD, linkage disequilibrium; MR, Mendelian randomization; MVMR, Multivariate Mendelian Randomization; OR, odds ratio; P53, tumor suppressor gene; PDZ protein, proteins containing the PDZ structural domain; pRb, Blocking unphosphorylated cell cycle progression phase; SNP, single-nucleotide polymorphism; WM, weighted mean.

You Wu, Haiyang Xia, and Feng Du contributed equally to this study.

1 | Introduction

Human papillomavirus (HPV) is recognized as the predominant sexually transmitted infection globally, which is a double-stranded circular DNA virus. Its replication occurs within the keratinocyte nucleus found in the skin or mucosal tissues [1]. According to estimated data, 79 million individuals in the United States have been identified as having contracted HPV and about 45.2% of males aged 18 to 59 years are affected by this virus [2, 3]. To date, more than 200 different HPV genotypes have been identified, at least 18 of which fall into the "high-risk" group and are a major contributor to cancer [4–6].

Observational studies provide evidence that HPV infection is the primary cause of cervical cancer [7]. Additional data indicates that infection with the HPV virus is linked to extrauterine cancers. These include malignancies affecting the oral and pharyngeal regions, larynx, esophagus, lungs, breasts, bladder, ovaries, colon, anus and anal canal, prostate, vulva, stomach, and skin [8–20]. While there is an established association between HPV virus infection and extrauterine tumors, the precise causative link between HPV and these malignancies remains unknown.

The current body of information regarding the causative link between HPV infection and the subsequent risk of related tumors primarily stems from observational research. However, it is essential to acknowledge that these studies are vulnerable to confounding bias and reverse causal relationships [8–20]. A statistical technique called Mendelian randomization (MR) utilizes genetic variations in the germline as instrumental variables (IVs) to find the causal effects of presumed risk factors. Due to the random assortment of genetic variations during conception, these variations remain unaffected by reverse causal relationships. Furthermore, in the absence of pleiotropy, which refers to the connection between genetic variations and diseases through alternative pathways, these variations can offer dependable assessments of disease susceptibility. MR analysis is becoming more prevalent in examining the possible effects of treatments on disease risk due to its ability to circumvent several constraints associated with conventional observational research. The genetic variety of an individual is determined and established from the moment of birth and remains constant throughout their lifetime. The utilization of MR emerges as a promising approach for deducing the causal association between exposure and outcomes [21, 22].

As of this writing, the existing repertoire of HPV types exceeds 200, although only a limited subset has been linked to the pathogenesis of human cancer. Among these, HPV16 and HPV18 have emerged as the prevailing culprits [23]. Hence, our primary emphasis revolves around examining and analyzing the E7 protein found in malignancies associated with HPV16 and HPV18. This study aims to confirm the previously described malignancies via a genetic perspective by employing two samples and multivariate MR methods. Additionally, a thorough investigation was done on tumors associated with the E7 protein of HPV16 and HPV18, including both malignant and benign tumors.

2 | Materials and Methods

2.1 | Study Design

Utilizing single nucleotide polymorphisms (SNPs) as the variable tool, the study employed a two-sample MR analysis to find the link between HPV and several other neoplasm [24]. To optimize the precision of the outcomes, it was essential to confirm three crucial assumptions throughout the procedure as depicted in Figure 1 [25]. For MR studies, since it relies on summary-level data obtained from publicly available genome-wide association study (GWAS), no additional ethical approval was deemed necessary. The reporting follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement of Mendelian Randomization (Supporting Information File S1).

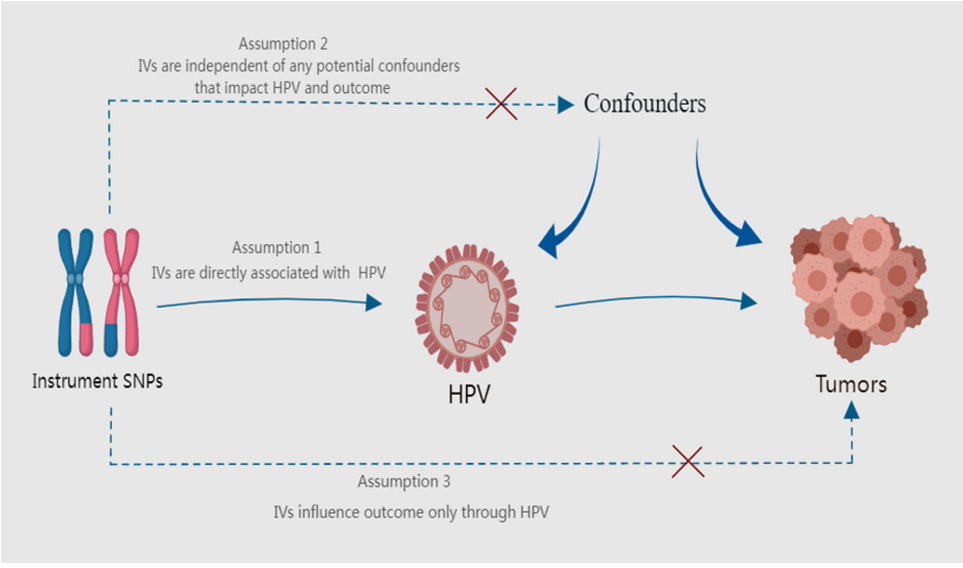

FIGURE 1 | Overview of the Mendelian randomization design. Abbreviation: IVs, instrumental variables.

2.2 | Summary Statistics of HPV16, HPV18, and Tumors (Benign and Malignant)

The genetic variants that exhibit a significant association with HPV infection were chosen from the OpenGWAS (https://gwas.mrcieu.ac.uk/; version: v7.5.5—2023-08-09) developed by Medical Research Council (MRC), Integrative Epidemiology Unit (IEU) [26]. The summary-level data included in this investigation consisted of two GWAS IDs: prot-c-2623_54_4 (representing the E7 protein of HPV16) and prot-c-2624_31_2 (representing the E7 protein of HPV18). The study population comprised a European population, including 338 individuals with chronic HPV infection and 1000 control individuals [27]. The summary statistics regarding the outcome variable were obtained from the MRC IEU OpenGWAS database. There was a comprehensive set of 31 outcomes, encompassing various types of cancers such as benign tumors of the brain, bronchus and lung, lip, uterus, oropharynx, lower limb bones, melanocytic naevi of the lip, rectum, pancreas, hemangioma, lymphangioma and the female sexual organs. Malignant tumors including oropharyngeal cancer, lung cancer (including non-small cell lung cancer), bladder cancer, colorectal cancer, bone, and articular cartilage tumors, breast cancer (including ER+ breast cancer), anus and anal canal cancer, pancreatic cancer, chronic lymphoblastic leukemia, secondary liver cancer, prostate cancer, female genital cancers, laryngeal cancer, skin cancer, ovarian cancer, esophageal cancer, vulva cancer, and gastric cancer. To establish a robust inference about the causative connections between HPV and related cancers, the aforementioned summary statistics were made available through the utilization of the TwoSampleMR (R program; version 0.5.7) in the IEU OpenGWAS database [28].

2.3 | Genetic Instrumental Variable Selection

IVs were employed to find the causal link between HPV and the associated neoplasms. A stringent quality review approach was undertaken to identify eligible instrumental factors for HPV16 and HPV18. Initially, we identified SNPs associated with particular HPV at a certain threshold of statistical significance ($p < 5 \times 10^{-5}$) throughout the whole genome as potential IVs for further MR study. In addition, to establish the independence of the instrumental factors associated with each exposed phenotype, we employ a clustering approach that relies on linkage disequilibrium (LD) to exclude SNPs exhibiting significant LD ($r^2 = 0.001$; kb = 10000). The clumping procedure was performed utilizing the reference panel of European origin from the 1000 Genomes Project, to assess the LD between SNPs. In addition, the SNPs exhibiting non-concordant alleles and palindromic SNPs with uncertain strands that could not be rectified throughout the harmonizing of the exposure and outcome data were excluded.

2.4 | Statistical Analysis

To evaluate the data, five MR methods were utilized, inverse variance weighting (IVW), weighted median, weighted mode, simple mode, and MR-Egger method. The IVW approach is distinguished by its utilization of the inverse of resultant variance as weight and its exclusion of the intercept. In an ideal scenario where all chosen genetic variants are supposed to be legitimate IVs without pleiotropy, the IVW method can yield unbiased estimations of causation [29]. Regarding MR-Egger, the presence of influential genetic factors might reduce its statistical ability. However, it can estimate the corrected causal impact and provide unbiased estimates, even when the IVs are not completely valid [30]. The weighted median technique may achieve an accurate and robust estimation of effects when a minimum of 50% of the information derived from reliable instruments is accessible and available [31]. The reliability of the weighted mode is contingent upon the validity of the largest selection of instruments exhibiting similar causal effects. Given the distinctive attributes of these methodologies, the primary reporting approach employed in this study was the IVW method, with the remaining four approaches serving as supplementary measures to IVW.

The causal relationship in MR analysis can only be described when all three essential assumptions are met. Effective sensitivity analysis is crucial, including pleiotropy testing and testing for heterogeneity across IVs. The initial approach involved the utilization of Cochran's Q-test to evaluate the heterogeneity of causal estimates, whereas MR Egger regression was utilized to identify heterogeneity. A $p < 0.05$ was indicative of statistically significant heterogeneity. If a statistically significant level of heterogeneity ($p < 0.05$) was detected, the random effects IVW technique was employed. To assess the possible pleiotropic effects of instrumental factors, we employed MR-Egger regression. The representation of the validity of the findings may be captured by the intercept term in MR-Egger regression. The $p < 0.05$ suggested the existence of horizontal pleiotropy. Lastly, the leave-one-out analysis was employed to validate the presence of outliers that significantly impact the outcome. This was achieved by systematically removing each SNP and applying the IVW approach to the obtained data afterward [30, 32].

To mitigate any ambiguity in our findings, we employed a MIVW approach to adjust for the influence of smoking and obesity (BMI) on the observed outcomes, similar to the approach of Grenville et al. [33–40]. The genetic variations closely related to smoking and BMI were obtained from the European population of IEU and included in multivariate MR analysis. Bonferroni correction was used as a conservative approach to address the issue of multiple testing. This correction calculated the significance level as 0.05 divided by the number of exposures ($N_{exposure}$; i.e., $P_{HPV16} < 0.0025$ (0.05/20), $P_{HPV18} < 0.0029$ (0.05/17)), when the Bonferroni corrected significance was above but the p-value was less than 0.05, the threshold was seen as indicative of supporting evidence for a probable correlation [41]. A two-sided p-value less than 0.05 was considered statistically significant. The R packages "TwoSampleMR" and "Mendelian Randomization" (v 0.5.7) were utilized to perform the necessary instrumental variable selection and quality control procedures.

3 | Results

3.1 | Selection of Instrumental Variables (IVs)

As there is no aggregated summary data set on nasopharyngeal cancer in European populations, we used the summary data for oropharyngeal carcinoma from Hispanic or Latin American

populations. All other summary data sets used in our study were from European populations. The number of cases ranged from 87 to 25509. The detailed information of GWASs and IVs we used are listed in Supporting Information Files S2 and S3.

3.2 | MR Estimates of HPV16 E7 Protein

The current study evaluated the causal association between the E7 protein of HPV16 infection and various malignancies among individuals of European heritage. Upon implementing the IVW methodology as our primary reporting approach, we identified a causal association between HPV16 infection and various cancers. The results of the study indicate that there was a positive association between oropharyngeal cancer (OR = 1.22, 95% CI: 1.03–1.45, $p = 0.024$), non-small cell lung cancer (OR = 1.14, 95% CI: 1–1.29, $p = 0.048$), breast cancer (OR = 1, 95% CI: 1–1.001, $p = 0.029$), bone and articular cartilage (OR = 1.5, 95% CI: 1.14–1.97, $p = 0.003$), bladder cancer (OR = 1, 95% CI: 0.99–1, $p = 0.033$), colorectal cancer (OR = 1.06, 95% CI: 1–1.12, $p = 0.047$), and anus and anal canal (OR = 0.74, 95% CI: 0.56–1, $p = 0.048$). The presence of benign tumors has been observed in various anatomical locations. Specifically, melanocytic naevi of the lip (OR = 1.36, 95% CI: 1.03–1.78, $p = 0.028$), brain (OR = 0.63, 95% CI: 0.46–0.87, $p = 0.004$), bronchus and lung (OR = 1.4, 95% CI: 1.1–1.79, $p = 0.007$), lip (OR = 0.79, 95% CI: 0.67–0.94, $p = 0.007$), uterus (OR = 1.11, 95% CI: 1–1.24, $p = 0.04$), mouth and pharynx (OR = 0.93, 95% CI: 0.86–1, $p = 0.044$), and long bones of the lower limb (OR = 1.17, 95% CI: 1.01–1.35, $p = 0.033$). Furthermore, the study revealed a positive, nonsignificant correlation between genetic susceptibility to HPV16 infection and the occurrence of laryngeal cancer and malignant neoplasm of the stomach. Also a negative correlation between genetic susceptibility to HPV16 infections and the occurrence of ovarian cancer, esophageal cancer, skin cancer, and malignant neoplasm of the vulva. Interestingly, it is essential to acknowledge that this particular finding did not achieve statistical significance either (Figure 2A and Supporting Information File S3: Table S1).

3.3 | MR Estimates of HPV18 E7 Protein

The IVW approach was chosen as the principal strategy to establish the causal link between HPV18 infection and its associated cancers. The results revealed a significant causal link between the infection caused by HPV18 and the development of malignancies. The outcomes indicated a statistically significant association with pancreatic cancer (OR = 1.48, 95% CI: 1.09–2, $p = 0.011$), chronic lymphocytic leukemia (OR = 0.72, 95% CI: 0.56–0.93, $p = 0.013$), lung adenocarcinoma (OR = 1.05, 95% CI: 1.01–1.09, $p = 0.008$), secondary malignant neoplasm of the liver (OR = 1, 95% CI: 1–1.001, $p = 0.049$), ER+ breast cancer (OR = 0.96, 95% CI: 0.92–1, $p = 0.049$), prostate cancer (OR = 1, 95% CI: 1–1.001, $p = 0.006$), and female genital cancer (OR = 1, 95% CI: 1–1.001, $p = 0.002$). The presence of benign tumors in several anatomical locations were also examined and the outcomes of the presence of these tumors were significant in the rectum (OR = 0.84, 95% CI: 0.77–0.93, $p = 0.0003$), pancreas (OR = 0.64, 95% CI: 0.46–0.89, $p = 0.007$), haemangioma and lymphangioma (OR = 0.82, 95% CI: 0.72–0.94, $p = 0.004$), and female genital organs (OR = 0.83, 95% CI: 0.71–0.98, $p = 0.027$). Similarly, the data also showed a positive but nonsignificant correlation between genetic susceptibility to HPV18 infection and the occurrence of laryngeal cancer, ovarian cancer, esophageal cancer, malignant neoplasm of the vulva, and malignancy of the stomach. However, it is important to note that only skin cancer exhibits a genetic predisposition to HPV18 infections, and this link is inversely related (Figure 2B and Supporting Information File S3: Table S2).

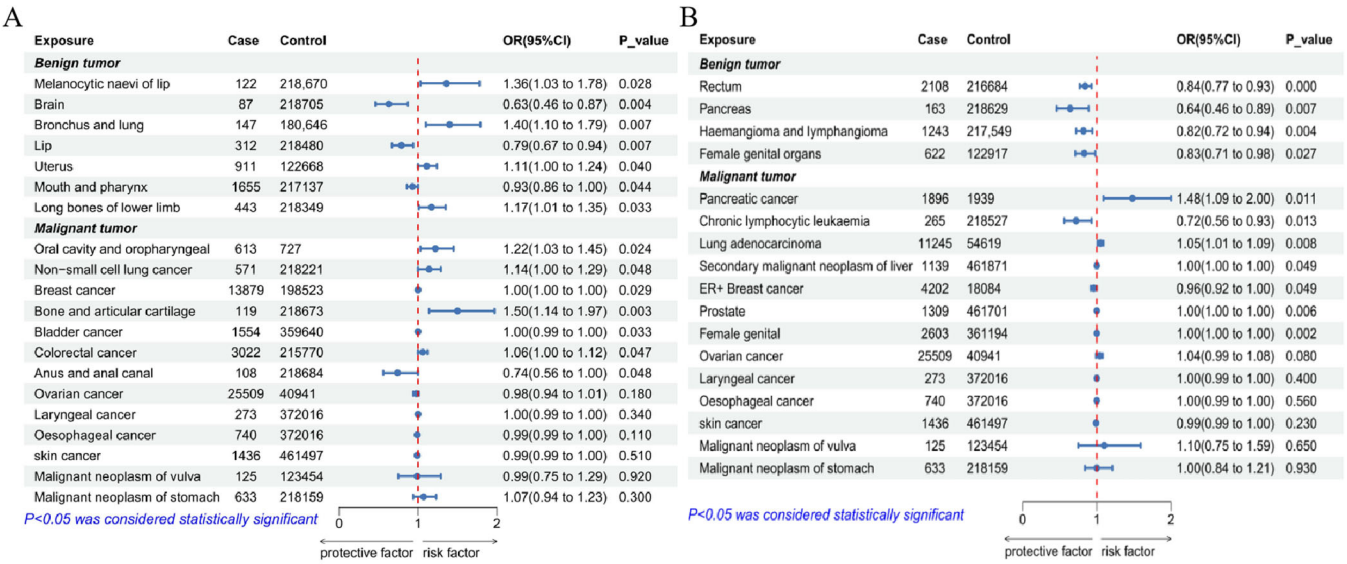

FIGURE 2 | Causal estimates of HPV infection on site-specific tumors by MR analysis. The causal effects of the E7 protein of HPV16 factors on site-specific tumors are depicted by forest plots (A). The causal effects of the E7 protein of HPV18 on site-specific tumors are shown by forest plots (B). The odds ratio (OR) was computed utilizing the random effects inverse variance weighted (IVW) approach, while the horizontal bars in the visual representation indicate the 95% confidence interval (CI).

3.4 | The Multivariable MR Estimates of HPV16 E7 Protein

The multivariable multiple regression analysis was conducted, adjusting for BMI and cigarette smoking. The results revealed a significant causal relationship between genetic susceptibility to HPV16 infection and the development of bladder cancer (OR = 1, 95% CI = 0.9–1, $p = 0.02$) and bone and articular cartilage carcinoma (OR = 1.44, 95% CI = 1.1–1.96, $p = 0.02$). The study observed associations between various types of benign tumors, specifically melanocytic naevi of the lip (OR = 1.43, 95% CI = 1.06–1.9, $p = 0.021$), brain (OR = 0.68, 95% CI = 0.47–0.97, $p = 0.032$), bronchus and lung (OR = 1.38, 95% CI = 1.05–1.8, $p = 0.023$), lip (OR = 0.75, 95% CI = 0.62–0.91, $p = 0.003$), and mouth and pharynx (OR = 0.9, 95% CI = 0.83–0.98, $p = 0.01$) in individuals of European ancestry residing in Germany. Although the remaining results did not achieve statistical significance, the findings indicate a positive association between genetic susceptibility to high-risk HPV (HPV16 & HPV18) infections and the development of colorectal cancer, laryngeal cancer, skin cancer, and malignant neoplasms of the oral cavity and oropharynx. Conversely, between genetic susceptibility to high-risk HPV (HPV16 & HPV18) infections and cancer associated with the vulva, stomach, anus, and anal canal, an inverse relation was observed (Figure 3, Supporting Information File S5: Table S1).

3.5 | Multivariable MR Estimates of HPV18 E7 Protein

Here we identified a causal association between genetic susceptibility to HPV18 infection and the development of secondary malignant neoplasms in the liver (OR = 1, 95% CI = 0.99–1, $p = 0.039$), prostate cancer (OR = 1, 95% CI = 0.99–1, $p = 0.009$), and female genital tumors (OR = 1, 95% CI = 0.9–1, $p = 0.019$). Additionally, the association between genetic susceptibility to HPV18 infection and the occurrence of benign tumors was identified in the pancreas (OR = 0.66, 95% CI = 0.45–0.98, $p = 0.037$), as well as haemangioma and lymphangioma (OR = 0.79, 95% CI = 0.69–0.91, $p = 0.001$). Also, there was a positive correlation between genetic susceptibility to HPV18 infections and the occurrence of lung adenocarcinoma, laryngeal cancer, pancreatic cancer, ovarian cancer, esophageal cancer, skin cancer, malignant neoplasm of the vulva, and stomach cancer. Conversely, there was an inverse relationship between genetic susceptibility to HPV-18 infections and the occurrence of chronic lymphocytic leukemia and ER+ breast cancer (Figure 3 and Supporting Information File S5: Table S2). Nevertheless, this data did not demonstrate statistical significance.

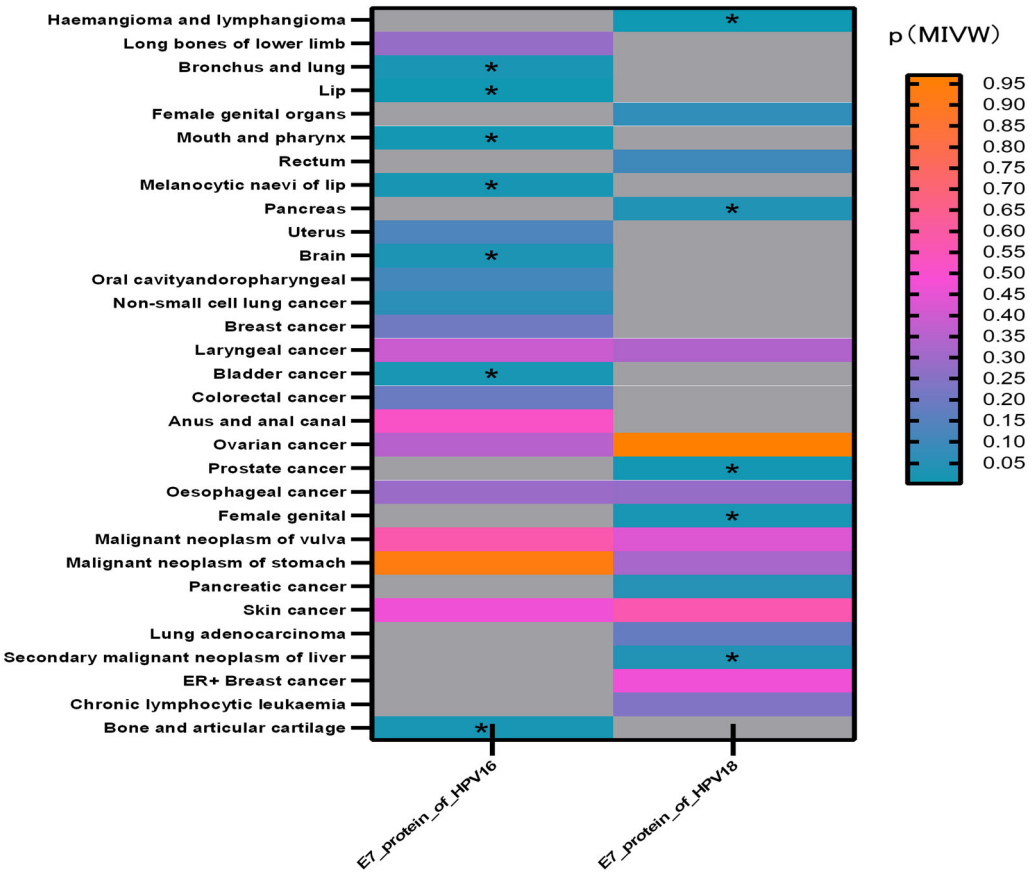

FIGURE 3 | Mendelian Randomization results show the causal relationship between HPV infection and site-specific tumors. Using colors in the inverse-variance weighted Mendelian randomization method is employed to visually represent the magnitude of the association *p*-value for causal effect estimations. *Represents the statistically significant associations.

3.6 | Sensitivity Analysis

Heterogeneity was observed in laryngeal cancer regarding the E7 protein of HPV18 (IVW: Cochran's $Q = 18.93$; $P_{heterogeneity} = 0.015$). Additional MR studies were conducted utilizing the random effects IVW approach, which resulted in the estimation of causal impact estimates. According to the MR-Egger intercept test, our analysis did not yield any statistically significant data on horizontal pleiotropy ($P_{pleiotropy} > 0.05$) (Supporting Information File S3: Tables S1 and S2). To find any possible outliers in the instrumental variable assessment of the causal effects of related malignancies on the risk of HPV16 and 18 E7 protein, leave-one-out analyses were done (Supporting Information File S4).

4 | Discussion

The current study conducted a comprehensive assessment of the correlation between genetically predicted high-risk HPV (HPV16 & HPV18) and various types of cancers. After employing multivariate MR to adjust for HPV-independent factors, including obesity and smoking, we revealed a causal link between genetically-predicted HPV16 infection and a range of benign and malignant tumors. This association heightened the risk of tumorigenesis, including benign tumors in specific areas such as the lip, brain, lung, and bronchus, as well as melanocytic naevi in the lip, mouth, and pharynx, and malignant tumors in the bone, articular cartilage, and bladder. These findings align with previous observational studies conducted on the subject. The genetic predisposition to HPV-18 has been found to elevate the susceptibility to malignant tumors in the female genital area, and a substantial body of observational research provides adequate evidence supporting this assertion [42, 43]. Additionally, it is correlated with the presence of benign pancreatic tumors, lymphangiomas, hemangiomas, secondary malignant liver neoplasms, and prostate cancer. The ramifications of these results are significant in the prevention and treatment of multisite tumors, encompassing both benign and malignant forms. The aforementioned findings underscore the significance of administering HPV vaccination to males as a preventive measure against bladder and prostate cancer.

HPV is a prevalent sexually transmitted infection, including a wide range of 200 distinct genotypes. Based on its propensity to cause cancer, it is frequently divided into high-risk and low-risk groups [44]. Low-risk variants of HPV, specifically HPV-6 and HPV-11, have a low propensity to progress into carcinoma. However, they cause recurrent pulmonary papillomas and genital warts. On the other hand, it is noteworthy that high-risk variants, namely HPV-16 and HPV-18, exhibit a statistically significant association with over 90% of cervical, anal, vulvar, and vaginal malignancies [45, 46]. This reflects the widespread pathogenicity of HPV. HPV is a micropapilloma DNA virus; all coding sequences are on a single strand of DNA, and genotyping is primarily based on homology within its capsid protein L1 (L1 Repeat Sequence) open reading frame (ORF) [47, 48]. The International Agency for Research on Cancer (IARC), in 1995, classified HPV-16 and HPV-18 as the primary etiological factors responsible for the emergence of cervical cancer [49]. The current study mainly explored the HPV-16 and HPV-18 associated tumors, and the results showed that HPV-18 caused female genital malignancies, which was consistent with the findings of IARC. A complete understanding of the molecular pathways behind HPV-induced cancers remains elusive. However, there is a prevailing belief that HPV is the primary cause of cervical cancer among females. This belief is supported by the presence of two significant tumor proteins, namely E6 and E7, which have been found to induce cell cycle arrest or uncontrolled cell growth by interacting with cellular E6-associated protein (E6AP) and tumor suppressor proteins, respectively. In addition, the E6 protein can also activate cellular telomerase (hTERT), causing an increase in cell proliferation, and ultimately tumor formation [50, 51]. In contrast, the E7 protein functions by suppressing protein phosphorylation and ubiquitination processes that facilitate cell division. Simultaneously, it triggers abnormal replication of centrosomes, resulting in polyploidy, genomic instability, and the development of tumors (Figure 4) [52].

The present study did not demonstrate the association of high-risk HPV (HPV16 & HPV18) infection with oropharyngeal, non-small cell lung, breast, colorectal, laryngeal, esophageal, and skin cancers. However, the odds ratio (OR) exceeding 1 suggests that high-risk HPV (HPV16 & HPV18) infection might be a risk factor for these types of cancer (Supporting Information File S4). During the implementation of the two-sample MR analysis, it was shown that laryngeal, non-small cell lung, breast, colorectal and anal canal, pancreatic, chronic lymphocytic leukemia, and lung adenocarcinomas exhibited significant associations with genetically predicted HPV16 and HPV18. However, when adjusting for variables such as smoking and BMI, the study's results revealed no discernible causal association between genetically predisposed high-risk HPV infections and the aforementioned cancers. This implies the presence of a confounding factor, wherein the occurrence of these malignancies is influenced by HPV infection and factors such as smoking and BMI. While MR analyses are generally considered less vulnerable to confounding and causal inference issues, the outcomes' reliability depends upon the specific MR analysis technique employed and the extent to which confounders related to causality are appropriately adjusted for. Nonetheless, our findings still demonstrate that high-risk HPV (HPV16 & HPV18) infection is associated with tumorigenesis at these specific sites.

This study possesses several strengths. Most studies on the correlation between HPV infection and the development of site-specific malignancies mainly consist of observational studies and literature reviews [8–20]. To our understanding, this study represents the first investigation of the relationship between high-risk HPV (HPV16 & HPV18) infection and neoplasms at specific anatomical sites, utilizing a substantial cohort derived from a population-based sample. The random distribution of genetic differences during gamete formation and fertilization made the MR method useful in reducing the impact of confounding factors and reverse causation. Furthermore, our study employed samples from distinct sources to assess exposures and outcomes using a two-sample MR approach. Considering the risks associated with smoking and obesity (BMI) in the development of diseases, we conducted MVMR analyses for each tumor type individually. This allowed us to elucidate the independent contribution of HPV in tumorigenesis.

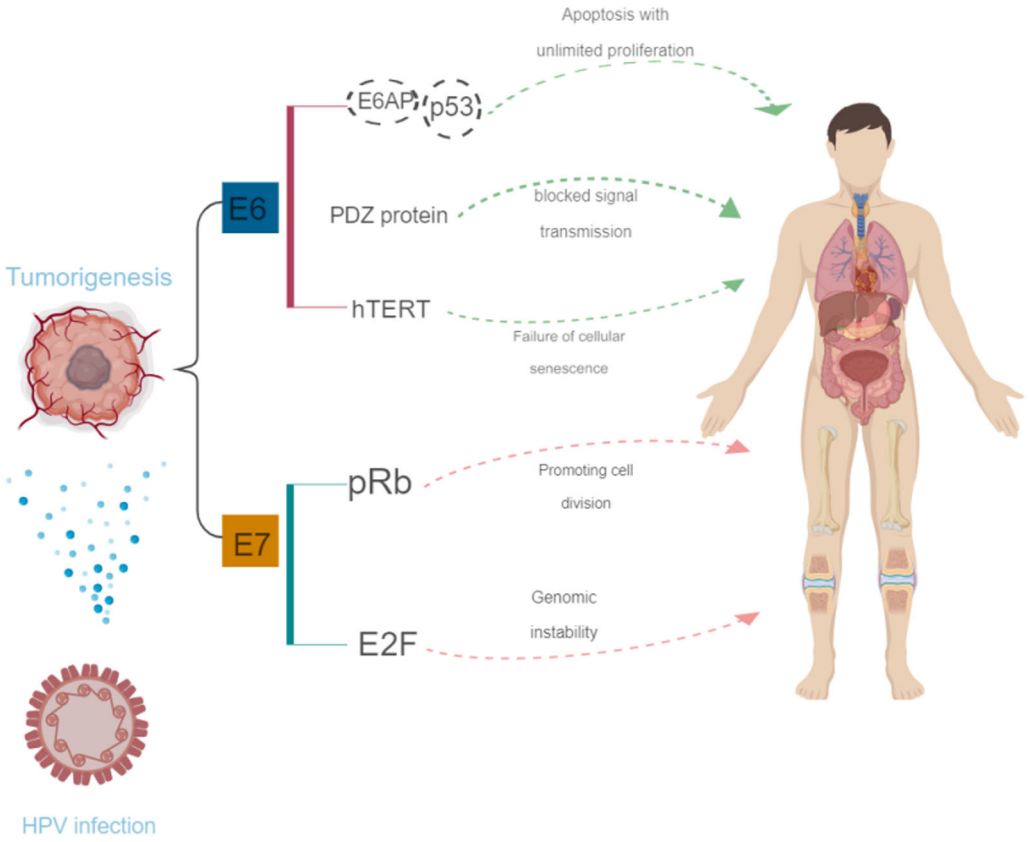

FIGURE 4 | Induction of carcinogenesis by HPV.

Subsequently, a range of supplementary examinations, including heterogeneity, multiple validity, and leave-one-out, were employed to assess the instrumental variables' robustness and increase the confidence of the results.

It is of the utmost importance to recognize that our study contains certain constraints. Initially, a subset of the HPV instrumental factors was eliminated throughout the process of screening for HPV to align the exposure and outcome data. Including rejected instrumental factors introduces a potential source of bias that might impact the outcomes of the research. Furthermore, our investigation primarily focused on high-risk HPV types 16 and 18, while neglecting the examination of other known types. This omission may result in the exclusion of subtypes that might potentially be linked to the development of tumors. Additionally, a significant limitation of our study is the inability to control for sexual activity as a potential confounder. Future research should aim to include comprehensive data on sexual activity and other relevant lifestyle factors to provide a more complete picture of the interplay between HPV infection and cancer risk. In the present research, it was observed that only patients with oropharyngeal cancer who were of Hispanic or Latino ethnicity were excluded. This resulted in mostly European population data, and it is crucial to admit that some demographic variables may have impacted our results, that differed between those of Hispanic or Latino descent and those of German descent. The variability of genetic polymorphisms among ethnic groups or races is a significant characteristic. Consequently, this leads to varying levels of susceptibility to HPV among different ethnic groups and regions [53]. It is important to recognize that our findings might not apply to other groups as a consequence. This necessitates careful interpretation of our findings and highlights the need for replication in diverse populations. Finally, Although MR following the principle of random allocation, may be confounding for non-genetic factors, such as lifestyle (smoking, alcohol consumption), that do not correctly reflect the association between exposure and outcome, which needs to include more non-genetic factors to better and comprehensively study the association of HPV with extra-cervical tumors.

This study provides evidence of the possible carcinogenic effects associated with HPV infection. The results and implications of the present research offer support for the implementation of public health interventions aimed at preventing HPV via vaccination and mitigating high-risk sexual practices. These initiatives have significant potential for alleviating the burden associated with cancer.

5 | Conclusion

We used European population-based study to extensively investigate the causal relationship between high-risk HPV (HPV16 & HPV18) and the risk of getting neoplasms at different anatomical locations using Mendelian randomization analysis. The findings of the research provide evidence that high-risk HPV (HPV16 & HPV18) is causally may associated with the development of both benign and malignant cancers, extending beyond cervical cancer. The results of this study might provide useful information for developing public health policy and for the management of malignancies.

Author Contributions

You Wu: data curation, writing – original draft, methodology, software, investigation. **Haiyang Xia:** methodology, software, data curation, writing – original draft. **Feng Du:** investigation, validation, data curation, writing – original draft. **Jian Zhang:** formal analysis. **Xulin Jiang:** formal analysis. **Jun Tang:** software, data curation, writing – review and editing. **Zhihong Li:** writing – review and editing, funding acquisition, supervision, resources, project administration.

Acknowledgments

The author extends sincere gratitude to the researchers and participants involved in the original GWASs (Genome-wide Association Studies) for their invaluable contributions to the collection and management of extensive data resources. The authors thank all the participants GWASs, UKB Biobank and FinnGen cohort, the institutions and their staff for providing data. Ezhou Science and Technology Bureau Basic and Talent Innovation Project: EZ01-007-20220006 and EZ01-007-20230125.

Ethics Statement

Our study used summary statistics from publicly available GWAS (https://gwas.mrcieu.ac.uk/), which followed the protocol approved by the respective institutional review boards, obtained informed consent from all participating studies, and did not require separate ethical approval.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in IEU GWAS database at https://gwas.mrcieu.ac.uk/.

Transparency Statement

The lead author Zhihong Li affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

1. J. Doorbar, W. Quint, L. Banks, et al., "The Biology and Life-Cycle of Human Papillomaviruses," *Vaccine* 30 (2012): F55–F70.

2. H. Factsheet, "Centers for Disease Control and Prevention," published January 23, 2014, http://www.cdc.gov/std/hpv/hpv-factsheet-march-2014.pdf.

3. J. J. Han, T. H. Beltran, J. W. Song, J. Klaric, and Y. S. Choi, "Prevalence of Genital Human Papillomavirus Infection and Human Papillomavirus Vaccination Rates Among US Adult Men," *JAMA Oncology* 3, no. 6 (2017): 810.

4. G. Y. F. Ho, R. Bierman, L. Beardsley, C. J. Chang, and R. D. Burk, "Natural History of Cervicovaginal Papillomavirus Infection in Young Women," *New England Journal of Medicine* 338, no. 7 (1998): 423–428.

5. N. Muñoz, F. X. Bosch, S. de Sanjosé, et al., "Epidemiologic Classification of Human Papillomavirus Types Associated With Cervical Cancer," *New England Journal of Medicine* 348, no. 6 (2003): 518–527.

6. E. M. de Villiers, "Cross-Roads in the Classification of Papillomaviruses," *Virology* 445, no. 1–2 (2013): 2–10.

7. L. Arribillaga, I. Echeverria, V. Belsue, et al., "Bivalent Therapeutic Vaccine Against HPV16/18 Genotypes Consisting of a Fusion Protein Between the Extra Domain A From Human Fibronectin and HPV16/18 E7 Viral Antigens," *Journal for Immunotherapy of Cancer* 8, no. 1 (2020): e000704.

8. A. K. Chaturvedi, E. A. Engels, R. M. Pfeiffer, et al., "Human Papillomavirus and Rising Oropharyngeal Cancer Incidence in the United States," *Journal of Clinical Oncology* 41, no. 17 (2023): 3081–3088.

9. M. C. Torrente, J. P. Rodrigo, M. Haigentz, Jr., et al., "Human Papillomavirus Infections in Laryngeal Cancer," *Head & Neck* 33, no. 4 (2011): 581–586.

10. R. Herrera-Goepfert, M. Lizano, S. Akiba, A. Carrillo-García, and M. Becker-D'Acosta, "Human Papilloma Virus and Esophageal Carcinoma in a Latin-American Region," *World Journal of Gastroenterology* 15, no. 25 (2009): 3142–3147.

11. S. A. Nadji, T. Mokhtari-Azad, M. Mahmoodi, et al., "Relationship Between Lung Cancer and Human Papillomavirus in North of Iran, Mazandaran Province," *Cancer Letters* 248, no. 1 (2007): 41–46.

12. C. H. Liu, C. Y. Liao, M. H. Yeh, and J. C. C. Wei, "Risk Role of Breast Cancer in Association With Human Papilloma Virus Among Female Population in Taiwan: A Nationwide Population-Based Cohort Study," *Healthcare* 10, no. 11 (2022): 2235.

13. N. Berrada, A. Al-Bouzidi, A. Ameur, et al., "Human Papillomavirus Detection in Moroccan Patients With Bladder Cancer," *Journal of Infection in Developing Countries* 7, no. 8 (2013): 586–592.

14. F. Atalay, C. Taskiran, M. Z. Taner, I. Pak, M. Or, and S. Tuncer, "Detection of Human Papillomavirus DNA and Genotyping in Patients With Epithelial Ovarian Carcinoma," *Journal of Obstetrics and Gynaecology Research* 33, no. 6 (2007): 823–828.

15. I. Khan, R. Harshithkumar, A. More, and A. Mukherjee, "Human Papilloma Virus: An Unraveled Enigma of Universal Burden of Malignancies," *Pathogens* 12, no. 4 (2023): 564.

16. L. C. Mendez, E. Hsieh, C. C. Earle, and S. Wong, "Synchronous Anal Canal Carcinoma in a Heterosexual Couple," *BMC Cancer* 18, no. 1 (2018): 884.

17. C. Moyret-Lalle, C. Marcais, J. Jacquemier, et al., "ras, p53 and HPV Status in Benign and Malignant Prostate Tumors," *International Journal of Cancer* 64, no. 2 (1995): 124–129.

18. G. L. Larsson, G. Helenius, S. Andersson, F. Elgh, B. Sorbe, and M. G. Karlsson, "Human Papillomavirus (HPV) and HPV 16–Variant Distribution in Vulvar Squamous Cell Carcinoma in Sweden," *International Journal of Gynecological Cancer* 22, no. 8 (2012): 1413–1419.

19. J. Baj, A. Forma, I. Dudek, et al., "The Involvement of Human Papilloma Virus in Gastrointestinal Cancers," *Cancers* 14, no. 11 (2022): 2607.

20. A. Iftner, S. J. Klug, C. Garbe, et al., "The Prevalence of Human Papillomavirus Genotypes in Nonmelanoma Skin Cancers of Non-immunosuppressed Individuals Identifies High-Risk Genital Types as Possible Risk Factors," *Cancer Research* 63, no. 21 (2003): 7515–7519.

21. G. Davey Smith and G. Hemani, "Mendelian Randomization: Genetic Anchors for Causal Inference in Epidemiological Studies," *Human Molecular Genetics* 23, no. R1 (2014): R89–R98.

22. R. C. Richmond and G. Davey Smith, "Mendelian Randomization: Concepts and Scope," *Cold Spring Harbor Perspectives in Medicine* 12, no. 1 (2022): a040501.

23. H. zur Hausen, "Papillomaviruses in the Causation of Human Cancers—A Brief Historical Account," *Virology* 384, no. 2 (2009): 260–265.

24. D. A. Lawlor, R. M. Harbord, J. A. C. Sterne, N. Timpson, and G. Davey Smith, "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology," *Statistics in Medicine* 27, no. 8 (2008): 1133–1163.

25. N. M. Davies, M. V. Holmes, and G. Davey Smith, "Reading Mendelian Randomisation Studies: A Guide, Glossary, and Checklist for Clinicians," *BMJ* 362 (2018): k601.

26. B. Elsworth, M. Lyon, T. Alexander, et al., "The MRC IEU OpenGWAS Data Infrastructure," preprint, bioRxiv, August 10, 2020, https://doi.org/10.1101/2020.08.10.244293.

27. K. Suhre, M. Arnold, A. M. Bhagwat, et al., "Connecting Genetic Risk to Disease End Points Through the Human Blood Plasma Proteome," *Nature Communications* 8 (2017): 14357.

28. G. Hemani, J. Zheng, B. Elsworth, et al., "The MR-Base Platform Supports Systematic Causal Inference Across the Human Phenome," *eLife* 7 (2018): e34408, https://doi.org/10.7554/eLife.34408.

29. J. Bowden, F. Del Greco M, C. Minelli, G. Davey Smith, N. Sheehan, and J. Thompson, "A Framework for the Investigation of Pleiotropy in Two-Sample Summary Data Mendelian Randomization," *Statistics in Medicine* 36, no. 11 (2017): 1783–1802.

30. J. Bowden, G. Davey Smith, and S. Burgess, "Mendelian Randomization With Invalid Instruments: Effect Estimation and Bias Detection Through Egger Regression," *International Journal of Epidemiology* 44, no. 2 (2015): 512–525.

31. J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess, "Consistent Estimation in Mendelian Randomization With Some Invalid Instruments Using a Weighted Median Estimator," *Genetic Epidemiology* 40, no. 4 (2016): 304–314.

32. M. Verbanck, C. Y. Chen, B. Neale, and R. Do, "Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred From Mendelian Randomization Between Complex Traits and Diseases," *Nature Genetics* 50, no. 5 (2018): 693–698.

33. Z. S. Grenville, U. Noor, M. His, et al., "Diet and BMI Correlate With Metabolite Patterns Associated With Aggressive Prostate Cancer," *Nutrients* 14, no. 16 (2022): 3306.

34. Z. Wu, S. Xie, F. Wang, et al., "BMI Changes and the Risk of Lung Cancer in Male Never-Smokers: A Prospective Cohort Study," *Cancer Medicine* 11, no. 5 (2022): 1336–1346.

35. J. Liu, Y. Jiang, H. Huang, et al., "BMI-1 Promotes Breast Cancer Proliferation and Metastasis Through Different Mechanisms in Different Subtypes," *Cancer science* 114, no. 2 (2023): 449–462.

36. Z. Levi, J. D. Kark, A. Shamiss, et al., "Body Mass Index and Socioeconomic Status Measured in Adolescence, Country of Origin, and the Incidence of Gastroesophageal Adenocarcinoma in a Cohort of 1 Million Men," *Cancer* 119, no. 23 (2013): 4086–4093.

37. B. C. Bade and C. S. Dela Cruz, "Lung Cancer 2020," *Clinics in Chest Medicine* 41, no. 1 (2020): 1–24.

38. R. Carreras-Torres, M. Johansson, V. Gaborieau, et al., "The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study," *JNCI: Journal of the National Cancer Institute* 109, no. 9 (2017): djx012.

39. S. Liu, S. Feng, F. Du, K. Zhang, and Y. Shen, "Association of Smoking, Alcohol, and Coffee Consumption With the Risk of Ovarian Cancer and Prognosis: A Mendelian Randomization Study," *BMC Cancer* 23, no. 1 (2023): 256.

40. K. Zu and E. Giovannucci, "Smoking and Aggressive Prostate Cancer: A Review of the Epidemiologic Evidence," *Cancer Causes & Control* 20, no. 10 (2009): 1799–1810.

41. S. C. Larsson, M. Traylor, R. Malik, M. Dichgans, S. Burgess, and H. S. Markus, "Modifiable Pathways in Alzheimer's Disease: Mendelian Randomisation Analysis," *BMJ* 359 (2017): j5375.

42. M. Stamenković, A. Knežević, I. Knežević, et al., "High-Risk Human Papilloma Virus Genotypes in Cervical Carcinoma of Serbian Women: Distribution and Association With Pathohistological Findings," *Biologicals* 44, no. 5 (2016): 412–416.

43. Y. Yang, J. Ren, and Q. Zhang, "Distribution of Human Papilloma Virus Type 16 E6/E7 Gene Mutation in Cervical Precancer or Cancer: A Case Control Study in Guizhou Province, China," *Journal of Medical Virology* 88, no. 2 (2016): 345–350.

44. M. Narisawa-Saito and T. Kiyono, "Basic Mechanisms of High-Risk Human Papillomavirus-Induced Carcinogenesis: Roles of E6 and E7 Proteins," *Cancer science* 98, no. 10 (2007): 1505–1511.

45. K. R. Jørgensen and J. B. Jensen, "Human Papillomavirus and Urinary Bladder Cancer Revisited," *APMIS* 128, no. 2 (2020): 72–79.

46. E. F. Dunne and I. U. Park, "HPV and HPV-Associated Diseases," *Infectious Disease Clinics of North America* 27, no. 4 (2013): 765–778.

47. E. M. de Villiers, "Papillomavirus and HPV Typing," *Clinics in Dermatology* 15, no. 2 (1997): 199–206, https://doi.org/10.1016/s0738-081x(96)00164-2.

48. E. M. de Villiers, C. Fauquet, T. R. Broker, H. U. Bernard, and H. Zur Hausen, "Classification of Papillomaviruses," *Virology* 324, no. 1 (2004): 17–27.

49. H. Vainio, E. Heseltine, and J. Wilbourn, "Priorities for Future IARC Monographs on the Evaluation of Carcinogenic Risks to Humans," *Environmental Health Perspectives* 102, no. 6–7 (1994): 590–591.

50. M. Scheffner and N. J. Whitaker, "Human Papillomavirus-Induced Carcinogenesis and the Ubiquitin-Proteasome System," *Seminars in Cancer Biology* 13, no. 1 (2003): 59–67.

51. F. Mantovani and L. Banks, "The Human Papillomavirus E6 Protein and Its Contribution to Malignant Progression," *Oncogene* 20, no. 54 (2001): 7874–7887.

52. K. Münger, J. R. Basile, S. Duensing, et al., "Biological Activities and Molecular Targets of the Human Papillomavirus E7 Oncoprotein," *Oncogene* 20, no. 54 (2001): 7888–7898.

53. M. Wang, W.-J. Kong, Y. Feng, et al., "Epidemiological, Clinical, and Histological Presentation of Celiac Disease in Northwest China," *World Journal of Gastroenterology* 28, no. 12 (2022): 1272–1283.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.