

# Polymorphisms in the Nonhomologous End-joining DNA Repair Pathway are Associated with HPV Integration in Cervical Dysplasia

Jennifer M. Geris<sup>1,2</sup>, E. Susan Amirian<sup>2,3</sup>, Deborah A. Marquez-Do<sup>3</sup>, Martial Guillaud<sup>4</sup>, Laura M. Dillon<sup>5</sup>, Michele Follen<sup>6</sup>, and Michael E. Scheurer<sup>1,2,3</sup>



## ABSTRACT

Previous evidence indicates that human papillomavirus (HPV) integration status may be associated with cervical cancer development and progression. However, host genetic variation within genes that may play important roles in the viral integration process is understudied. The aim of this study was to examine the association between HPV16 and HPV18 viral integration status and SNPs in nonhomologous-end-joining (NHEJ) DNA repair pathway genes on cervical dysplasia. Women enrolled in two large trials of optical technologies for cervical cancer detection and positive for HPV16 or HPV18 were selected for HPV integration analysis and genotyping. Associations between SNPs and cytology (normal, low-grade, or high-grade lesions) were evaluated. Among women with cervical dysplasia, polytomous logistic regression models were used to evaluate the effect of each SNP on viral integration status. Of the 710 women evaluated [149 high-grade squamous intraepithelial lesion (HSIL), 251; low-grade squamous intraepithelial lesion (LSIL), 310 normal], 395 (55.6%) were positive for HPV16 and 192 (27%) were positive for HPV18. Tag-SNPs

in 13 DNA repair genes, including *RAD50*, *WRN*, and *XRCC4*, were significantly associated with cervical dysplasia. HPV16 integration status was differential across cervical cytology, but overall, most participants had a mix of both episomal and integrated HPV16. Four tag-SNPs in the *XRCC4* gene were found to be significantly associated with HPV16 integration status. Our findings indicate that host genetic variation in NHEJ DNA repair pathway genes, specifically *XRCC4*, are significantly associated with HPV integration, and that these genes may play an important role in determining cervical cancer development and progression.

**Prevention Relevance:** HPV integration in premalignant lesions and is thought to be an important driver of carcinogenesis. However, it is unclear what factors promote integration. The use of targeted genotyping among women presenting with cervical dysplasia has the potential to be an effective tool in assessing the likelihood of progression to cancer.

## Introduction

Cervical cancer is the fourth most common cancer diagnosed among women globally, with more than 600,000 new cases and 340,000 deaths annually (1). The primary causal agent for cervical cancer is human papillomavirus (HPV), and it has been estimated that women have an 80% lifetime cumulative incidence of HPV infection (2). Two high-risk HPV types, 16 and 18, account for approximately 75%–80%

of cervical cancers in the United States, Canada, and Europe. Furthermore, HPV16, which is the most prevalent type, is detected in more than 50% of high-grade squamous intraepithelial lesions (HSIL; refs. 3, 4).

Papillomaviruses, which have double-stranded circular DNA, are capable of integrating into the host cell genome. In order to linearize before integration, the circular HPV DNA “breaks” at the E2 gene and several viral genes, including E4, E5, part of E2, and part of L2, lose function in this linearization process (Fig. 1; refs. 5–7). The integration of the HPV genome into the host cell genome is believed to be somewhat random with regard to where in the host cell’s genome it occurs, but there does appear to be a propensity toward integration at fragile sites in the host’s genome (8, 9). Unfortunately, several important details of HPV genome integration remain unknown, and no studies to date have investigated how host genetic variation in key pathways may influence the ability of the virus to complete the integration process. As integration is thought to be a seminal event in the ability of HPV to initiate the carcinogenic process, it is vital that the host–virus interactions that impact viral integration be elucidated.

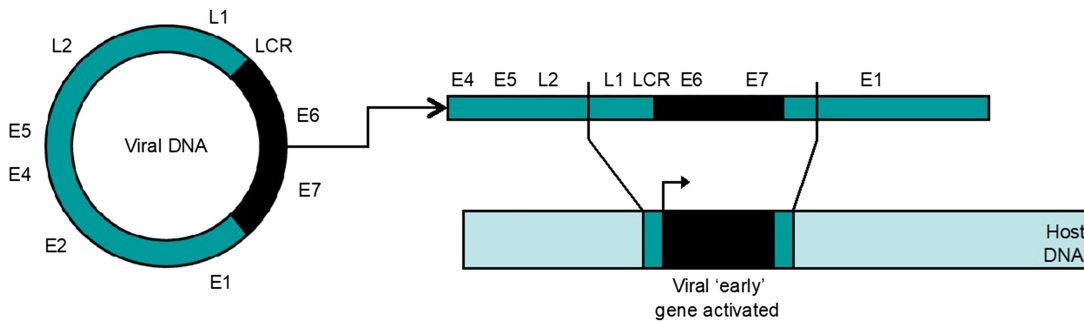
<sup>1</sup>Department of Medicine, Baylor College of Medicine, Houston, Texas. <sup>2</sup>Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas. <sup>3</sup>Department of Pediatrics, Baylor College of Medicine, Houston, Texas. <sup>4</sup>Department of Cancer Imaging, British Columbia Cancer Research Centre, Vancouver, British Columbia. <sup>5</sup>Department of Diagnostic and Biomedical Sciences, UTHealth School of Dentistry, Houston, Texas. <sup>6</sup>EmblemHealth, New York, New York.

**Corresponding Author:** Jennifer M. Geris, Baylor College of Medicine, 6620 Main Street, Houston, TX 77030. Phone: 713-798-4902; E-mail: Jennifer.Geris@bcm.edu

Cancer Prev Res 2023;16:461–70

doi: 10.1158/1940-6207.CAPR-23-0051

©2023 American Association for Cancer Research



**Figure 1.** HPV integration into the host genome. HPV integration begins with the linearization of the circular double-stranded DNA at the E2 open reading frame (ORF). During this process, several viral genes lose function, including E4, E5, part of E2, and part of L2. The loss of function at E2 leads to overexpression of the E6 and E7 genes, allowing uncontrolled growth and malignant transformation.

Because integration consists of the insertion of foreign (viral) DNA into host cell genomes, polymorphisms in host DNA repair pathway genes may potentially be very influential in the ability of HPV to integrate. The double-strand break repair process is a proposed method by which viruses may integrate into the host cell genome. Humans have evolved two ways to repair DNA damage from double-strand breaks: homologous recombination and nonhomologous end-joining (NHEJ). The NHEJ mechanism is the likely candidate for repair facilitating viral integration because it does not rely on the presence of homologous sequences for template copy. Hence, repair can lead to the incorporation of foreign DNA and is often accompanied by a loss of surrounding genetic material (10, 11). The initial steps in recognizing double-strand breaks is carried out by a complex of proteins encoded by the *NBS1*, *MRE11*, and *RAD50* genes (Fig. 2; ref. 12). The repair process is then mediated by the protein complex of KU70, KU80, DNAPK, XRCC4, and LIG4.

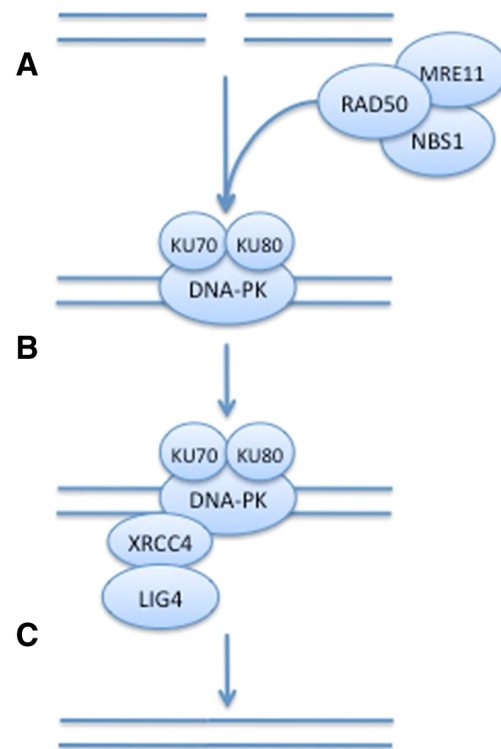
Although little is understood about the factors that cause some HPV-exposed women to become HPV-infected and develop cancerous lesions while other exposed women are able to clear the infection rapidly with no sequelae, previous studies have not, to our knowledge, examined how host genetic factors may influence one’s susceptibility to HPV viral integration and subsequent HPV-associated carcinogenesis. It is, however, well known that high-risk HPV integration into the host genome is able to disrupt numerous signal transduction pathways, which can lead to uncontrolled cellular proliferation. In this study, we sought to examine the association between both NHEJ pathway SNPs and HPV16 and HPV18 viral integration status on cervical dysplasia, comparing women with normal cytology to those with low-grade squamous intraepithelial lesions (LSIL) or high-grade squamous intraepithelial lesions (HSIL).

## Materials and Methods

### Study population and data collection

This study included a subset of women from two large trials evaluating optical technologies for cervical cancer detection enrolling at The University of Texas MD Anderson Cancer Center (MDACC, Houston, TX) and Lyndon Baines Johnson

(LBJ) General Hospital (Houston, TX), and at the British Columbia Cancer Agency (BCCA, Vancouver, Canada). The study population and subject characteristics for the parent trials have been described in detail elsewhere (13–16). Briefly, participants were classified into either the diagnostic (high-risk) population (defined by history of abnormal Papanicolaou smears), or as the screening (low-risk) population (history of



**Figure 2.** NHEJ double-strand break DNA repair process. **A**, End-binding and tethering: recognition of double-strand breaks is carried out by complex of proteins from *NBS1*, *MRE11*, and *RAD50* genes. **B**, End processing: damaged or mismatched nucleotides are removed by DNA nucleases, resulting in the potential loss of genetic material. **C**, Ligation: the repair process is then mediated by the protein complex of KU70, KU80, DNAPK, XRCC4, and DNA ligase IV (LIG4).

negative Papanicolaou smears and no cervical treatments). Study procedures included the completion of a risk factor questionnaire; a complete medical history; full physical and gynecologic examinations; regular and ThinPrep (Cytoc Corp.) Papanicolaou smears; cervical cultures; specimens for HPV typing; pan-colposcopy of the vulva, vagina, and cervix; spectroscopic measurements of the cervix; and corresponding biopsies. The overall participation rate among women who met all eligibility criteria was 81%. The research protocol was approved by the Institutional Review Boards (IRB) at each institution, and each participant gave written informed consent prior to enrollment in the study. This study was conducted in accordance with U.S. Common Rule.

For this study, women who were positive for the presence of HPV16 and/or HPV18 by our PCR HPV typing assay (described below) were selected for HPV integration analysis and genotyping. Because the HPV integration assays are type-specific, and HPV16 and HPV18 account for the majority (>75%) of cervical cancer cases, we chose to include these two types in our study.

#### Confirmation of disease status

As part of the parent trials, all women underwent a Papanicolaou smear for cytologic testing, followed by colposcopy and biopsies of regions of interest (for those with abnormalities) and representative normal cervix (in both groups). Cytology specimens were sent to a reference laboratory (LabCorp) for interpretation. Biopsies were then scored by three independent pathologists at MDACC and BCCA, and a consensus diagnosis was obtained (17). The worst diagnosis for each patient was obtained by considering the most serious diagnosis of the cytology or histology across all biopsied lesions.

#### Disease definition

The 2001 Bethesda Classification System was used to define disease outcome in this study (18). HSIL were defined as those having a cytologic diagnosis of moderate dysplasia or a histologic diagnosis of cervical intraepithelial neoplasia (CIN) 2, severe dysplasia (CIN3) or carcinoma *in situ* (CIS). LSIL were defined as those having a cytologic diagnosis of HPV-associated changes or a histologic diagnosis of mild dysplasia (CIN1). Samples negative for dysplasia by both histology and cytology were defined as normal. Samples with a diagnosis of atypical squamous cells of undetermined significance (ASCUS) were grouped into a separate category due to potential histologic importance in showing a transition from a normal to an abnormal state. For the purposes of our statistical analyses, the following pathologic groupings were used: normal (no cytologic or histologic abnormalities); LSIL (ASCUS or LSIL on cytology or CIN1 on histology); and HSIL (HSIL on cytology or CIN2, CIN3, or CIS on histology).

#### HPV typing

Details of HPV DNA typing for HPV16 and HPV18 have been reported elsewhere (19). Briefly, viral DNA was extracted from cervical cytobrush specimens using a commercially

available kit (Qiagen DNA Mini Kit, Qiagen) and analyzed for HPV L1 gene consensus sequences (20) by PCR, followed by specific typing with HPV16 and HPV18 probes.

#### SNP selection and genotyping

Genomic DNA was extracted from cervical cytobrush samples using a commercially available kit (Qiagen DNA Mini, Qiagen) according to manufacturer's directions. Genes in the NHEJ DNA repair pathway include: *ATM*, *LIG4*, *MRE11A*, *NBN*, *NHEJ1*, *POLB*, *PRKDC*, *RAD50*, *WRN*, *XRCC4*, *XRCC5*, *XRCC6*, and *XRCC6BP1*. Tagging SNPs in these genes were identified using the MultiPop-TagSelect algorithm in the Genome Variation Server (21) using an  $r^2$  threshold of 0.80. SNPs with minor allele frequencies of <10% were not included in the analysis. On the basis of these criteria, 449 tag SNPs were selected for analysis (Supplementary Table S1 for the full list). In addition, we genotyped a set ( $n = 176$ ) of well characterized ancestry informative markers (22) for use in adjusting for population stratification in our analyses. Genotyping was conducted using the Illumina GoldenGate platform (Illumina). Quality control strategies included the genotyping of internal positive control samples, the use of no template controls, and replicates for 10% of the samples. Further, SNPs with a call rate across all samples <95% and individuals with call rates <85% were excluded from the genetic analyses.

#### HPV integration assays

Integration assays were performed according to the procedures described by Peitsaro and colleagues (23) and Huang and colleagues (24). Briefly, a qRT-PCR assay was designed for HPV16 and HPV18 separately. Primers and probes (Supplementary Table S2) for the E2 and E6 genes were designed using full-length gene information for HPV16 and HPV18 available in NCBI (GenBank accession numbers: HPV16: NC\_001526; HPV18: LC636309.1). For each HPV type, the E2 and E6 targets were assayed in a multiplex PCR, with the E2 sequence tagged by FAM and the E6 sequence by HEX. Standard curves ( $1 \times 10^8$ ,  $1 \times 10^6$ ,  $1 \times 10^5$ ,  $1 \times 10^4$ ,  $1 \times 10^3$ ,  $1 \times 10^2$ ,  $1 \times 10^1$ , and  $1 \times 10^0$  copies/ $\mu$ L) for both E2 and E6 were included on each plate using DNA extracted from plasmids (RRIDs: Addgene\_22457, Addgene\_8642, Addgene\_10876, Addgene\_10850; Addgene) containing full length type-specific E2 or E6. Each amplification experiment was performed in the iCycler iQ real-time PCR instrument (Bio-Rad Laboratories). A final volume of 25  $\mu$ L was used containing 20 ng of cDNA template, 12.5  $\mu$ L of iQ SYBR Green Supermix (Bio-Rad Laboratories), 1.25  $\mu$ L containing 0.2 micromoles of either HPV16 or HPV18 forward and reverse primers, and water. The reaction was subjected to denaturation at 95°C for 2 minutes followed by 60 cycles of denaturation at 95°C for 5 seconds, annealing at 56°C (HPV16) or 51.2°C (HPV18) for 15 seconds, and elongation at 72°C for 45 seconds. Fluorescence data were specified for collection at the end of the elongation step in each cycle. The absolute number of copies of full-length E2 and/or E6 was calculated against the standard curve for each gene product. Calls regarding integration status were made on the basis of the ratio of

copy numbers of HPV E2 to E6 genes: copy numbers of E2 and E6 that were equal (E2:E6 = 1) were considered episomal, E2 copies less than E6 (E2:E6 < 1) were considered mixed, and having only E6 copies present (E2 = 0) was defined as fully integrated.

**Statistical analyses**

Population characteristics were examined overall and by level of cervical dysplasia (normal, LSIL, and HSIL).  $\chi^2$  and Fisher exact tests were conducted using an additive genetic model (based on the number of minor alleles) to determine which SNPs were significantly associated with normal cytology, LSIL, or HSIL. *P* values were then adjusted for multiple comparisons using the FDR method of Benjamini and Hochberg (25).  $\chi^2$  tests were also utilized to assess differences in HPV16 or HPV18 integration status by cervical dysplasia. Among participants with cervical dysplasia, polytomous logistic regression was used to evaluate the association between each SNP (under an additive genetic model) and viral integration status (episomal, mixed, or fully integrated), separately for HPV16 and HPV18, controlling for population stratification. Percent ancestry (European, African, and Asian) was determined using Structure version 2.3 and included in all modeling procedures as continuous variables. All analyses were conducted in SAS version 9.2 and STATA version 12.

**Data availability**

The data generated in this study are available upon request from the corresponding author.

**Results**

Of the 1,784 subjects available from the parent study, 946 (53%) were identified as HPV16 and/or HPV18 positive, 793 (84%) were selected for genotyping based on sample availability, and 765 (96%) of those were successfully genotyped. Of these, 55 individuals (7%) with call rates <85% were excluded from the genetic analyses presented here. The distributions of selected study population characteristics (*n* = 710) are presented in **Table 1**. Overall, 310 women (43.7%) had normal cervical histology, 251 (35.4%) had LSIL, and 149 (21.0%) had HSIL. Approximately 58% of the study participants were non-Hispanic white, and the average age was 38.8 ( $\pm$ 12.0) years. Women with HSIL were significantly younger than women with no lesion or LSIL (*P* < 0.001). Smoking status differed by disease groups, as a greater proportion of women with HSIL (26.8%) and LSIL (22.3%) were current smokers than women with normal histology (11.9%; *P* < 0.001). There were no differences in body mass index (BMI) or employment status between disease groups.

HPV16 and HPV18 integration status by level of cervical dysplasia are presented in **Table 2**. Overall, 55.6% and 27.0% of women were positive for HPV16 and HPV18, respectively. Among all women tested (*n* = 710), 105 (14.8%) were positive for both HPV16 and HPV18. Among those positive for HPV16, most participants (61.0%) had a mix of episomal and integrated HPV DNA, and integration status differed across level of cervical dysplasia (*P* < 0.001). In contrast, among women positive for HPV18, there was a similar proportion of

**Table 1.** Population characteristics by level of cervical dysplasia.

	Total N = 710	HSIL N = 149	LSIL N = 251	Normal N = 310	<i>P</i>
Age at diagnosis (SD), years	38.8 (12.0)	33.3 (9.9)	36.2 (11.1)	43.5 (11.8)	<0.001
Race/Ethnicity					0.005
Asian	45 (6.3%)	12 (8.1%)	14 (5.6%)	19 (6.1%)	
Black	105 (14.8%)	17 (11.4%)	40 (15.9%)	48 (15.5%)	
Hispanic	131 (18.5%)	15 (10.1%)	62 (24.7%)	54 (17.4%)	
Native American	3 (0.4%)	2 (1.3%)	0 (0.0%)	1 (0.3%)	
Other	11 (1.5%)	5 (3.4%)	1 (0.4%)	5 (1.6%)	
White	415 (58.5%)	98 (65.8%)	134 (53.4%)	183 (59.0%)	
Marital status					<0.001
Never married	174 (24.5%)	57 (38.3%)	68 (27.1%)	49 (15.8%)	
Married	342 (48.2%)	49 (32.9%)	108 (43.0%)	185 (59.7%)	
Living together	54 (7.6%)	19 (12.8%)	21 (8.4%)	14 (4.5%)	
Divorced	129 (18.2%)	22 (14.8%)	53 (21.1%)	54 (17.4%)	
Widowed	10 (1.4%)	1 (0.7%)	1 (0.4%)	8 (2.6%)	
Unknown	1 (0.1%)	1 (0.7%)	0 (0.0%)	0 (0.0%)	
Employment					0.35
Currently employed	394 (55.5%)	74 (49.7%)	147 (58.6%)	173 (55.8%)	
Other	313 (44.1%)	74 (49.7%)	104 (41.4%)	135 (43.5%)	
Unknown	3 (0.4%)	1 (0.7%)	0 (0.0%)	2 (0.6%)	
Smoking status					<0.001
Current	133 (18.7%)	40 (26.8%)	56 (22.3%)	37 (11.9%)	
Former	141 (19.9%)	24 (16.1%)	34 (13.5%)	83 (26.8%)	
Never	435 (61.3%)	85 (57.0%)	161 (64.1%)	189 (61.0%)	
Unknown	1 (0.1%)	0 (0.0%)	0 (0.0%)	1 (0.3%)	
BMI (SD)	24.9 (18.7)	24.7 (11.9)	25.8 (16.4)	24.3 (22.7)	0.66

Downloaded from <http://aacrjournals.org/cancerpreventionresearch/article-pdf/16/8/461/3350953/461.pdf> by guest on 02 December 2024

**Table 2.** HPV16 and HPV18 integration by level of cervical dysplasia.

	Total N = 710	HSIL N = 149	LSIL N = 251	Normal N = 310	P
HPV16 <sup>-</sup>	315 (44.4%)	35 (23.5%)	126 (50.2%)	154 (49.7%)	<0.001
HPV16 <sup>+</sup>	395 (55.6%)	114 (76.5%)	125 (49.8%)	156 (50.3%)	
HPV16 integration status <sup>a</sup>					
Episomal	77 (19.5%)	30 (26.3%)	26 (20.8%)	21 (13.5%)	
Mixed	241 (61.0%)	75 (65.8%)	72 (57.6%)	94 (60.3%)	
Integrated	77 (19.5%)	9 (7.9%)	27 (21.6%)	41 (26.3%)	
HPV18 <sup>-</sup>	518 (73.0%)	123 (82.6%)	196 (78.1%)	199 (64.2%)	<0.001
HPV18 <sup>+</sup>	192 (27.0%)	26 (17.4%)	55 (21.9%)	111 (35.8%)	
HPV18 Integration Status <sup>a</sup>					
Episomal	12 (6.3%)	4 (15.4%)	4 (7.2%)	4 (3.6%)	
Mixed	92 (47.9%)	17 (65.4%)	25 (45.5%)	50 (45.0%)	
Integrated	88 (45.8%)	5 (19.2%)	26 (47.3%)	57 (51.4%)	

Note: Columns are presented as n and %, unless otherwise indicated. P values calculated by  $\chi^2$  statistic.  
<sup>a</sup>Column n and % among HPV<sup>+</sup> participants.

participants with exclusively integrated HPV DNA (45.8%) and a mix of episomal and integrated HPV DNA (47.9%). Integration status of HPV18 also differed by level of cervical dysplasia (P < 0.001).

Overall, a total of 338 tag-SNPs in 13 key NHEJ genes were significantly associated with cervical dysplasia status, including SNPs in the *XRCC4*, *WRN*, and *NBN* genes, after corrections for multiple comparisons (FDR q ≤ 0.05; **Table 3**). Specifically, there were 77 significant SNPs (22.8%) in the *XRCC4* gene and 47 (13.9%) in *WRN*. Among HPV16-positive individuals, 299 tag-SNPs were significantly associated with cervical dysplasia status. Of these SNPs, 57 (19.1%) were in *XRCC4* and 29 (9.7%) were in *MRE11A*. Among HPV18-positive individuals, none of the tag-SNPs reached statistical significance after controlling for multiple comparisons, likely due to inadequate statistical power.

The same set of tag-SNPs was examined for associations with viral integration status. The top hits associated with HPV16 integration status were in the *XRCC4* gene (**Table 4**). Specifically, SNPs at *rs7718284* [OR, 2.74; 95% confidence interval (CI), 1.66–4.53], *rs17205699* (2.72, 1.62–4.56), *rs1193693* (2.61, 1.52–4.12), and *rs4266384* (2.48, 1.48–4.15) were associated with nearly 2.5 times greater odds of integrated HPV DNA compared with episomal DNA. One of the top hits associated with HPV18 integration status, *WRN rs4733221*, was associated with nearly five times greater odds of mixed (OR, 5.25; 95% CI, 1.36–20.35) or exclusively integrated (4.95; 1.26–19.43) HPV DNA compared to episomal DNA (**Table 4**). The three other top hits involved SNPs in the *XRCC4* gene. One of the top hits associated with HPV18 integration status was *PRKCH rs1111108* (integrated HPV18 OR, 0.07; 95% CI, 0.01–0.49); the three other top hits involved SNPs in the *XRCC4* gene.

## Discussion

Persistent HPV infection can lead to progression to high-grade dysplasia and invasive carcinoma; however, HPV infection alone is not sufficient to drive the malignant transformation. In fact, high-risk HPV types can often be found in women

with normal cytology, and most of these women will not develop carcinoma as a result of their infection. It is likely that viral and host factors play a role in the development of cervical cancer, and yet, host genetic variants have been understudied as risk factors for progression. The few genetic association studies performed have focused on host immune genes related to HPV infection or to cervical cancer without HPV infection status known (26–28). Ours is the first study, to our knowledge, to examine host genetics in relation to HPV integration in cervical dysplasia. Our finding that polymorphisms in the *XRCC4* gene are strongly associated with HPV integration status is of key importance and may lend insight into why some women have a

**Table 3.** Top tSNPs significantly associated with cervical dysplasia.

Genes	Freq.	Percent
Overall		
<i>ATM</i>	31	9.17%
<i>LIG4</i>	8	2.37%
<i>MRE11A</i>	31	9.17%
<i>NBN</i>	33	9.76%
<i>NHEJ1</i>	13	3.85%
<i>POLB</i>	7	2.07%
<i>PRKDC</i>	28	8.28%
<i>RAD50</i>	15	4.44%
<i>WRN</i>	47	13.91%
<i>XRCC4</i>	77	22.78%
<i>XRCC5</i>	30	8.88%
<i>XRCC6</i>	8	2.37%
<i>XRCC6BPI</i>	10	2.96%
Total	338	100%
HPV16 <sup>+</sup> women		
<i>XRCC4</i>	57	19.1%
<i>MRE11A</i>	29	9.7%
<i>ATM</i>	28	9.4%
<i>WRN</i>	28	9.4%
<i>NBN</i>	25	8.4%
<i>XRCC5</i>	25	8.4%

Note: Frequency of tSNPs by gene associated with cervical dysplasia with FDR q ≤ 0.05 overall and among women who were HPV16<sup>+</sup>.

**Table 4.** Top SNPs associated with viral integration status among HPV16<sup>+</sup> and HPV18<sup>+</sup> women.

Gene and SNP ID	Mixed HPV OR (95% CI)	Integrated HPV OR (95% CI)
HPV16		
XRCC4 rs2731849	0.55 (0.35–0.85)	0.38 (0.22–0.66)
XRCC4 rs7718284	1.76 (1.18–2.63)	2.74 (1.66–4.53)
XRCC4 rs17205699	1.91 (1.23–2.95)	2.72 (1.62–4.56)
XRCC4 rs1193693	1.85 (1.23–2.80)	2.61 (1.58–4.31)
XRCC4 rs4266384	1.64 (1.07–2.49)	2.48 (1.48–4.15)
HPV18		
WRN rs4733221	5.25 (1.36–20.35)	4.95 (1.26–19.43)
XRCC4 rs301279	0.11 (0.02–0.55)	0.15 (0.03–0.74)
XRCC4 rs6889850	0.07 (0.01–0.52)	0.14 (0.02–0.94)
XRCC4 rs6864659	0.11 (0.01–1.04)	0.03 (0.003–0.34)

Note: Top SNPs associated with mixed or integrated HPV among HPV16<sup>+</sup> and HPV18<sup>+</sup> women. Top SNPs based on an additive genetic model. Polytomous logistic regression was used to estimate ORs and 95% CIs, adjusted for ancestry informative markers to control for population stratification. Reference group was those with episomal DNA specific to HPV type.

higher proportion of integrated virus compared with others, which may, in turn, explain why some go on to develop cancer while others do not.

There is growing interest in how these integration events occur and the resulting types of genomic status for HPV. In exclusively integrated HPV, two types of integration patterns have been described. In type 1, one copy of HPV DNA is inserted into cellular DNA; and in type 2, concatemeric integrants are methylated to silence the viral genome, and E1/E2 is not expressed (29, 30). Alternatively, there have been situations where there is a mixture of both integrated and episomal HPV genomes, as we describe in this study. There is debate on how to interpret “mixed” integration, as there is some evidence that suggests that these are rather virus-human hybrid episomes (30). However, further research is needed to support this hypothesis.

The NHEJ pathway is key for the repair of dsDNA breaks when homologous template sequences are not available to guide the process. Instead, short homologous sequences in single-stranded overhangs at the break are used for the repair. Papillomaviruses exploit this process to insert their own DNA into the host's when these types of double-strand breaks occur (12, 31). Because the homology of sequences is incomplete, this insertion and repair process can lead to loss of host genetic material and/or interruption of coding sequences that could also affect the malignant transformation (32, 33).

NHEJ proceeds in three defined steps: end-binding and tethering, end processing, and ligation. Specific gene products are responsible for carrying out these three steps (Fig. 2). XRCC4 is the key protein that enables interaction of DNA ligase IV (LIG4) to damaged DNA in the final ligation step of repair. Mouse models have shown that knock-out of XRCC4 and even certain mutations in the gene are lethal to embryos (34). Other studies have shown that mutations in XRCC4 can lead to developmental inhibition and immunodeficiency in

humans (35). In addition, polymorphisms in XRCC4 are associated with an increased risk of cancer, especially prostate, lung, and bladder cancers (36).

The three genes with the most variants were XRCC4 (5q13–q14), WRN (8p12–11.2), and NBN (8q21–q24). The WRN gene is responsible for producing the Werner protein, which functions as a helicase and exonuclease, unwinding the DNA and removing abnormal DNA nucleotides. In contrast, the NBN gene makes proteins that interact with the protein complex of MRE11A and RAD50 in the end-binding and tethering step of repair. However, top SNPs were found within the XRCC4 gene for both HPV16- and HPV18-positive individuals. Specifically, we report that women with SNPs at rs7718284, rs17205699, rs4266384, or rs1193693 had nearly 2.5 times greater odds of integrated HPV16. This finding is in agreement with a recent report by Gupta and colleagues, who found in the context of cervical cancer, specific haplotypes of XRCC4 (rs28360071, rs28360317, rs6869366, and rs2075685) were nearly three times more prevalent among cervical cancer cases compared with healthy controls (OR: 3.08, 1.25–7.55; ref. 37). However, HPV infection status among cases and controls was not known. Currently, no GWAS or genetic studies have been conducted on HPV and cervical dysplasia. A study by Bowden and colleagues conducted a GWAS in the context of CIN, but HPV status was not reported. However, the authors report six independent variants associated with CIN3 and invasive cervical cancer, which included novel loci rs10175462 (PAX8) and rs27069 (CLPTM1L; ref. 28). In the context of HPV infection in cervical cancer development, a GWAS by Takeuchi and colleagues identified significant SNPs in the promoter region of the ARRDC3 gene, which may be involved in the infectious entry of HPV into the cell (38). Other genetic studies performed to date have focused on HLA and the relation to HPV infection. Specifically, a protective effect has been noted for the HLA DRB\*1301-DQB1\*0603 haplotype, while a positive association with HLA B7/DQB\*0301 and HLA-DQA1 has been shown (28, 39). In a paper by Wang and colleagues, XRCC1 Q399R polymorphism was associated with HPV persistence, while the EX01 T349M, CYBA 3' untranslated region (UTR) and FANCA polymorphisms were associated with risk of progression to CIN3 or ICC (40). Similarly, previous studies that examined host genetic variants in DNA repair-related genes have also reported significant associations with cervical cancer and HPV persistence (41, 42). However, ours is the first to present an association in the context of HPV integration and cervical dysplasia. Together these findings indicate that host genetic variation in NHEJ DNA repair pathway genes, especially XRCC4, are significantly associated with HPV integration.

We found integration events highest among women with normal cytology compared with those with LSIL and HSIL. This finding was unexpected, as data has shown integration increases from precancer to cancer (5, 43). One explanation is age at diagnosis differed significantly across histology groups, where women with diagnosed with HSIL tended to be younger

(mean age = 33.3 years) compared with women with normal cytology (43.5 years). Chromosomal instability is associated with HPV integration and has also shown to increase with age (44). Therefore, the likelihood of HPV integration may increase with increasing age.

The current study has a number of strengths. First, we had a large number of HPV16/18-positive samples from women representing the spectrum of cervical dysplasia (normal, LSIL, and HSIL). The samples were systematically collected and processed as part of the parent trials. Second, we included all genes known to function in the NHEJ DNA repair pathway, and we used a set of tagging SNPs for each included gene to ensure full coverage of the pathway while minimizing the number of statistical comparisons performed. Third, our study included a diverse population of women representing various racial/ethnic backgrounds, making the study more generalizable. To overcome any potential confounding by population stratification, we included a set of well-characterized ancestry informative markers to adjust for ancestry in our genetic analyses. Fourth, we utilized a robust and established assay to determine the viral integration status specific to HPV16 and HPV18 of the samples that were included in the study.

In spite of these strengths, there are also limitations to our study. First, due to the type-specificity of the integration assay and a lack of HPV typing beyond types 16 and 18, we were unable to examine a broader range of types. However, these two types represent approximately 75% of all invasive cervical cancers that occur. Second, the DNA used for genotyping was extracted from cervical cells collected as part of the study procedures. Although not ideal, it has been shown in other studies that the use of target tissue as a source of germline DNA for genotyping is acceptable unless the genes being examined are known to be mutated in the cancer tissue (45). As we examined the effects of these SNPs on integration in normal and dysplastic samples, rather than invasive carcinomas, we feel that this choice of DNA source is appropriate. Lastly, while we controlled for genetic ancestry in our analyses, the possibility of residual confounding due to population stratification still exists. However, when we restricted the analyses to women with at least 80% European ancestry, the results held.

This is the first study to evaluate the association between polymorphisms in the NHEJ pathway and HPV16 and HPV18 viral integration status on cervical dysplasia. Our results suggest that most women positive for HPV16 and HPV18 had a mix of episomal and integrated HPV DNA, and integration status differed across level of cervical dysplasia. Furthermore, our findings also indicate that host genetic variation in genes in

the NHEJ DNA repair pathway are significantly associated with integration of HPV into the host genome, and that these genes may play an important role in determining cervical cancer development and progression. Future work is needed to examine potential interactions between HPV integration status and other polymorphisms in DNA repair genes on cervical cancer risk. Overall, these findings may be of importance in future public health strategies that target detection and prevention of cervical dysplasia.

### Authors' Disclosures

E. Amirian reports other support from McKesson outside the submitted work. M.E. Scheurer reports grants from NIH during the conduct of the study. No disclosures were reported by the other authors.

### Authors' Contributions

**J.M. Geris:** Data curation, formal analysis, investigation, visualization, methodology, writing—original draft, writing—review and editing. **E.S. Amirian:** Conceptualization, formal analysis, investigation, methodology, writing—review and editing. **D.A. Marquez-Do:** Conceptualization, resources, data curation, investigation, methodology, writing—review and editing. **M. Guillaud:** Conceptualization, resources, data curation, supervision, investigation, methodology, writing—review and editing. **L.M. Dillon:** Conceptualization, resources, investigation, project administration, writing—review and editing. **M. Follen:** Conceptualization, resources, supervision, funding acquisition, investigation, writing—review and editing. **M.E. Scheurer:** Conceptualization, resources, data curation, supervision, funding acquisition, investigation, methodology, project administration, writing—review and editing.

### Acknowledgments

This work was funded in part by the NIH (R03CA143965, principal investigator: M. Scheurer; K07CA131505, principal investigator: M. Scheurer; P01CA82710, principal investigator: M. Follen). J.M. Geris was funded by a Research Training Award from the Cancer Prevention & Research Institute of Texas (CPRIT) for the Systems Epidemiology of Cancer Training (SECT) Program (RP210037; principal investigator: A. Thrift). We would also like to acknowledge the contributions of our late colleague, Dr. Karen Adler-Storthz. She was an important contributor to this work, and her guidance and friendship will be missed.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

### Note

Supplementary data for this article are available at Cancer Prevention Research Online (<http://cancerprevres.aacrjournals.org/>).

Received February 9, 2023; revised March 6, 2023; accepted May 19, 2023; published first May 22, 2023.

### References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
- Chesson HW, Dunne EF, Hariri S, Markowitz LE. The estimated lifetime probability of acquiring human papillomavirus in the United States. *Sex Transm Dis* 2014;41:660–4.
- Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst* 1995;87:796–802.
- Arends MJ, Donaldson YK, Duvall E, Wyllie AH, Bird CC. HPV in full thickness cervical biopsies: high prevalence in CIN 2 and CIN 3 detected by a sensitive PCR method. *J Pathol* 1991;165:301–9.

5. Arias-Pulido H, Peyton CL, Joste NE, Vargas H, Wheeler CM. Human papillomavirus Type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. *J Clin Microbiol* 2006;44:1755–62.
6. Das BC, Sharma JK, Gopalakrishna V, Luthra UK. Analysis by polymerase chain reaction of the physical state of human papillomavirus type 16 DNA in cervical preneoplastic and neoplastic lesions. *J Gen Virol* 1992;73:2327–36.
7. Hong S, Laimins LA. Regulation of the life cycle of HPVs by differentiation and the DNA damage response. *Future Microbiology* 2013;8:1547–57.
8. Thorland EC, Myers SL, Persing DH, Sarkar G, McGovern RM, Gostout BS, et al. Human papillomavirus Type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res* 2000;60:5916–21.
9. Xu B, Chotewutmontri S, Wolf S, Klos U, Schmitz M, Dürst M, et al. Multiplex identification of human papillomavirus 16 DNA integration sites in cervical carcinomas. *PLoS One* 2013;8:e66693.
10. Haber JE. Partners and pathways repairing a double-strand break. *Trends Genet* 2000;16:259–64.
11. Lieber MR. The mechanism of human nonhomologous DNA end joining. *J Biol Chem* 2008;283:1–5.
12. Van Gent DC, Hoesjmakers JHJ, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2001;2:196–206.
13. Shuhatovich OM, Sharman MP, Mirabal YN, Earle NR, Follen M, Basen-Engquist K. Participant recruitment and motivation for participation in optical technology for cervical cancer screening research trials. *Gynecol Oncol* 2005;99(3 Suppl 1):S226–31.
14. Pham B, Earle N, Rabel K, Follen M, Scheurer ME. Maximizing the diversity of participants in a Phase II clinical trial of optical technologies to detect cervical neoplasia. *Gynecol Oncol* 2007;107:S208–S14.
15. Buys TPH, Cantor SB, Guillaud M, Adler-Storthz K, Cox DD, Okolo C, et al. Optical technologies and molecular imaging for cervical neoplasia: a program project update. *Gender Medicine* 2012;9:S7–S24.
16. Pham B, Rhodes H, Milbourne A, Adler-Storthz K, Follen M, Scheurer ME. Epidemiologic differentiation of diagnostic and screening populations for the assessment of cervical dysplasia using optical technologies. *Gender Medicine* 2012;9:S36–47.
17. Malpica A, Maticic JP, Niekirk DV, Crum CP, Staerke GA, Yamal JM, et al. Kappa statistics to measure interrater and intrarater agreement for 1790 cervical biopsy specimens among twelve pathologists: qualitative histopathologic analysis and methodologic issues. *Gynecol Oncol* 2005;99(3 Suppl 1):S38–52.
18. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, et al. The 2001 Bethesda system: terminology for reporting results of cervical cytology. *JAMA* 2002;287:2114–9.
19. Scheurer ME, Tortolero-Luna G, Guillaud M, Follen M, Chen Z, Dillon LM, et al. Correlation of human papillomavirus type 16 and human papillomavirus type 18 e7 messenger RNA levels with degree of cervical dysplasia. *Cancer Epidemiol Biomarkers Prev* 2005;14:1948–52.
20. Manos MM, Shin TY, Wright DK, Lewis AI, Broker TR, Wolinsky SM, et al. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* 1989;7:209–14.
21. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–20.
22. Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat* 2008;29:648–58.
23. Peitsaro P, Johansson B, Syrjänen S. Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique. *J Clin Microbiol* 2002;40:886–91.
24. Huang LW, Chao SL, Lee BH. Integration of human papillomavirus type-16 and type-18 is a very early event in cervical carcinogenesis. *J Clin Pathol* 2008;61:627–31.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodological)* 1995;57:289–300.
26. Adebamowo SN, Adeyemo AA, Rotimi CN, Olaniyan O, Offiong R, Adebamowo CA. Genome-wide association study of prevalent and persistent cervical high-risk human papillomavirus (HPV) infection. *BMC Med Genet* 2020;21:231.
27. Espinoza H, Ha KT, Pham TT, Espinoza JL. Genetic predisposition to persistent human papillomavirus-infection and virus-induced cancers. *Microorganisms* 2021;9:2092.
28. Bowden SJ, Bodinier B, Kalliala I, Zuber V, Vuckovic D, Doulgeraki T, et al. Genetic variation in cervical preinvasive and invasive disease: a genome-wide association study. *Lancet Oncol* 2021;22:548–57.
29. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. *PLoS Pathog* 2017;13:e1006211.
30. Morgan I, Dinardo L, Windle B. Integration of human papillomavirus genomes in head and neck cancer: is it time to consider a paradigm shift? *Viruses* 2017;9:208.
31. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 2001;27:247–54.
32. Pett MR, Alazawi WOF, Roberts I, Downen S, Smith DI, Stanley MA, et al. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus Type 16 in cervical keratinocytes. *Cancer Res* 2004;64:1359–68.
33. Winder DM, Pett MR, Foster N, Shivji MK, Herdman MT, Stanley MA, et al. An increase in DNA double-strand breaks, induced by Ku70 depletion, is associated with human papillomavirus 16 episome loss and de novo viral integration events. *J Pathol* 2007;213:27–34.
34. Li H, Vogel H, Holcomb VB, Gu Y, Hasty P. Deletion of Ku70, Ku80, or both causes early aging without substantially increased cancer. *Mol Cell Biol* 2007;27:8205–14.
35. Andres SN, Vergnes A, Ristic D, Wyman C, Modesti M, Junop M. A human XRCC4–XLF complex bridges DNA. *Nucleic Acids Res* 2012;40:1868–78.
36. Shao N, Jiang WY, Qiao D, Zhang SG, Wu Y, Zhang XX, et al. An updated meta-analysis of XRCC4 polymorphisms and cancer risk based on 31 case-control studies. *Cancer Biomark* 2012;12:37–47.
37. Gupta MK, Kushwah AS, Singh R, Banerjee M. Genotypic analysis of XRCC4 and susceptibility to cervical cancer. *Br J Biomed Sci* 2020;77:7–12.
38. Takeuchi F, Kukimoto I, Li Z, Li S, Li N, Hu Z, et al. Genome-wide association study of cervical cancer suggests a role for ARRDC3 gene in human papillomavirus infection. *Hum Mol Genet* 2019;28:341–8.
39. Wang SS, Hildesheim A. Chapter 5: viral and host factors in human papillomavirus persistence and progression. *J Natl Cancer Inst Monogr* 2003:35–40.
40. Wang SS, Bratti MC, Rodriguez AC, Herrero R, Burk RD, Porras C, et al. Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer. *J Infect Dis* 2009;199:20–30.
41. Wang SS, Gonzalez P, Yu K, Porras C, Li Q, Safaeian M, et al. Common genetic variants and risk for HPV persistence and progression to cervical cancer. *PLoS One* 2010;5:e8667.



42. Joo J, Omae Y, Hitomi Y, Park B, Shin H-J, Yoon K-A, et al. The association of integration patterns of human papilloma virus and single nucleotide polymorphisms on immune- or DNA repair-related genes in cervical cancer patients. *Sci Rep* 2019;9:13132.
43. Hudelist G, Manavi M, Pischinger KI, Watkins-Riedel T, Singer CF, Kubista E, et al. Physical state and expression of HPV DNA in benign and dysplastic cervical tissue: different levels of viral integration are correlated with lesion grade. *Gynecol Oncol* 2004;92:873–80.
44. Bodelon C, Vinokurova S, Sampson JN, den Boon JA, Walker JL, Horswill MA, et al. Chromosomal copy number alterations and HPV integration in cervical precancer and invasive cancer. *Carcinogenesis* 2016;37:188–96.
45. Koushik A, Ghosh A, Duarte-Franco E, Forest P, Voyer H, Matlaszewski G, et al. The p53 codon 72 polymorphism and risk of high-grade cervical intraepithelial neoplasia. *Cancer Detect Prev* 2005;29:307–16.

