



Article

# Genetic and Epigenetic Variations of HPV52 in Cervical Precancer

Katharine J. Bee <sup>1,2,†</sup>, Ana Gradissimo <sup>1,†</sup>, Zigui Chen <sup>1,3</sup> , Ariana Harari <sup>1</sup>, Mark Schiffman <sup>4</sup>, Tina Raine-Bennett <sup>5</sup>, Philip E. Castle <sup>4,6,7</sup>, Megan Clarke <sup>4</sup>, Nicolas Wentzensen <sup>4</sup> and Robert D. Burk <sup>1,6,8,\*</sup>

- <sup>1</sup> Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY 10461, USA; katharine.bee@dbv-technologies.com (K.J.B.); ana.gradissimo@einsteinmed.org (A.G.); zigui.chen@cuhk.edu.hk (Z.C.); ariana.harari@merck.com (A.H.)
- <sup>2</sup> DBV Technologies, 92120 Montrouge, France
- <sup>3</sup> Department of Microbiology, The Chinese University of Hong Kong, Hong Kong, China
- <sup>4</sup> Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; schiffmm@exchange.nih.gov (M.S.); castlep@mail.nih.gov (P.E.C.); megan.clarke@nih.gov (M.C.); wentzenn@mail.nih.gov (N.W.)
- <sup>5</sup> Division of Research, Kaiser Permanente Northern California, Oakland, CA 94612, USA; Tina.R.Raine-Bennett@kp.org
- <sup>6</sup> Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA
- <sup>7</sup> Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Rockville, MD 20850, USA
- <sup>8</sup> Microbiology & Immunology, and Obstetrics, Gynecology & Women's Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA
- \* Correspondence: robert.burk@einsteinmed.org; Tel.: +1-718-430-3720
- † Co-authors contributed equally.



**Citation:** Bee, K.J.; Gradissimo, A.; Chen, Z.; Harari, A.; Schiffman, M.; Raine-Bennett, T.; Castle, P.E.; Clarke, M.; Wentzensen, N.; Burk, R.D. Genetic and Epigenetic Variations of HPV52 in Cervical Precancer. *Int. J. Mol. Sci.* **2021**, *22*, 6463. <https://doi.org/10.3390/ijms22126463>

Academic Editor: Marta del Pino

Received: 24 May 2021  
Accepted: 11 June 2021  
Published: 16 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The goal of this study was to identify human papillomavirus (HPV) type 52 genetic and epigenetic changes associated with high-grade cervical precancer and cancer. Patients were selected from the HPV Persistence and Progression (PaP) cohort, a cervical cancer screening program at Kaiser Permanente Northern California (KPNC). We performed a nested case-control study of 89 HPV52-positive women, including 50 cases with predominantly cervical intraepithelial neoplasia grade 3 (CIN3) and 39 controls without evidence of abnormalities. We conducted methylation analyses using Illumina sequencing and viral whole genome Sanger sequencing. Of the 24 CpG sites examined, increased methylation at CpG site 5615 in HPV52 L1 region was the most significantly associated with CIN3, with a difference in median methylation of 17.9% (odds ratio (OR) = 4.8, 95% confidence interval (CI) = 1.9–11.8) and an area under the curve of 0.73 (AUC; 95% CI = 0.62–0.83). Complete genomic sequencing of HPV52 isolates revealed associations between SNPs present in sublineage C2 and a higher risk of CIN3, with ORs ranging from 2.8 to 3.3. This study identified genetic and epigenetic HPV52 variants associated with high risk for cervical precancer, improving the potential for early diagnosis of cervical neoplasia caused by HPV52.

**Keywords:** HPV52; papillomavirus; methylation; next generation sequencing; evolution; phylogeny; precancer

## 1. Introduction

Cervical cancer is the fourth most common type of cancer affecting women worldwide and the fourth highest cause of death from cancer in women [1,2]. The highest disease burden occurs in low- and middle-income countries, where more than 90% of deaths occur due to poor access to screening and treatment [3]. In fact, in many areas of Africa, cervical cancer is the leading cause of cancer-related deaths [2]. Practically all cases of cervical cancer are caused by infection with the human papillomavirus (HPV). Of the over 200 types of HPV curated, at least 12 are considered to be high-risk (HR-HPV) [4]. Two

of the high-risk types, HPV16 and HPV18, are responsible for roughly 70% of cases of cervical cancer, with HPV31, 33, 35, 45, 52, and 58 constituting the majority of the remaining cases [5,6]. The distribution of these HR-HPVs varies geographically across continents, particularly HPV52 and HPV58, whose prevalence is notably higher in cases of invasive cervical cancer in Asia [5,6], though also frequently detected in Brazil, according to a recent report [7]. For example, HPV52 is the fourth most frequent HPV type detected in cervical cancer cases in China [8]. Many countries in Africa and Asia have yet to introduce the recently FDA-approved nonavalent HPV vaccine that includes HPV52 [9], and until the vaccine is implemented, these regions will continue to encounter a high burden of HPV52-related cervix cancer deaths [10–19]. Given that HPV52 is a major circulating HPV type in various countries in Eastern Asia, where the cervical cancer incidence rates remain elevated, the genomic characterization of this HPV type is of epidemiological importance, with implications for cervical cancer screening strategies. In addition, HPV52 has also been detected in Bowen disease of the palm [20] and penile cancer [21], indicating an oncogenic potential at other epithelial sites.

Most HR-HPV types associated with cervical cancer are phylogenetically closely related, belonging to the species groups *Alphapapillomavirus* 9 (Alpha9, HPV16-related) or *Alphapapillomavirus* 7 (Alpha7, HPV18-related) [22]. HPV52 falls within the Alpha9 species group. HPV types are characterized by their nucleotide sequence similarity and differ by more than 10% in the L1 open reading frame by definition [23]. HPV types are further grouped into variant lineages and sublineages when the pairwise nucleotide sequences of their complete genomes differ by 1–10% or 0.5–1%, respectively [24].

The majority of HPV infections clear (i.e., become undetectable by commonly used HPV tests) naturally within the first two years after exposure to the virus [25]. However, approximately 5% of HR-HPV infections will progress to cervical intraepithelial neoplasia grade 3 (CIN3), the most specific histopathologic diagnosis of precancer [26]. Of these cases, up to one-third will eventually develop into invasive cervical cancer [27]. Although cytological screening (Pap test) has greatly reduced the disease burden in high-resource countries, it has a lower sensitivity than testing for HR-HPV. The high sensitivity as well as the strong negative predictive value of HR-HPV testing have led to changes in screening guidelines by including HPV primary screening [28]. The U.S. FDA has approved HPV tests for cervical cancer prevention that do not require cytology. As a result of screening and early detection, ablation or excision of precancerous lesions can prevent cervical cancer. However, these screening methods are still unable to differentiate between benign HR-HPV infections that will clear and those that will progress to cancer, making it difficult to effectively manage women who test positive for HR-HPV infection [29]. Therefore, identifying biomarkers that differentiate between more or less risky HR-HPV infections is highly desirable [30].

A potential biomarker of HR-HPV infections associated with cervix precancer/cancer is HPV DNA methylation [31]. Methylation is the covalent modification of the DNA base cytosine that occurs classically at CG dinucleotides in which the cytosine is 5' to the guanine (CpG site). Previous studies of cases of HPV16 infection have shown that viral methylation levels at CpG sites in the L1 region were significantly higher in patients with CIN3 or CIN3+ (i.e., CIN3, adenocarcinoma in situ (AIS) or squamous cell cancer (SCC)) compared to those women with HPV16 infections with <CIN2 (for a review, see [31–34]). Similarly, highly elevated levels of viral methylation were seen in women with CIN3+ infected with HPV18, -31, or -45 when compared to women < CIN2 [35]. Therefore, increased levels of methylation associated with cervical precancer may serve as a predictive or diagnostic biomarker for risk of developing cervical cancer in HPV-positive women.

HR-HPVs have between 80 and 150 CpG sites across the genome, many of which are highly conserved across species types [31]. Increased methylation at two CpG sites in the L1 region were among those with the highest association between methylation levels and disease status; these two sites show very high topological conservation across the HR-HPVs belonging to the Alpha9 species (HPV16, -31, -33, -35, -52, and -58) [31,36]. The high

levels of conservation as well as the strong correlation of methylation and disease status at these sites suggest a potential functional role for methylation in disease pathogenesis. Additionally, sequence variations among sublineages may have evolved through niche and host adaptation with higher oncogenic potential including changes that may create or eliminate CpG sites, affecting the potential for methylation of the viral genome.

We sought to expand our understanding of HPV52 at the genomic level. Therefore, we report on the levels of viral methylation in patients positive for infection with HPV52 at selective CpG sites across the L1, L2, E7, E1, and URR (upstream regulatory region) regions on the basis of conservation across related types. In addition, we sequenced the complete viral genome of HPV52-positive case and control samples to identify possible variants that may be more strongly associated with CIN3.

## 2. Results

### 2.1. CpG Methylation Determined by Pyrosequencing

HPV isolates from 89 women infected with HPV52 but no other HR types were selected from previously characterized cervical samples from the PaP cohort for analysis of viral methylation levels in patients with CIN3 and controls (<CIN2). Out of the 24 CpG sites analyzed across the genome, 22 CpG sites had equal or higher median levels of HPV52 methylation in women diagnosed with CIN3 as compared to control HPV52-positive women without histological or cytological abnormalities (Table 1).

**Table 1.** Median methylation values of cases and controls by pyrosequencing.

Region Name	Genome Position	Median		Difference	Man-Whitney (Wilcoxon Rank Sum) <i>p</i> Value	ROC Curve	95% CI Interval	Univariate Regression High HPV Methylation vs. Low/Medium Methylation		
		Control ( <i>n</i> = 39)	Case ( <i>n</i> = 50)					AUC	OR	95% CI
E7	c_754	1.40	1.80	0.40	0.010910	0.6583	(0.54, 0.78)	4.64	(1.88, 11.47)	0.000886
	c_776	13.20	13.30	0.10	0.542481	0.5379	(0.42, 0.66)	2.36	(0.95, 5.89)	0.064880
	c_779	1.20	1.20	0.00	0.771323	0.5181	(0.39, 0.65)	0.50	(0.18, 1.39)	0.182842
	c_787	5.90	6.40	0.50	0.141489	0.5916	(0.47, 0.71)	1.99	(0.82, 4.81)	0.125934
E1	c_1122	1.85	1.55	−0.30	0.319005	0.4405	(0.32, 0.56)	0.61	(0.23, 1.59)	0.313096
	c_1141	6.20	7.55	1.35	0.057360	0.6187	(0.50, 0.74)	1.85	(0.76, 4.46)	0.173033
L2	c_4274	8.00	13.50	5.50	0.001171	0.7013	(0.59, 0.82)	2.44	(1.01, 5.86)	0.046613
	c_4283	18.50	17.20	−1.30	0.003156	0.6831	(0.57, 0.80)	2.25	(0.94, 5.41)	0.070045
	c_4292	10.90	19.90	9.00	0.001699	0.6946	(0.58, 0.81)	2.44	(1.01, 5.86)	0.046613
	c_4919	2.40	3.00	0.60	0.085078	0.6062	(0.49, 0.72)	2.25	(0.94, 5.41)	0.070045
	c_4926	4.30	5.25	0.95	0.251942	0.5710	(0.45, 0.69)	1.50	(0.62, 3.63)	0.368990
	c_4931	4.40	4.70	0.30	0.295395	0.5649	(0.45, 0.68)	1.56	(0.63, 3.84)	0.333509
	c_4949	32.60	39.50	6.90	0.371808	0.5554	(0.43, 0.68)	1.38	(0.57, 3.35)	0.478098
L1	c_5606	18.00	33.20	15.20	0.001393	0.6982	(0.58, 0.81)	3.38	(1.39, 8.17)	0.007037
	c_5609	15.00	25.90	10.90	0.007199	0.6667	(0.55, 0.79)	2.64	(1.10, 6.36)	0.030232
	c_5615	30.20	48.10	17.90	0.000257	0.7267	(0.62, 0.83)	4.78	(1.94, 11.80)	0.000684
	c_5621	5.00	8.05	3.05	0.001651	0.6951	(0.58, 0.81)	2.86	(1.19, 6.90)	0.019101
	c_5648	26.80	32.15	5.35	0.034620	0.631	(0.51, 0.75)	2.08	(0.86, 5.00)	0.102620
	c_5655	35.90	51.20	15.30	0.001799	0.6936	(0.58, 0.80)	3.67	(1.51, 8.92)	0.004099
	c_7098	3.10	3.90	0.80	0.024831	0.6397	(0.52, 0.76)	1.99	(0.82, 4.81)	0.125934
	c_7111	5.90	10.30	4.40	0.016845	0.6489	(0.53, 0.77)	2.76	(1.14, 6.68)	0.024124
URR	c_7119	3.40	5.70	2.30	0.032968	0.6327	(0.51, 0.76)	3.55	(1.46, 8.65)	0.005258
	c_7557	2.50	2.50	0.00	0.885024	0.5096	(0.37, 0.64)	0.81	(0.30, 2.19)	0.680344
	c_7563	2.80	3.05	0.25	0.423051	0.553	(0.42, 0.68)	1.71	(0.68, 2.01)	0.579475

Note: Pyrosequencing was conducted for cases (*n* = 50) and controls (*n* = 39) across 24 CpG loci, as described in the Materials and Methods section. Methylated loci are grouped to their location within the HPV52 genome.

The median methylation levels ranged from 1.2 to 35.9% for controls ( $n = 39$ ) and from 1.2 to 51.2% for cases ( $n = 50$ ). Overall, the L1 region showed the highest median levels of methylation compared to other analyzed regions of the viral genome in both cases and controls. The median differences in methylation for cases compared to controls ranged from  $-1.3$  to 17.9% across all sites, and  $p$ -values ranged from 0.89 to 0.00026 (Table 1). The calculated odds ratios (ORs) ranged from 0.5 to 4.8. The CpG site 5615, located in L1, was the most significant site associated with CIN3, with a difference in median methylation of 17.9% (OR = 4.8; 95% confidence interval (CI) = 1.9–11.8) and had an area under the curve (AUC) of 0.73 (95% CI = 0.62–0.83). The majority of CpG sites with the most significant differences in median methylation levels between cases and controls clustered in the 5' ends of both L2 and L1 regions.

Two single nucleotide polymorphisms were identified at sites 7111 and 7112 in the L1 region in six samples (C7111A and G7112A); both eliminated a CpG site. When these samples were removed from the statistical analyses, the difference in median methylation of cases compared to controls changed slightly, from 4.4% to 4.3%. No other single nucleotide polymorphisms (SNPs) were found that affected the presence of a CpG site among those sites that were analyzed for methylation.

## 2.2. CpG Methylation Determined by Next-Generation Sequencing (NGS)

In addition to pyrosequencing, methylation levels of bisulfite-converted viral DNA were also analyzed using NGS for 6 CpG sites in the L1 region that are highly conserved across multiple HR-HPVs [31,37]. For this analysis, 88 samples were analyzed; one sample was of insufficient quantity. The median levels of methylation ranged from 5.3 to 48.2% for controls and from 6.3 to 52.9% for cases (Table 2). A comparison of the pyrosequencing and NGS analyses revealed that the results from the two methods were correlated, with R squared values for each CpG site ranging from 0.571 to 0.773 (Figure S1). The CpG site at position 5615 had the largest difference between case and control samples, with median methylation values of 40.9% in controls and 46.8% in cases, with a OR of 3.67 (95% CI = 1.51–8.92) (Table 2), similar to methylation determined by pyrosequencing (see above).

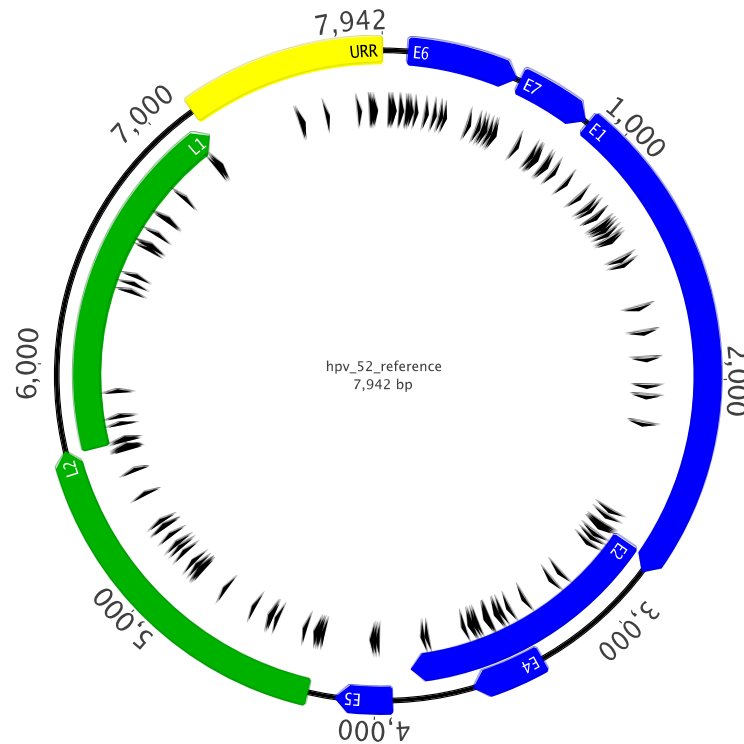
**Table 2.** Median methylation values of cases and controls by next-generation sequencing.

Region Name	Genome Position	Median		Difference	Man-Whitney (Wilcoxon Rank Sum)	ROC Curve	95% CI Interval	Univariate Regression High HPV Methylation vs. Low/Medium Methylation		
		Control ( $n = 38$ )	Case ( $n = 50$ )		$p$ Value	AUC	OR	95% CI	$p$ Value	
L1	c_5606	27.36	30.45	3.09	0.005258	0.6731	(0.56, 0.79)	3.38	(1.39, 8.17)	0.007037
	c_5609	18.84	22.11	3.27	0.003606	0.6805	(0.56, 0.80)	1.92	(0.80, 4.61)	0.146627
	c_5615	40.90	46.76	5.86	0.003330	0.6821	(0.57, 0.80)	3.67	(1.51, 8.92)	0.004099
	c_5621	5.34	6.26	0.92	0.052489	0.6203	(0.50, 0.74)	1.92	(0.80, 4.61)	0.146627
	c_5648	32.21	37.07	4.86	0.003512	0.681	(0.57, 0.80)	2.86	(1.19, 6.90)	0.019101
	c_5655	48.22	52.92	4.70	0.005895	0.6708	(0.55, 0.79)	2.86	(1.19, 6.90)	0.019101

## 2.3. Sanger Sequencing Analysis of Complete HPV52 Genomes

To examine possible associations of HPV52 variant lineages and SNPs with disease status and methylation levels, we obtained complete HPV52 genomic sequences for 83 of 89 samples, and partial sequences were obtained for four samples (Figure 1). The genomic sequence of an additional 23 HPV52 samples from a previous study (Qv) were added to the analyses for a total of 106 complete and 4 partial genomes [36]. A total of 781 nucleotide sites across the full 7942 bp genome (9.8%) were variable when compared to the HPV52 reference genome (data not shown). The most variable region was the URR with

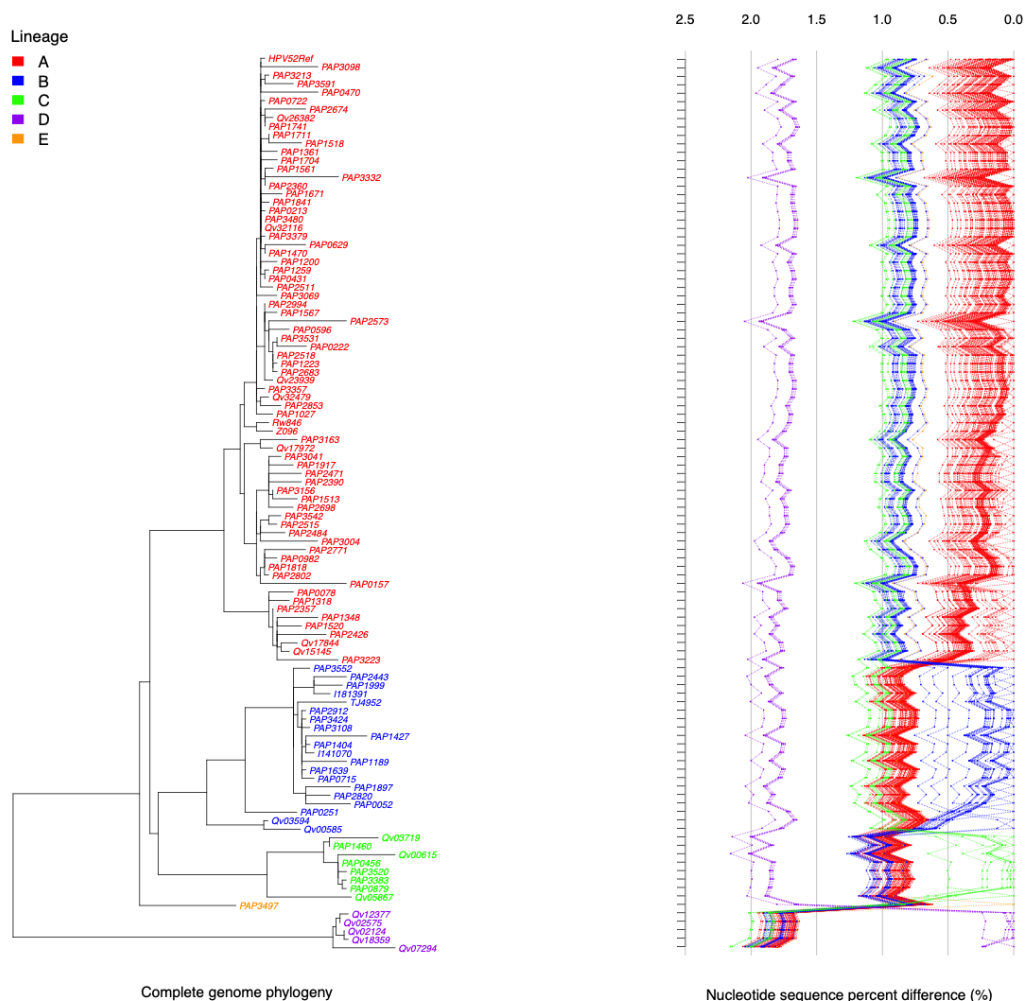
13.7% overall nucleotide diversity, followed by the noncoding region 1 (NCR1) located between the E2 and E5 genes with 13.3% diversity. In addition to SNPs, various insertions and deletions were seen throughout the genome, the majority of which were located in the URR.



**Figure 1.** Distribution of CpG sites across HPV52 reference genome. The complete genomic sequence of the HPV52 genome (GenBank ID 397038) was extracted from the PaVE database (<http://pave.niaid.nih.gov>; accessed on 4 March 2021). CpG sites are marked with a black arrowhead. Early genes are shown in blue, late genes are shown in green, and the upstream regulatory region (URR) is shown in yellow. This image was created and annotated using Geneious v6.1.7 (<http://geneious.com>).

#### 2.4. HPV52 Phylogenetic Analysis of 106 Isolates

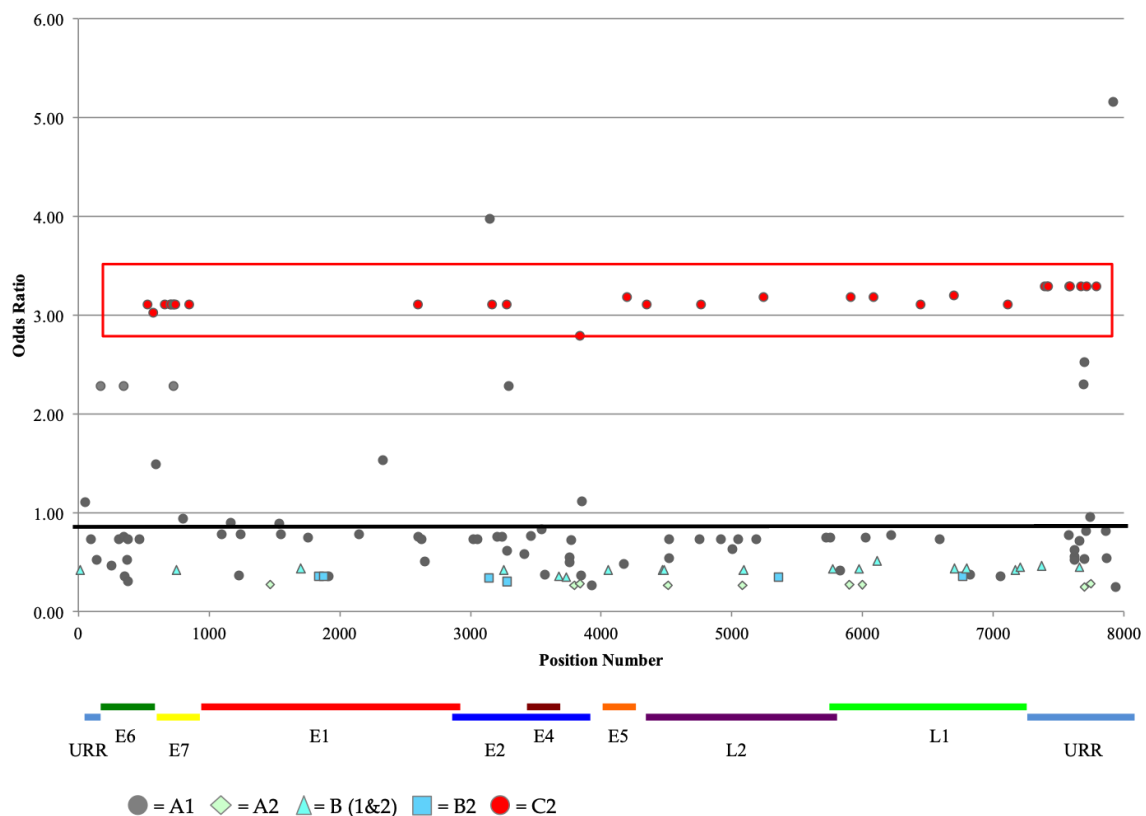
On the basis of the alignment of the complete genomic sequence of 106 HPV52 isolates (the four isolates for which only partial sequence was obtained were omitted from these analyses), we assigned variant lineages and sublineages using an alphanumeric system, with the “A1” clade containing the reference genome as previously described [36]. The phylogenetic tree generated from the complete nucleotide sequences clustered HPV52 isolates into a total of five lineages, including nine sublineages, with the majority of isolates belonging to lineage A ( $n = 72$ ) (Figure 2). One lineage, E, and two sublineages, B3 and C1, consisted of only one isolate each. The maximum nucleotide pairwise difference between the most dissimilar isolates was 2.2%, seen between isolates in the C and D lineages.



**Figure 2.** HPV52 phylogenetic tree topology and pairwise comparisons of individual complete genomes. The phylogenetic tree was inferred from global alignment of the complete nucleotide sequences from 106 HPV52 genomes. Distinct variant lineages (A, B, C, D, and E) were classified according to the topology and nucleotide sequence differences from >1% to <10% and are represented according to the color code shown at the top left of the figure. The percent nucleotide differences for each isolate compared to all other isolates are shown on the right. Lines indicate values for each comparison of a given isolate with the comparison to self as indicated by the 0% difference point to the far right.

### 2.5. Association of Variant Lineage and Disease Status

To determine if there was an association between HPV52 SNPs and disease outcome, the OR values for each SNP position were calculated and plotted as a viral genome wide SNP analysis (VWAS) (Figure 3). The OR values appeared stratified, displaying horizontal clusters of SNPs with approximately equal ORs. Upon further analysis, these clusters constituted the SNPs diagnostic of a sublineage, consistent with previous findings [38–40]. Of all the sublineages, the C2 sublineage had the highest range of ORs (2.8–3.3) associated with CIN3 (Figure 3, see red rectangular box). Of the five samples in the C2 sublineage, four were diagnosed with CIN3 and one had normal cytology.



**Figure 3.** HPV52 viral genome-wide association study. Odds ratios for cases of CIN3 compared to HPV52 control infections were calculated for each SNP that occurred at least twice from the genomic sequence of 87 HPV52s. The red box represents variants associated with the C2 sublineage.

### 2.6. Analysis of Nucleotide Variations at CpG Sites

Of the 781 variable nucleotides found throughout the genome, 125 (16.0%) were located within a CpG site based on the HPV52 reference sequence. Of the 131 CpG sites in the reference sequence, 56 of these CpG sites were disrupted by a nucleotide change in one of the HPV52 isolates. An additional 69 SNPs resulted in the creation of a CpG site not found in the reference genome. Of the 125 SNPs that either created or destroyed CpG sites, 80 were found among all members of B2 sublineage.

## 3. Discussion

The results of the present study reflect the importance of understanding how both genetic and epigenetic variations of HR-HPVs are risk factors for the pathogenicity of a given infection. The association of cervical cancer and infection with HR-HPV has been well established, and it has become possible to distinguish increasingly refined levels of genomic diversity responsible for this association. For example, it was initially determined that HR-HPVs belong predominantly to the Alpha9 and Alpha7 clades. It is now possible to define the magnitude of association with cervical precancer/cancer of not only clade and species, but also lineages, sublineages, and even specific nucleotides [34,39–41]. An additional method of analyzing association of HR-HPV infection and cervical cancer is the use of methylation levels as a biomarker of disease severity. Results of this study add to the understanding of the oncogenic qualities of HPV52 associated with both genetic variation and epigenetic modification of the viral genome.

A primary aim of our study was to analyze the association of disease outcome with levels of viral methylation to further validate the use of this epigenetic modification as a biomarker for identifying at-risk individuals [31,34,37,41–44]. Our data indicate that multiple CpG sites across the HPV52 genome show differential methylation levels associated

with CIN3 compared to HPV52-positive infections without high-grade disease, particularly within the L1 region. Results achieved through pyrosequencing were confirmed for several of the CpG sites by NGS, in accordance with previous results focused on CpG methylation determined by NGS [32,33]. Our results support previous findings in which elevated levels of viral methylation were associated with high-grade cervical neoplasia and cancer (for review, see [31,37,41,43,45]). Taken together, these data add to the growing body of knowledge that increased HPV methylation may serve as a biomarker to identify women with precancerous/cancerous lesions, particularly in the viral L1 region [46,47]. Moreover, HR-HPV DNA methylation has also been shown to be predictive in women infected with HIV (which harbor multiple HR-HPV-type infections), as it could be used to discriminate active HR-HPV infections while suggesting the “causal” HPV type responsible for the detected cervical abnormalities [48].

Another goal of this study was to analyze the relationship between HPV52 sequence variation and disease outcome. We identified a cluster of SNPs diagnostic of the HPV52 C2 sublineage that was associated with an increased risk of CIN3. Of interest was the fact that lineage E and sublineages B2/B3 and C2 contained the L83V amino acid substitution previously suggested to alter the ability of E6 to bind and degrade p53, inducing stronger antiapoptotic signals [49–52]. These findings are consistent with a higher pathogenicity of HPV52 sublineage C2. In line with our findings that HPV52 lineage C is associated with CIN3, an increased prevalence of cervical neoplasia and this lineage was also identified in a Taiwanese population [53], as well as in a study from Korea [52]. However, other studies that have analyzed associations of HPV52 variant lineages and cervical precancer/cancer have included few or no samples belonging to the C lineage [54–59]. Additionally, most similar studies have only examined sequence variations in small regions of the genome, most often fragments of E6, E7, L1, and URR, making it difficult to definitively assign lineage and sublineage classification to the viral isolate [52–57,60]. It is interesting to note that many of the SNPs diagnostic of the C2 sublineage clustered in the E7 and URR portions of the genome, suggesting a potential functional role for these regions that might confer greater pathogenicity in members of this sublineage. Moreover, the dominance of A1 sublineage in our study is consistent with the distribution of this particular sublineage in the Americas and Europe [7,53,58,61].

Although the mechanism by which viral methylation results in an increased risk of developing cervical precancer and cancer is currently unknown, a number of possibilities exist to explain the association. It is possible that HR-HPV CpG methylation influences gene expression and cell cycle disruption [62]. Alternatively, CpG methylation may be a marker of cell dedifferentiation associated with precancer and the changes in viral gene expression [63]. High levels of conservation seen at CpG sites across phylogenetically related HPV types as well as the correlation between viral methylation and cervical disease status suggest that, in fact, these sites may play an important functional role. Levels of methylation may be impacted by the number of CpG sites in the HPV genome, a factor that changes as a result of sequence variations between and within types that create and destroy CpG sites. Future studies should consider HPV sequence variation and its influence on number and topology of CpG sites and level of CpG methylation.

The current report has strengths and weaknesses. Some of the strengths are related to the population-based selection of cases and controls with HPV52 and that HPV52 was the only HR-HPV type present in the sample. Nevertheless, the study had limited number of cases and controls and did not include sufficient numbers of various population groups that might be more susceptible to HPV52 transformation. Thus, larger studies amongst other population groups are needed to validate the genomic and epigenetic contributions of HPV52-associated cervix cancer risks.



## 4. Materials and Methods

### 4.1. Study Design and Population

Patients were selected from the HPV Persistence and Progression (PaP) cohort, which include over 55,000 women who were enrolled in the cervical cancer screening program at Kaiser Permanente Northern California (KPNC). The cohort was created by banking residual, waste cervical specimens from women aged 25 and older who tested positive for hybrid capture 2 (HC2, Qiagen, Georgetown, MD, USA) during screening. Women with collected specimens were mailed opt-out consent letters and could respond if they wished not to have their specimens and information included in the study for HPV-related biomarkers for cervical precancer and cancer. Approximately 8% elected to opt-out and their specimens were destroyed. The study protocol was reviewed and approved by the Kaiser Foundation Research Institute and the National Cancer Institute (NCI) Institutional Review Boards. Patient cervical samples were screened using either conventional or liquid-based cytology and were tested for HPV by HC2, as previously described [64]. Samples positive for HC2 were analyzed for HPV type by PCR, as reported [64,65]. The current study consisted of a nested case–control population from the PaP cohort containing a total of 89 women 30 years of age or older that tested positive for HPV52 infection without other carcinogenic HR-HPV infections at the start of the study; 50 women were confirmed to have precancer by histology (49 CIN3 and 1 AIS) and 39 women were negative for intraepithelial lesions or malignancy (NILM) on cytologic exam and served as controls.

### 4.2. Sodium Bisulfite Treatment

DNA was treated with sodium bisulfite using the EpiTect Plus DNA Bisulfite Kit (Qiagen, Georgetown, MD, USA) as per the manufacturer's protocol, which allows for the downstream conversion of unmethylated cytosine nucleotides to thymine and the preservation of methylated cytosines in the subsequent PCR reaction. Briefly, 15  $\mu$ L of DNA was incubated with freshly prepared sodium bisulfite for approximately 5 h, cycling between 95 and 60 °C incubations, and held at 20 °C overnight. Following conversion, DNA samples were desulphonated and eluted in 25  $\mu$ L of elution buffer. Bisulfite-converted DNA was stored at –20 °C or used immediately for PCR.

### 4.3. Pyrosequencing

Pyrosequencing assays were designed for the quantitation of methylated cytosines at 24 CpG sites across the HPV genome (located in viral genes E7, E1, L2, L1, and the URR) (see Table 1). Sites were chosen on the basis of a high level of conservation among HPV types belonging to the Alpha9 clade and therefore closely related to HPV52 [36]. Bisulfite-converted DNA was amplified by PCR using the PyroMark PCR kit (Qiagen, Georgetown, MD, USA). HPV52-specific primers were designed to amplify a  $\approx$ 150 base pair (bp) region surrounding the CpG site (s). For each primer set, the reverse primer was 5'-biotinylated to permit downstream pyrosequencing of the PCR products. PCR amplification of HPV bisulfite-converted DNA was performed under the following conditions: initial denaturation at 95 °C for 15 min, 45 cycles of 94 °C 30 s, 55 °C 30 s, and 72 °C 30 s, and a final extension for 10 min at 72 °C. For some assays, the 55 °C annealing temperature was adjusted for optimization. PCR products were separated by electrophoresis on a 3% agarose gel. An aliquot of successful PCR products (a single band of anticipated size) was sent for pyrosequencing on a PSQ96 ID (Qiagen, Georgetown, MD, USA) at the Epigenomics Core Facility, Albert Einstein College of Medicine, Bronx, NY, USA.

### 4.4. Next-Generation Sequencing (NGS)

DNA was bisulfite-converted and purified as described above. Methylation-sensitive PCR primers were designed using MethPrimer [66] and synthesized by IDT (Integrated DNA Technologies, Coralville, IA, USA), as previously reported [32]. A region of 153 bp containing a total of 6 CpG sites in L1 was amplified by PCR. For all samples, a unique 8 bp DNA Hamming barcode [67] flanked by a 3 bp (5') and 4 bp (3') padding sequence was

introduced to both the 5' and 3' ends of the amplicon by the PCR primers. PCR products were analyzed by gel electrophoresis to confirm the presence of the predicted fragment size and to determine relative concentrations. Barcoded PCR products were pooled at approximately equal concentrations and purified by DNA electro-elution and isopropanol precipitation (QIAquick Gel Extraction Kit, Qiagen). Library preparation of the purified PCR products was performed using KAPA LTP Library Kit (Kapa Biosystems, Wilmington, MA, USA), and the pooled library was submitted for paired-end sequencing on an Illumina HiSeq2000 (Illumina Inc., San Diego, CA, USA) at the Epigenomics Core Facility, Albert Einstein College of Medicine, Bronx, NY, USA.

#### 4.5. Bioinformatic Analysis of NGS Reads

Prior to analysis for methylation quantification, sequencing files were first demultiplexed on the basis of the unique barcodes using scripts generated *in-house* in order to obtain Illumina paired-end reads for each sample [32]. Reads were aligned with the HPV52 reference genome by Bowtie v0.12.9 [68], and the methylation status of each CpG site was determined by Bismark v0.7.7 [69] using the default quality score parameter set to Q30. The formula for determining the percent methylation was the ratio of the number of C reads divided by the number of C + T reads.

#### 4.6. Sequencing of HPV52 Complete Genomes

The complete 8Kb genome of HPV52 was PCR-amplified in 2 to 3 overlapping fragments using type-specific primers with an equal mix of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA) and Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, USA) as previously described [36]. Successful PCR products, as determined by gel electrophoresis to be of anticipated size, were isolated using the Quick-Step 2 PCR Purification Kit (Edge BioSystems, Gaithersburg, MD, USA) or QIAquick Gel Extraction Kit (Qiagen, Georgetown, MD, USA) and submitted for Sanger sequencing at the Genomics Core Facility, Albert Einstein College of Medicine, Bronx, NY. Additional sequencing primers were applied to “walk” across each overlapping fragment. The complete genomes were assembled through mapping the obtained electropherogram to the HPV52 reference sequence (GenBank ID 397038).

#### 4.7. Construction of Phylogentic Tree

The nucleotide sequences of the complete genomes were aligned to the reference sequence using the program MAFFT v6.846 [70]. Maximum likelihood (ML) trees were constructed using RAxML v7.2.8.27 [71] with optimized parameters on the basis of the aligned complete sequences linearized at the first “A” of the start codon of the E1 ORF. Bayesian trees were constructed using MrBayes v3.1.2 with 10,000,000 cycles for the Markov chain Monte Carlo (MCMC) algorithm [72]. Single-nucleotide polymorphisms (SNPs) within the HPV52 lineages/sublineages were identified from the global sequence alignments using MEGA5 v5 [73], which was used to calculate mean nucleotide differences and standard errors within and between sublineages. The p-distance method in MEGA5 v5 was used to calculate the genome–genome pairwise differences from the above alignments.

#### 4.8. Statistical Analyses

For both pyrosequencing and NGS results, median methylation values for each CpG site were compared between cases and controls using the nonparametric Mann–Whitney *U* test. OR values were computed using methylation levels stratified at the highest tertile in controls, and the AUC estimates were calculated for individual CpG sites. Methylation values obtained by pyrosequencing and NGS were plotted and evaluated using a linear regression model. R squared values to evaluate the correlation between pyrosequencing and NGS for each CpG site were calculated in Microsoft Excel 2008.

Differences in nucleotides at SNP positions and methylation at CpG sites among cases and controls were examined. Logistic regression was performed on identified SNP

positions comparing HPV52-positive subjects with CIN3 to the HPV52-positive control subjects (<CIN2). Only SNP positions with >2% variant frequency were analyzed. The reference group at each position was determined by setting the most prevalent nucleotide as the reference. The ORs and 95% CI were calculated, and results were plotted as OR values (see Figure 3). Differences in HPV52 lineage and CIN3 were examined using logistic regression models, comparing A lineage to B lineage and within a specific lineage (i.e., A1 vs. A2, A2 vs. B2, etc.). Other lineages were not included in these analyses due to small numbers of samples. All analyses were performed using SAS v9.3 (SAS Institute, Cary, NC, USA) with a  $p$ -value < 0.05 as statistically significant.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijms22126463/s1>.

**Author Contributions:** Conceptualization, M.S., P.E.C., N.W. and R.D.B.; methodology, K.J.B., A.G., Z.C., A.H. and R.D.B.; bioinformatics, Z.C., A.G. and A.H.; analysis, K.J.B., A.G., M.C. and R.D.B.; data curation, T.R.-B.; writing—original draft preparation, K.J.B. and A.G.; writing—review and editing, A.G., Z.C., P.E.C., M.S., N.W. and R.D.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported in part by the National Cancer Institute (grant number CA78527); the Einstein Cancer Research Center (grant number P30CA013330) from the National Cancer Institute; and the Einstein-Montefiore Center for AIDS funded by the NIH (grant number AI-51519) and by the Intramural Research Program of the NIH, NCI, and by Kaiser Permanente Northern California (KPNC). K.J.B. was supported by a grant titled The Bronx Einstein Training in Teaching and Research (BETTR) (grant number GM102779) from the NIH.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the KAISER FOUNDATION RESEARCH INSTITUTE and the NATIONAL CANCER INSTITUTE (IRB# 00001045, continuing review approved 24 February 2021).

**Informed Consent Statement:** Participants were included in the study unless they chose the “opt-out” option as described in the Material and Methods section.

**Data Availability Statement:** The data presented in this study are openly available in GenBank, accession numbers: MZ374362-MZ374448.

**Acknowledgments:** The authors would like to thank Barbara Birshtein for her critical review of the manuscript.

**Conflicts of Interest:** Castle has received HPV tests and assays for research at a reduced or no cost from Roche, Becton Dickinson, Cepheid, and Arbor Vita Corporation. Other authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

**Disclaimer:** The opinions expressed by the authors are their own and this material should not be interpreted as representing the official viewpoint of the U.S. Department Health and Human Services, the National Institutes of Health, or the National Cancer Institute.

## Abbreviations

HPV	Human papillomavirus
HR-HPV	High-risk human papillomavirus
Alpha9	<i>Alphapapillomavirus</i> species 9 (HPV16-related)
Alpha7	<i>Alphapapillomavirus</i> species 7 (HPV18-related)
CIN3	Cervical intraepithelial neoplasia grade 3
CpG site	CG dinucleotide
AIS	Adenocarcinoma in situ
SCC	Squamous cell cancer
URR	Upstream regulatory region
OR	Odds ratio

CI	Confidence interval
AUC	Area under the curve
SNP	Single nucleotide polymorphism
NGS	Next-generation sequencing
NCR1	Noncoding region 1
VWAS	Viral genome wide association study
PaP cohort	HPV Persistence and Progression
KPNC	Kaiser Permanente Northern California
NCI	National Cancer Institute
HC2	Hybrid capture 2
NILM	Negative for intraepithelial lesions or malignancy
bp	Base pair
ML	Maximum likelihood
MCMC	Markov chain Monte Carlo

## References

- Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. Global cancer statistics, 2012. *CA Cancer J. Clin.* **2015**, *65*, 87–108. [\[CrossRef\]](#)
- Arbyn, M.; Weiderpass, E.; Bruni, L.; de Sanjose, S.; Saraiya, M.; Ferlay, J.; Bray, F. Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis. *Lancet Glob. Health* **2020**, *8*, e191–e203. [\[CrossRef\]](#)
- Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359–E386. [\[CrossRef\]](#)
- Bouvard, V.; Baan, R.; Straif, K.; Grosse, Y.; Secretan, B.; El Ghissassi, F.; Benbrahim-Tallaa, L.; Guha, N.; Freeman, C.; Galichet, L.; et al. A review of human carcinogens—Part B: Biological agents. *Lancet Oncol.* **2009**, *10*, 321–322. [\[CrossRef\]](#)
- Smith, J.S.; Lindsay, L.; Hoots, B.; Keys, J.; Franceschi, S.; Winer, R.; Clifford, G.M. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: A meta-analysis update. *Int. J. Cancer* **2007**, *121*, 621–632. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, N.; Franceschi, S.; Howell-Jones, R.; Snijders, P.J.; Clifford, G.M. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int. J. Cancer* **2011**, *128*, 927–935. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lorenzi, A.T.; Fregnani, J.H.; Villa, L.L.; Sichero, L.; Nunes, E.M.; Longatto-Filho, A. Diversity of human papillomavirus typing among women population living in rural and remote areas of Brazilian territory. *Papillomavirus Res.* **2019**, *8*, 100186. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dong, L.; Li, T.; Li, L.; Wang, M.Z.; Wu, Z.; Cui, J.; Liu, B.; Zhang, X.; Qiao, Y.; Chen, W. Clustering patterns of type-type combination in multiple genotypes infections of human papillomavirus in cervical adenocarcinoma. *J. Med. Virol.* **2019**, *91*, 2001–2008. [\[CrossRef\]](#)
- Toh, Z.Q.; Kosasih, J.; Russell, F.M.; Garland, S.M.; Mulholland, E.K.; Licciardi, P.V. Recombinant human papillomavirus nonavalent vaccine in the prevention of cancers caused by human papillomavirus. *Infect. Drug Resist.* **2019**, *12*, 1951–1967. [\[CrossRef\]](#)
- Luo, G.; Sun, X.; Li, M.; Liu, T.; Hu, G.; He, Y.; Mao, L.; Yan, L.; Xie, L.; Zou, H.; et al. Cervical human papillomavirus among women in Guangdong, China 2008-2017: Implication for screening and vaccination. *J. Med. Virol.* **2019**, *91*, 1856–1865. [\[CrossRef\]](#)
- Pity, I.S.; Abdo, H.M.; Goreal, A.A. Human Papillomavirus Genotyping among Different Cervical Smears in Duhok/Iraq. *Asian Pac. J. Cancer Prev.* **2019**, *20*, 2059–2064. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vyankandondera, J.; Wambua, S.; Irungu, E.; Mandaliya, K.; Temmerman, M.; Ryan, C.; Mohamed, Y.; Vanden Broeck, D.; Verhelst, R.; Chersich, M.F.; et al. Type-Specific Human Papillomavirus Prevalence, Incident Cases, Persistence, and Associated Pregnancy Outcomes Among HIV-Infected Women in Kenya. *Sex Transm. Dis.* **2019**, *46*, 532–539. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, J.; Tang, D.; Wang, K.; Wang, J.; Zhang, Z.; Chen, Y.; Zhang, X.; Ma, C. HPV genotype prevalence and distribution during 2009–2018 in Xinjiang, China: Baseline surveys prior to mass HPV vaccination. *BMC Womens Health* **2019**, *19*, 90. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhu, B.; Liu, Y.; Zuo, T.; Cui, X.; Li, M.; Zhang, J.; Yu, H.; Piao, H. The prevalence, trends, and geographical distribution of human papillomavirus infection in China: The pooled analysis of 1.7 million women. *Cancer Med.* **2019**, *8*, 5373–5385. [\[CrossRef\]](#)
- Li, S.; Ye, M.; Chen, Y.; Gong, Q.; Mei, B. Genetic variation of E6 and E7 genes of human papillomavirus 52 from Central China. *J. Med. Virol.* **2021**, *93*, 3849–3856. [\[CrossRef\]](#) [\[PubMed\]](#)
- McHome, B.L.; Kjaer, S.K.; Manongi, R.; Swai, P.; Waldstroem, M.; Iftner, T.; Wu, C.; Mwaiselage, J.; Rasch, V. HPV types, cervical high-grade lesions and risk factors for oncogenic human papillomavirus infection among 3416 Tanzanian women. *Sex Transm. Infect.* **2021**, *97*, 56–62. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sweet, K.; Bosire, C.; Sanusi, B.; Sherrod, C.J.; Kwatampora, J.; Waweru, W.; Mugo, N.; Kimani, J.; Ting, J.; Clark, J.; et al. Prevalence, incidence, and distribution of human papillomavirus types in female sex workers in Kenya. *Int. J. STD AIDS* **2020**, *31*, 109–118. [\[CrossRef\]](#)

18. Tounkara, F.K.; Teguede, I.; Guedou, F.A.; Goma-Matsetse, E.; Kone, A.; Behanzin, L.; Traore, S.; Aza-Gnandji, M.; Keita, B.; Guenoun, J.; et al. Human papillomavirus genotype distribution and factors associated among female sex workers in West Africa. *PLoS ONE* **2020**, *15*, e0242711. [[CrossRef](#)]
19. Wu, W.; Song, L.; Yang, Y.; Wang, J.; Liu, H.; Zhang, L. Exploring the dynamics and interplay of human papillomavirus and cervical tumorigenesis by integrating biological data into a mathematical model. *BMC Bioinform.* **2020**, *21*, 152. [[CrossRef](#)]
20. Murao, K.; Tetsutani, M.; Ishigami, T.; Kubo, Y.; Arase, S. Bowen disease of the palm associated with human papillomavirus 52. *Clin. Exp. Dermatol.* **2013**, *38*, 489–491. [[CrossRef](#)]
21. Martinez-Bailon, C.; Mantilla-Morales, A.; Mendez-Matias, G.; Alvarado-Cabrero, I.; Maldonado-Rodriguez, R.; Quintero-Becerra, J.; Arias-Flores, R.; Pina-Sanchez, P. Human papillomavirus genotypes and P16INK4A expression in squamous penile carcinoma in Mexican patients. *BMC Infect. Dis.* **2019**, *19*, 1068. [[CrossRef](#)]
22. Schiffman, M.; Herrero, R.; Desalle, R.; Hildesheim, A.; Wacholder, S.; Rodriguez, A.C.; Bratti, M.C.; Sherman, M.E.; Morales, J.; Guillen, D.; et al. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* **2005**, *337*, 76–84. [[CrossRef](#)]
23. Bernard, H.U.; Burk, R.D.; Chen, Z.; van Doorslaer, K.; Hausen, H.; de Villiers, E.M. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **2010**, *401*, 70–79. [[CrossRef](#)]
24. Burk, R.D.; Harari, A.; Chen, Z. Human papillomavirus genome variants. *Virology* **2013**, *445*, 232–243. [[CrossRef](#)] [[PubMed](#)]
25. Ho, G.Y.; Bierman, R.; Beardsley, L.; Chang, C.J.; Burk, R.D. Natural history of cervicovaginal papillomavirus infection in young women. *N. Engl. J. Med.* **1998**, *338*, 423–428. [[CrossRef](#)] [[PubMed](#)]
26. Schiffman, M.; Wentzensen, N. Human papillomavirus infection and the multistage carcinogenesis of cervical cancer. *Cancer Epidemiol. Biomark. Prev.* **2013**, *22*, 553–560. [[CrossRef](#)] [[PubMed](#)]
27. McCredie, M.R.; Sharples, K.J.; Paul, C.; Baranyai, J.; Medley, G.; Jones, R.W.; Skegg, D.C. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: A retrospective cohort study. *Lancet Oncol.* **2008**, *9*, 425–434. [[CrossRef](#)]
28. Fontham, E.T.H.; Wolf, A.M.D.; Church, T.R.; Etzioni, R.; Flowers, C.R.; Herzig, A.; Guerra, C.E.; Oeffinger, K.C.; Shih, Y.T.; Walter, L.C.; et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. *CA Cancer J. Clin.* **2020**, *70*, 321–346. [[CrossRef](#)]
29. Wentzensen, N.; Schiffman, M.; Palmer, T.; Arbyn, M. Triage of HPV positive women in cervical cancer screening. *J. Clin. Virol.* **2016**, *76* (Suppl. 1), S49–S55. [[CrossRef](#)] [[PubMed](#)]
30. Dasari, S.; Wudayagiri, R.; Valluru, L. Cervical cancer: Biomarkers for diagnosis and treatment. *Clin. Chim. Acta* **2015**, *445*, 7–11. [[CrossRef](#)]
31. Clarke, M.A.; Wentzensen, N.; Mirabello, L.; Ghosh, A.; Wacholder, S.; Harari, A.; Lorincz, A.; Schiffman, M.; Burk, R.D. Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiol. Biomark. Prev.* **2012**, *21*, 2125–2137. [[CrossRef](#)] [[PubMed](#)]
32. Frimer, M.; Sun, C.; McAndrew, T.; Smith, B.; Harari, A.; Chen, Z.; Mirabello, L.; Wentzensen, N.; Goldberg, G.L.; Rodriguez, A.C.; et al. HPV16 CpG methyl-haplotypes are associated with cervix precancer and cancer in the Guanacaste natural history study. *Gynecol. Oncol.* **2015**, *138*, 94–100. [[CrossRef](#)]
33. Mirabello, L.; Frimer, M.; Harari, A.; McAndrew, T.; Smith, B.; Chen, Z.; Wentzensen, N.; Wacholder, S.; Castle, P.E.; Raine-Bennett, T.; et al. HPV16 methyl-haplotypes determined by a novel next-generation sequencing method are associated with cervical precancer. *Int. J. Cancer* **2015**, *136*, E146–E153. [[CrossRef](#)] [[PubMed](#)]
34. Lorincz, A.T.; Brentnall, A.R.; Scibior-Bentkowska, D.; Reuter, C.; Banwait, R.; Cadman, L.; Austin, J.; Cuzick, J.; Vasiljevic, N. Validation of a DNA methylation HPV triage classifier in a screening sample. *Int. J. Cancer* **2016**, *138*, 2745–2751. [[CrossRef](#)]
35. Wentzensen, N.; Sun, C.; Ghosh, A.; Kinney, W.; Mirabello, L.; Wacholder, S.; Shaber, R.; LaMere, B.; Clarke, M.; Lorincz, A.T.; et al. Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J. Natl. Cancer Inst.* **2012**, *104*, 1738–1749. [[CrossRef](#)]
36. Chen, Z.; Schiffman, M.; Herrero, R.; Desalle, R.; Anastos, K.; Segondy, M.; Sahasrabudde, V.V.; Gravitt, P.E.; Hsing, A.W.; Burk, R.D. Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. *PLoS ONE* **2011**, *6*, e20183. [[CrossRef](#)]
37. Clarke, M.A.; Gradissimo, A.; Schiffman, M.; Lam, J.; Sollecito, C.C.; Fetterman, B.; Lorey, T.; Poitras, N.; Raine-Bennett, T.R.; Castle, P.E.; et al. Human Papillomavirus DNA Methylation as a Biomarker for Cervical Precancer: Consistency across 12 Genotypes and Potential Impact on Management of HPV-Positive Women. *Clin. Cancer Res.* **2018**, *24*, 2194–2202. [[CrossRef](#)]
38. Smith, B.; Chen, Z.; Reimers, L.; van Doorslaer, K.; Schiffman, M.; Desalle, R.; Herrero, R.; Yu, K.; Wacholder, S.; Wang, T.; et al. Sequence imputation of HPV16 genomes for genetic association studies. *PLoS ONE* **2011**, *6*, e21375. [[CrossRef](#)]
39. Mirabello, L.; Yeager, M.; Yu, K.; Clifford, G.M.; Xiao, Y.; Zhu, B.; Cullen, M.; Boland, J.F.; Wentzensen, N.; Nelson, C.W.; et al. HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell* **2017**, *170*, 1164–1174 e1166. [[CrossRef](#)] [[PubMed](#)]
40. Pinheiro, M.; Gage, J.C.; Clifford, G.M.; Demarco, M.; Cheung, L.C.; Chen, Z.; Yeager, M.; Cullen, M.; Boland, J.F.; Chen, X.; et al. Association of HPV35 with cervical carcinogenesis among women of African ancestry: Evidence of viral-host interaction with implications for disease intervention. *Int. J. Cancer* **2020**, *147*, 2677–2686. [[CrossRef](#)] [[PubMed](#)]
41. Brentnall, A.R.; Vasiljevic, N.; Scibior-Bentkowska, D.; Cadman, L.; Austin, J.; Cuzick, J.; Lorincz, A.T. HPV33 DNA methylation measurement improves cervical pre-cancer risk estimation of an HPV16, HPV18, HPV31 and EPB41L3 methylation classifier. *Cancer Biomark.* **2015**, *15*, 669–675. [[CrossRef](#)]

42. Louvanto, K.; Franco, E.L.; Ramanakumar, A.V.; Vasiljevic, N.; Scibior-Bentkowska, D.; Koushik, A.; Cuzick, J.; Coutlee, F.; Lorincz, A.T.; Biomarkers of Cervical Cancer Risk Study Team. Methylation of viral and host genes and severity of cervical lesions associated with human papillomavirus type 16. *Int. J. Cancer* **2015**, *136*, E638–E645. [[CrossRef](#)]
43. Ramirez, A.T.; Sanchez, G.I.; Nedjai, B.; Agudelo, M.C.; Brentnall, A.R.; Cuschieri, K.; Castaneda, K.M.; Cuzick, J.; Lorincz, A.T.; Group, A.-U.-C.T. Effective methylation triage of HPV positive women with abnormal cytology in a middle-income country. *Int. J. Cancer* **2021**, *148*, 1383–1393. [[CrossRef](#)] [[PubMed](#)]
44. Gillio-Tos, A.; Fiano, V.; Grasso, C.; Trevisan, M.; Gori, S.; Mongia, A.; De Marco, L.; Ronco, G.; New Technologies for Cervical Cancer Screening Working Group. Assessment of viral methylation levels for high risk HPV types by newly designed consensus primers PCR and pyrosequencing. *PLoS ONE* **2018**, *13*, e0194619. [[CrossRef](#)] [[PubMed](#)]
45. von Knebel Doeberitz, M.; Prigge, E.S. Role of DNA methylation in HPV associated lesions. *Papillomavirus Res.* **2019**, *7*, 180–183. [[CrossRef](#)] [[PubMed](#)]
46. Murakami, I.; Fujii, T.; Dan, K.; Saito, M.; Ohno, A.; Iwata, T.; Aoki, D. Methylation of human papillomavirus-52 and -58 is a candidate biomarker in cervical neoplasia. *J. Clin. Virol.* **2013**, *58*, 149–154. [[CrossRef](#)] [[PubMed](#)]
47. Hsu, Y.W.; Huang, R.L.; Su, P.H.; Chen, Y.C.; Wang, H.C.; Liao, C.C.; Lai, H.C. Genotype-specific methylation of HPV in cervical intraepithelial neoplasia. *J. Gynecol. Oncol.* **2017**, *28*, e56. [[CrossRef](#)] [[PubMed](#)]
48. Gradissimo, A.; Lam, J.; Attonito, J.D.; Palefsky, J.; Massad, L.S.; Xie, X.; Eltoun, I.E.; Rahangdale, L.; Fischl, M.A.; Anastos, K.; et al. Methylation of High-Risk Human Papillomavirus Genomes Are Associated with Cervical Precancer in HIV-Positive Women. *Cancer Epidemiol. Biomark. Prev.* **2018**, *27*, 1407–1415. [[CrossRef](#)]
49. Aho, J.; Hankins, C.; Tremblay, C.; Lang, F.; Forest, P.; Pourreaux, K.; Rouah, F.; Coutlee, F.; The Canadian Women's HIV Study Group. Molecular analysis of human papillomavirus type 52 isolates detected in the genital tract of human immunodeficiency virus-seropositive and -seronegative women. *J. Infect. Dis.* **2003**, *188*, 1517–1527. [[CrossRef](#)]
50. Van Doorslaer, K.; Burk, R.D. Evolution of human papillomavirus carcinogenicity. *Adv. Virus Res.* **2010**, *77*, 41–62. [[CrossRef](#)]
51. Sun, Z.; Lu, Z.; Liu, J.; Wang, G.; Zhou, W.; Yang, L.; Liu, C.; Wang, B.; Ruan, Q. Genetic variations of E6 and long control region of human papillomavirus type 16 from patients with cervical lesion in Liaoning, China. *BMC Cancer* **2013**, *13*, 459. [[CrossRef](#)]
52. Choi, Y.J.; Ki, E.Y.; Zhang, C.; Ho, W.C.; Lee, S.J.; Jeong, M.J.; Chan, P.K.; Park, J.S. Analysis of Sequence Variation and Risk Association of Human Papillomavirus 52 Variants Circulating in Korea. *PLoS ONE* **2016**, *11*, e0168178. [[CrossRef](#)]
53. Chang, Y.J.; Chen, H.C.; Lee, B.H.; You, S.L.; Lin, C.Y.; Pan, M.H.; Chou, Y.C.; Hsieh, C.Y.; Chen, Y.M.; Cheng, Y.J.; et al. Unique variants of human papillomavirus genotypes 52 and 58 and risk of cervical neoplasia. *Int. J. Cancer* **2011**, *129*, 965–973. [[CrossRef](#)]
54. Sun, Z.; Lu, Z.; Liu, J.; Wang, G.; Zhou, W.; Yang, L.; Liu, C.; Ruan, Q. Genomic polymorphism of human papillomavirus type 52 in women from Northeast China. *Int. J. Mol. Sci.* **2012**, *13*, 14962–14972. [[CrossRef](#)] [[PubMed](#)]
55. Formentin, A.; Archambault, J.; Koushik, A.; Richardson, H.; Brassard, P.; Franco, E.L.; Coutlee, F. Human papillomavirus type 52 polymorphism and high-grade lesions of the uterine cervix. *Int. J. Cancer* **2013**, *132*, 1821–1830. [[CrossRef](#)] [[PubMed](#)]
56. Ishizaki, A.; Matsushita, K.; Hoang, H.T.; Agdamag, D.M.; Nguyen, C.H.; Tran, V.T.; Sasagawa, T.; Saikawa, K.; Lihana, R.; Pham, H.V.; et al. E6 and E7 variants of human papillomavirus-16 and -52 in Japan, the Philippines, and Vietnam. *J. Med. Virol.* **2013**, *85*, 1069–1076. [[CrossRef](#)] [[PubMed](#)]
57. Ding, X.; Liu, Z.; Su, J.; Yan, D.; Sun, W.; Zeng, Z. Human papillomavirus type-specific prevalence in women referred for colposcopic examination in Beijing. *J. Med. Virol.* **2014**, *86*, 1937–1943. [[CrossRef](#)]
58. Zhang, C.; Park, J.S.; Grce, M.; Hibbitts, S.; Palefsky, J.M.; Konno, R.; Smith-McCune, K.K.; Giovannelli, L.; Chu, T.Y.; Picconi, M.A.; et al. Geographical distribution and risk association of human papillomavirus genotype 52-variant lineages. *J. Infect. Dis.* **2014**, *210*, 1600–1604. [[CrossRef](#)]
59. Tenjimbayashi, Y.; Onuki, M.; Hirose, Y.; Mori, S.; Ishii, Y.; Takeuchi, T.; Tasaka, N.; Satoh, T.; Morisada, T.; Iwata, T.; et al. Whole-genome analysis of human papillomavirus genotypes 52 and 58 isolated from Japanese women with cervical intraepithelial neoplasia and invasive cervical cancer. *Infect. Agent Cancer* **2017**, *12*, 44. [[CrossRef](#)] [[PubMed](#)]
60. Frati, E.R.; Bianchi, S.; Amendola, A.; Colzani, D.; Petrelli, F.; Zehender, G.; Tanzi, E. Genetic characterization of variants of HPV16, HPV18 and HPV52 circulating in Italy among general and highrisk populations. *Mol. Med. Rep.* **2020**, *21*, 894–902. [[CrossRef](#)] [[PubMed](#)]
61. Raiol, T.; Wyant, P.S.; de Amorim, R.M.; Cerqueira, D.M.; Milanezi, N.; Brigido Mde, M.; Sichero, L.; Martins, C.R. Genetic variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58 in central Brazil. *J. Med. Virol.* **2009**, *81*, 685–692. [[CrossRef](#)] [[PubMed](#)]
62. Rosl, F.; Arab, A.; Klevenz, B.; zur Hausen, H. The effect of DNA methylation on gene regulation of human papillomaviruses. *J. Gen. Virol.* **1993**, *74*, 791–801. [[CrossRef](#)]
63. Doorbar, J.; Egawa, N.; Griffin, H.; Kranjec, C.; Murakami, I. Human papillomavirus molecular biology and disease association. *Rev. Med. Virol.* **2015**, *25* (Suppl. 1), 2–23. [[CrossRef](#)]
64. Castle, P.E.; Shaber, R.; LaMere, B.J.; Kinney, W.; Fetterma, B.; Poitras, N.; Lorey, T.; Schiffman, M.; Dunne, A.; Ostolaza, J.M.; et al. Human Papillomavirus (HPV) Genotypes in Women with Cervical Precancer and Cancer at Kaiser Permanente Northern California. *Cancer Epidemiol. Biomark. Prev.* **2011**, *20*, 946–953. [[CrossRef](#)]
65. Castle, P.E.; Schiffman, M.; Gravitt, P.E.; Kendall, H.; Fishman, S.; Dong, H.; Hildesheim, A.; Herrero, R.; Bratti, M.C.; Sherman, M.E.; et al. Comparisons of HPV DNA detection by MY09/11 PCR methods. *J. Med. Virol.* **2002**, *68*, 417–423. [[CrossRef](#)] [[PubMed](#)]

66. Li, L.C.; Dahiya, R. MethPrimer: Designing primers for methylation PCRs. *Bioinformatics* **2002**, *18*, 1427–1431. [[CrossRef](#)] [[PubMed](#)]
67. Hamady, M.; Walker, J.J.; Harris, J.K.; Gold, N.J.; Knight, R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* **2008**, *5*, 235–237. [[CrossRef](#)]
68. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [[CrossRef](#)]
69. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, *27*, 1571–1572. [[CrossRef](#)]
70. Katoh, K.; Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **2010**, *26*, 1899–1900. [[CrossRef](#)]
71. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)] [[PubMed](#)]
72. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)] [[PubMed](#)]
73. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]