Human papillomavirus 16 positive cervical cancer in Guatemala: The D2 and D3 sublineages differ in integration rate and age of diagnosis

Hong Lou[1*], Joseph F. Boland[1*], Edmundo Torres-Gonzalez[2*], Anaseidy Albanez[3*], Weiyin Zhou[1], Mia Steinberg[1], Lena Diaw[2], Jason Mitchell[1], David Roberson[1], Michael Cullen[1], Lisa Garland[1], Sara Bass[1], Robert Burk[5] Meredith Yeager[1], Nicolas Wentzensen[4], Mark Schiffman[4], Enrique Alvirez[6], Eduardo Gharzouzi[7], Lisa Mirabello[4], Michael Dean[2#]

[1]Cancer Genetics Research Laboratory, Division of Cancer Epidemiology and Genetics, Leidos Biomedical Research Inc.; National Laboratory for Cancer Research, Gaithersburg, MD USA

[2]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD USA

[3]Instituto de Cancerologia, Guatemala City, Guatemala

[4]Laboratory of Cancer Genetics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD USA

[5]Departments of Pediatrics, Microbiology and Immunology, and Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY, USA

[6]Hospital Central Universitario "Dr. Antonio M Pineda", Barquisimeto, Lara State, Venezuela

[7]Integra Cancer Institute, Guatemala City, Guatemala

*Contributed equally

#Correspondence to Michael Dean, Ph.D., ATC, Room 134D, Laboratory of Translational Genomics, National Cancer Institute, Gaithersburg, MD 20877 USA.
Tel: 240-760-6484, FAX: 301-402-3134 (email: deanm@mail.nih.gov)

Running Title: HPV16 integration in cervical cancer
Key Words: HPV16, integration, cervical cancer, PIK3CA gene, human papillomavirus, IRF4 gene

1

The authors have no conflicts of interest to report.

Word count: 2733 words, 3 tables, 4 figures

Abstract

Human papillomavirus (HPV) 16 displays substantial sequence variation; four HPV16 lineages (A, B, C, D) have been described, as well as multiple sub-lineages. To identify molecular events associated with HPV16 carcinogenesis we evaluated viral variation, the integration of HPV16, and somatic mutation in 96 cervical cancer samples from Guatemala. A total of 64% (60/94) of the samples had integrated HPV16 sequences, and integration was associated with an earlier age of diagnosis (P=0.0007) and pre-menopausal disease. HPV16 integration sites were broadly distributed in the genome but in one tumor, HPV16 integrated into the promoter of the interferon regulatory factor 4 (*IRF4*) gene, which plays an important role in the regulation of the interferon response to viral infection. The HPV16 D2 and D3 sub-lineages were found in 23% and 30% of the tumors, respectively and were significantly associated with adenocarcinoma. D2-positive tumors had a higher rate of integration (P=0.011), earlier age of diagnosis (P=0.012), and a lower rate of somatic mutation (P=0.03). Whereas D3-positive tumors are less likely to integrate, have later age-of- diagnosis, and a higher rate of somatic mutation. In conclusion, Guatemalan cervical tumors have a high frequency of the very high-risk HPV16 D2 and D3 sub-lineages and cervical cancer patients with these variants of HPV16 differ in histology, age of- diagnosis, integration, and somatic mutation frequency. In summary, related lineages of HPV16 have different features of oncogenicity.

Introduction

High-risk human papillomavirus (HR-HPV) infection is responsible for nearly all cervical cancer (1). However, for unknown reasons, only <5% of these common HPV infections will progress to cervical precancer, which could lead to invasive cervical cancer (2). Other factors such as viral integration, viral variation, somatic mutations, genetic susceptibility, and host-virus interactions contribute to carcinogenesis (3-7). There are thirteen HPV types considered carcinogenic by IARC, with the most frequent being HPV16 (8). HPV16 is found in over 50% of the cervical cancers worldwide but displays substantial regional variation (9,10). HPV16 can be classified into four evolutionary-defined lineages (A, B, C, D), and 16 sublineages (A1-4, B1-4, C1-4, D1-4) (11), that vary in geographic distribution and disease risks (12-15). HPV16 sublineages differ by 0.5-1% of the viral genome, and the A4, D2 and D3 sublineages are associated with 9-28-fold higher risk of cancer and 10-130-fold risk of adenocarcinoma (15).

The oncogenicity of HPV is largely due to the E6 and E7 viral oncoproteins, essential to the life cycle of the virus (16). During an HPV infection, several transcripts are produced including an unspliced transcript producing E6 protein and an E6* spliced transcript generating the viral oncoprotein E7 (17,18). Recent studies show that HPV16 shifts to express predominantly the spliced, E7-producing transcript upon integration and that the E7 sequence is very highly conserved in cancers (3,7).

HPV DNA integration occurs in many sites in the human genome, with some preference for transcribed regions and common fragile sites (19-22). Integration into the genome of cancer cells can involve the deletion of all or portions of the viral E2 and E1 genes leading to deregulated expression of E6 and E7 viral oncogenes, which accelerates cell cycle entry, apoptosis, chromosome instability, and evasion of apoptosis (23). Inactivation of cellular tumor suppressor genes or activation of proto-oncogenes is rarely a direct consequence of HPV DNA integration (3,4,24-26); however, integration can induce chromosome instability (26,27). The rate of HPV integration is variable by HPV type, ranging from 55 to 95% of cervical tumors, with the highest rates observed for HPV18 and 45 (3,28). Integration is not required for HPV16

3

to cause invasive cancer, and episomal-only tumors are found as well as tumors with both integrated and episomal HPV16 sequences (29).

Our previous study demonstrated that up to 33% of cervical cancers from Latin America carry *PIK3CA* mutations, mainly at E542K and E545K, two specific hot spots in the helical domain of the *PIK3CA* gene (30). These mutations are part of a spectrum of mutations caused by the APOBEC family of enzymes activated by viral infection (31,32). Also, there are at least 14 cellular genes frequently mutated in cervical tumors, including *EP300*, *PTEN*, *HLA-B*, *KRAS*, and *FBXW7* (3,4,30,31,33).

The goal of this study was to further dissect the molecular events involved in cervical cancer in HPV16-positive tumors. Using whole virus sequencing, viral DNA capture' and integration analysis, and somatic sequencing we have characterized the integration status of the major HPV16 sublineages.

Materials and Methods

Sample Collection

Ninety-six HPV16-positive samples from Guatemala were selected from a total of 298 cervical tumors collected between 2011 and 2013 at the Instituto de Cancerologia (INCAN), the central reference hospital for adult cancer therapy in the country. Selected samples were confirmed HPV16 positive with at least 1 ug of tumor DNA for sequence capture. All patients, between 25- and 85-year-old, presenting for confirmation of the diagnosis of cervical cancer were invited to participate. Written informed consent was obtained from the patients. Protocols were approved by the Ethical Review Committees of the local hospitals in Guatemala City, Guatemala, and the Office of Human Research Subjects, NIH, in accordance with the Belmont Report. Cervical tumor biopsy tissues were immediately stored in RNAlater (QIAGEN) at -20°C until extraction; blood samples were stored at -20°C and a questionnaire on reproductive history, smoking, oral contraceptive use, and family cancer history was administered.

DNA and RNA Extraction, and HPV Genotyping

DNA and RNA were extracted from the cervical cancer tissues (5-10 mg) using the AllPrep DNA/RNA Micro Kit (QIAGEN) as described by the manufacturer. HPV type was determined by touch-down PCR using degenerate Broad-Spectrum (BS) GP5+ and GP6+ primer amplification and Sanger sequencing (Schmitt M. et al. J Clin Microbiol. 2008; 46(3):1050–9). The detailed method for the determination of HPV genotyping is described (30).

Quantitative RT-real-time PCR

One microgram of total RNA was reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche) with the oligonucleotide $(dT)_{18}$ primer according to the manufacturer's instructions (30,34). Real-time PCR was performed for E6 and E7 HPV transcripts, using gene and type-specific primers, in the presence of SYBR green. Relative expression levels of E6 and E7 were normalized using the *β-actin* gene as described (35). The E6 unspliced/spliced ratio was determined using the following formula: E6 mRNA value (unspliced) divided by (E6*/E7 mRNA value minus E6 mRNA value).

Determination of *PIK3CA* and other somatic mutations

Somatic mutations were identified using exome and ultra-deep targeted sequencing as described in (30).

HPV DNA Capture using the NimbleGen Array

DNA preparation: 200 ng genomic DNA was purified using Agencourt AMPure XP Reagent (Beckman Coulter, Inc.) according to the manufacturer's protocol. HPV library preparation: full-length HPV genomes were amplified pre-hybridization by ligation-mediated PCR consisting of one reaction containing 20 µl library DNA, 25 µl 2x KAPA HiFi HotStart ReadyMix, and 5 µl 10x Library Amplification Primer Mix. PCR was subjected to the following cycling conditions: 98°C for 45 s, followed by 5 cycles of 98°C for 15 s, 60°C for 30s, 72°C for 30 s. The last step was an extension at 72°C for 1 minute. The amplicons were purified by Agencourt AMPure XP Reagent (Beckman Coulter, Inc.) according to the KAPA provided protocol. Liquid Phase Sequence Capture: HPV DNA capture was performed with NimbleGen's SeqCap EZ Custom Library containing probes for all HR HPV types (Roche NimbleGen, Inc.,) according to the manufacturer's protocol. The DNA mixture was hybridized at 47°C for 64 to 72 hours and washing and recovery of captured DNA were performed as described in NimbleGen SeqCap EZ Library SR Protocol. The captured DNA was amplified by ligation mediated PCR (Beckman Coulter, Inc.) according to NimbleGen SeqCap EZ Library SR Protocol. The captured DNA was quantified via Kapa's Library Quantification Kit for Ion Torrent (Kapa Biosystems) on the LightCycler 480 (Roche, Indianapolis, IN, USA). All sequencing was performed using an Ion Torrent Proton Sequencer using. The sequence captured both episomal and integrated HPV16 DNA. The integrated reads map to both human sequence and HPV16 sequence and determine the breakpoint for HPV integration.

HPV16 genome sequencing and sub-lineage determination

We used a custom Thermo Fisher Ion Torrent AmpliSeq HPV16 panel approach to amplify the entire 7906 bp HPV16 genome as previously described (36). HPV16 variant lineage assignment was based on the maximum likelihood tree topology constructed in MEGA, including 16 HPV16 variant sublineage reference sequences (10), and lineage assignments were confirmed with SNP patterns. The detailed methodology was described previously (15).

6

Confirmation of IRF4 integration by PCR, cloning, and sequencing

HPV16 integration into the *IRF4* gene was confirmed by PCR using GoTaq Long PCR Master Mix kit (Promega U.S.), with 5'-AAGCATGTCAGACACGCAGA- 3'as a forward primer, and 5'- GGAAGACGGAGGAATGGTCC- 3'as a reverse primer. DNA samples from blood and tumor tissues were assessed; normal donor blood DNA was used as a negative control (Figure S1A).

Statistical Analysis

Unpaired T-test, Pearson's chi-squared test, and Fisher's exact test statistical analyses were performed using GraphPad Prism version 7 for Windows; $P<0.05$ was regarded as significant.

Results

Analysis of HPV16 sequence and integration

To determine the genomic characteristics of HPV16 positive tumor samples from an understudied population, we selected 94 HPV positive tumor samples from 298 subjects in an ongoing hospital-based case study of advanced cervical cancer (94% stage IIB or higher) in Guatemala. The selection criteria were HPV16 positive by amplification and sequencing of a region of the L1 gene, and adequate DNA (1 ug) for sequence capture. Both squamous cell carcinoma (SCC) and adenocarcinoma cases were included. HPV DNA was captured and reads mapped to both the human and HPV16 genome and overlapping amplicon sequencing of the virus was also performed (30,36). Complete HPV16 sequences were aligned, and phylogenetic analyses were used to determine HPV16 sublineages (15). HPV integration sites were determined from HPV-human junction reads based on captured HPV DNA and targeted human gene sequencing was previously reported (30). Samples with high E6 and E7 coverage and low coverage in multiple E2 and E1 gene amplicons in the amplicon sequencing were determined to have integration without episomal DNA based on HPV16 genome capture.

A total of 64% (60/94) of the tumors had integrated HPV16 with equal numbers having both episomal DNA and integration (30/60) and only integrated HPV16 (30/60). (**Figure 1**). There was a significant decrease in the frequency of integration (82% to 48%, P=0.03) with increasing age of diagnosis and with post-menopausal disease (P=0.009) **(Table 1)**. The median age at diagnosis for patients with tumors with HPV16 integration was younger (median=48, SD=12) compared to patients with tumors without integration (median=57, SD=13) (P=0.002, Mann-Whitney test) (**Figure 2A**). Integration was not significantly correlated with histological subtype (adenocarcinoma or squamous cell carcinoma), the number of pregnancies, smoking, cooking method, oral contraceptive use, or age at menarche (**Figure 1, Table 1**).

8

Table 1. Demographic and clinicopathologic characteristics of cervical cancer and HPV16 integration

Variable	No. of patients	Integration No.	Frequency (%)	P-value	P-value for trend
	96	62	65%		
Age (years)					
<40	17	14	82%	0.14	0.025
40-49	29	20	69%		
≥50-60	21	14	67%		
>60	23	11	48%		
Type of pathology					
SCC	74	47	64%	0.67	
AC	6	3	50%		
No. of pregnancies					
<3	10	7	70%	0.56	
≥3-5	38	27	71%		
≥6-14	40	24	60%		
Smoking status					
Yes	5	4	80%	0.66	
No	85	56	66%		
Cooking method					
Wood*	68	46	68%	0.73	
Gas	22	14	64%		
Contraceptive use					
Yes	12	10	83%	0.32	
No	77	49	64%		
Age Menarche (years)					
<13	26	13	50%	0.076	
≥13	63	44	70%		
Menopausal Status					
No	42	33	79%	0.0086	
Yes	49	25	51%		

SCC, squamous cell carcinoma, AC, adenocarcinoma. * included subjects cooking with both wood and gas.

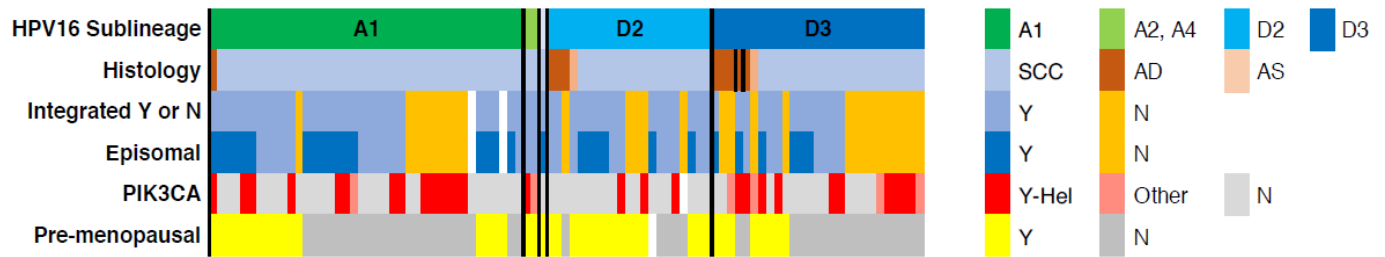**Figure 1. Characteristics of HPV16 Cervical Tumors in Guatemala.**

Figure 1. The landscape of HPV16 Cervical Tumors in Guatemala. Characteristics of HPV16 tumors are shown including HPV16 sublineage, histology, menopausal status, HPV integration status and PIK3CA mutation status. SCC, squamous cell carcinoma; AC, adenocarcinoma, ASC, adenosquamous; Pre, pre-menopausal; Men, post-menopausal; Int, HPV integrated only; Int/Ep, integrated and episomal; Ep, Episomal virus only; Mut-H, *PIK3CA* helical domain mutation; Mut, any *PIK3CA* mutation; WT, wild type for *PIK3CA*.

## Location of HPV16 integration sites

To confirm that there were not specific or recurrent sites of HPV16 integration, we mapped integration location. In the 60 patients with integration, a dispersed distribution of sites in the human genome was observed with sequences localized to promoter, intron, exon, and intergenic regions with no two integrations in the same gene (**Table S1, S2**). Integration in genes not previously reported includes the promoter of *EGR3*, *IRF4*, and *PLA2G4E*, and exons of *LOC100506022*, *FOXK1*, *GAPDH,* and *TBC1D24* (**Table S3**). The *IRF4* integration was confirmed using PCR, cloning, and sequencing (**Figure S1A, S1B**). Integration into promoter regions occurred predominantly in D2 or D3 sublineage-containing tumors and exon regions for A1-positive tumors, but this was not significant (**Table S3**).

Somatic mutations

In HPV integrated samples, there were fewer mutations in *PIK3CA* (29% in integrated vs. 53% in unintegrated tumors, P=0.02). This was also true for mutations in known cervical cancer driver genes including *PTEN*, *TP53*, *STK11 HRAS*, *MAPK1*, and *FBXW7* (45% vs. 68%, P=0.03). There was also a correlation between age of diagnosis and somatic mutation status, with a higher frequency of *PIK3CA* mutations in older onset patients (**Figure 2B**).

Figure 2. HPV integration and age of diagnosis.
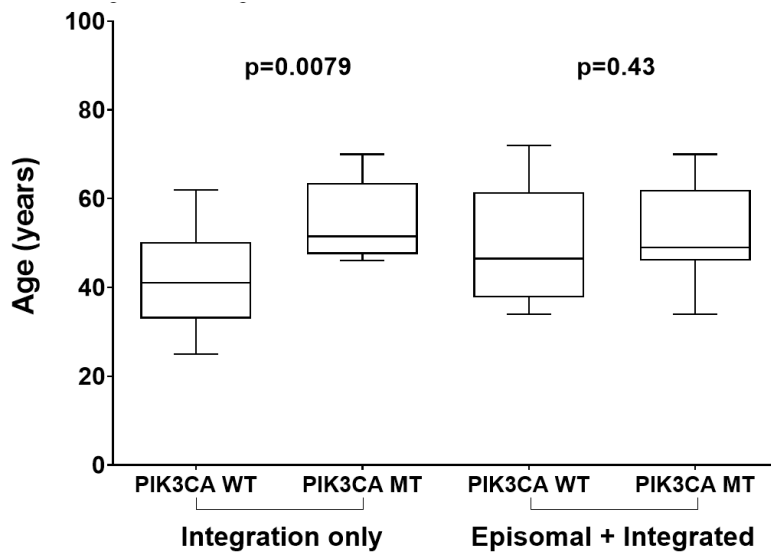
A.

B.
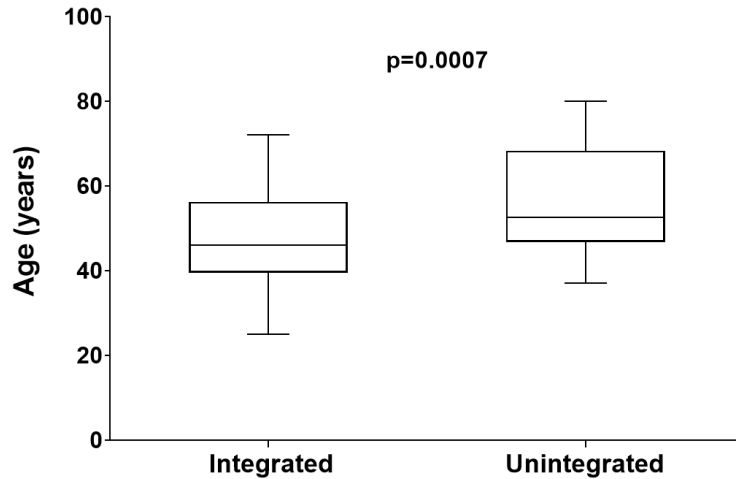
Figure 2. HPV integration and age of diagnosis. A, the age of diagnosis of patients with integrated HPV (includes integrated only and integrated and episomal) is compared against unintegrated samples. B, the age of diagnosis of integration only and episomal in integrated samples is compared by *PIK3CA* mutation status.

HPV16 Sub-lineage frequency and variants

To understand the role of viral variation we analyzed HPV16 sublineages and clinical and molecular characteristics of the patients. The sub-lineage A1 (44%, 40/91) was the most frequent followed by D3 (30%, 27/91), D2 (23%, 21/91) and only infrequent numbers of A2, A4, and D1 were observed (**Table 2**). Adenocarcinomas had almost exclusively A4, D2 and D3 viruses (11/12, 92%), whereas only 37/77 SCC tumors (48%) were positive for A4, D2, or D3 (P=0.002) (**Table 2**).

Table 2. Frequency of HPV16 sub-lineages and histologic types

Sublineage	Total No. (n=88)	Frequency	Squamous cell carcinoma	Adenocarcinoma	Freq. Adenocarcinoma
A1	40	45%	39	1	3%
D2	21	24%	17	4	19%
D3	27	31%	19	7	27%
	88		75	12	

Not shown are single samples of A2, A4, and D1 all squamous cell carcinoma.

Adenocarcinoma includes adenosquamous carcinoma.

Analysis of rare variants in HPV16 sequences identified only 1 variant (F57V) in the E7 gene in 91 samples as compared to 9 variants (Q10E, H31Y, H31N, I34R, E36Q, R55W, S78C, C118R, and T156A) in the E6 gene (**Table S4**). This is consistent with hypovariation in the E7 gene in cancers described in our previous study (7).

To identify differences associated with the more common HPV16 A1, D2, and D3 sublineages, we analyzed the data by age-of-diagnosis, integration frequency, and presence of a *PIK3CA* or other somatic mutation. Patients with the D2 and A1 sublineages had a significantly higher integration frequency (76%, 16/21 and 78%, 31/40, respectively) than those with D3 (44%, 12/27; P=0.01) (**Figure 3A**). Also, patients with D2 had a significantly younger age of diagnosis than A1 and D3-containing patients and are more often pre-menopausal (P=0.017) (**Figure 3B**). D2-containing tumors had a significantly lower rate of somatic mutation in

PIK3CA than D3 (P=0.007) (**Table 3**), however, this was not significant for all somatic driver mutations (data not shown).

Table 3. Correlation between HPV16 sublineage, age, and *PIK3CA* mutation

	Number	Mean Age	Premen.	Men.	P value	PIK3CA Status		P value	
						WT	All MT		MT (%)
A1	40	52±14	18	22		25	15		38%
D2	21	45±11	16	5	0.003	17	3	0.014	14%
D3	27	54±12	8	19		13	14		52%

PIK3CA status is the presence (Mutation, Mut) or absence (wild-type, WT) of mutations in this gene.P-values are comparisons of D2 and D3.

Figure 3. HPV16 sublineage, integration and menopausal status
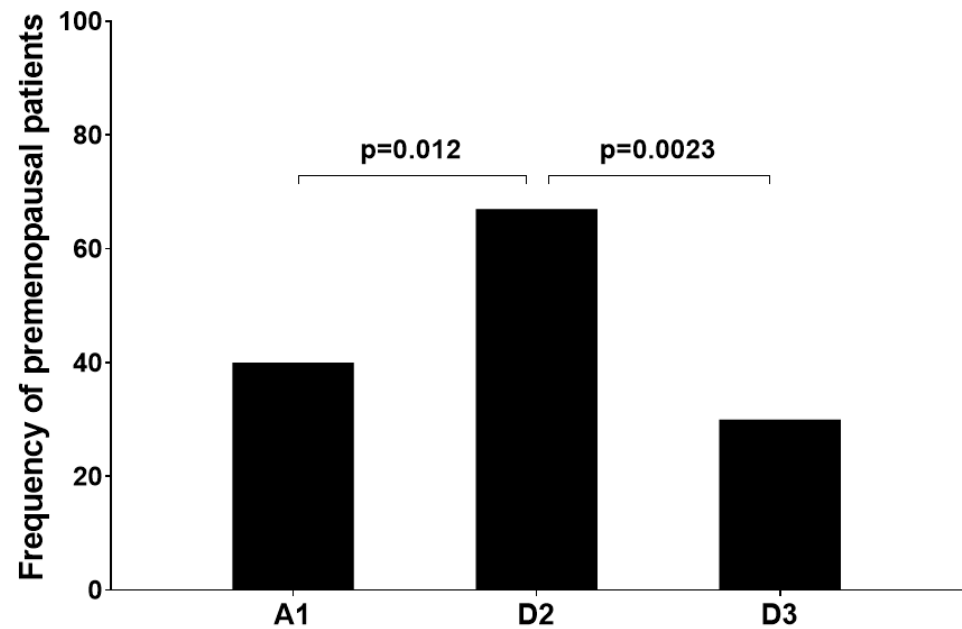
A.

14
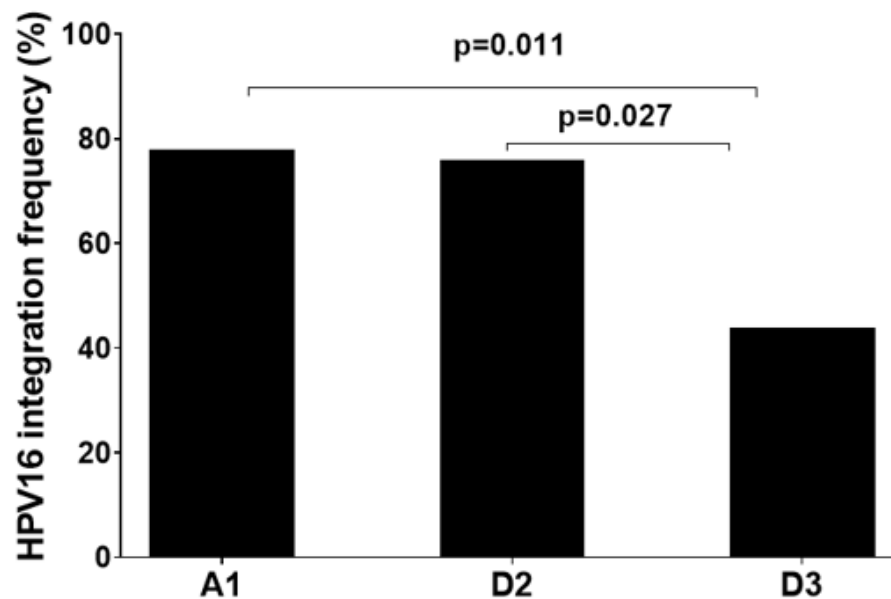
B.

Figure 3. HPV16 sublineage, integration and menopausal status. A, the frequency of integration is displayed by HPV16 sublineage. B, The frequency of pre-menopausal diagnosis is displayed by the HPV16 sublineage.

Because we observed significant differences between the D2 and D3 sublineages, we examined the lineage-defining amino acid changes between these sublineages. There are 11 amino acid differences between D2 and D3 as compared with the reference A1. These alterations involve 5 proteins, E1, E2, E5, L1, and L2, but 5/11 (45%) alterations are in the E2 gene (**Table S5**). The A1, D2 and D3 E2 proteins show five amino acid positions that are altered at residues 136, 142, 157, 211 and 221 (**Table S6**).

Discussion

To comprehensively study integration events in HPV16 positive cervical tumors, we used HPV DNA capture and HPV whole-genome sequencing on 96 advanced-stage cervical tumors from Guatemala. We analyzed the HPV16 sequence, sublineage distribution, and viral integration frequency and sites of viral insertion. The integration frequency we observed, 65% of tumors, is consistent with other reports (24,28,37), and HPV16 integration was associated with a significantly younger age of diagnosis, as seen previously (28,38). The location of integrated HPV16 was dispersed throughout the human genome, but in selected tumors may disrupt potential cancer-related genes such as *FHIT*, *GLI2*, and *IRF4*. Our sample has a significant fraction of D2 and D3 sublineage-containing tumors allowing us to distinguish differences in age of diagnosis, integration and somatic mutation in these two highly carcinogenic forms of HPV16.

Recent progress in determining HPV whole-genome sequences at high-throughput from clinical samples have revealed extensive diversity in HPV16 genome sequences and allowed for comprehensive classification of HPV16 variants and sublineages (10,36). The A4, D2, and D3 sublineages have been found to have a higher risk for cancer, particularly for adenocarcinoma (15). There were only 8 D2 and 4 D3 HPV16 samples in the TCGA study, providing little information on the molecular details of these cancers. Our study provides information on age-of-diagnosis, histology and somatic mutations for 21 additional D2 and 27 D3-containing tumors. This sample size, although modest, was sufficient to determine that D2 and D3-tumors are distinct regarding integration (lower in D3), age-of- diagnosis (earlier in D2) and percentage of *PIK3CA* mutations (higher in D3). Combining our data with two other studies shows that adenocarcinoma is consistently higher in A4, D2, and D3 compared to A1 (**Figure 4A**). Also, the

16

integration rate was lower in HPV16 A1 and HPV31, intermediate in HPV16 D2 and D3 and

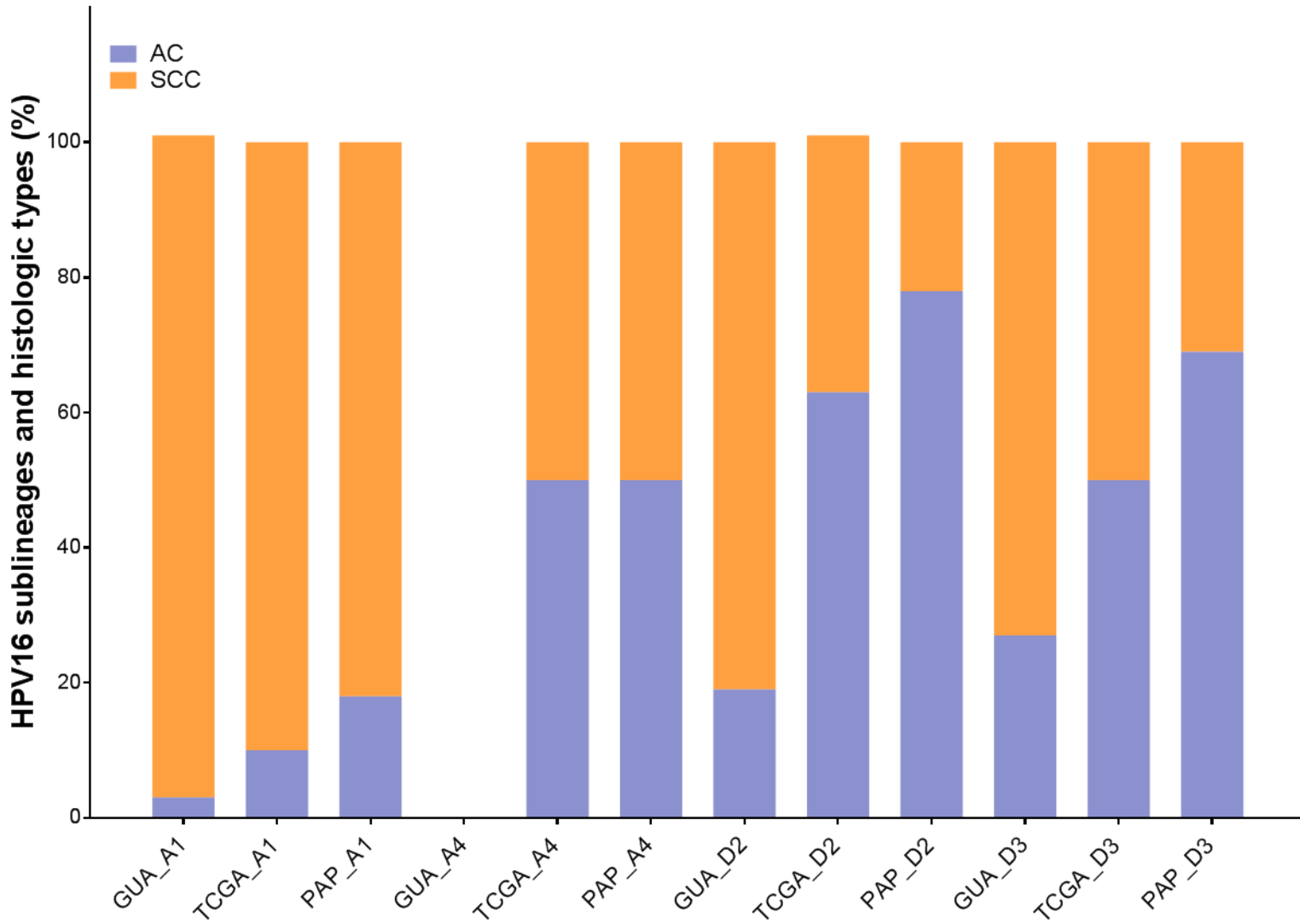high in HPV18 and HPV45 when surveying other reports (**Figure 4B**).
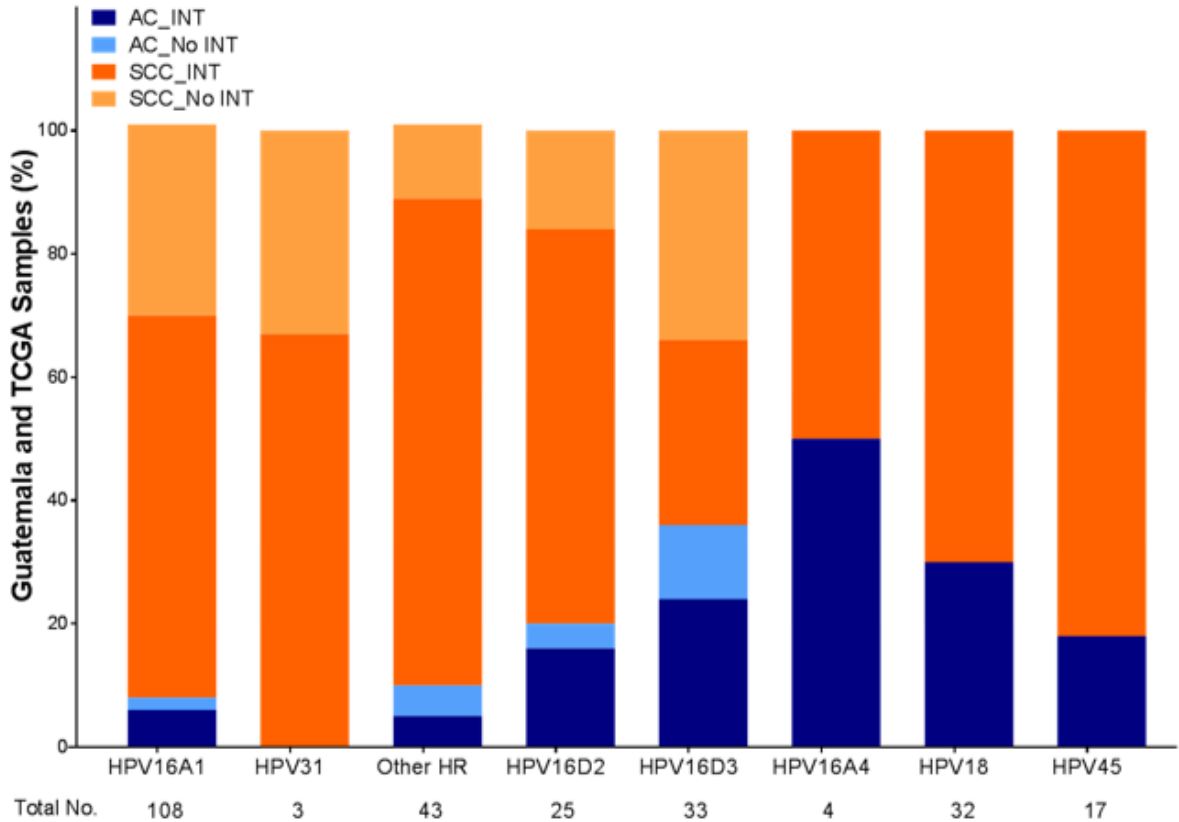
Figure 4   A

Figure 4B

Figure 4. Comparison of adenocarcinoma percentage in Guatemala and other studies. A, The percentage of adenocarcinoma (AC) versus squamous cell carcinoma (SCC) is shown by HPV16 sublineage in the current study compared to that of TCGA (3) or the PAP cohort (15). B, combined integration and histology status of US and Guatemalan tumors for HPV16 sublineages, HPV18, HPV31, and HPV45.

The biological basis for the dramatic difference in carcinogenicity, especially for adenocarcinoma of the D2 and D3 sublineages, compared to A1 is important to understand. While there are no differences in E7 between A1 and D2 or D3, there are 3 amino acid differences in the E6 gene and functional studies indicate that these changes increase transforming ability (39). There are also a considerable number of amino acid differences (3-16) between A1 and D2 and D3 in other viral proteins, with the most variable E1, E2, and L1. Because the early genes E2, E1, E5 are required for viral replication, transcription, cellular transformation, initiation of neoplasia and maintenance of the viral genome in infected cells, they represent the best candidates for the difference in carcinogenicity. Nevertheless, because of the high correlation of changes throughout the genome, it is hard to infer causality of specific mutations and/or genes.

Our finding of substantial differences in the age of diagnosis, integration and somatic mutation between D2 and D3 tumors indicates that a careful comparison of these two sublineages would be informative. Almost half of the amino acid alterations between D2 and D3. are in the E2 gene, a transcription factor that binds to the viral origin of replication and the major promoter for E6 and E7 oncogene expression. These six amino acid changes at five sites are in the E2 transactivation domain (H136Y, E142D, L157M, and L157I) and the hinge domain (I211T, A221T). We propose that the E2 gene is a candidate for a gene that contributes to the higher rate of cervical cancer and adenocarcinoma seen with HPV16 D2 and D3, but direct functional analyses would be required to demonstrate this and non-coding variants are also likely to contribute to functional differences (7).

Conclusions

In conclusion, we have established that a high proportion of HPV16 isolates in Guatemala are sub-lineage D2 and D3, known to have high cancer and adenocarcinoma risk. D2-positive patients have a significantly higher integration frequency, significantly younger age and significantly lower *PIK3CA* mutation frequency than D3. Further studies of these isolates and tumors may yield information on the mechanism of adenocarcinoma formation and the role of integration in cervical carcinogenesis.

Acknowledgments

Contributions

MD and LM contributed to study conception and design, WZ, MS, JM, DM, SB, and JB to development of methodology, AA, EA, and EG to acquisition of data and samples, MD, LM, and ET-G to analysis and interpretation of data, HL, RB, MW, NW, MS, LM, and MD to writing, review or revision of the manuscript.

References

1. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer **2002**;2:342-50

2. Schiffman M, Doorbar J, Wentzensen N, de Sanjose S, Fakhry C, Monk BJ, *et al.* Carcinogenic human papillomavirus infection. Nat Rev Dis Primers **2016**;2:16086

3. Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H, *et al.* Integrated genomic and molecular characterization of cervical cancer. Nature **2017**;543:378-84

4. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, *et al.* Landscape of genomic alterations in cervical carcinomas. Nature **2014**;506:371-5

5. Chen D, Juko-Pecirep I, Hammer J, Ivansson E, Enroth S, Gustavsson I, *et al.* Genome-wide association study of susceptibility loci for cervical cancer. J Natl Cancer Inst **2013**;105:624-33

6. Shi Y, Li L, Hu Z, Li S, Wang S, Liu J, *et al.* A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. Nat Genet **2013**;45:918-22

7. Mirabello L, Yeager M, Yu K, Clifford GM, Xiao Y, Zhu B, *et al.* HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. Cell **2017**;170:1164-74 e6

8. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med **2003**;348:518-27

9. Clifford GM, Tenet V, Georges D, Alemany L, Pavon MA, Chen Z, *et al.* Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: Whole viral genome sequences from 7116 HPV16-positive women. Papillomavirus Res **2019**;7:67-74

10. Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. Virology **2013**;445:232-43

11. Mirabello L, Clarke MA, Nelson CW, Dean M, Wentzensen N, Yeager M, *et al.* The Intersection of HPV Epidemiology, Genomics and Mechanistic Studies of HPV-Mediated Carcinogenesis. Viruses **2018**;10

12. Cornet I, Gheit T, Iannacone MR, Vignat J, Sylla BS, Del Mistro A, *et al.* HPV16 genetic variation and the development of cervical cancer worldwide. Br J Cancer **2013**;108:240-4

13. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, *et al.* Human papillomavirus type 16 and 18 variants: race-related distribution and persistence. J Natl Cancer Inst **2006**;98:1045-52

14. Chen Z, Schiffman M, Herrero R, Desalle R, Anastos K, Segondy M, *et al.* Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. PLoS One **2011**;6:e20183

15. Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, *et al.* HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. J Natl Cancer Inst **2016**;108

16. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer **2010**;10:550-60

17.   Olmedo-Nieva L, Munoz-Bello JO, Contreras-Paredes A, Lizano M. The Role of E6 Spliced Isoforms (E6*) in Human Papillomavirus-Induced Carcinogenesis. Viruses **2018**;10

18.   Tang S, Tao M, McCoy JP, Jr., Zheng ZM. The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. J Virol **2006**;80:4249-63

19.   Popescu NC, DiPaolo JA. Preferential sites for viral integration on mammalian genome. Cancer Genet Cytogenet **1989**;42:157-71

20.   Thorland EC, Myers SL, Gostout BS, Smith DI. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. Oncogene **2003**;22:1225-37

21.   Wentzensen N, Vinokurova S, von Knebel Doeberitz M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res **2004**;64:3878-84

22.   Bodelon C, Untereiner ME, Machiela MJ, Vinokurova S, Wentzensen N. Genomic characterization of viral integration sites in HPV-related cancers. International Journal of Cancer **2016**;139:2001-11

23.   McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. PLoS Pathog **2017**;13:e1006211

24.   Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, *et al.* Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. Nat Genet **2015**;47:158-63

25.   Liu Y, Zhang C, Gao W, Wang L, Pan Y, Gao Y, *et al.* Genome-wide profiling of the human papillomavirus DNA integration in cervical intraepithelial neoplasia and normal cervical epithelium by HPV capture technology. Sci Rep **2016**;6:35427

26.   Zhao JW, Fang F, Guo Y, Zhu TL, Yu YY, Kong FF, *et al.* HPV16 integration probably contributes to cervical oncogenesis through interrupting tumor suppressor genes and inducing chromosome instability. J Exp Clin Cancer Res **2016**;35:180

27.   Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, *et al.* Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res **2014**;24:185-99

28.   Vinokurova S, Wentzensen N, Kraus I, Klaes R, Driesch C, Melsheimer P, *et al.* Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res **2008**;68:307-13

29.   Jeon S, Allen-Hoffmann BL, Lambert PF. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. J Virol **1995**;69:2989-97

30.   Lou H, Villagran G, Boland JF, Im KM, Polo S, Zhou W, *et al.* Genome Analysis of Latin American Cervical Cancer: Frequent Activation of the PIK3CA Pathway. Clin Cancer Res **2015**;21:5360-70

31.   Litwin TR, Clarke MA, Dean M, Wentzensen N. Somatic Host Cell Alterations in HPV Carcinogenesis. Viruses **2017**;9

32.   Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov **2015**;5:704-12

33. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, *et al.* A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell **2018**;33:690-705 e9

34. Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, *et al.* Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. Proc Natl Acad Sci U S A **2009**;106:7933-8

35. Rosty C, Sheffer M, Tsafrir D, Stransky N, Tsafrir I, Peter M, *et al.* Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma. Oncogene **2005**;24:7094-104

36. Cullen M, Boland JF, Schiffman M, Zhang X, Wentzensen N, Yang Q, *et al.* Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection. Papillomavirus Res **2015**;1:3-11

37. Melsheimer P, Vinokurova S, Wentzensen N, Bastert G, von Knebel Doeberitz M. DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. Clinical cancer research : an official journal of the American Association for Cancer Research **2004**;10:3059-63

38. Calvete O, Garcia-Pavia P, Domínguez F, Bougeard G, Kunze K, Braeuninger A, *et al.* The wide spectrum of POT1 gene variants correlates with multiple cancer types. European Journal Of Human Genetics **2017**;25:1278

39. Richard C, Lanner C, Naryzhny SN, Sherman L, Lee H, Lambert PF, *et al.* The immortalizing and transforming ability of two common human papillomavirus 16 E6 variants with different prevalences in cervical cancer. Oncogene **2010**;29:3435-45