Contents lists available at ScienceDirect

Papillomavirus Research

HPV16 variant analysis in primary and recurrent CIN2/3 lesions demonstrates presence of the same consensus variant

Pascal van der Weele[a,b], Audrey J. King[a], Chris J.L.M. Meijer[b], Renske D.M. Steenbergen[b,*]

[a] National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Antonie van Leeuwenhoeklaan 9, Bilthoven, the Netherlands
[b] Amsterdam UMC, Vrije Universiteit, Department of Pathology, Cancer Center Amsterdam, De Boelelaan 1117, Amsterdam, the Netherlands

ABSTRACT

Introduction: Recurrent cervical intraepithelial lesions (rCIN2/3) after treatment of CIN2/3 occur in 5–15% of cases. rCIN2/3 can result from incomplete resection of CIN2/3, where the same HPV type and variant remains present. rCIN2/3 could also occur following a new infection with a different HPV variant of the same HPV type as the initial lesion. This study investigates HPV16 consensus variants in paired HPV16 positive scrapes from baseline CIN2/3 and rCIN2/3 lesions.
Methods: Paired HPV16 positive cervical scrapes of women with CIN2/3 at baseline and rCIN2/3 6 or 12 months after treatment were selected for whole-genome amplification and Illumina sequencing. Sequences were compared and nucleotide changes over time were characterized.
Results: From 14 paired samples, 10 had identical consensus variants in baseline CIN2/3 and rCIN2/3. Four paired samples showed one to three nucleotide variations at recurrent disease compared to baseline.
Conclusion: Identical or nearly identical HPV16 consensus variants were found in scrapes of paired HPV16 positive baseline CIN2/3 and rCIN2/3 lesions after treatment, suggesting no need for HPV variant analysis when the same HPV type is found in both lesions. These results argue for either incomplete excision of baseline CIN2/3 or inability of clearance of the original HPV infection.

1. Introduction

Persistent infection with an oncogenic human papillomavirus (HPV) type is an essential requirement for the development of cervical cancer [1]. However, most HPV infections are cleared by the immune system within one to two years after initial infection [2]. Of the currently known oncogenic HPV types, HPV16 and HPV18 cause around 70% of all cervical cancer cases worldwide [3].

The progression from initial HPV infection to cervical cancer occurs via precursor lesions (cervical intraepithelial neoplasia, CIN grade 1 to 3) and may take decades [4,5]. Treatment of CIN2/3 lesions is performed by ablative or excisional treatment. Women treated for CIN2/3 lesions have a 5–15% risk of developing recurrent high-grade lesions (rCIN2/3) within two years post-treatment [6,7]. Consequently, women undergoing treatment for CIN2/3 are closely monitored in the post-treatment period, before they return to a regular screening routine [7,8]. In the Netherlands women are tested by cytology and HPV co-testing at 6, 12 and 24 months post-treatment [9]. The 12 months visit can be omitted when the 6-month visit shows absence of HPV and normal cytology. After three consecutive negative co-tests, the woman

is referred back to screening programme [7,9].

rCIN2/3 represents a heterogeneous group of lesions, consisting of either residual CIN2/3 or a new CIN2/3 lesion. Residual CIN2/3 is a possible consequence of incomplete excision of the original CIN2/3 lesion, characterized by the same HPV type in the rCIN2/3 as in the baseline lesion. A new CIN2/3 lesion would occur from a newly acquired HPV infection by a different type (type switch), or an infection with a different variant of the same HPV type present in the baseline lesion [10]. Post-treatment surveillance should ideally differentiate residual from incident lesions, as women with residual lesions are in need of immediate treatment. Women with incident lesions may benefit from a more conservative approach due to a lower cancer risk [11].

In a recent multicenter post-treatment surveillance study, most baseline CIN2/3, rCIN3 and a subset of rCIN2 harbored HPV16 by genotyping [11]. From a clinical perspective, this poses the question whether the rCIN2/3 was caused by a newly acquired HPV16 infection, or by the same infection detected at baseline, which resulted in CIN2/3. Here, we employ whole genome next-generation sequencing of HPV16 to identify and compare consensus variants in paired baseline CIN2/3 and rCIN2/3 cervical scrapes.

* Corresponding author.
E-mail address: r.steenbergen@amsterdamumc.nl (R.D.M. Steenbergen).

Table 1
Characteristics of patients included in this study.

Study number	Histology at 6 or 12 months post treatment	Months post-disease diagnosis	Nucleotide changes
10	CIN2	7	3723 T > G
203	CIN2	5	–
402	CIN3	14	–
410	CIN3	11	–
422	CIN3	12	3800C > G, 7702 G > C
638	CIN3	6	–
648	CIN3	6	–
665	CIN3	7	–
669	CIN2	5	3800C > G
672	CIN2	12	–
680	CIN3	6	–
681	CIN2	6	–
834	CIN3	8	–
872	CIN2	12	3375C > T, 3575C > T, 5306C > T

Analysis was performed on HPV16 positive cervical scrapes for both baseline CIN2/3 and rCIN2/3.

2. Materials and methods

2.1. Sample selection

Cervical scrapes from women (aged 18 + years) with CIN2/3 derived from a multicenter study (SIMONATH) which has been described earlier [11,12] and who were scheduled for LLETZ (Large Loop Excision of the Transformation Zone) treatment of CIN2/3 were included. Baseline scrapes were obtained between two weeks before up to immediately before LLLETZ. In addition, preceding scrapes from women with a CIN2/3 found in the LLETZ material, without an additional biopsy could also be included. An additional scrape was taken from patients prior to treatment. For this study, only baseline CIN2/3 and rCIN2/3 scrapes were tested. Women treated for HPV16 positive baseline CIN2/3, and with HPV16 positive rCIN2/3 at six or 12-month follow-up were selected. Based on sample availability, a total of 14 HPV16 positive scrape pairs were tested, corresponding to 14 baseline CIN2/3 and 14 rCIN2/3 (6 rCIN2, 8 rCIN3), as shown in Table 1. The SIMONATH study was approved by the Medical Ethical Committee (METC) of VUmc (2009/285) and was registered in the Dutch Trial Registry (NTR1964).

2.2. HPV DNA detection

HPV detection and genotyping have been described previously for this study [11,12]. In short, total DNA was isolated from cervical scrapes using the Microlab start platform (Hamilton Robotics, Switzerland) with magnetic beads (Macherey-Nagel, Germany) according to the manufacturer's protocol. HPV DNA was amplified using the GP5+/ 6 + PCR, followed by detection via enzyme immunoassay readout and genotyping via an in-house reverse line blot [13].

2.3. HPV16 whole genome amplification

Complete HPV16 genomes were amplified in ten fragments from selected samples using primers displayed in Table 2. PCR's were performed using AmpliTaq Gold (Thermo Fisher Scientific, United States) according to the manufacturer's protocol. Cycling conditions consisted of an initial incubation of 15 min at 95°, followed by 38 cycles of alternating 95° for 15 s, 55° for 30 s and 72° for 90 s, followed by a final elongation step at 72° for 10 min. Amplicon integrity was checked using the Lonza FlashGel (Lonza, Switzerland) system. If no product was

observed, it was assumed the PCR performed at too little efficiency to observe on gel. Sample dsDNA concentrations were quantitated using the Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific, United States) according to the manufacturer's protocol to facilitate equimolar pooling of PCR products.

2.4. Illumina NGS analysis

Samples were submitted to BaseClear for Illumina HiSeq PE125 sequencing. Raw sequencing data was subjected to trimmomatic 0.36 for quality and adapter trimming (SLIDINGWINDOW:5:25; MINLEN:35), followed by FastQC 0.11.6 and MultiQC 1.3 for quality checking. Trimmed sequences were assembled to a HPV16 (K02718) reference genome using Bowtie2 2.3.4. Assembly files (.sam) were converted to bam and indexed using samtools 1.6. Consensus sequences were extracted from bam files using samtools and seqtk 1.2. Variant calling files (vcf) were generated using Lofreq to assess the presence of variants at heterogeneous positions. The cutoff for minority variants was set at 0.5%. Finally, bed files were extracted using bedtools 2.27.1 to facilitate the generation of coverage plots in R.

2.5. Alignment and phylogeny

Consensus sequences were aligned using MUSCLE 3.8.1551. Maximum likelihood phylogenetic inference was performed using IQ-tree 1.5.5. The model finder option (-m MF) was used to identify the best fit model for this study, resulting in the HKY + F + I model. The final alignment was bootstrapped using IQ-tree's ultrafast bootstrapping option (-bb 1000). Alignments were visualized using FigTree 1.4.3.

3. Results

3.1. Sample selection, amplification and sequencing

For this study, fourteen women treated for HPV16 positive CIN2/3, and with HPV16 positive rCIN2/3 at follow-up were selected (Table 1). Amplification of the whole HPV viral genome via overlapping PCR fragments was successful in all cervical scrapes. Subsequent ultra-deep sequencing of the cervical scrapes resulted in very high genome-wide coverage (Fig. S1), with a pooled average coverage per genome position of 112.287 (median: 104.130, minimum: 2045, maximum: 250.000), allowing for reliable assessment of infection variants.

3.2. Sequence comparison and characterization of nucleotide changes

Comparison of ultra-deep sequencing results of all fourteen paired samples showed (near) identical consensus variants at baseline and at recurrent disease, as shown in a maximum likelihood plot of the consensus sequence data in Fig. 1. Out of fourteen infections included in this study, ten are identical at the consensus level for both baseline and recurrent disease. In three patients, a single nucleotide polymorphism (SNP) was found at recurrent disease, at nucleotides 3723 and twice at 3800, respectively. In the remaining patient, three SNPs were detected at recurrent disease, at nucleotides 3375, 3575 and 5306. The sequence variations are listed in Table 1. A comparison of variant sequence, variant counts at each variable position and trinucleotide context, is presented in Table 3.

4. Discussion

In this study, we have investigated HPV16 variants in paired cervical scrapes of baseline CIN2/3 and rCIN2/3 at six or 12-month follow-up. Consensus variant analysis suggested that baseline CIN2/3 and rCIN2/3 are (near) identical in all cases. From a clinical perspective, this implies that the infection causing the initial CIN2/3 lesion was either not completely removed during treatment and resulted in a

Table 2
Primer sequences used in this study with references of origin.

Primers developed for this study are marked with *.

Forward	Sequence 5′-3′	Reference	Reverse	Sequence 5′-3′	Reference
F7869	GGTTACACATTTACAAGCAAC	*	R1312	ACATGGTGTTTCAGTCTCATGGC	[14]
F6835	CTGTGCAAAATAACCTTAACTGC	[14]	R162	GCAGCTCTGTGCATAACTGTG	[14]
F5492	TATAACTGACCAAGCTCCTTCA	[15]	R6599	TTATTGTGGCCCTGTGC	[16]
F6201	GAACACTGGGGCAAAGGATC	[17]	R6890	GAATTCATAGAATGTATGTATGTC	[14]
F3701	CGTCTACATGGCATTGGAC	[18]	R5024	AAGCAGGGTCTACAACTTTAAC	[14]
F4930	AACTAGTAGCACACCCATACCA	[14]	R5725	CGTGCAACATATTCATCCGT	[17]
F2529	CAATTTAAGAAATGCATTGGATGG	[14]	R3551	GTCTGGCTCTGATCTTGGTC	[14]
F3387	GTCAGGTAATATTATGTCCTACA	[14]	R4321	TGCAGAACGTTTGTGTCGCATT	[14]
F901	ACGGGATGTAATGGATGG	*	R1780	ATCATACACATTGGAGACACA	[18]
F1832	CAATGTGTATGATGATAGAGCC	[14]	R2915	AATAGTCTATATGGTCACGTAGG	[14]

recurrent lesion, or was caused by a novel infection with a (nearly) identical HPV16 variant. Our findings imply that in clinical practice, conventional PCR and genotyping is sufficient to detect type switches, since in the case of rCIN with the same HPV type as baseline CIN, both are likely caused by the same HPV variant.

Our comparison of consensus sequences showed that 10 out of 14 patients have identical HPV16 consensus genomes at baseline CIN2/3 and rCIN2/3. For these patients, the most parsimonious explanation is that the infection causing the initial CIN2/3 lesion also caused the rCIN2/3 lesion. The remaining four patients had consensus sequences that were nearly identical (1 and 3 nucleotides difference).

Previous studies have suggested that whole genome sequences differing ≥ 2 nucleotides could be considered unique variants, while sequences differing < 2 nucleotides cannot be reliably discriminated from each other [14]. In two patients, the rCIN2/3 positive HPV16 infection showed a one nucleotide difference compared to the HPV16 infection identified at baseline CIN2/3. Two more infections were found in which CIN2/3 and rCIN2/3 differed by two or three nucleotides. Considering the conservation of the HPV genome over time and the diversity of variants circulating in populations, the variants in the

rCIN2/3 lesions most likely originated from the variants causing the initial CIN2/3. By definition, some, or all of these infections could be reinfections with unique variants, however, considering the plethora of different HPV16 variants circulating in populations [14,15] it is improbable that one would be repeatedly infected with so closely related variants.

Assessment of variants at the variable positions showed that in five out six cases, the consensus nucleotide at rCIN was already present as a minority at baseline CIN. In addition, five out of six SNPs are possible apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) mutations (C > G and C > T in a tCn trinucleotide motif) [16]. APOBEC mutations are suggested to accumulate naturally as a host response against viruses, and have been shown to occur on the HPV genome [14,17,18]. This could be particularly relevant for the patient with three nucleotide changes between CIN2/3 and rCIN2/3 (872), as all three changes are possible APOBEC mutations. Combined, these findings suggest that the variable nucleotides originated as mutation events, but the effect of treatment on their prevalence cannot be assessed, since this study lacks the appropriate controls and power for such an analysis.
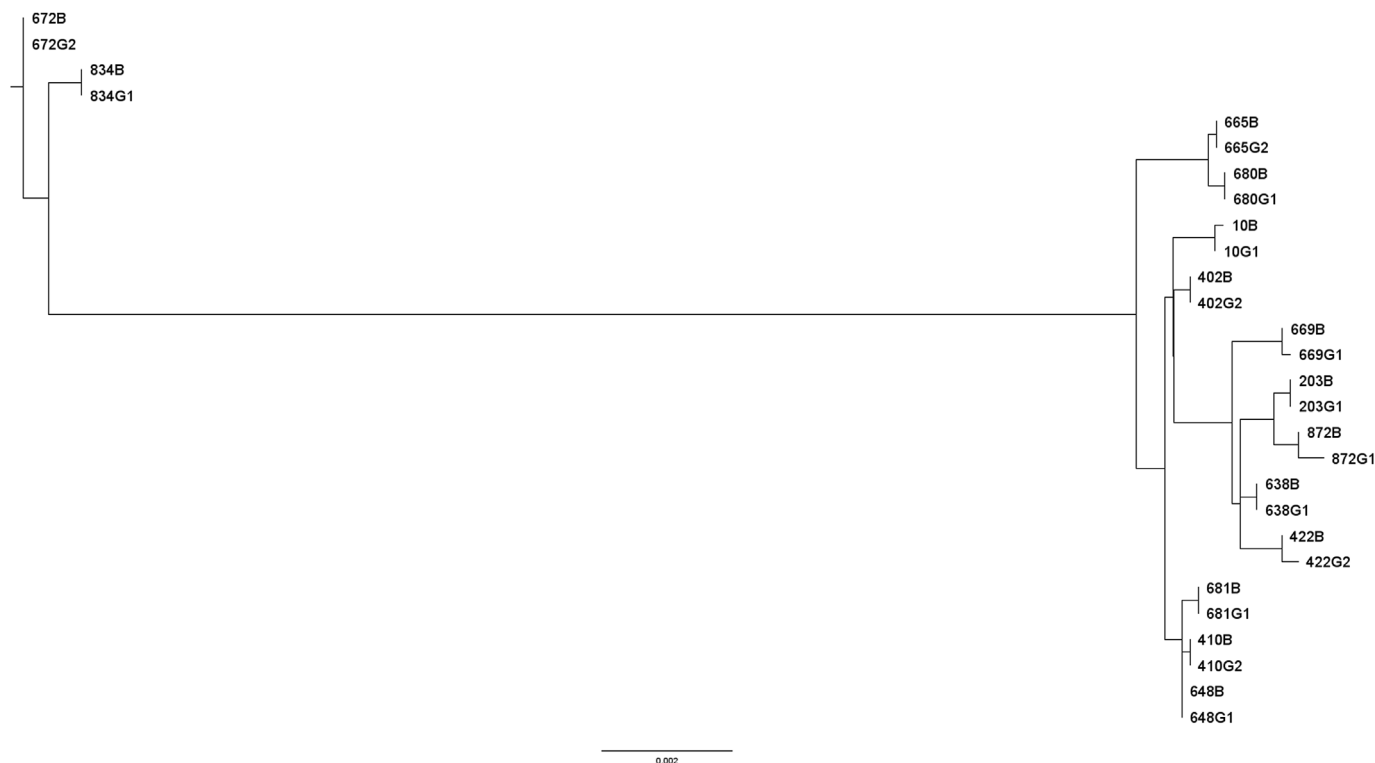

Fig. 1. Maximum likelihood tree of consensus sequences obtained in this study. Baseline (B) and follow-up (G1/2) samples are shown to cluster close or identical to each other. X-axis shows genetic distance between samples.

Table 3
Nucleotide counts at variable positions between baseline CIN and rCIN. In addition, the trinucleotide context in which each variation occurs is given.

Sample	Nucleotide position	Total coverage	Major nucleotide	Coverage	Minor nucleotide	Coverage	Trinucleotide context
10B	3723	193,552	T	169,840	G	23,467	gTa
10G1	3723	167,524	G	167,242	–	–	gGa
422B	3800	163,308	C	146,166	G	16,826	tCa
422G2	3800	153,655	G	146,319	C	7123	tGa
669B	3800	102,591	C	102,279	–	–	tCa
669G1	3800	138,622	G	131,277	C	7121	tGa
872B	3375	99,262	C	56,321	T	42,874	tCc
872G1	3375	105,095	T	93,927	C	11,124	tTc
872B	3576	92,831	C	50,678	T	42,041	tCa
872G1	3576	106,969	T	96,364	C	10,557	tTa
872B	5306	111,061	C	59,651	T	50,363	tCa
872G1	5306	118,822	T	98,347	C	20,445	tTa

Overall, most variations were observed in the *E2* gene. The SNP at position 3375 is in the *E2* hinge region, which is hypervariable between HPV types, and is not clearly associated with any function [19]. Position 3575 has been described as an integration site of HPV16 [20], although it is unknown how a nucleotide shift at this position affects this. Variations at positions 3723 and 3800 lead to changes in the DNA binding domain of *E2* [19], and 3800C > G has been identified in a cervical cancer case in India [21]. In *L2* 5306C > T was found, although this position does not encode any known epitopes [22]. URR position 7702 does not encode for any known promoter sites [23]. Currently it is unclear what the exact role is of any of the nucleotide changes observed between baseline CIN2/3 and rCIN2/3, but the fact that most are found in *E2*, which is associated with *E6/E7* regulation [19] is striking and warrants close monitoring of these infections.

While the method employed in this study generates ultra-high resolution sequencing data across most of the HPV16 genome, upon observing Fig. S1, it becomes apparent that there are some coverage dips around nucleotide positions 1800, 4100, 7000 and 7800. This suggests sub-optimal primer design, although coverage does not drop below 2000 on average. Further optimization could yield a more equal genome wide result, although this is not required for this study. For one infection (669), coverage dropped to 1–10 × across a 600bp fragment for the rCIN sample, while at these positions (positions 4251–4871) a 100 × coverage was obtained for the baseline sample. This could cause discrepancies between baseline CIN2/3 and rCIN2/3. However, as seen in Fig. 1, samples from this infection differ by only one nucleotide, at position 3800, where both samples have > 100.000 coverage, ruling out that the difference in sequence is caused by low coverage. Potentially, this could be caused by the presence of both episomal and integrated virus in the CIN lesion at baseline. It may be speculated that following resection, the episomal fraction may have been cleared, leaving only the integrated fraction in the recurrent lesion, or vice versa. We can also not exclude that the low coverage is potentially caused by a deletion in the genome of this specific variant. The method employed in this study could potentially detect deletions within amplicons, but only if the deleted area does not overlap with a primer site.

The results from this study focus primarily on the comparison of consensus sequences to identify causative variants for CIN and rCIN. However, there could be a change in the distribution of minority variants. Although a detailed assessment of the effects of treatment on minority variant diversity would be of interest, this is beyond the scope of the present study.

In summary, we showed that when paired scrapes from women with baseline CIN2/3 and post-treatment rCIN2/3 at 6 or 12 months follow-up are both HPV 16 positive, the same variant is responsible for CIN2/3 and rCIN2/3. Consequentially, in clinical practice no need exists for variant analysis when the same HPV type is found at baseline and at recurrent disease. Our findings suggest possible inadequacies in treatment procedures enforced upon detection of CIN2/3 or persistence of the original HPV infection.

Conflicts of interest

RDMS and CJLMM are minority shareholders of Self-Screen B.V., a spin-off company of VUmc, of which CJLMM is part-time director since September 2017. Self-screen holds patents on methylation markers and HPV detection. CJLMM has received speakers' fee from Qiagen and SPMSD/Merck, served occasionally on the scientific advisory board (expert meeting) of Qiagen, SPMSD/Merck and GSK. CJLMM has a very small number of shares in Qiagen, and was minority shareholder of Diassay until April 2016. CJLMM has been co-investigator on a SPMSD sponsored trial, for which his institute received research funding. PvdW and AJK have no conflict of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pvr.2019.04.008.

References

[1] J.M. Walboomers, et al., Human papillomavirus is a necessary cause of invasive cervical cancer worldwide, J. Pathol. 189 (1) (1999) 12–19.
[2] A.F. Rositch, et al., Patterns of persistent genital human papillomavirus infection among women worldwide: a literature review and meta-analysis, Int. J. Cancer 133 (6) (2013) 1271–1285.
[3] S. de Sanjose, et al., Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, Lancet Oncol. 11 (11) (2010) 1048–1056.
[4] M.A. Vink, et al., Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data, Am. J. Epidemiol. 178 (7) (2013) 1161–1169.
[5] R.D. Steenbergen, et al., Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions, Nat. Rev. Canc. 14 (6) (2014) 395–405.
[6] J. Melnikow, et al., Cervical intraepithelial neoplasia outcomes after treatment: long-term follow-up from the British Columbia Cohort Study, J. Natl. Cancer Inst. 101 (10) (2009) 721–728.
[7] M. Kocken, et al., Risk of recurrent high-grade cervical intraepithelial neoplasia after successful treatment: a long-term multi-cohort study, Lancet Oncol. 12 (5) (2011) 441–450.
[8] M. Kocken, et al., High-risk human papillomavirus testing versus cytology in predicting post-treatment disease in women treated for high-grade cervical disease: a systematic review and meta-analysis, Gynecol. Oncol. 125 (2) (2012) 500–507.
[9] Integraal Kankercentrum Nederland. *Richtlijn CIN, AIS EN VAIN*, (2015) Available from: https://www.oncoline.nl/cin-ais-en-vain.
[10] M.C. Bleeker, C.J. Meijer, J. Berkhof, Follow-up after treatment for cervical intraepithelial neoplasia, BMJ 345 (2012) e7186.

[11] M.H. Uijterwaal, et al., Performance of CADM1/MAL-methylation analysis for monitoring of women treated for high-grade CIN, Gynecol. Oncol. 143 (1) (2016) 135–142.

[12] N.J. Polman, et al., Good performance of p16/ki-67 dual-stained cytology for surveillance of women treated for high-grade CIN, Int. J. Cancer 140 (2) (2017) 423–430.

[13] A.J. van den Brule, et al., GP5+/6+ PCR followed by reverse line blot analysis enables rapid and high-throughput identification of human papillomavirus genotypes, J. Clin. Microbiol. 40 (3) (2002) 779–787.

[14] L. Mirabello, et al., HPV16 E7 genetic conservation is critical to carcinogenesis, Cell 170 (6) (2017) 1164–1174 e6.

[15] P. van der Weele, C. Meijer, A.J. King, Whole-genome sequencing and variant analysis of human papillomavirus 16 infections, J. Virol. 91 (19) (2017).

[16] S.A. Roberts, et al., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers, Nat. Genet. 45 (9) (2013) 970–976.

[17] Cancer Genome Atlas Research Network, et al., Integrated genomic and molecular characterization of cervical cancer, Nature 543 (7645) (2017) 378–384.

[18] M.M. Gaglia, K. Munger, More than just oncogenes: mechanisms of tumorigenesis by human viruses, Curr Opin Virol 32 (2018) 48–59.

[19] A.A. McBride, The papillomavirus E2 proteins, Virology 445 (1–2) (2013) 57–79.

[20] Y. Liu, et al., Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology, Oncotarget 7 (5) (2016) 5852–5864.

[21] B. Bhattacharjee, et al., Characterization of sequence variations within HPV16 isolates among Indian women: prediction of causal role of rare non-synonymous variations within intact isolates in cervical cancer pathogenesis, Virology 377 (1) (2008) 143–150.

[22] J.W. Wang, R.B. Roden, L2, the minor capsid protein of papillomavirus, Virology 445 (1–2) (2013) 175–186.

[23] M.J. Lace, et al., Upstream regulatory region alterations found in human papillomavirus type 16 (HPV-16) isolates from cervical carcinomas increase transcription, ori function, and HPV immortalization capacity in culture, J. Virol. 83 (15) (2009) 7457–7466.