

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

High-level of viral genomic diversity in cervical cancers: A Brazilian study on human papillomavirus type 16



Cristina Mendes de Oliveira ^{a,b,*}, Ignacio G. Bravo ^b, Nathália Caroline Santiago e Souza ^a, Maria Luiza Nogueira Dias Genta ^c, José Humberto Tavares Guerreiro Fregnani ^d, Maricy Tacla ^e, Jesus Paula Carvalho ^c, Adhemar Longatto-Filho ^{f,g,h,i}, José Eduardo Levi ^a

^aLaboratório de Virologia, Instituto de Medicina Tropical, Universidade de São Paulo, São Paulo, Brazil

^bInfections and Cancer Laboratory, Catalan Institute of Oncology, Barcelona, Spain

^cInstituto do Câncer do Estado de São Paulo (ICESP), Faculdade de Medicina, São Paulo, Brazil

^dTeaching and Research Institute, Barretos Cancer Hospital, Barretos, São Paulo, Brazil

^eGynecology Department, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

^fMolecular Oncology Research Center, Hospital de Câncer de Barretos, Barretos, São Paulo, Brazil

^gLaboratory of Medical Investigation (LIM) 14, Faculty of Medicine, University of São Paulo, São Paulo, Brazil

^hLife and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal

ⁱICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal

ARTICLE INFO

Article history:

Received 25 March 2015

Received in revised form 2 July 2015

Accepted 4 July 2015

Available online 6 July 2015

Keywords:

HPV16 diversity

Intralesion

Cervical cancer

ABSTRACT

Invasive cervical cancer (ICC) is the third most frequent cancer among women worldwide and is associated with persistent infection by carcinogenic human papillomaviruses (HPVs). The combination of large populations of viral progeny and decades of sustained infection may allow for the generation of intra-patient diversity, in spite of the assumedly low mutation rates of PVs. While the natural history of chronic HPV infections has been comprehensively described, within-host viral diversity remains largely unexplored.

In this study we have applied next generation sequencing to the analysis of intra-host genetic diversity in ten ICC and one condyloma cases associated to single HPV16 infection. We retrieved from all cases near full-length genomic sequences.

All samples analyzed contained polymorphic sites, ranging from 3 to 125 polymorphic positions per genome, and the median probability of a viral genome picked at random to be identical to the consensus sequence in the lesion was only 40%. We have also identified two independent putative duplication events in two samples, spanning the L2 and the L1 gene, respectively. Finally, we have identified with good support a chimera of human and viral DNA.

We propose that viral diversity generated during HPVs chronic infection may be fueled by innate and adaptive immune pressures. Further research will be needed to understand the dynamics of viral DNA variability, differentially in benign and malignant lesions, as well as in tissues with differential intensity of immune surveillance. Finally, the impact of intralesion viral diversity on the long-term oncogenic potential may deserve closer attention.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Papillomaviruses (PVs) are a family of small non-enveloped viruses, with a circular double-stranded DNA genome, that infect epithelia in a wide range of vertebrates and probably all amniotes

* Corresponding author at: Laboratório de Virologia, LIM 52, Instituto de Medicina Tropical da Universidade de São Paulo, Rua Dr Enéas de Carvalho Aguiar 470, 2o andar, São Paulo, SP CEP 05403-000, Brazil.

E-mail address: cristina.oliveira1@gmail.com (C.M.d. Oliveira).

(Bravo et al., 2010). Human PVs (HPVs) are the most studied members in the family, because of their medical importance. They are related to benign proliferative disorders, such as skin warts, epidermodysplasias or *condylomata acuminata*, and also to malignant neoplasms of the cervix, anus, vagina, vulva, penis and oropharynx (Cubie, 2013). To date, more than 180 HPV types have been fully sequenced (<http://pave.niaid.nih.gov/>). HPVs are not monophyletic and belong instead into five genera, with different types showing differential association with different diseases (Bernard et al., 2010; Bravo et al., 2010; de Villiers, 2013).

Infection by most HPVs is asymptomatic (Antonsson et al., 2000), but 13 HPV genotypes are classified by the International Agency for Research on Cancer as oncogenic: HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 (Bouvard et al., 2009). Carcinogenic potential is not evenly distributed among HPVs, and HPV16 alone is responsible for more than 50% of the invasive cervical cancer (ICC) cases worldwide (de Sanjose et al., 2010; Li et al., 2011), and is the most prevalent HPV in other squamous carcinomas associated to HPVs infections (De Vuyst et al., 2009; de Sanjose et al., 2014), as well as in women with normal cytology (Bruni et al., 2010). In South-East Brazil, 2.3% of women on the screening population were positive for HPV16 (Levi et al., 2014), whereas 75% of the ICC cases from the same region were associated with this HPV (de Oliveira et al., 2013). Genetic diversity among HPV16 isolates can reach 2%, and the trend is towards full-genome sequencing of the isolates (Chen et al., 2005, 2009, 2013; Smith et al., 2011).

For rapidly evolving viruses such as HCV or HIV, a wealth of data has been generated during follow-up studies to address the question of the amount and generation pace of intra-host genetic diversity. For PVs infections, however, the common view about these viruses has precluded experimental research about mutation and evolutionary rates. Most studies have addressed changes in the most prevalent, “consensus”, sequence during chronic PVs infections based on Sanger sequencing. This technology may however not be appropriate to capture the dynamics of slowly evolving viruses, such as PVs. Very few reports have applied deep sequencing technologies to gain insight into the accumulation of genetic diversity during the course of a PV infection. To help bridging this gap, the aim of the present study is to analyze by means of next-generation sequencing a number of ICC cases associated to HPV16 single infections.

2. Materials and methods

2.1. Clinical specimens and total DNA extraction

Ten invasive cervical cancers containing HPV16 DNA as single infection, determined after Linear Array HPV Genotyping (Roche Molecular Diagnostics, Pleasanton, USA) or PapilloCheck (Greiner Bio-One GmbH, Frickenhausen, Germany) tests were selected for the present study. The ICC cases originated from women participating in a previous study (de Oliveira et al., 2013), and the condylo-ma originated from a woman attended at the Gynecological Department from Hospital das Clínicas – Universidade de São Paulo. The study was approved by the Brazilian National Research Ethical Committee (CONEP n°158.490).

Total DNA was extracted from fresh tissue using QIAamp DNA Mini Kit (Qiagen, Gaithersburg, USA) or NucleoSpin Tissue kit (Macherey–Nagel GmbH&Co, Germany), according to manufacturer's instructions.

2.2. HPV16 E6 and E2 real time PCR

To determine the HPV16 physical status a multiplex real-time PCR for simultaneous detection of the E6 and E2 regions were used. Primers and probes were modified from (Canadas et al., 2010), as follows: primer E6F 5'ACCGTTAGTATAAAAAGCAGACATTTTMT3', primer E6R 5'GCTCCTGTGGGTCCTGRAAC3', probe E6 CAL Fluor Orange 560 BHQplus – CACAAAAGAGAAGTCAA – BHQ1, primer E2F 5'GCRACGAAGTATCCTCTCTGAA3', primer E2R 5'AAGGCGACGGCTTTGGKAT3' and probe E2 FAM – TAYTAGGCAGCACTTGGCCAACCA – BHQ1. Real-time PCR was performed in a final volume of 25 µL, containing 5 µL of DNA, 1x qPCR Low Mg²⁺ (Next Generation ECO HotStart qPCR kit), 200 nM of each primer and

probe. PCR was carried out according to the protocol: for 95 °C for 2 min, 45 cycles of 95 °C for 15 s, 43 °C for 30 s 60 °C for 45 s, in a 7500 FAST Real Time Systems (Applied Biosystems). In each run, water was included as negative control, an HPV16 complete genome plasmid as episomal control, and SiHa DNA as HPV16 DNA integrated control. Physical state was determined by comparing the CtE6/CtE2 ratio of each sample with the ratio observed in the plasmid control. Samples negative for E2 region were considered as integrated, samples with the same CtE6/CtE2 ratio of the plasmid as pure episomal form and samples with a CtE6/CtE2 ratio lower than the plasmid as mixed (both forms episomal and integrated).

2.3. HPV16-specific long PCR

PCR was performed in 50 µL consisting of 1x long PCR buffer (Fermentas GmbH&Co, Germany), 1.5 mM MgCl₂, 0.2 mM of each dNTP, 0.2 µM of each primer, 1.25 U of long PCR enzyme mix (Fermentas GmbH&Co, Germany) and 100–300 ng of template DNA, generating an amplicon of 7935pb. The primers used (forward-21 5'AAACTAAGGGCGTAACCGA3' and reverse-7862 5'GGTTTGCACACCCATGT3') were previously described (Sun et al., 2012). Positions corresponding to 5'-end nucleotide of primers in the HPV16 reference K02718 were annotated following primer orientation.

PCR cycling conditions were: 94 °C for 2 min, 10 cycles of (95 °C for 15 s, 60 °C for 30 s and 68 °C for 7 min), followed by 25 cycles of (95 °C for 15 s, 60 °C for 30 s and 68 °C for 7 min with an increment of 5 s by cycle), plus a final extension of 68 °C for 10 min.

Long PCR products were separated by electrophoresis and the approximately 8Kbp band was purified using Invisorb® Fragment Clean Up (Invitek, Berlin, Germany), according to manufacturer's instructions. The purified PCR product was then submitted to deep sequencing on the Ion Torrent platform (Life technologies, CA, USA).

2.4. Deep sequencing

Libraries were generated using 100 ng of the long PCR product and the Ion Shear™ fragmentation method (Life technologies, CA, USA). Adapters with barcode allowing for multiplexing were ligated and purified using the E-Gel® SizeSelected™ system (Life technologies, CA, USA). Samples were quantified using Ion Library Quantification Kit (Life technologies, CA, USA) and 1–5 libraries from different samples were multiplexed. Samples were submitted to emulsion PCR using the Ion PGM™ Template 200 kit (Life technologies, CA, USA) and finally to sequencing reaction using Ion PGM™ 200 Sequencing kit (Life technologies, CA, USA). Sequencing was performed in the Ion 314™ chip (Life technologies, CA, USA) according to the manufacturer's instructions.

2.5. Deep sequence data analyses

Fastq files were extracted from the Ion Torrent SFF files, clipped and submitted to quality control analyses using FastQC High Throughput Sequence QC Report software v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), for trimming according to phred scores. Reads were mapped against the HPV16 reference sequence K02718 using the TMAP v0.2.3 software. The mapping results were visualized with IGV viewer (Thorvaldsdottir et al., 2013). Consensus fasta sequences were generated using SAMtools v0.1.19 (Li et al., 2009).

2.6. Read depth analysis

A slide-window analysis was performed to verify differences in the read depth coverage between and/or within the samples, using

a window size of 100 pb and a step size of 20 pb. The median coverage value for each window was determined for each sample, and normalized by the overall median sample coverage.

2.7. Phylogenetic analysis

Fasta consensus sequences obtained from deep sequencing and a group of 20 HPV16 complete genome sequences retrieved from GenBank were aligned with Muscle (Edgar, 2004) and edited with Se-AL v2.0a11 (available at <http://tree.bio.ed.ac.uk/software/sea/>). Phylogenetic reconstructions were performed by maximum likelihood (ML) criterion using RAXML version 7.4.8 (Stamatakis et al., 2005). The best fit model for the ML reconstructions were defined comparing the Akaike Information Criterion (AIC). For the nucleotide sequences we tested the GTR + Γ model without partitions, with one partition per gene and with three partitions per gene. The preferred model was GTR + Γ without partitions, which was used to infer a ML tree. Node support was evaluated using 5000 bootstraps cycles.

2.8. De novo assembly

Regarding intra-host analysis, the reads generated by the Ion Torrent were submitted to *de novo* assembly using the CLC genomics workbench software (Qiagen, Gaithersburg, USA). All contigs obtained were submitted to Blast (<http://www.ncbi.nlm.nih.gov/>) to verify whether it was HPV16 or human DNA. Variant calling was also performed using CLC genomics workbench software (Qiagen, Gaithersburg, USA) setting the quality phred to 30, minimum coverage 50, minimum frequency to call a SNP of 1% and at least two reads showing the SNP. The definition of the threshold value was based on DNA polymerases proofreading activity, the deep sequencing protocol substitution error rate and in a previous report (Kukimoto et al., 2013). The enzyme mix used for the long PCR is a unique blend of two thermostable high fidelity DNA polymerases with proofreading activity. The two enzymes generate long PCR product with three times higher fidelity compared to Taq DNA polymerase, according to the manufacturer. The deep sequencing approach used in the present study has a substitution error rate estimated in 0.07% (Bragg et al., 2013). Besides that, a previous report by Kukimoto and colleagues analyzing HPV16 episomes present in W12 cells and determining a threshold value of 0.5% (Kukimoto et al., 2013). For the present analysis, we decided to be more stringent and used 1%. Finally, the presence of SNPs and indels obtained by the analysis were verified by visual inspection of the mapped reads.

2.9. GenBank accession number

KP212150–KP212159 and KP874716.

3. Results

3.1. HPV16 physical status

Based on the E6/E2 real time PCR results, we observed both forms of the HPV16 genome, episomal and integrated, in six ICC samples. In the remaining four ICC samples, as well as in the condyloma, we observed exclusively episomal forms (Table 1).

3.2. HPV16-specific long PCR

Using an HPV16-specific long PCR protocol we generated an amplicon of ca. 8 Kbp from ten clinical ICC samples. In addition to the expected ca 8 Kbp band, both larger (that were retained in

well of the agarose gel) and shorter amplicons were also produced. From one of the samples, these additional bands were individually excised from the agarose gel, purified and submitted to another HPV16-specific PCR targeting the E7 ORF (Walboomers and Meijer, 1997) that confirmed all bands originated from HPV16 genome (data not shown).

3.3. Resequencing

Our HPV16-specific long PCR was able to amplify 7817 bp of HPV16 genome for all samples from position 37 to 7853 of the HPV16 reference genome, therefore missing to amplify 89 bp from the LCR region. HPV16 near full-length and the internal genetic variation from eleven clinical samples were determined using Ion Torrent deep sequencing by assembly and comparison against the HPV16 reference sequence K02718. For each sample, the assembly algorithm obtained only one contiguous sequence. The sequence length of the assembled sequences ranged from 7832 to 7901. Table 1 shows the results of this analysis.

The maximum likelihood phylogenetic tree recovered with high branch support, the four major clades representing the HPV16 variants and sub-lineages: A (European and E(As)), B (African-1), C (African-2) and D (North America and Asian-American 1 and 2), as previous reported (Chen et al., 2005; Burk et al., 2013), (Fig. 1). Among the eleven samples analyzed, six were classified as HPV16_A1, one as HPV16_A2 and four as HPV16_D3. It was observed a correlation between HPV16 variant and tumor histological type ($p = 0.047$), with squamous cell carcinomas being preferentially associated to HPV16_A, whereas adenocarcinoma cases were preferentially associated to HPV16_D.

Concerning the ICC samples, pairwise distance of nucleotide sequences showed high degree of conservation, with the maximum distance being 0.018. The overall mean genetic diversity was 0.01 for nucleotide sequence. Comparing to the HPV16 reference sequence changes in 190 (2.4%) nucleotide sites were observed. At the amino acid level, changes in 67 (2.8%) sites were observed: the E2 gene showed the highest proportion of amino acid changes (5.5%), followed by E5 (4.8%), whereas L1 displayed the lowest (1.7%).

Read depth analysis showed an increase in sequencing coverage in the region spanning the L1 gene of one sample (CNA20), and in the L2 region for another one (CNA55). These findings are compatible with a duplication event (Fig. 2).

3.4. Intra-host HPV16 genetic variability

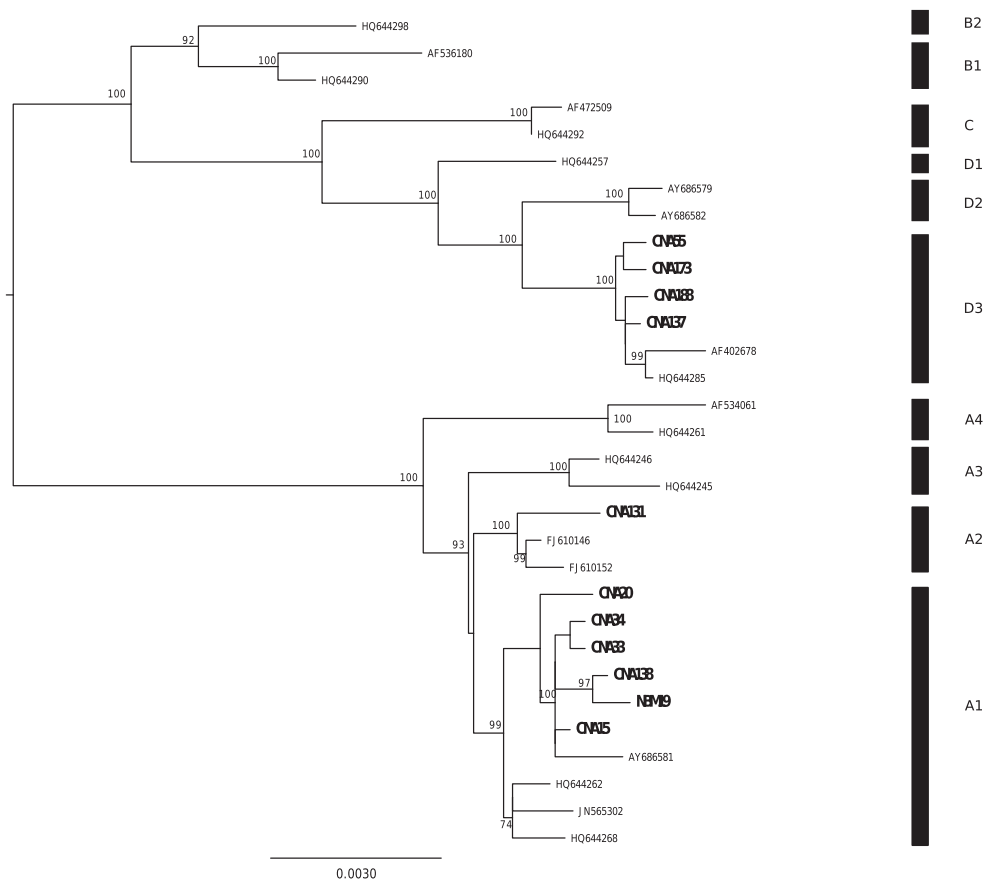
In order to analyze the intra-host genetic variation, we performed a *de novo* assembly without a reference sequence, instead of using any HPV16 genome as scaffold for the assembly of the gained reads. This assembly generated only one contiguous sequence for four of the samples (NBM19, CNA15, CNA34 and CNA138), which were confirmed as being only HPV16. Assemblies for the remaining seven samples generated three to eight contiguous sequences, ranging from 553 to 8164 bp in length; some of them correspond to HPV16 and some to human DNA. Samples showing more than one contiguous sequence corresponding to HPV16 were analyzed and in all cases only one HPV16 lineage was observed in each sample. Findings of multiple contigs were thus an artifact of the assembly algorithms. Interestingly, in one sample (CNA20) one out of six contiguous sequences contained a mixture of HPV16 and human genome, compatible with a chimeric fragment containing both viral and cellular DNA. This contiguous sequence, with 671 bp length, was generated by the assembly of 3186 reads, with an average coverage of 435 is therefore extremely unlikely to be an artifact. The first breakpoint was covered by 220 reads and the second breakpoint by 430 reads.

Table 1

Clinical characteristics, HPV16 physical status and resequencing mapping characteristics of near full-length HPV16 genome sequences from deep sequencing.

Sample ID	Histology	FIGO stage	Age at diagnosis	HPV16 physical status	Sequence length	Number of reads	Assembled reads	Mean coverage/nt	HPV16 sub lineage	Diversity index	GenBank accession number
CNA15	SCC	IV	40	Episomal	7850	49061	94.7%	660.7	A1	0.393	KP212150
CNA20	SCC	IB	32	Mixed	7901	210773	76.4%	2079.7	A1	0.513	KP212151
CNA33	SCC	IB	47	Mixed	7901	89247	79.5%	979.8	A1	0.091	KP212152
CNA34	SCC	IB	73	Mixed	7888	87408	94.5%	1127.1	A1	0.791	KP212153
CNA55	ADC	IIB	53	Mixed	7854	15940	94.8%	191.0	D3	0.741	KP212154
CNA131	SCC	IIIA	58	Episomal	7857	29450	94.4%	351.8	A2	0.175	KP212155
CNA137	ADC	IB	42	Mixed	7900	258065	79%	2832.0	D3	0.0000791	KP212156
CNA138	ADC	IB	40	Mixed	7893	176520	98.3%	2494.7	A1	0.890	KP212157
CNA173	ADC	IB	32	Episomal	7832	22923	88.3%	253.1	D3	0.473	KP212158
CNA188	ADC	Unk	32	Episomal	7876	44136	89.3%	525.4	D3	0.181	KP212159
NBM19	Condyloma	-	Unk	Episomal	7877	38800	89.3%	600.0	A1	0.387	KP874716

ADC, adenocarcinoma; SCC, squamous cell carcinoma; mixed, both forms episomal and integrated; Unk, unknown.

**Fig. 1.** HPV16 complete genome unrooted maximum likelihood phylogenetic tree inferred with the study samples (bold) and the HPV16 genome references using GTR + 4Γ nucleotide substitution model. Support after 1000 bootstrap cycles is shown for each node. The phylogenetic tree recovered with high branch support, the four major clades representing the HPV16 variants and sub-lineages.

The first 17 nucleotides correspond to HPV16 L1 region, nucleotides from 18 to 609 to a 17q12 region of human genome and the last 69 nucleotides correspond to HPV16 E2 region.

Concerning the variant calling analysis, all SNPs or indels mapped to homopolymer regions (three or more repetitions) were disregarded, as homopolymer stretches can induce erroneous base calls on the Ion Torrent platform. All samples contained polymorphic sites, albeit with different degrees (Fig. 3). Sample CNA138 showed the lowest degree of variability, with only three polymorphic sites (0.04% of the total positions sequenced), and only one of them representing more than 5% of the reads in that position. Conversely, sample CNA137 showed the highest degree of

intra-host variability, with 125 polymorphic sites (1.6% of the total positions sequenced), and 31 of them representing more than 10% of the reads in that position.

Samples showed polymorphic sites in all viral genomic regions (Figs. S1–S9), with exception of CNA138, which harbored polymorphisms only in E1 and L2 regions. No specific profile of mutations was identified. However, four samples concentrated almost 40% of polymorphisms in certain positions: CNA33 showed an admixed population in the E1 region (positions 903, 1086 and 1113), CNA131 (position 4157) and CNA188 (position 4182) in the non-coding region between E5 and L2 genes, and CNA137 in the L2 region (position 4597) and LCR (position 7796).

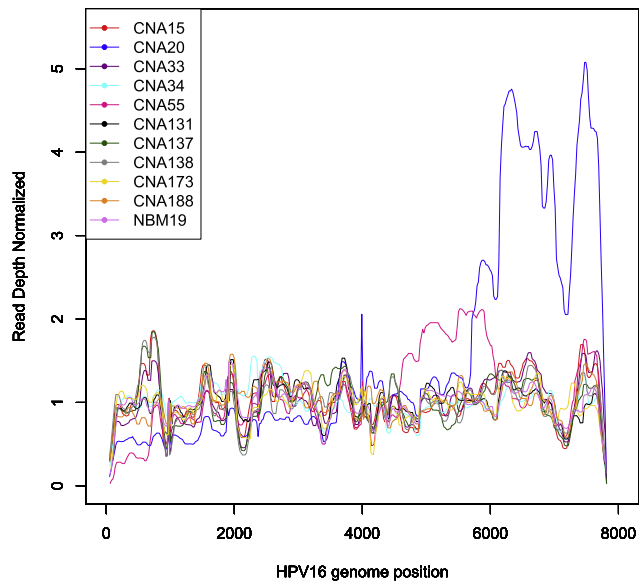


Fig. 2. Slide-window analysis, window size 100 bp and step 20 bp, of the read depth of ten invasive cervical cancer cases and one condyloma. The median coverage value for each window was determined for each sample, and normalized by the overall median sample coverage. Read depth analysis shows an increase in sequencing coverage in the region spanning the L1 gene of one sample (CNA20), and in the L2 region for another one (CNA55).

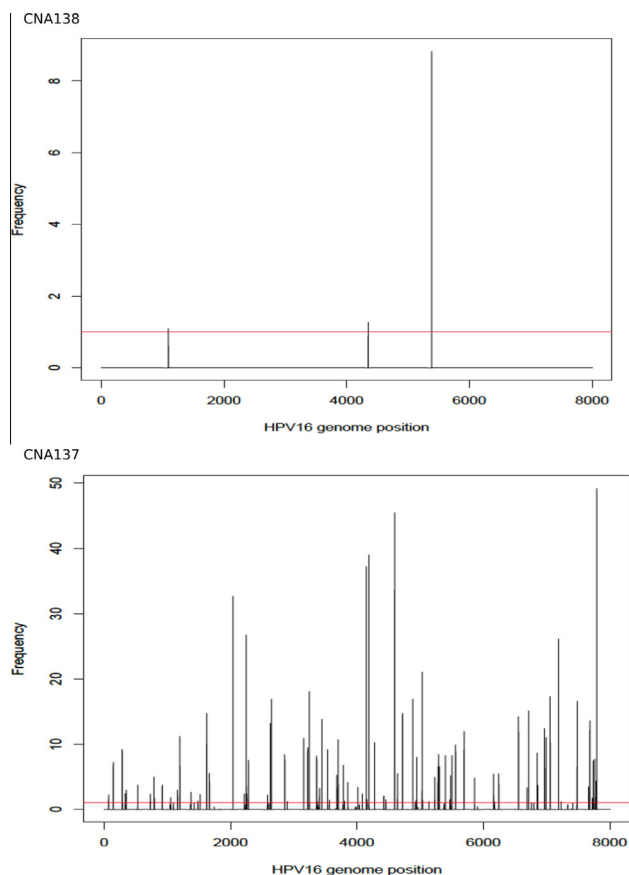


Fig. 3. Representation of intra-host HPV16 genetic variability from two invasive cervical cancer cases. The near full-length HPV16 genomes from clinical samples were aligned to their *de novo* assembled genome, and mutation frequency at each nucleotide position is presented. The red line indicates the threshold value of 1%. Sample CNA138 showed the lowest degree of variability, with only three polymorphic sites and sample CNA137 showed the highest degree of intra-host variability, with 125 polymorphic sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For each sample, we calculated a diversity index, as the probability of a randomly chosen genome to be identical to the consensus genome, assuming complete independence between sites (Table 1). The median value of this diversity index for our samples was 0.39, ranging between $7.9e-4$ and 0.89. No correlation was observed between the diversity index and the age at diagnosis ($p = 0.267$), FIGO stage ($p = 0.304$), tumor histological type, *i.e.* squamous vs adenocarcinoma ($p = 0.841$) or HPV16 physical status, *i.e.* integrated vs mixed forms ($p = 0.429$).

4. Discussion

In the present study, we have analyzed HPV16 near full-length genome from eleven clinical samples (ten ICC cases and one condyloma) using an HPV16-specific long PCR method followed by Ion Torrent next generation sequencing, a technology able to generate three–four orders of magnitude more information than Sanger sequencing technology (Liu et al., 2012). This information density provided by the deep sequencing approach is fundamental for the study of viral diversity, as it allows the simultaneous detection of minor variants. The resolution power makes these techniques the current standard tool for intra-host viral diversity analyses. This technology did not allow recovering long-distance haplotypes, and therefore we cannot estimate a number of potential coexisting viral variants present in the same sample. Nonetheless, we were able to approximate the total amount of diversity among the viral genomes present in each lesion by combining the information about the individual frequencies of each single nucleotide polymorphism identified, under the assumption of independence between sites.

Seven out of eleven HPV16 sequences from cervical samples analyzed in the present study were classified as variant A. This variant distribution is in agreement with previous Brazilian studies, which observed that 34% (Gheit et al., 2006) to 65.8% (Freitas et al., 2014) of the HPV16 isolates are classified as HPV16_A variant. We could ascribe the SNPs in the different isolates here sequenced to the differential variant signatures (Smith et al., 2011).

Viral genomes in the HPV infected cells in the mucosal epithelial basal layer remain essentially episomal. During a productive infection, such episomal forms constitute also the viral genomes packed into virions and released with keratinocyte shedding (Doorbar, 2006). However, progression of the malignisation process often concurs with integration of viral sequences into the cellular genome as well as with loss of the episomal forms (Wentzensen et al., 2004; Xu et al., 2013). Thus, preneoplastic lesions as well as invasive carcinomas usually contain both episomal and integrated HPV genomes, and a fraction (20–40%) of ICC harbors exclusively integrated viral genomes (Canadas et al., 2010; Aromseree et al., 2014). In our samples, 45% of the ICC cases were identified to harbor HPV16 genomes exclusively in the episomal form. The E6/E2 real time PCR used to assess the HPV physical status is based on the differential detection of the E6 and E2 genes (Canadas et al., 2010; Valmary-Degano et al., 2013), since HPV16 integration often entails the deletion of the E2 gene, as observed in SiHa cells. In certain cases, however, the target region excised and lost during the HPV integration process does not span the E2 gene (Wang et al., 2013; Akagi et al., 2014). Therefore, we cannot exclude the possibility that some of the samples identified to exclusively carry episomal HPV16 genomes could also contain integrated forms, as long as the integration has not resulted in the loss of the E2 gene. Indeed, our results show that we are able to identify non-standard integration events. One of the contiguous sequences obtained from the sample containing L1 duplication, displayed a fragment of human DNA sequence flanked by viral L1

and E2 sequences. The breakpoints of integrated HPV can be anywhere in the late region (5' breakpoint) and E1–E2–E5 region (3' breakpoint) (Dall et al., 2008; Xu et al., 2013). Integration sites of HPV16 in cervical intraepithelial neoplasia are more frequent in the L1 gene followed by E1 gene (Wang et al., 2013); in ICC samples although HPV breakpoints are reported preferentially in the viral E2 gene, high resolution method showed that viral breakpoints are distributed across the viral genome (Akagi et al., 2014). Our interpretation is that this contiguous sequence could represent integrated viruses that we were able to detect. The integration site was in the 17q12 within the IRZF3 gene that can participate in the regulation of apoptotic process and Bcl-2 expression (Rebollo et al., 2001).

One of the main findings of our study is that intralesion papillomavirus diversity can be high, showing globally that the median probability of a viral genome picked at random to be identical to the consensus sequence amounted to barely 40%. Such accumulation of diversity Intra-host HPV genetic diversity is poorly studied, but deserves more attention, as it could be a hidden factor contributing to different outcomes after infections by the same HPV type. It is commonly assumed that PVs evolve slowly because they use the high-fidelity cellular polymerases during replication (Doorbar et al., 2012). However, the human genome encodes for several DNA polymerases, differing largely on fidelity and substrate preferences (Loeb and Monnat, 2008). It is not clear which cellular enzyme/s replicate the viral DNA, and it is also not known whether the shift in replication mechanics of PVs mentioned above concurs with a shift in the cellular enzyme(s) in charge of viral replication. Our results show that viral intralesion diversity levels did not differ between samples carrying exclusively episomal viral genomes and samples carrying both episomal and integrated forms. The contribution of the integration event in viral diversity accumulation may therefore be minor. Next-generation sequencing data for cell-culture conditions have shown that HPV16 episomes present in the W12 cell line did not accumulate sequence variation above 0.5% (Kukimoto et al., 2013). However, using real clinical samples the same authors reported that above 40% of the clinical samples contained polymorphic sites reaching frequencies up to 5%, thus suggesting the potential for intra-host generation of diversity (Kukimoto et al., 2013). In our study, we have further detected polymorphic sites in all samples submitted to deep sequencing, with values for the less prevalent alleles amounting above 40%. Thus, viral genomic diversity accumulating during the course of a natural infection contrasts sharply with the absence of variation generated during *in vitro* subculture of the W12 episome-containing cells.

Our long PCR is able to detect both HPV16 in episomal form, as well as the HPV16 DNA integrated into the human genome, provided there is more than one copy of the virus integrated. Besides the anticipated near full-length genome, the long PCR generated additional, smaller and larger amplicons that we confirmed to originate from HPV16. Generation of multiple amplicons using HPV16-specific long PCR has been described in previous studies (Kukimoto et al., 2013). Although we cannot exclude the possibility of these bands to be an artifact of the long PCR, the replication mechanics of PV genome may also allow for the generation of such incomplete genomes. PV replication occurs bidirectionally generating episomes in the early stages of the infection, and switches to rolling circle mode during the productive stages of the infection yielding concatemers, in a process that requires homologous recombination (Flores and Lambert, 1997; Orav et al., 2013). It is unclear how genomic concatemers are excised and circularized, and viral genomes carrying deletions and/or insertions may be generated as defective by products, together with the intact full-length circular genomes (Bravo and Felez-Sanchez, 2015). On the one hand, we did not observe any

deletion event in any of the samples analyzed, even in the samples known to contain integrated viral genomes, and similar findings have been reported in other laboratories (Kukimoto et al., 2013). On the other hand, two of our ICC samples showed a significant increase in read depth close to L1 and L2, respectively, compatible with a duplication event of a genome fragment spanning these regions. Similar results of partial duplications of the viral genome have been reported, independently, by different laboratories: analyses of three independent cases of cancer developed after a long period of recurrent respiratory papillomatosis associated to HPV6 or HPV11, reported duplications in the HPV regulatory region in the tumor sample that were absent in the benign samples from the same patients (Byrne et al., 1987; DiLorenzo et al., 1992; Yuan et al., 2012). Thus, duplication events in the papillomavirus genome may have an impact in the infection phenotype, as they can deregulate the equilibria between early and late genes expression, exacerbating viral replication as well as cellular duplication.

It could be claimed that intralesion viral diversity here described may rather reflect an original infection by several very-closely-related HPV16 genotypes. However, our current understanding of HPV-associated disease suggests that individual lesions are associated to individual infection events (Murray et al., 1971; Quint et al., 2012). Instead, we propose here that at least a fraction of such changes accumulated during infection may arise from selection pressures associated to immune response, underlying the high frequencies of above 40% for certain polymorphisms in a number of samples, such as CNA33, CNA131 and CNA137. Regarding innate immune response, the APOBEC3 internal mutators have been shown to target PV DNA, introducing directional C>T changes (Vartanian et al., 2008; Wang et al., 2014). The APOBEC3 locus has undergone a large expansion in the human lineage (Munk et al., 2012), which may partially account for the broad diversity of human PVs. Regarding adaptive immune response, polymorphisms observed in the E6 gene could be a result of an immune selective pressure. Many linear epitopes have already been described in this region (Dillner, 1990) and also could be related to the affinity of the interactions of E6 protein with E6-AP (Watts et al., 2002) and p53 (Beckman et al., 1994). Polymorphisms in the E1 gene could be related to the DNA binding domain and present in different HPV16 sub-lineages, but do not affect the replication activity (Kukimoto et al., 2013). Other humoral epitopes have also been described in the E2 region (Dillner, 1990) and could drive immune selection. Finally, an additional factor non-related to immune response and that may have also fueled accumulation of viral diversity in these lesions could be a cancer-related defect in p53 functioning. Indeed, mutations ablating p53 activity are a common feature of multiple human cancers, and underlie mutation accumulation in different loci in advanced cancer stages. However, HPV16 diversity in the condylomatous benign lesion analyzed did not differ from the median of the viral diversity detected in carcinomatous samples (0.387 vs 0.433, respectively). The individual contribution of p53 mutations to intralesion viral diversity remains thus unsolved.

Our results show that accumulation of HPV16 intralesion diversity is a common feature in malignant and benign lesions induced by this virus. We propose that such intra-host diversity is generated during the course of a chronic infection, either under mutational pressures and drift, or also possibly fueled by immune selection pressures, either innate – e.g. in response to APOBEC3 directional mutation – or adaptive – e.g. in response to humoral or cellular directed responses. Since immune restrictions differ between tissues, we anticipate that infections by the same viral genotype may evolve to varying degrees of intralesion diversity. Future larger studies of intralesion HPV genetic diversity, as well

as, research on the natural history of the infection by HPVs will be needed to monitor generation of viral diversity in asymptomatic, productive, benign, premalignant and malignant infections. Finally, the possible role of oncovirus intralesion diversity generated during chronic infections should be explored as a differential factor for increased oncogenic potential.

Acknowledgments

C.M.O. is funded by Grants # 2011/24035-2 and # 2012/23290-1, São Paulo Research Foundation (FAPESP).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.meegid.2015.07.002>.

References

- Akagi, K., Li, J., Broutian, T.R., et al., 2014. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* 24, 185–199.
- Antonsson, A., Forslund, O., Ekberg, H., Sterner, G., Hansson, B.G., 2000. The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensal nature of these viruses. *J. Virol.* 74, 11636–11641.
- Aromseree, S., Chaiwongkot, A., Ekalaksananan, T., Kongyingyoes, B., Patarapadungkit, N., Pientong, C., 2014. The three most common human papillomavirus oncogenic types and their integration state in Thai women with cervical precancerous lesions and carcinomas. *J. Med. Virol.* 86, 1911–1919.
- Beckman, G., Birgander, R., Sjalander, A., Saha, N., Holmberg, P.A., Kivela, A., Beckman, L., 1994. Is p53 polymorphism maintained by natural selection? *Hum. Hered.* 44, 266–270.
- Bernard, H.U., Burk, R.D., Chen, Z., van Doorslaer, K., zur Hausen, H., de Villiers, E.M., 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 401, 70–79.
- Bouvard, V., Baan, R., Straif, K., et al., 2009. A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* 10, 321–322.
- Bragg, L.M., Stone, G., Butler, M.K., Hugenholtz, P., Tyson, G.W., 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 9, e1003031.
- Bravo, I.G., de Sanjose, S., Gottschling, M., 2010. The clinical importance of understanding the evolution of papillomaviruses. *Trends Microbiol.* 18, 432–438.
- Bravo, I.G., Felez-Sanchez, M., 2015. Papillomaviruses: Viral evolution, cancer and evolutionary medicine. *Evol. Med. Public Health* 2015, 32–51.
- Bruni, L., Diaz, M., Castellsague, X., Ferrer, E., Bosch, F.X., de Sanjose, S., 2010. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. *J. Infect. Dis.* 202, 1789–1799.
- Burk, R.D., Harari, A., Chen, Z., 2013. Human papillomavirus genome variants. *Virology* 445, 232–243.
- Byrne, J.C., Tsao, M.S., Fraser, R.S., Howley, P.M., 1987. Human papillomavirus-11 DNA in a patient with chronic laryngotracheobronchial papillomatosis and metastatic squamous-cell carcinoma of the lung. *N. Engl. J. Med.* 317, 873–878.
- Canadas, M.P., Darwich, L., Sirera, G., Cirigliano, V., Bofill, M., Clotet, B., Videla, S., 2010. New molecular method for the detection of human papillomavirus type 16 integration. *Clin. Microbiol. Infect.* 16, 836–842.
- Chen, Z., DeSalle, R., Schiffman, M., Herrero, R., Burk, R.D., 2009. Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. *J. Virol.* 83, 1443–1455.
- Chen, Z., Schiffman, M., Herrero, R., DeSalle, R., Anastos, K., Segondy, M., Sahasrabudde, V.V., Gravitt, P.E., Hsing, A.W., Burk, R.D., 2013. Evolution and taxonomic classification of alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70. *PLoS ONE* 8, e72565.
- Chen, Z., Terai, M., Fu, L., Herrero, R., DeSalle, R., Burk, R.D., 2005. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. *J. Virol.* 79, 7014–7023.
- Cubie, H.A., 2013. Diseases associated with human papillomavirus infection. *Virology* 445, 21–34.
- Dall, K.L., Scarpini, C.G., Roberts, I., Winder, D.M., Stanley, M.A., Muralidhar, B., Herdman, M.T., Pett, M.R., Coleman, N., 2008. Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. *Cancer Res.* 68, 8249–8259.
- de Oliveira, C.M., Fregnani, J.H., Carvalho, J.P., Longatto-Filho, A., Levi, J.E., 2013. Human papillomavirus genotypes distribution in 175 invasive cervical cancer cases from Brazil. *BMC Cancer* 13, 357.
- de Sanjose, S., Bruni, L., Alemany, L., 2014. HPV in genital cancers (at the exception of cervical cancer) and anal cancers. *Presse Med.* 43, e423–e428.
- de Sanjose, S., Quint, W.G., Alemany, L., et al., 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol.* 11, 1048–1056.
- de Villiers, E.M., 2013. Cross-roads in the classification of papillomaviruses. *Virology* 445, 2–10.
- De Vuyst, H., Clifford, G.M., Nascimento, M.C., Madeleine, M.M., Franceschi, S., 2009. Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: a meta-analysis. *Int. J. Cancer* 124, 1626–1636.
- Dillner, J., 1990. Mapping of linear epitopes of human papillomavirus type 16: the E1, E2, E4, E5, E6 and E7 open reading frames. *Int. J. Cancer* 46, 703–711.
- DiLorenzo, T.P., Tamsen, A., Abramson, A.L., Steinberg, B.M., 1992. Human papillomavirus type 6a DNA in the lung carcinoma of a patient with recurrent laryngeal papillomatosis is characterized by a partial duplication. *J. Gen. Virol.* 73 (Pt 2), 423–428.
- Doorbar, J., 2006. Molecular biology of human papillomavirus infection and cervical cancer. *Clin. Sci. (Lond.)* 110, 525–541.
- Doorbar, J., Quint, W., Banks, L., Bravo, I.G., Stoler, M., Broker, T.R., Stanley, M.A., 2012. The biology and life-cycle of human papillomaviruses. *Vaccine* 30 (Suppl 5), F55–70.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Flores, E.R., Lambert, P.F., 1997. Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle. *J. Virol.* 71, 7167–7179.
- Freitas, L.B., Chen, Z., Muqui, E.F., Boldrini, N.A., Miranda, A.E., Spano, L.C., Burk, R.D., 2014. Human papillomavirus 16 non-European variants are preferentially associated with high-grade cervical lesions. *PLoS ONE* 9, e100746.
- Gheit, T., Simoes, R.T., Tommasino, M., Donadi, E.A., Goncalves, M.A., 2006. HPV16 variants in squamous intraepithelial lesions in human immunodeficiency virus-negative and -positive Brazilian women. *Viral Immunol.* 19, 340–345.
- Kukimoto, I., Maehama, T., Sekizuka, T., et al., 2013. Genetic variation of human papillomavirus type 16 in individual clinical specimens revealed by deep sequencing. *PLoS ONE* 8, e80583.
- Levi, J.E., Longatto-Filho, A., Eluf-Neto, J., Rodrigues, C.L., Oliveira, C.M., Carloni, A.C., Lorenzi, A.T., Tacla, M., Fregnani, J.H., Ab'Saber, A.M., Scapulatoto, C., Villa, L.L., 2014. Evaluation of HPV molecular tests in primary screening for cervical cancer in Brazil. *Open J. Obstetr. Gynecol.* 4, 470–478.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, N., Franceschi, S., Howell-Jones, R., Snijders, P.J., Clifford, G.M., 2011. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int. J. Cancer* 128, 927–935.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012, 251364.
- Loeb, L.A., Monnat Jr., R.J., 2008. DNA polymerases and human disease. *Nat. Rev. Genet.* 9, 594–604.
- Munk, C., Willemsen, A., Bravo, I.G., 2012. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol. Biol.* 12, 71.
- Murray, R.F., Hobbs, J., Payne, B., 1971. Possible clonal origin of common warts (*Verruca vulgaris*). *Nature* 232, 51–52.
- Orav, M., Henno, L., Isok-Paas, H., Geimanen, J., Ustav, M., Ustav, E., 2013. Recombination-dependent oligomerization of human papillomavirus genomes upon transient DNA replication. *J. Virol.* 87, 12051–12068.
- Quint, W., Jenkins, D., Molijn, A., et al., 2012. One virus, one lesion—individual components of CIN lesions contain a specific HPV type. *J. Pathol.* 227, 62–71.
- Rebollo, A., Ayllon, V., Fleischer, A., Martinez, C.A., Zaballos, A., 2001. The association of Aiolos transcription factor and Bcl-xL is involved in the control of apoptosis. *J. Immunol.* 167, 6366–6373.
- Smith, B., Chen, Z., Reimers, L., et al., 2011. Sequence imputation of HPV16 genomes for genetic association studies. *PLoS ONE* 6, e21375.
- Stamatakis, A., Ludwig, T., Meier, H., 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21, 456–463.
- Sun, M., Gao, L., Liu, Y., et al., 2012. Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China. *PLoS ONE* 7, e36577.
- Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 14, 178–192.
- Valmary-Degano, S., Jacquin, E., Pretet, J.L., Monnier, F., Girardo, B., Arbez-Gindre, F., Joly, M., Bosses, J.F., Kantelip, B., Mougou, C., 2013. Signature patterns of human papillomavirus type 16 in invasive anal carcinoma. *Hum. Pathol.* 44, 992–1002.
- Vartanian, J.P., Guetard, D., Henry, M., Wain-Hobson, S., 2008. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* 320, 230–233.
- Walboomers, J.M., Meijer, C.J., 1997. Do HPV-negative cervical carcinomas exist? *J. Pathol.* 181, 253–254.

- Wang, L., Dai, S.Z., Chu, H.J., Cui, H.F., Xu, X.Y., 2013. Integration sites and genotype distributions of human papillomavirus in cervical intraepithelial neoplasia. *Asian Pac. J. Cancer Prev.* 14, 3837–3841.
- Wang, Z., Wakae, K., Kitamura, K., Aoyama, S., Liu, G., Koura, M., Monjurul, A.M., Kukimoto, I., Muramatsu, M., 2014. APOBEC3 deaminases induce hypermutation in human papillomavirus 16 DNA upon beta interferon stimulation. *J. Virol.* 88, 1308–1317.
- Watts, K.J., Thompson, C.H., Cossart, Y.E., Rose, B.R., 2002. Sequence variation and physical state of human papillomavirus type 16 cervical cancer isolates from Australia and New Caledonia. *Int. J. Cancer* 97, 868–874.
- Wentzensen, N., Vinokurova, S., von Knebel Doeberitz, M., 2004. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res.* 64, 3878–3884.
- Xu, B., Chotewutmontri, S., Wolf, S., Klos, U., Schmitz, M., Durst, M., Schwarz, E., 2013. Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS ONE* 8, e66693.
- Yuan, H., Myers, S., Wang, J., et al., 2012. Use of reprogrammed cells to identify therapy for respiratory papillomatosis. *N. Engl. J. Med.* 367, 1220–1227.