Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Genomics

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

Original Article

Characterization of HPV integration, viral gene expression and *E6E7* alternative transcripts by RNA-Seq: A descriptive study in invasive cervical cancer

Ayslan C. Brant[a,b,1], Albert N. Menezes[c,1], Shayany P. Felix[a], Liz M. de Almeida[d], Michael Sammeth[e], Miguel A.M. Moreira[a,*]

[a] *Genetics Program, Instituto Nacional de Câncer (INCA), Rio de Janeiro, Brazil*
[b] *Post-Graduate Program in Genetics, Universidade Federal do Rio de Janeiro (UFRJ), Brazil*
[c] *Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, UK*
[d] *Department of Population Research, Instituto Nacional de Câncer (INCA), Rio de Janeiro, Brazil*
[e] *Department of Bioinformatics, Transcriptomics and Functional Genomics, Federal University of Rio de Janeiro (UFRJ), Brazil*

ARTICLE INFO

Keywords:
HPV expression
HPV genome integration
E6E7 alternative transcripts
Alternative splicing
Cervical cancer

ABSTRACT

Scarce data are available on the expression of papillomavirus genome and the frequency of alternatively spliced *E6E7* mRNAs in invasive cervical cancer. We carried out a comprehensive characterization of HPV expression by RNA-Seq analysis in 22 invasive cervical cancer with HPV16 or HPV18, characterizing the presence of integrated/episomal viral DNA, the integration sites in human genome and the proportion of alternative splicing products of *E6* and *E7* genes. The expression patterns suggested the presence of episomal and/or integrated viral DNA, with integration detected in most tumors, frequently occurring within human genes in HPV18+ and in intergenic regions in HPV16+ tumors. Alternative splicing of *E6E7* transcripts showed *E6*I* as the most frequent isoform for both viral types, followed by *E6*II* and *E6/E7* (unspliced) transcripts in HPV16+, and by *E6/E7* in HPV18+ tumors. Previously described *E6*VI* and *E6*V* transcript isoforms for HPV16, and *E6*X* for HPV18, were rare or not detected.

1. Introduction

Cervical cancer is the fourth most frequent cancer in women worldwide, with approximately 570,000 new cases estimated for 2018 [1]. The presence of high-risk human papillomavirus (HR-HPV), mainly HPV16 and HPV18 viral types, is a predetermined, albeit insufficient, event for developing cervical cancer [2]. HPV is a non-enveloped virus with an icosahedral capsid and an approximately 8Kb circular DNA genome comprising three regions: (i) the early expressed region composed of six relevant genes for viral replication (*E1*, *E2*, *E4*, *E5*, *E6* and *E7*), (ii) the late expressed region, with two genes (*L1* and *L2*) coding for capsid proteins, and (iii) a non-coding region (Long Control Region - LCR) containing the origin of viral DNA replication and the early promoter that controls early gene translation [3,4]. During the initial phase of HPV infection, the E2 viral protein modulates early gene expression [5], but when infected cells migrate to the upper layers of the epithelium, early genes become downregulated and late genes overexpressed, allowing the release of new virions [6,7].

Along the carcinogenic process, HPV DNA usually integrates into the host genome, mainly by splitting its genome within the *E1* or *E2* regions [8]. This rupture results in loss of *E2* function and increased expression and translation of the *E6* and *E7* genes that code for the major viral oncoproteins, subsequently leading to inactivation of TP53 and pRB, the key proteins for cell cycle control [9,10]. Besides HPV genome integration, *de novo* methylation of the E2 binding sites at LCR can also modulate the transcription of *E6* and *E7* [11].

Although partial evidence for the presence of episomal viral genomes in invasive cancer has been reported, previous studies have shown that viral DNA is predominantly integrated into the host genome [12,13]. Integration may occur in different chromosomes and regions of the human genome [14,15], although several reports showed that it preferentially occurs at fragile sites, transcriptionally active regions and those recurrently involved in translocation events [16–21]. Analysis of chimeric mRNAs containing viral and human genomic transcripts

* Corresponding author at: Genetics Program, Instituto Nacional de Câncer, André Cavalcanti 37, Rio de Janeiro, RJ 20231-050, Brazil.
E-mail addresses: lalmeida@inca.gov.br (L.M. de Almeida), micha@sammeth.net (M. Sammeth), miguelm@inca.gov.br (M.A.M. Moreira).
[1] These authors contributed equally to the present work.

https://doi.org/10.1016/j.ygeno.2018.12.008
Received 19 November 2018; Received in revised form 6 December 2018; Accepted 7 December 2018

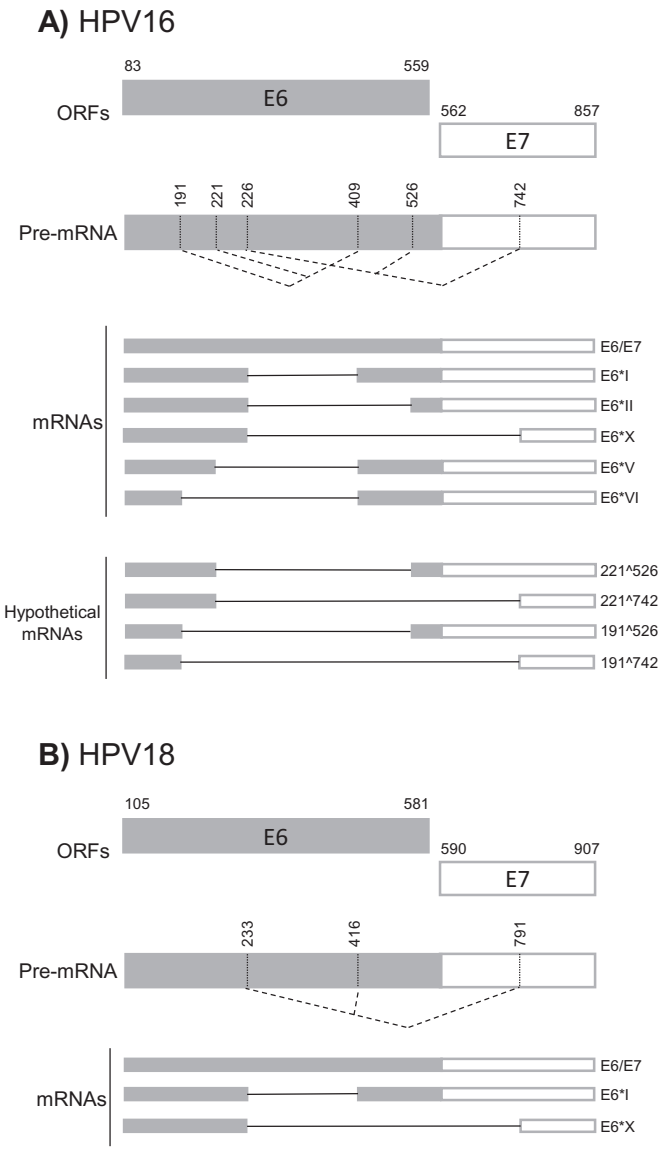

Fig. 1. Representation of *E6E7* alternative mRNAs for HPV16 and HPV18. The *E6* and *E7* open read frames (ORFs) are respectively represented in gray and white boxes, the first and the last genomic position of each ORF are indicated. Dashed lines at the pre-mRNA represent the splicing events between donor splicing site (5′ss) and acceptor splicing site (3′ss) of previously reported mRNAs. (A) HPV16 *E6E7* pre-mRNA contains three 5′ss (at positions 191, 221 and 226) and three 3′ss (at positions 409, 526 and 742). Until now six mRNAs products originated from *E6E7* pre-mRNA were reported, being five spliced products (*E6*I*, *E6*II*, *E6*X*, *E6*V* and *E6*VI*) and one unspliced (*E6/E7*). Additional four mRNAs can be hypothesized, with the splicing junctions: 221^526, 221^742, 191^526 and 191^742. (B) The HPV18 *E6E7* pre-mRNA contains one 5′ss (nt 233) and two 3′ss (nt 416 and 791), being reported three alternative mRNAs: *E6/E7* (unspliced), *E6*I* and *E6*X*.

allows for the characterization of integration sites [22].

In HR-HPVs, *E6* and *E7* genes are transcribed in a single, polycistronic pre-mRNA. However, within *E6* ORF there is a region that alternatively function as an intron or an exon, and the removal of this region (intron 1) from the pre-mRNA by splicing is necessary for E7 protein translation, while the retention is required for E6 translation [4,23]. Previous reports showed a single donor splicing site (5′ss) at genomic position 233, and two acceptor splicing sites (3′ss) at genomic positions 416 and 791 inside the *E6* and *E7* region of HPV18 pre-mRNA [24,25]. Conversely, in HPV16 pre-mRNA, three alternative 5′ss have been found at genomic positions 191, 221, and 226, and three 3′ss at genomic positions 409, 526, 742 [4,24]. In HPV16+ cells, a total of six *E6E7* alternative transcripts and their respective splicing junctions (in parentheses) have been reported: *E6*I* (226^409), *E6*II* (226^526), *E6*X* (226^742), *E6*V* (221^409), *E6*VI* (191^409) in addition to the *E6/E7* transcript containing the full *E6* and *E7* ORFs in the mature mRNA [24,26] (Fig. 1A). On the other hand, only three alternative transcripts have been so far reported in HPV18+ cells, *E6*I* (233^416), *E6*X* (233^791) and the *E6/E7* unspliced transcript [23,25,27] (Fig. 1B). Based on the presence of the three functional 5′ss and the three functional 3′ss in HPV16 *E6E7* pre-mRNA, the presence of another four alternative transcripts with splicing junctions, 191^526, 191^742, 221^526 and 221^742, may be hypothesized (Fig. 1A).

Scarce data are available on the expression pattern of the viral genome and the frequency of alternatively spliced viral mRNA in invasive cancer samples associated to HPV16 or HPV18 infection because most findings have been reported in cell lines, cervical pre-neoplastic lesions and a limited number of invasive cancers [24–26,28]. In this study, we carried out an exploratory and comprehensive characterization of HPV expression by RNA-Seq analysis in cervical cancer samples with single HPV16 and HPV18 infections. This approach allowed us to characterize the expression patterns of the viral genome with respect to the presence of integrated/episomal viral DNA and to estimate the proportion of the different splicing products encompassing the *E6* and *E7* ORFs.

2. Materials and methods

2.1. Samples and nucleic acid isolation

Twenty-two biopsies of invasive cervical carcinomas from patients admitted to Instituto Nacional de Câncer (INCA, Brazil), comprising 12 samples of HPV16+ tumors and 10 samples of HPV18+ tumors, were analyzed (Additional File 1). Samples were randomly selected from a set of previously characterized tumors with single HPV16 and HPV18 infections [29]. All biopsies were collected before treatment and stored in RNAlater (Qiagen) at −80 °C. Additionally, the CaSki cervical cancer cell line (HPV16+), grown in RPMI with 10% fetal bovine serum (Gibco) and 5% CO_2, was also analyzed. DNA and RNA were isolated from each sample and CaSki cells with Qiagen Allprep DNA/RNA mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. DNA and RNA were quantified and stored at −20 °C and −80 °C, respectively.

2.2. High throughput mRNA sequencing

Total RNA (0.1 to 4 μg) was used for library preparation with TruSeq RNA Sample Prep Kit (Illumina, U.S.A.) according to the manufacturer's recommendations. Libraries were run in an Illumina HiSeq 2500 platform, with paired-end sequencing strategy.

2.3. Sequence data preprocessing, quality control and read mapping

Fastq files, comprising demultiplexed reads, were obtained with Casava 1.8 [30] allowing a single barcode mismatch. Reads were quality filtered by mean phred score (< 20) and length (< 30 nucleotides) using PRINSEQ [31]. Filtered reads were mapped to HPV reference sequences (HPV16 - GenBank K02718.1 and HPV18 - AY262282.1) with BowTie 2 [32]. Nucleotide coverage per position was estimated with GATK DepthOfCoverage [33] and coverage maps were plotted with R [34]. Samples with < 500 reads mapped to the HPV reference were discarded from further analyses (Additional File 2).

2.4. HPV gene expression analysis (heatmap)

To certify that the amount of mature mRNAs coding for complete proteins was accurately estimated and to avoid regions shared by

alternative transcripts, the expression of two or three regions of larger genes (*E6*, *E7*, *E1*, *E2*, *L1* and *L2*) was separately estimated (Additional File 3). For each gene, the partial region with the lowest mean coverage among samples was selected as representative of ORF expression. These regions, together with the complete *E4* and *E5* ORFs, were subsequently used as reference for mapping and counting reads with Subread aligner [35]. However, when a complete ORF of a given gene was not covered by reads in at least 80% of its length or uncovered regions encompassed the first 100 bp in its 3′ partition, count estimates for the given gene in the sample was assumed to be 0 (zero).

Read count values were used to build a matrix (sample *versus* gene), and values were normalized by estimating the ratio between the read count of each ORF and the total number of HPV reads of a sample. Log_2 of these ratios were calculated and used to generate a heatmap in R [34]; when count = 0 (zero) or considered to be zero, the ratio was assumed to be 0.00001.

2.5. Analysis of alternative transcripts of E6 and E7

In order to verify the presence and quantify of alternative *E6* and *E7* transcripts, reads were mapped with BowTie 2 [32] to references of manually constructed alternative transcripts. These sequences were created based on all possible *E6* and *E7* splicing junctions in coding region of HPV16 (191^409, 191^526, 191^742, 221^409, 221^526, 221^742, 226^409, 226^526 and 226^742) and HPV18 (233^416 and 233^731) (Additional File 4). Additionally, a concatenated nucleotide sequence comprising the entire *E6* and *E7* genes was used as reference for the unspliced *E6/E7* transcript.

To verifying the occurrence of each transcript resulting from alternative splicing and avoiding methodological mapping artefacts, the usage of each splicing junction was estimated from the number of reads covering each junction. Splicing junction reads (SJRs) were identified by (i) at least six contiguous nucleotides flanking each border of an exon/exon junction in a given read (Additional File 5), and (ii) a maximum of three mismatches along a read. Only splicing junctions covered by at least six SJRs were considered true for each sample [22]. The number of reads covering the 5′ exon/intron junctions overlapping nucleotides 226/227 of HPV16 and 233/234 of HPV18 was used for quantifying expression of *E6/E7* unspliced transcripts; reads were considered true following the same criteria for SJRs.

2.6. Identification of viral integration sites in the human genome

A *de novo* transcriptome assembly was created with Trinity [36] for identifying integration sites. A two-step approach was used for identifying chimeric transcripts sharing human and HPV sequences on assembled contigs. Firstly, contigs were compared with their corresponding HPV reference sequence with Blast tools in the stingray@galaxy platform [37], and contigs with alignment scores with e-values $< 1 \times 10^{-6}$ and identity score $\geq 98\%$ were selected. Subsequently, these selected contigs were aligned against the human genome (GRCh38) with Blast online (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and used as reference for mapping RNA-Seq reads with BowTie 2 [32]. Read coverage of each chimeric sequence was estimated based on the number of reads overlapping at least six nucleotides in both directions (Additional File 5) at the borders of each HPV/human junction.

To validate the presence of the selected chimeric transcripts (Additional File 6), RT-PCR (reverse transcription-polymerase chain reaction) assays were carried out using primers aligning the HPV and human genomic regions of chimeric contigs (Additional File 7). RNA was treated with RQ1 RNase-Free DNase (Promega) and cDNA was synthesized with SuperScript® II Reverse Transcriptase (Thermo Scientific) using random primers. Each reaction was performed in a final volume of 25 µL containing 0.2 mM of each dNTP, 25 pmol of each primer, 1 U of Platinum *Taq I* DNA Polymerase (Life Technologies), 1 × reaction buffer and 2 mM $MgCl_2$. Reaction cycling condition was 95 °C for 5 min followed by 35 cycles of 95 °C for 30 s, annealing temperature (see Additional File 7) for 30 s, extension at 72 °C for 40 s, and final extension at 72 °C for 5 min. Amplicons were purified with GFX PCR DNA and Gel Band Purification kit (GE Healthcare) and sequenced in an ABI Prism® 3130xl platform.

2.7. Analysis of E1/E2 integrity

Disruptions of the HPV genome resulting from integration were also characterized by PCR amplifications of *E1* and *E2* ORFs in HPV16+ and HPV18+ tumor DNA as previously reported [38,39]. The presence of amplified products was checked by electrophoresis in 1.5% ultrapure agarose gels. The integrity of each region was assumed when all PCR amplicons were obtained for a given sample, while lack of amplification of any PCR product was indicative of viral DNA integration [40–42]. Due to possible PCR bias, all PCR reactions without the expected amplicon were repeated with the respective primers pair.

3. Results

Analyzes of RNA-Seq data showed a median of 44,377,920 quality filtered reads (Phred score > 20) from 12 HPV16+ tumor samples and the CaSki cell line, and 47,967,862.5 from 10 HPV18+ tumors. The majority of HPV16+ reads (77.4% to 83.8%) mapped to the human genome (GRCh38), while 0% to 0.125% mapped to the HPV16 genomic reference (K02718.1) (Additional File 2). Two samples (HPV16_8 and HPV16_11) were discarded from further analyses due to their low number of HPV-mapped reads. On the other hand, 80.7% to 85.6% of HPV18+ samples mapped to the human genome (GRCh38), and 0.004% to 0.042% to the HPV18 genomic reference (AY262282.1) (Additional File 2).

3.1. HPV gene expression patterns

HPV16 and HPV18 expression patterns in tumors and CaSki cells are illustrated in heatmaps (Fig. 2A and B, respectively) based on read counts of selected regions with potentially translatable mRNAs of complete *E6*, *E7*, *E1*, *E2*, *E4*, *E5*, *L2* and *L1* ORFs. Regions with overlapping ORFs (like between *E2* and *E4*) and regions present in multiple splicing isoforms were disconsidered.

In most of samples of either HPV type, *E7* showed the highest expression with respect to other genes (Fig. 2). On the other hand, the *E6* expression pattern differed between HPV16+ and HPV18+ samples, being highly variable among the former but uniform among the latter. Heatmaps showed two different sets of HPV16+ samples with respect to *E2* and *E4* expression (Fig. 2A) and two sets of HPV18+ samples with respect to *E2* expression (Fig. 2B). On the other hand, heatmaps did not show association between sample sets, tumor histology or viral lineage (Additional File 1).

E5 was found to be expressed in 4/11 HPV16+ samples and 7/10 HPV18+ samples. Low expression of late genes (*L1* and/or *L2*) was detected in 4/11 HPV16+ and 3/10 HPV18+ samples (Fig. 2 and Additional File 3).

3.2. HPV genome integration

A PCR-based strategy for detecting HPV genome disruption in *E1* or *E2* ORFs indicated that 6/11 HPV16+ and 9/10 HPV18+ samples were disrupted in tumors and CaSki cells (Table 1, and Additional File 8). On the other hand, RNA-Seq data analysis showed evidence of chimeric transcripts in 8/11 HPV16+, 9/10 HPV18+ samples and CaSki cells, indicating HPV integration into the host genome. A comparison between these two approaches showed discordant estimates of viral disruption in four HPV16+ samples (HPV16_1, HPV16_5, HPV16_7, HPV16_12) and CaSki cells.

The identification of chimeric transcripts provided information on

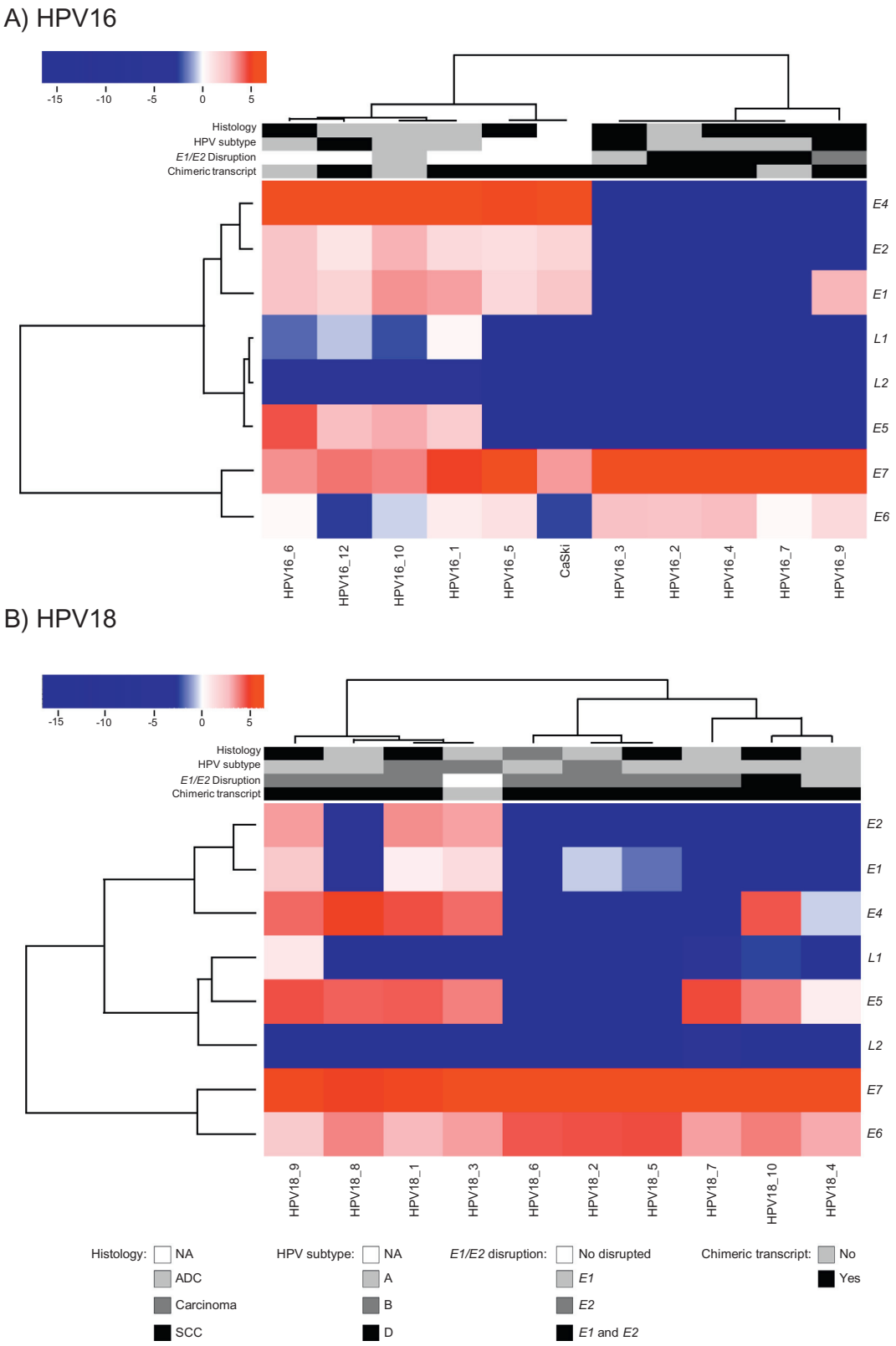

Fig. 2. Heatmaps of HPV genome expression in invasive cervical cancer infected by HPV16 (A) or HPV18 (B). The viral genome expression was estimated by the number of reads counted for a partial region (for *E6*, *E7*, *E1*, *E2*, *L1* and *L2*) or the full (for *E4* and *E5*) ORFs. The heatmaps were correlated with: (1) the tumor histological classification, as adenocarcinoma (ADC), Squamous Cell Carcinoma (SCC) or carcinoma; (2) the HPV lineage (A, B or D); (3) presence of disruption at *E1* and/or *E2* viral genes by PCR-based methodology; and (4) presence of HPV/human chimeric transcripts detected by RNA-Seq analysis. NA – not available.

Table 1
Integration analyses of HPV16+ or HPV18+ in invasive cervical samples by identification of chimeric transcripts using RNA-Seq or by PCR-based methodology (*E1* or *E2* disruption).

Samples	Chimeric transcripts					*E1* or *E2* disruption
	HPV breakpoint (ORF)	Human breakpoint	Locus (Gene)	Coverage [a]	PCR/Sanger [b]	
HPV16_1	880 (*E1*)	150,577,491	1q21.2 (*MCL1*)	679	+	–
HPV16_2	880 (*E1*)	206,950,242	1q32.1	33	+	*E1*, *E2*
HPV16_3	880 (*E1*)	138,685,418	4q31.1	361	+	*E1*
HPV16_4	880 (*E1*)	33,714,736	20q11.22 (*PXMP4*)	280	+	*E1*, *E2*
HPV16_5	880 (*E1*)	73,145,556	13q22.1	191	+	–
	3637 (*E2*)	89,010,656	Xq21.31	39	+	
HPV16_6	ND	ND	ND	ND	ND	–
HPV16_7	ND	ND	ND	ND	ND	*E1*, *E2*
HPV16_9	880 (*E1*)	102,768,320	11q22.2	618	+	*E2*
HPV16_10	ND	ND	ND	ND	ND	*E1*
HPV16_12	4189: 4215 (*L2*)	41,480,536: 41,480,562	9p11.2	3	–	–
CaSki	3727: 3729 (*E2*)	45,691,386: 45,691,388	6p21.1	4041	+	–
HPV18_1	929 (*E1*)	58,225,822	15q21.3 (*ALDH1A2*)	233	+	*E2*
HPV18_2	929 (*E1*)	189,889,340	3q28 (*TP63*)	230	+	*E2*
HPV18_3	ND	ND	ND	ND	ND	–
HPV18_4	673: 680 (*E7*)	102,952,168: 102,952,175	11q22.2 (*MMP13*)	209	–	*E1*
HPV18_5	929 (*E1*)	230,037,531	2q36.3 (*SLC16A14*)	989	+	*E2*
HPV18_6	3053: 3062 (*E2*)	49,758,595: 49,758,598	20q13.13	334	+	*E2*
HPV18_7[c]	929 (*E1*)	61,520,063	14q23.1 (*PRKCH*)	955	+	*E2*
	3854 (*E2*)	61,519,708	14q23.1 (*PRKCH*)	691	+	
HPV18_8	929 (*E1*)	11,259,003	3p25.3 (*HRH1*)	128	+	*E2*
HPV18_9[c]	929 (*E1*)	68,561,998	14q24.1 (*RAD51B*)	30	+	*E2*
	5117 (*L2*)	68,562,438	14q24.1 (*RAD51B*)	25	+	
HPV18_10[d]	929 (*E1*)	50,024,114	2p16.3 (*NRXN1*)	103	+	*E1*, *E2*
	929 (*E1*)	49,943,791	2p16.3 (*NRXN1*)	178	+	

ND – Not determined.
[a] Chimeric transcript coverage was obtained from the number of reads covering the integration junction (HPV/human - see method session).
[b] Validation of chimeric transcripts by RT-PCR and Sanger sequencing. Electropherograms shown at Additional File 9.
[c] Two integration junctions in a same chimeric transcript.
[d] Two independents chimeric transcripts.

the site of viral integration (Table 1 and Additional File 6). Integration took place within human genes in eight HPV18+ samples and two HPV16+ samples, and in intergenic regions in seven samples, six HPV16+ and one HPV18+. In most samples, single chimeric transcripts were found. In two samples (HPV18_7 and HPV18_9), the single chimeric transcripts showed HPV-flanked human sequences, indicating that two closeby integration events had occurred within the same gene in each sample. More than one chimeric transcript was found in two samples, HPV16_5 and HPV18_10. HPV16_5 was the only sample with chimeric transcripts resulting from two integration events in different human chromosomes (13 and the X chromosome), while HPV18_10 showed two chimeric transcripts with different fragments of the *NRXN1* gene. Finally, in two other samples, HPV16_12 and HPV18_4, chimeric transcripts could not be confirmed by RT-PCR (reverse transcription-polymerase chain reaction; Table 1 and Additional File 9).

Chimeric transcripts contained different HPV genomic regions at the HPV/human borderline. In six HVP16+ samples, nucleotide 880 of the viral genome, located in the *E1* gene, was recurrently found at the HPV/human borderline, while in HPV18+ samples, nucleotide 929, also within *E1*, was the most frequent viral position at the HPV/human borderline (in 7 samples; Table 1). In other samples herein analyzed, positions located within *E2*, *E7* and *L2* were found at the HPV/human borderline.

3.3. Alternative splicing along E6 and E7 ORFs

In all HPV16+ and all HPV18+ samples, the alternative *E6*I* transcript (with respectively splicing junctions 226ˆ409 and 233ˆ416) was the most highly expressed (Tables 2 and 3, respectively), ranging from 72.02% to 89.24% in HPV16+ samples, and from 72.06% to 94.88% in HPV18+ samples. In HPV16+ samples, the second and third most highly expressed transcripts were *E6*II* (with 226ˆ526 splicing junction) and the unspliced *E6/E7* transcript. *E6*II* was the second most highly expressed transcript in seven samples (HPV16_1, HPV16_2, HPV16_5, HPV16_7, HPV16_9, HPV16_10, HPV16_12) and CaSki cells, although with a tenfold lower expression than the *E6*I* transcript. *E6/E7* was the second most highly expressed transcript in three samples (HPV16_3, HPV16_4, HPV16_6). Transcripts with the 226ˆ742 junction (*E6*X*), with more than six splicing junction reads (SJRs – see methods), were found in three samples (HPV16_1, HPV16_4, HPV16_7) and CaSki cells, while *E6*VI* (with the 191ˆ409 junction) was only found in CaSki cells with more than six SJRs. Transcripts with junction 221ˆ409 (*E6*V*), although present in three samples, were not covered by at least six SJRs, while the four hypothetical transcripts with junctions 191ˆ526, 191ˆ742, 221ˆ526 and 221ˆ742 were not detected in any sample (Table 2).

In HPV18+ samples the *E6/E7* unspliced transcript was the second most highly expressed in all samples, ranging from 5.12% to 31.42% (Table 3), while transcripts with splicing junction 233ˆ791 (*E6*X*) were not detected.

4. Discussion

In this study we analyzed the expression of the HPV genome in cervical cancer associated with single HPV16 and HPV18 infections. The proportion of HPV RNA-Seq reads varied considerably between samples, a finding that might reflect different proportions of tumor and normal cells in each of them, a variable not taken in consideration in this study. This might explain the low number of HPV reads in HPV16_8 and HPV16_11 samples (with 25 and 17 reads, respectively). On the other hand, differences in expression levels may also reflect actual differences in the amount of HPV mRNA reads.

PCR-assays and RNA-Seq analysis showed discordant estimates of the presence of integrated and episomal viral DNA. In fact, both

Table 2
Number of reads covering splicing junctions and the frequency (%, between parentheses) of *E6E7* alternative transcripts for HPV16+ samples.

Splicing junctions	HPV16_1	HPV16_2	HPV16_3	HPV16_4	HPV16_5	HPV16_6	HPV16_7	HPV16_9	HPV16_10	HPV16_12	CaSki
191ˆ409 (*E6*VI*)	1	0	0	0	1	0	0	0	0	0	7
	(0,07)	(0)	(0)	(0)	(0,09)	(0)	(0)	(0)	(0)	(0)	(0,09)
191ˆ526	0	0	0	0	0	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
191ˆ742	0	0	0	0	0	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
221ˆ409 (*E6*V*)	0	0	1	0	0	0	1	0	0	0	5
	(0)	(0)	(0,17)	(0)	(0)	(0)	(0,04)	(0)	(0)	(0)	(0,07)
221ˆ526	0	0	0	0	0	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
221ˆ742	0	0	0	0	0	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
226ˆ409 (*E6*I*)	1317	160	523	853	950	157	1991	626	578	163	6431
	(87,05)	(87,43)	(87,17)	(79,87)	(84,15)	(72,02)	(89,24)	(88,29)	(77,07)	(79,13)	(85,96)
226ˆ526 (*E6*II*)	133	16	22	85	112	25	170	63	123	24	837
	(8,79)	(8,74)	(3,67)	(7,96)	(9,92)	(11,47)	(7,62)	(8,89)	(16,40)	(11,65)	(11,19)
226ˆ742 (*E6*X*)	13	0	0	12	2	3	41	4	2	5	90
	(0,86)	(0)	(0)	(1,12)	(0,18)	(1,38)	(1,84)	(0,56)	(0,27)	(2,43)	(1,20)
E6/E7[a]	49	7	54	118	64	33	28	16	47	14	111
	(3,24)	(3,83)	(9,00)	(11,05)	(5,67)	(15,14)	(1,26)	(2,26)	(6,27)	(6,80)	(1,48)
Total	1513	183	600	1068	1129	218	2231	709	750	206	7481

[a] Number of reads covering the exon-intron juxtaposition (nucleotides 226/227) are shown for *E6/E7* unspliced transcripts.

methods are limited because they do not allow the identification of the simultaneous presence of episomal and integrated viral DNA, and are incapable of identifying integration of multiple copies of the viral genome. On the other hand, disruptions of viral DNA outside the *E1* or *E2* ORFs cannot be detected by the PCR-based approach. A comparison between these approaches indicated that the identification of chimeric reads by RNA-Seq provided a more reliable estimate of viral integration. The identification of chimeric transcripts without *E1* or *E2* disruption detected by PCR (in four HPV16+ samples), might indicate either simultaneous presence of episomal and integrated viral genomes or integration of multiple HPV copies (in tandem or with disruption at different sites).

The chimeric transcripts showed that the donor splicing sites in *E1*, at position 880 of HPV16 and 929 of HPV18, were recurrently present at the boundaries between HPV/human sequences. These findings did not indicate that disruption of the viral genome occurred at these positions; it rather indicated that the chimeric transcript resulted from a splicing event between viral sites 880 or 929 with a nearby acceptor splicing site in the human genome. Altogether, our data corroborated that these positions in HPV16 or HPV18 were strong donor splicing sites, unrestricted by the HPV genomic context, and used by these viral types to produce mature mRNAs for E1ˆE4 protein translation [28,43,44]. These findings also indicated that the rupture of the viral genome occurred at positions downstream of these sites, as previously reported [45,46].

Chimeric transcripts showed that integration occurred in different chromosomes, most frequently at intergenic regions by HPV16, and within genes by HPV18 (Table 1). Some reports showed that integrations frequently occurred inside or next to genes associated with oncogenesis [22,47], while, in this study, integration occurred within 10 different human genes, some of which previously reported, and other ones considered to be tumor suppressors (*TP63* and *RAD51B*) [18,46,48–51].

The methods herein carried out pointed to the exclusive presence of episomal DNA in a single HPV16+ sample (HPV16_6). This was because (i) its expression pattern showed transcripts of all viral genes, (ii) PCR assays did not provide evidence of *E1/E2* disruption, and (iii) chimeric transcripts were not detected. Heatmap analysis (Fig. 2) pointed to the presence of episomal HPV genomes in HPV16_12, HPV16_10 and HPV16_1 samples because RNA-Seq reads covered all viral genes, although the presence of chimeric transcripts in HPV16_1 and HPV16_12 suggested simultaneous presence of integrated viral DNA. In HPV16_10, PCR assays suggested disruption of the viral genome although chimeric transcripts could not be detected, a reason why viral integration could not be demonstrated. In the remaining seven samples, the expression pattern of the viral genes, presence of chimeric transcripts and/or disruption of the viral genome at *E1/E2* were consistent with the exclusive presence of an integrated viral genome. It should be stressed, however, that the simultaneous presence of episomal and integrated viral genomes could not be distinguished from multiple copy, tandem integrations [52–55] because expression of all viral genes and presence of chimeric transcripts would be detectable in both cases.

Analyzes of HPV18+ tumors were consistent in detecting viral integration in 9/10 samples (Table 1), pointing to a higher frequency of integration of HPV18 with respect to HPV16, as previously reported in almost all HPV18+ tumors and approximately 80% of HPV16+ tumors [8,56]. All samples, except HPV18_3, were concordant in presenting

Table 3
Number of reads covering splicing junctions and the frequency (%, between parentheses) of *E6E7* alternative transcripts for HPV18+ samples.

Splicing junctions	HPV18_1	HPV18_2	HPV18_3	HPV18_4	HPV18_5	HPV18_6	HPV18_7	HPV18_8	HPV18_9	HPV18_10
233ˆ416 (*E6*I*)	1189	330	181	204	1581	559	2614	467	97	405
	(86,28)	(77,65)	(80,80)	(94,88)	(76,97)	(74,63)	(88,70)	(68,58)	(92,38)	(72,06)
233ˆ791 (*E6*X*)	0	0	0	0	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
E6/E7[a]	189	95	43	11	473	190	333	214	8	157
	(13,72)	(22,35)	(19,20)	(5,12)	(23,03)	(25,37)	(11,30)	(31,42)	(7,62)	(27,94)
Total	1378	425	224	215	2054	749	2947	681	105	562

[a] Number of reads covering the exon-intron juxtaposition (nucleotides 233/234) are shown for *E6/E7* unspliced transcripts.

chimeric transcripts and *E1/E2* disruption detected by PCR assays. Samples HPV18_6, HPV18_2 and HPV18_5 showed an exclusive expression pattern of viral integration, suggesting *E1* rupture in HPV18_6 and *E2/E4* rupture in the two other samples. Conversely, in the remaining samples, *E5* expression was detected despite *E1/E2* disruption. Discordance between expression patterns and disruptions might be a result of viral integration within human genes, consequently to which the presence of *E5* reads probably resulted by throughout transcription from a human gene promoter to the viral integrated genome (Additional File 10). HPV18_3 was the only HPV18+ sample lacking chimeric transcripts as well as *E1/E2* disruption, suggesting that the viral genome was exclusively present as an episome, without expression of late genes.

Analyses of alternative splicing isoforms comprising the *E6* and *E7* genes (Tables 2 and 3) showed that *E6*I* was the most abundant transcript in HPV16 and HPV18, resulting from a splicing event that allowed *E7* translation. This finding coincided with expression patterns (Fig. 2) showing that *E7* was more highly expressed than *E6*.

In most HPV16 samples, the proportion of *E6*I* was approximately tenfold higher than *E6*II*, as previously reported [23,26,57] but we also found two other transcripts, *E6*VI* (191^409) and *E6*V* (221^409) previously reported by Ajiro et al. [24]. Transcript *E6*VI* was found in CaSki cells, and with less than six SJRs in two samples (HPV16_1 and HPV16_5), while transcript *E6*V* was found only in CaSki cells, and with < 6 SJRs in other two samples (in HPV16_3 and HPV16_7). CaSki cells were the only ones with > 6 SJR for *E6*VI* and *E6*V* and with the highest number of sequenced reads (Additional File 2), suggesting that these transcripts were weakly expressed. In HPV18+ samples, transcript *E6*X*, identified by Ajiro and Zheng [25], was not detected (Table 3). Our findings suggest that there is no difference in the proportion of different isoforms of *E6E7* transcripts among tumoral samples containing episomal and/or integrated HPV genome, in agreement with previous works [58,59].

5. Conclusion

The present approach allowed the characterization of HPV16 and HPV18 expression patterns in invasive cervical tumors and identified potential carriers of viral episomes. HPV18 was found to be recurrently integrated within genes. With respect to alternative splicing isoforms encompassing the major viral oncogenes (*E6* and *E7*), the *E6*I* transcript, allowing E7 translation, was the most frequently expressed in all HPV16+ and HPV18+ samples, while other alternative transcripts showed low levels of expression or could not be detected; their identification requiring deep RNA-Seq read coverage.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2018.12.008.

Ethics approval and consent to participate

This work was approved by the Institutional Research Ethics Committee of Instituto Nacional de Câncer and by Comissão Nacional de Ética em Pesquisa (CONEP – Brazil; C.A.A.E. 53,398,416.0.0000.5274). All patients signed an informed consent authorizing access and use of clinical data for scientific purposes.

Availability of data and materials

Data generated for this work (RNA-Seq reads - fastq files- of each samples, fasta files of chimeric transcripts and HPV16 and HPV18 reference sequences for *E6E7* alternative transcripts) were deposited in Gene Expression Omnibus (GEO - http://www.ncbi.nlm.nih.gov/geo) with access number GSE91065.

Disclosure of interests

The authors declare that they do not have competing interests.

Funding

This work was supported by National Institute for Cancer Control (INCT para Controle do Câncer; http://www.inct-cancer-control.com.br), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil; grant numbers: 305873/2014-8 and 573806/2008-0), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ, Brazil; grant number: E26/170.026/2008), Ministry of Health, Brazil, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil), and Pan-American Health Organization (PAHO).

Authors' contributions

Conceived and designed the experiments (A.C.B., M.S. and M.A.M.M.); prepared NGS libraries (A.C.B.); carried out bioinformatics analyses (A.C.B. and A.N.M.); isolated nucleic acids, carried out validation of chimeric transcripts and characterized HPV disruption (A.C.B. and S.P.F.); identified HPV genotypes and lineages and participated in the collection of tumor samples (S.P.F.); coordinated the collection of tumor samples and clinical data (L.M.A.); wrote the manuscript (A.C.B., A.N.M and M.A.M.M.); all authors approved the final version of the manuscript.

Acknowledgements

This work is part of A.C.B.´s Ph.D thesis pursued at the Genetics Post-Graduate Program (PGGEN) of the Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil. The authors are grateful to Carolina Furtado for NGS sequencing and valuable discussion on RNA-Seq strategies and to Hector N. Seuanez for a critical revision of the manuscript.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. (2018), https://doi.org/10.3322/caac.21492.

[2] N. Muñoz, F.X. Bosch, S. de Sanjosé, R. Herrero, X. Castellsagué, K.V. Shah, P.J.F. Snijders, C.J.L.M. Meijer, International Agency for Research on Cancer Multicenter Cervical Cancer Study Group, Epidemiologic classification of human papillomavirus types associated with cervical cancer, N. Engl. J. Med. 348 (2003) 518–527, https://doi.org/10.1056/NEJMoa021641.

[3] M. Tommasino, The human papillomavirus family and its role in carcinogenesis, Semin. Cancer Biol. 26 (2014) 13–21, https://doi.org/10.1016/j.semcancer.2013.11.002.

[4] Z.-M. Zheng, C.C. Baker, Papillomavirus genome structure, expression, and post-transcriptional regulation, Front. Biosci. 11 (2006) 2286–2302.

[5] G. Steger, S. Corbach, Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein, J. Virol. 71 (1997) 50–58.

[6] S. Cheng, D.C. Schmidt-Grimminger, T. Murant, T.R. Broker, L.T. Chow, Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes, Genes Dev. 9 (1995) 2335–2349.

[7] M. Hummel, J.B. Hudson, L.A. Laimins, Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes, J. Virol. 66 (1992) 6070–6080.

[8] C.B.J. Woodman, S.I. Collins, L.S. Young, The natural history of cervical HPV infection: unresolved issues, Nat. Rev. Cancer 7 (2007) 11–22, https://doi.org/10.1038/nrc2050.

[9] M. Scheffner, B.A. Werness, J.M. Huibregtse, A.J. Levine, P.M. Howley, The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53, Cell 63 (1990) 1129–1136.

[10] S.N. Boyer, D.E. Wazer, V. Band, E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway, Cancer Res. 56 (1996) 4620–4624.

[11] A. Chaiwongkot, S. Vinokurova, C. Pientong, T. Ekalaksananan, B. Kongyingyoes, P. Kleebkaow, B. Chumworathayi, N. Patarapadungkit, M. Reuschenbach, M. von Knebel Doeberitz, Differential methylation of E2 binding sites in episomal and

integrated HPV 16 genomes in preinvasive and invasive cervical lesions, Int. J. Cancer 132 (2013) 2087–2094, https://doi.org/10.1002/ijc.27906.
[12] P. Das, A. Thomas, S. Kannan, K. Deodhar, S.K. Shrivastava, U. Mahantshetty, R. Mulherkar, Human papillomavirus (HPV) genome status & cervical cancer outcome - a retrospective study, Indian J. Med. Res. 142 (2015) 525–532, https://doi.org/10.4103/0971-5916.171276.
[13] Y. Xue, D. Lim, L. Zhi, P. He, J.-P. Abastado, F. Thierry, Loss of HPV16 E2 protein expression without disruption of the E2 ORF correlates with carcinogenic progression, Open Virol J. 6 (2012) 163–172, https://doi.org/10.2174/1874357901206010163.
[14] H.M. Walline, C.M. Goudsmit, J.B. McHugh, A.L. Tang, J.H. Owen, B.T. Teh, E. McKean, T.W. Glover, M.P. Graham, M.E. Prince, D.B. Chepeha, S.B. Chinn, R.L. Ferris, S.M. Gollin, T.K. Hoffmann, H. Bier, R. Brakenhoff, C.R. Bradford, T.E. Carey, University of Michigan Head and Neck Specialized Program of Research Excellence (SPORE) program, Integration of high-risk human papillomavirus into cellular cancer-related genes in head and neck cancer cell lines, Head Neck 39 (2017) 840–852, https://doi.org/10.1002/hed.24729.
[15] M. Parfenov, C.S. Pedamallu, N. Gehlenborg, S.S. Freeman, L. Danilova, C.A. Bristow, S. Lee, A.G. Hadjipanayis, E.V. Ivanova, M.D. Wilkerson, A. Protopopov, L. Yang, S. Seth, X. Song, J. Tang, X. Ren, J. Zhang, A. Pantazi, N. Santoso, A.W. Xu, H. Mahadeshwar, D.A. Wheeler, R.I. Haddad, J. Jung, A.I. Ojesina, N. Issaeva, W.G. Yarbrough, D.N. Hayes, J.R. Grandis, A.K. El-Naggar, M. Meyerson, P.J. Park, L. Chin, J.G. Seidman, P.S. Hammerman, R. Kucherlapati, Cancer Genome Atlas Network, Characterization of HPV and host genome interactions in primary head and neck cancers, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 15544–15549, https://doi.org/10.1073/pnas.1416074111.
[16] M.J. Ferber, E.C. Thorland, A.A.T.P. Brink, A.K. Rapp, L.A. Phillips, R. McGovern, B.S. Gostout, T.H. Cheung, T.K.H. Chung, W.Y. Fu, D.I. Smith, Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma, Oncogene 22 (2003) 7233–7242, https://doi.org/10.1038/sj.onc.1207006.
[17] M. Peter, N. Stransky, J. Couturier, P. Hupé, E. Barillot, P. de Cremoux, P. Cottu, F. Radvanyi, X. Sastre-Garau, Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma, J. Pathol. 221 (2010) 320–330, https://doi.org/10.1002/path.2713.
[18] R. Zhang, C. Shen, L. Zhao, J. Wang, M. McCrae, X. Chen, F. Lu, Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration contributes to HPV-related cervical carcinogenesis, Int. J. Cancer 138 (2016) 1163–1174, https://doi.org/10.1002/ijc.29872.
[19] I. Kraus, C. Driesch, S. Vinokurova, E. Hovig, A. Schneider, M. von Knebel Doeberitz, M. Dürst, The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes, Cancer Res. 68 (2008) 2514–2522, https://doi.org/10.1158/0008-5472.CAN-07-2776.
[20] N.C. Popescu, J.A. Dipaolo, Preferential sites for viral integration on mammalian genome, Cancer Genet. Cytogenet. 42 (1989) 157–171.
[21] E.C. Thorland, S.L. Myers, B.S. Gostout, D.I. Smith, Common fragile sites are preferential targets for HPV16 integrations in cervical tumors, Oncogene 22 (2003) 1225–1237, https://doi.org/10.1038/sj.onc.1206170.
[22] A.I. Ojesina, L. Lichtenstein, S.S. Freeman, C.S. Pedamallu, I. Imaz-Rosshandler, T.J. Pugh, A.D. Cherniack, L. Ambrogio, K. Cibulskis, B. Bertelsen, S. Romero-Cordoba, V. Treviño, K. Vazquez-Santillan, A.S. Guadarrama, A.A. Wright, M.W. Rosenberg, F. Duke, B. Kaplan, R. Wang, E. Nickerson, H.M. Walline, M.S. Lawrence, C. Stewart, S.L. Carter, A. McKenna, I.P. Rodriguez-Sanchez, M. Espinosa-Castilla, K. Woie, L. Bjorge, E. Wik, M.K. Halle, E.A. Hoivik, C. Krakstad, N.B. Gabiño, G.S. Gómez-Macías, L.D. Valdez-Chapa, M.L. Garza-Rodríguez, G. Maytorena, J. Vazquez, C. Rodea, A. Cravioto, M.L. Cortes, H. Greulich, C.P. Crum, D.S. Neuberg, A. Hidalgo-Miranda, C.R. Escareno, L.A. Akslen, T.E. Carey, O.K. Vintermyr, S.B. Gabriel, H.A. Barrera-Saldaña, J. Melendez-Zajgla, G. Getz, H.B. Salvesen, M. Meyerson, Landscape of genomic alterations in cervical carcinomas, Nature 506 (2014) 371–375, https://doi.org/10.1038/nature12881.
[23] Z.-M. Zheng, M. Tao, K. Yamanegi, S. Bodaghi, W. Xiao, Splicing of a cap-proximal human Papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5′ cap, J. Mol. Biol. 337 (2004) 1091–1108, https://doi.org/10.1016/j.jmb.2004.02.023.
[24] M. Ajiro, R. Jia, L. Zhang, X. Liu, Z.-M. Zheng, Intron definition and a branch site adenosine at nt 385 control RNA splicing of HPV16 E6*I and E7 expression, PLoS ONE 7 (2012) e46412, https://doi.org/10.1371/journal.pone.0046412.
[25] M. Ajiro, Z.-M. Zheng, E6˜E7, a novel splice isoform protein of human papillomavirus 16, stabilizes viral E6 and E7 oncoproteins via HSP90 and GRP78, MBio 6 (2015), https://doi.org/10.1128/mBio.02068-14 (e02068–02014).
[26] S. Tang, M. Tao, J.P. McCoy, Z.-M. Zheng, 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. 80 (2006) 4249–4263, https://doi.org/10.1128/JVI.80.9.4249-4263.2006.
[27] Z.-M. Zheng, Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression, J. Biomed. Sci. 11 (2004) 278–294, https://doi.org/10.1159/000077096.
[28] K. Lin, X. Lu, J. Chen, R. Zou, L. Zhang, X. Xue, E6-associated transcription patterns in human papilloma virus 16-positive cervical tissues, Oncol. Lett. 9 (2015) 478–482, https://doi.org/10.3892/ol.2014.2698.
[29] J.P.C.B. Vidal, S.P. Felix, C.B.P. Chaves, P. Patury, V.F. Franco, E.A. de Morais, N.A. de Carvalho, A.C.L. Carvalho, O.F.A. Neto, L.M.T.M. Vieira, F.M. Correa, L.F.L. Martins, A. Negrão, L.M. de Almeida, M.A.M. Moreira, Genetic diversity of HPV16 and HPV18 in Brazilian patients with invasive cervical cancer, J. Med. Virol. (2015), https://doi.org/10.1002/jmv.24458.
[30] Illumina, CASAVA v1.8.2 User Guide, (2011).
[31] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, Bioinformatics 27 (2011) 863–864, https://doi.org/10.1093/bioinformatics/btr026.
[32] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.
[33] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. Depristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303, https://doi.org/10.1101/gr.107524.110.
[34] R. Core Team, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, (2008).
[35] Y. Liao, G.K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote, Nucleic Acids Res. 41 (2013) e108, https://doi.org/10.1093/nar/gkt214.
[36] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data, Nat. Biotechnol. 29 (2011) 644–652, https://doi.org/10.1038/nbt.1883.
[37] G. Wagner, R. Jardim, D.A. Tschoeke, D.R. Loureiro, K.A.C.S. Ocaña, A.C.B. Ribeiro, V.E. Emmel, C.M. Probst, A.N. Pitaluga, E.C. Grisard, M.C. Cavalcanti, M.L.M. Campos, M. Mattoso, A.M.R. Dávila, STINGRAY: system for integrated genomic resources and analysis, BMC Res. Notes 7 (2014) 132, https://doi.org/10.1186/1756-0500-7-132.
[38] S.M. Amaro-Filho, C.B. Pereira Chaves, S.P. Felix, D.L. Basto, L.M. de Almeida, M.A.M. Moreira, HPV DNA methylation at the early promoter and E1/E2 integrity: a comparison between HPV16, HPV18 and HPV45 in cervical cancer, Papillomavirus Res. 5 (2018) 172–179, https://doi.org/10.1016/j.pvr.2018.04.002.
[39] S.M.A. Filho, N. Bertoni, A.C. Brant, J.P.C.B. Vidal, S.P. Felix, S.M.B. Cavalcanti, F.N. Carestiato, L.F.L. Martins, L.M. de Almeida, M.A.M. Moreira, Methylation at 3′LCR of HPV16 can be affected by patient age and disruption of E1 or E2 genes, Virus Res. 232 (2017) 48–53, https://doi.org/10.1016/j.virusres.2017.01.022.
[40] M. del R. González-Losa, M. Puerto-Solis, J. Tenorio Ruiz, A.I. Rosado-López, O. Hau-Aviles, G. Ayora-Talavera, I. Cisneros-Cutz, L. Conde-Ferráez, Analysis of E2 gene integrity in HPV16 and HPV58 viruses isolated from women with cervical pathology, Mem. Inst. Oswaldo Cruz 111 (2016) 770–773, https://doi.org/10.1590/0074-02760160269.
[41] S.D. Vernon, E.R. Unger, D.L. Miller, D.R. Lee, W.C. Reeves, Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer, Int. J. Cancer 74 (1997) 50–56.
[42] S.I. Collins, C. Constandinou-Williams, K. Wen, L.S. Young, S. Roberts, P.G. Murray, C.B.J. Woodman, Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study, Cancer Res. 69 (2009) 3828–3832, https://doi.org/10.1158/0008-5472.CAN-08-3099.
[43] J. Chen, Y. Xue, M. Poidinger, T. Lim, S.H. Chew, C.L. Pang, J.-P. Abastado, F. Thierry, Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements during carcinogenic progression, Virology 462–463 (2014) 14–24, https://doi.org/10.1016/j.virol.2014.05.026.
[44] X. Wang, C. Meyers, H.-K. Wang, L.T. Chow, Z.-M. Zheng, Construction of a full transcription map of human papillomavirus type 18 during productive viral infection, J. Virol. 85 (2011) 8080–8092, https://doi.org/10.1128/JVI.00670-11.
[45] R. Jackson, B.A. Rosa, S. Lameiras, S. Cuninghame, J. Bernard, W.B. Floriano, P.F. Lambert, A. Nicolas, I. Zehbe, Functional variants of human papillomavirus type 16 demonstrate host genome integration and transcriptional alterations corresponding to their unique cancer epidemiology, BMC Genomics 17 (2016) 851, https://doi.org/10.1186/s12864-016-3203-3.
[46] M. Schmitz, C. Driesch, K. Beer-Grondke, L. Jansen, I.B. Runnebaum, M. Dürst, Loss of gene function as a consequence of human papillomavirus DNA integration, Int. J. Cancer 131 (2012) E593–E602, https://doi.org/10.1002/ijc.27433.
[47] K. Akagi, J. Li, T.R. Broutian, H. Padilla-Nash, W. Xiao, B. Jiang, J.W. Rocco, T.N. Teknos, B. Kumar, D. Wangsa, D. He, T. Ried, D.E. Symer, M.L. Gillison, Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability, Genome Res. (2013), https://doi.org/10.1101/gr.164806.113 gr.164806.113.
[48] M.-K. Diao, C.-Y. Liu, H.-W. Liu, J.-T. Li, F. Li, M.M. Mehryar, Y.-J. Wang, S.-B. Zhan, Y. Zhou, R. Zhong, Y. Zeng, Integrated HPV genomes tend to integrate in gene desert areas in the CaSki, HeLa, and SiHa cervical cancer cell lines, Life Sci. 127 (2015) 46–52, https://doi.org/10.1016/j.lfs.2015.01.039.
[49] Z. Hu, D. Zhu, W. Wang, W. Li, W. Jia, X. Zeng, W. Ding, L. Yu, X. Wang, L. Wang, H. Shen, C. Zhang, H. Liu, X. Liu, Y. Zhao, X. Fang, S. Li, W. Chen, T. Tang, A. Fu, Z. Wang, G. Chen, Q. Gao, S. Li, L. Xi, C. Wang, S. Liao, X. Ma, P. Wu, K. Li, S. Wang, J. Zhou, J. Wang, X. Xu, H. Wang, D. Ma, Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism, Nat. Genet. 47 (2015) 158–163, https://doi.org/10.1038/ng.3178.
[50] B. Xu, S. Chotewutmontri, S. Wolf, U. Klos, M. Schmitz, M. Dürst, E. Schwarz, Multiplex identification of human papillomavirus 16 DNA integration sites in cervical carcinomas, PLoS ONE 8 (2013), https://doi.org/10.1371/journal.pone.0066693.
[51] H.M. Walline, C.M. Komarck, J.B. McHugh, E.L. Bellile, J.C. Brenner, M.E. Prince, E.L. McKean, D.B. Chepeha, G.T. Wolf, F.P. Worden, C.R. Bradford, T.E. Carey, Genomic integration of high-risk HPV alters gene expression in oropharyngeal squamous cell carcinoma, Mol. Cancer Res. 14 (2016) 941–952, https://doi.org/10.

1158/1541-7786.MCR-16-0105.
[52] C.C. Baker, W.C. Phelps, V. Lindgren, M.J. Braun, M.A. Gonda, P.M. Howley, Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines, J. Virol. 61 (1987) 962–971.
[53] M. Wagatsuma, K. Hashimoto, T. Matsukura, Analysis of integrated human papillomavirus type 16 DNA in cervical cancers: amplification of viral sequences together with cellular flanking sequences, J. Virol. 64 (1990) 813–821.
[54] M. Cricca, A.M. Morselli-Labate, S. Venturoli, S. Ambretti, G.A. Gentilomi, G. Gallinella, S. Costa, M. Musiani, M. Zerbini, Viral DNA load, physical status and E2/E6 ratio as markers to grade HPV16 positive women for high-grade cervical lesions, Gynecol. Oncol. 106 (2007) 549–557, https://doi.org/10.1016/j.ygyno.2007.05.004.
[55] J.D. Meissner, Nucleotide sequences and further characterization of human papillomavirus DNA present in the CaSki, SiHa and HeLa cervical carcinoma cell lines, J. Gen. Virol. 80 (Pt 7) (1999) 1725–1733, https://doi.org/10.1099/0022-1317-80-7-1725.
[56] M. Pett, N. Coleman, Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? J. Pathol. 212 (2007) 356–367, https://doi.org/10.1002/path.2192.
[57] L. Sherman, N. Alloul, I. Golan, M. Durst, A. Baram, Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers, Int. J. Cancer 50 (1992) 356–364.
[58] N. Häfner, C. Driesch, M. Gajda, L. Jansen, R. Kirchmayr, I.B. Runnebaum, M. Dürst, Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts, Oncogene 27 (2008) 1610–1617, https://doi.org/10.1038/sj.onc.1210791.
[59] J.L.K. Cheung, T.-H. Cheung, M.Y. Yu, P.K.S. Chan, Virological characteristics of cervical cancers carrying pure episomal form of HPV16 genome, Gynecol. Oncol. 131 (2013) 374–379, https://doi.org/10.1016/j.ygyno.2013.08.026.