Genomics 112 (2020) 2942–2948

Contents lists available at ScienceDirect

Genomics

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

Preferential expression of a HPV genotype in invasive cervical carcinomas infected by multiple genotypes

A.C. Brant[a,b], A.N. Menezes[c], S.P. Felix[a], L.M. Almeida[d], M.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] Cancer Genetics and Evolution Laboratory, Cancer Research UK, Institute of Genetics & Molecular Medicine, The University of Edinburgh, UK
[d] Department of Population Research, Instituto Nacional de Câncer (INCA), Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:
Cervical cancer
HPV expression
Multiple HPV infection
HPV coinfection

ABSTRACT

Multiple infections by HPV genotypes are frequently detected in HPV+ cervical lesions but the interaction between each viral genotype during carcinogenesis is poorly understood. Here we carried out a comprehensive study to characterize the multiple HPV expression and integration by RNA-seq analyses of 19 invasive cervical carcinomas coinfected by several HPV genotypes. Analysis of tumor DNA by a hybridization assay indicated multiple infections ranging from 2 to 6 different HPV genotypes. RNA-seq analysis showed that a single HPV genotype was preferentially expressed. Finally, the search for HPV/human chimeric transcripts indicated integration from preferentially expressed genotypes. In conclusion, the present study indicated that, in invasive cervical carcinomas infected by multiple HPV genotypes, one HPV was preferentially expressed, supporting the hypothesis that a single HPV genotype was associated with cancer development.

1. Introduction

Human papillomavirus (HPV) infection is the main cause of cervical cancer, the fourth most frequent cancer in women worldwide [1,2]. HPV is a small non-enveloped, double stranded DNA virus with a genome of approximately 8 kb. It usually contains six early expressed genes (*E1*, *E2*, *E4*, *E5*, *E6* and *E7*) responsible for viral DNA replication and gene expression and two late expressed genes (*L1* and *L2*) coding for capsid proteins during its productive viral cycle [3]. *E6* and *E7*, the major HPV oncogenes, are transcribed as a bicistronic pre-mRNA [4]. Splicing of intron 1 (located at the *E6* ORF) is required to produce the *E6** transcript responsible for E7 translation, while intron 1 retention (forming *E6E7* mRNA) is necessary for E6 translation [5]. The HPV genome remains as an episome during the productive cycle; however, along the carcinogenic process, it usually integrates into the host cell genome, mainly through disruption at *E1* or *E2* coding regions [6]. Integration into the host genome can occur at any region although some reports indicated that it preferentially occurs at fragile sites and transcriptionally active regions [7–9].

According to International HPV Reference Center from Karolinska Institutet (https://ki.se/en/labmed/international-hpv-reference-center) more than 200 different HPV genotypes have been identified but the most relevant genotypes causing anogenital high-grade lesions and cancer belong to the *Alphapapillomavirus* (α-papillomaviruses), a genus comprising low-risk (LR-HPV) and high-risk (HR-HPV) HPVs for cancer development. Multiple infection by different HPV genotypes is frequently reported in women without lesions and in cervical lesions, corresponding to 20–50% of all HPV+ cervical infections [10,11]. Coinfection by multiple HPV genotypes is considered a risk factor for invasive cervical carcinogenesis, increasing 4.1-fold the risk for cancer development [12]. It is assumed that infections by different HPV genotypes occurs independently, with limited interaction between them [13]. The carcinogenic process associated with multiple HPV infections is not well understood; it is unknown whether all genotypes contribute equally to cervical cancer development, and whether simultaneous infections occur randomly or preferentially associated with specific genotypes [12]. Additionally, there is no information on interaction between HPV genotypes and the pattern of viral expression in carcinomas with multiple infections.

To understanding multiple HPV infection in the biology of cervical cancer, we carried out an exploratory analysis of the viral transcriptome of 19 cases of invasive cervical carcinomas coinfected by several HPV genotypes. By analyzing RNA-seq data we tested whether (i) all viral genotypes detected in each biopsy expressed their genes, (ii) all genotypes present in each biopsy expressed transcripts required for translation of the major E6 and E7 viral oncoproteins, and (iii) all viral

* Corresponding author at: Genetics Program, Instituto Nacional de Câncer, André Cavalcanti 37, Rio de Janeiro, RJ 20231-050, Brazil.
E-mail address: miguelm@inca.gov.br (M.A.M. Moreira).

https://doi.org/10.1016/j.ygeno.2020.05.009
Received 6 March 2020; Received in revised form 17 April 2020; Accepted 7 May 2020
Available online 11 May 2020
0888-7543/

genotypes in multiple infections were integrated in the host genome.

2. Material and methods

2.1. Samples selection and nucleic acid isolation

Nineteen biopsies of invasive cervical carcinomas (HPVco_1 to HPVco_19), previously characterized by presenting multiple HPV in coinfections, were selected from a sample set from patients admitted to the Instituto Nacional de Câncer - INCA, Rio de Janeiro - Brazil for cancer treatment [14]. Biopsies were collected before treatment, stored in *RNAlater* (Qiagen) at −80 °C. DNA and RNA were isolated from the same biopsy fragments with *Qiagen Allprep DNA/RNA mini kit* (Qiagen) according to the manufacturer's instructions. Following isolation, DNA and RNA were quantified by spectrophotometry and stored at −20 °C and − 80 °C, respectively.

2.2. Identification of HPV genotypes

Identification of HPV genotypes in each sample was carried out with the *High + Low PapillomaStrip Hybridization Kit* (Operon) following the manufacturer's instructions. This method allows the identification of 37 different HPV genotypes by reverse hybridization, comprising 19 HR-HPVs (genotypes: 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 69, 73 and 82) and 18 LR-HPVs (genotypes: 6, 11, 40, 42, 43, 44, 54, 61, 62, 67, 70, 71, 72, 74, 81, 83, 84 and 91).

2.3. High throughput mRNA sequencing and data preprocessing

Each library was prepared with total RNA (0.1 to 4 μg) with *TruSeq RNA Sample Prep Kit* (Illumina) according to the manufacturer's recommendations. Libraries were run, with paired-end sequencing strategy, in a HiSeq 2500 platform (Illumina). Casava 1.8 (Illumina) was used for obtaining Fastq files, comprising demultiplexed reads and allowing a single barcode mismatch. Low quality reads, with mean phred scores < 20 and lengths < 30 bp were filtered out using PRINSEQ. [15] as previously reported [16].

2.4. Identification of HPV genotype expression by read sequencing

To identify which HPV genotype was expressed in each sample, quality filtered reads were mapped against reference genomes from 64 α-HPVs genotypes deposited in GenBank (Additional Files 1). Mapping was performed with Bowtie2 (version 2.2.2) [17] with the default setting. Mapped reads to each genotype reference sequence were checked using Integrative Genomics Viewer (IGV) [18], and the coverage per nucleotide was estimated with GATK DepthOfCoverage [19]. Coverage per base was subsequently used for estimating the mean depth coverage for each HPV genotype and its breadth of coverage (the extension of HPV genome coverage by sequence reads) using R [20]. Depth coverage was plotted against breadth coverage, in base pairs, for each sample. This allowed us to simultaneously compare expression by depth and extension of each virus genotype infecting each patient.

2.5. E6E7 alternative transcripts

The presence and quantification of *E6E7* alternative transcripts of the HPV coinfecting genotypes were analyzed following a strategy previously developed by our group [16]. Sequenced reads were mapped against reference sequence of each *E6E7* alternative transcript, manually constructed according to previously known splicing junctions: 226^409, 226^526, 226^742 for HPV16 [21,22]; 233^416, 233^791 for HPV18 [5,23,24]; 210^413 for HPV31 [25,26]; 231^509, 231^785 for HPV33 [27]; 228^419 for HPV35 [28]; 231^420 for HPV39 [28]; 230^413 for HPV45 [29]; 224^502 for HPV52 [29]; 232^510 for HPV58 [30]; and 232^415 for HPV68 [29]. The expression of *E6/E7* unspliced transcript and *E6** alternative transcripts were also quantified as previously reported taking into account the number of splicing junction reads (SJRs) [16].

2.6. Integration analysis by identification of HPV/human chimeric transcripts

To identify integration sites in the human genome and disruption sites within the viral genome, we used a *de novo* transcriptome assembly approach for identifying likely HPV/human chimeric transcripts, as previously described [16]. To validate each assembled chimeric transcript (Additional File 2), RT-PCR assays were carried out using specific primer pairs annealing at human and viral regions (Additional File 3). RNA samples were treated with *RQ1 RNase-Free DNase* (Promega) and cDNA was synthesized *with SuperScript® II Reverse Transcriptase* (Thermo Scientific) using random primers. PCR amplification reactions were performed in final volumes of 25 μl containing 0.2 mM of each dNTP, 25 pmol of each primer, 1 U *Platinum Taq I DNA Polymerase* (Life Technologies), 1× reaction buffer and 2 mM of $MgCl_2$. Cycling conditions were 95 °C for 5 min followed by 35 cycles at 95 °C for 30 s, annealing temperature (see additional File 3) for 30 s, extension at 72 °C for 40 s, and final extension at 72 °C for 5 min. Amplicons were analyzed by gel electrophoresis in 1.8% ultrapure agarose gels (Life Technologies), purified with *GFX PCR DNA and Gel Band Purification kit* (GE Healthcare), and subsequently sequenced in an ABI Prism® 3130xl platform.

3. Results

3.1. Identification of HPV genotypes coinfecting invasive cervical carcinomas

Hybridization analyses of 19 cervical tumors with previously detected [14] multiple HPV genotypes allowed us to identify viral genotypes in each sample. Presence of HPV coinfection ranged from two to six different HPV genotypes (Table 1). Sample HPVco_15 was the only one infected by six different HPVs. HPV16 was detected in all samples but one (HPVco_12), and HPV18 was the second genotype more frequently detected. Low-risk HPV genotypes (HPV42, HPV54 and HPV61) were also detected in coinfection with HR-HPVs in four samples (HPVco_3, HPVco_4, HPVco_5 and HPVco_7).

Table 1
Identification of infected HPV genotypes by hybridization assay.

Samples	Genotypes
HPVco_1	16, 18
HPVco_2	16, 18, 45
HPVco_3	16, 45, 58, 54[a]
HPVco_4	16, 18, 54[a], 61[a]
HPVco_5	16, 18, 33, 54[a]
HPVco_6	16, 18
HPVco_7	16, 42[a]
HPVco_8	16, 45
HPVco_9	16, 45
HPVco_10	16, 18
HPVco_11	16, 39
HPVco_12	31, 35, 39
HPVco_13	16, 18, 31
HPVco_14	16, 18, 33, 45
HPVco_15	16, 18, 31, 33, 35, 52
HPVco_16	16, 31, 33, 39, 68
HPVco_17	16, 18, 31, 33
HPVco_18	16, 18, 31, 33, 39
HPVco_19	16, 18

[a] LR-HPV genotypes.

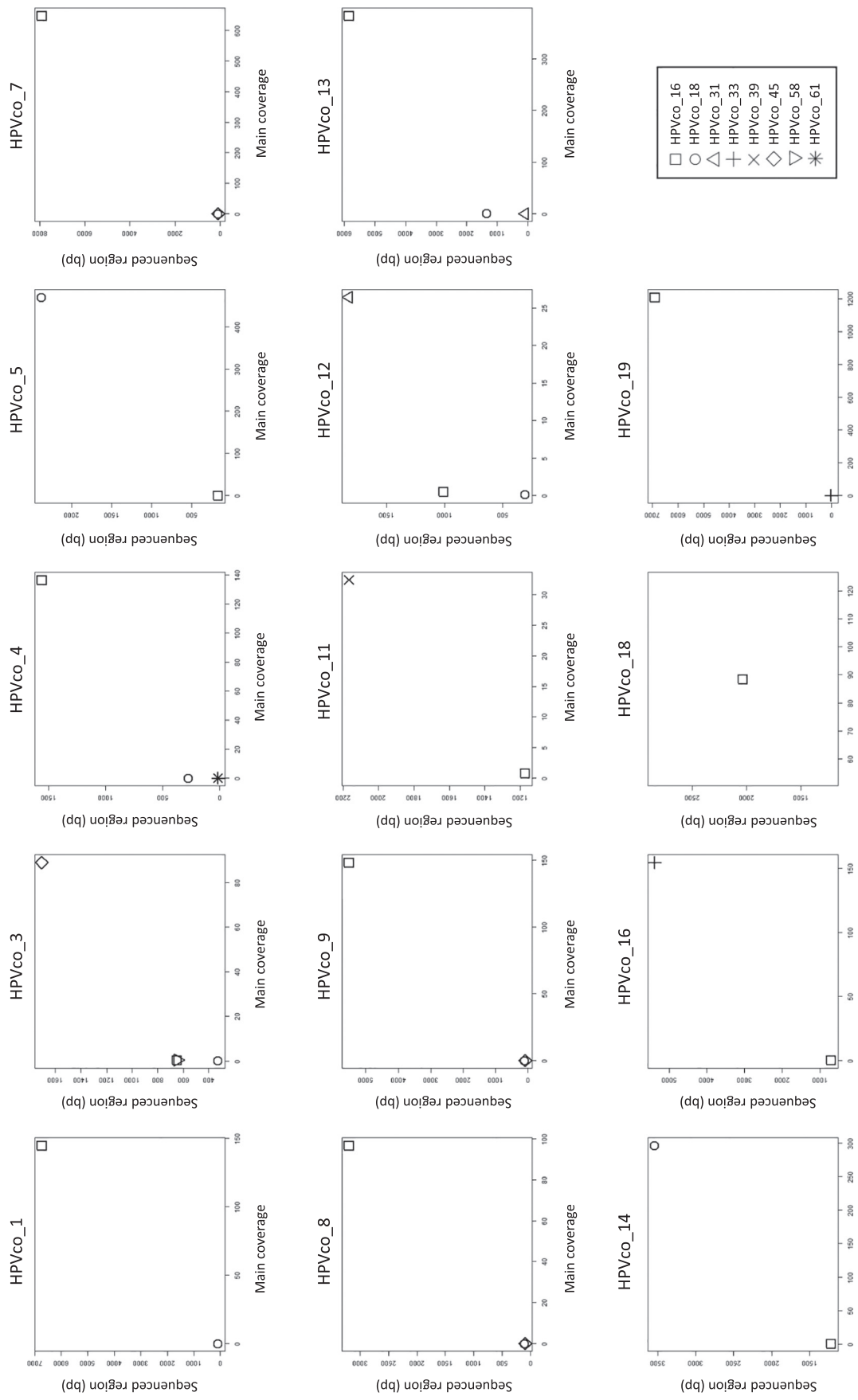

Fig. 1. Plot of depth coverage against breadth coverage (in nucleotides) per sample. At multiple infected samples, despite presence of multiple HPV genotype, just a single genotype presented higher genome expression.

3.2. Expression of genotypes coinfecting cervical cancers

A mean of 86,600,386 quality filtered reads (Phred score > 20) were obtained after high throughput sequencing per tumor sample. More than 80% of sequenced reads per sample mapped to the human genome (GRCh38), while less than 0.1% of reads mapped to the reference sequence of α-HPV genotypes (Additional File 1). In five samples (HPVco_2, HPVco_6, HPVco_10, HPVco_15 and HPVco_17) a small number of HPVs reads (< 600) were detected, and these samples were excluded from RNA-seq analyses (Additional File 4).

To identify which HPV genotype was expressing its genome, RNA-seq reads were mapped to α-HPVs reference sequences. Two factors were taken in consideration (i) the depth coverage of the sequenced viral genome, and (ii) the breadth of coverage (Fig. 1 and Additional File 5). Transcripts of a single HPV genotype were detected in one sample (HPVco_18), and two to four genotypes in the remaining 13 samples (Additional File 5). Despite that more than one viral genotype were detected in these samples, one genotype presented a higher expression (with depth coverage higher than 20× and a simultaneous breadth of coverage higher than 20%) (Fig. 1). Additionally, in seven samples (HPVco_3, HPVco_4, HPVco_7, HPVco_8, HPVco_9, HPVco_12 and HPVco_16), RNA-seq showed mRNA reads of HPV genotypes undetected by hybridization assays, all of them with low depth coverage and low breadth coverage (Additional File 5).

3.3. Detection of E6E7 splicing isoforms

Splicing events of the primary mRNA from the early promoter region of HR-HPVs are crucial for translation of viral proteins. In HR-HPVs, the splicing event removes a segment of the *E6* open reading frame required for producing the mature transcript for *E7* translation. On the other hand, mRNAs carrying the complete *E6* ORF are required for E6 translation. To verify whether the mRNAs detected by RNA-seq were capable of translating E6 and E7, RNA-seq data were analyzed for the presence of splicing junction reads (SJR).

Splicing analyses of *E6E7* transcripts allowed the detection of SJRs, with one genotype presenting a higher number of SJRs than the other coinfecting genotypes (Table 2); this genotype showing the highest sequence coverages (Fig. 1). In most cases, a higher number of SJRs was found in the spliced *E6*I* transcript than in unspliced *E6/E7* transcripts, except for HPV39 (HPVco_11) (Table 2).

3.4. Viral genome integration

As integration of the viral genome to the host genome is a common event in cervical cancer, we searched for chimeric reads (HPV/human) to determine whether a specific HPV genotype or more than one genotype were integrated in each sample.

The presence of HPV/human chimeric transcripts was found in 12 of 14 samples (Table 3). Integration of two different HPV genotypes was found in a single sample (HPVco_8) where HPV16 and HPV45 were integrated at different human loci. The identification of chimeric transcripts provided information about the site of integration in the host genome (Table 3 and Additional File 6). Integration took place in human genes in six samples, and in intergenic regions in six others. A single chimeric transcript was found in most samples, although more than one chimeric transcript, indicating multiple integration events, were found in four samples (HPVco_5, HPVco_8, HPVco_9 and HPVco_18), in nearby chromosome regions, except for HPVco_8. The chimeric transcripts of HPVco_3, HPVco_7 and HPVco_8 (for HPV45) were not independently confirmed by RT-PCR (Table 3).

The majority of chimeric transcripts showed *E1* or *E2* sequences at their borderlines. Some chimeric transcripts showed that the limits between HPV/human borders occurred at the viral genomic positions 880 for HPV16, and 929 for HPV18 and HPV45 (sites located in the *E1* gene), indicating that mature transcripts were generated by a splicing process of chimeric mRNA [16,31,32], where the viral genomic positions were donor splicing sites. For samples HPVco_5 and HPVco_7 chimeric transcripts suggested disruption in *E7* and within the intergenic region between *E5* and *L2*, respectively (Table 3).

4. Discussion

Although multiple HPV infections are frequently reported in normal cervical tissue, pre-neoplastic lesions and cervical cancer [10,11], the association between the presence of different HPV genotypes and cancer development is poorly understood. Most studies were focused on epidemiological data, with multiple HPV infections based on detection of HPV DNA [13,33–39]. In this work, multiple infections in cervical cancers were characterized by analyzing the expression of viral genotypes by RNA-seq, an approach allowing the identification and quantification of HPV genotype expression.

The identification of infective HPV genotypes in cancer biopsies by reverse hybridization detected the presence of two to six different genotypes (Table 1). Despite the identification of multiple HPV infections, expression analysis showed a preferential expression of a single genotype (Fig. 1). Lack of detection of transcripts from all HPV genotypes previously detected by reverse hybridization pointed to the following alternatives: (i) presence of viral genomes that remained non-expressed in tumors, (ii) detected genotypes without their corresponding transcripts were actually virions that did not express their genes. On the other hand, in seven samples (HPVco_3, HPVco_4, HPVco_7, HPVco_8, HPVco_9, HPVco_12 and HPVco_16) RNA from HPV genotypes undetected by hybridization assays was present (Additional File 5). This could be explained by the genomic similarity between some HPV genotypes, which can cause missmapping during the computational analysis.

Overexpression of *E6* and *E7* is an important step for tumor development [40] and HPV DNA integration is associated to overexpression of these two oncogenes [41–43]. When searching for *E6E7* alternative transcripts from previously identified HPV genotypes, the HPV genotype with the highest genome expression was also the one with the highest production of *E6E7* alternative transcripts (Table 2), mainly of the *E6*I* transcript, responsible for *E7* translation. Altogether, these findings pointed to the predominance of a specific HPV genotype in the development of cervical cancers with multiple infections. Conversely, a recent publication analyzing the expression of E6 protein in HPV16/18 coinfection showed the simultaneous expression of E6, indicating that both viruses might contribute to cervical lesions [44].

In agreement with our previous results, HPV genotype integrated in the host genome was the highly expressed or the only one expressed. In a single tumor biopsy (HPVco_8, see Table 3), transcriptome analysis showed the integration of two genotypes (HPV16 and HPV45). As the presence of HPV45/human chimeric transcripts was supported by few reads, it is possible that the integrated HPV45 was transcribed from a cell promoter with low transcriptional activity [16].

Information regarding the infection by multiple HPV genotypes is necessary to better understand the populational infection pattern and, consequently, find out a better strategy for diagnosis, prevention and eventually for vaccination. To our knowledge this is the first study to analyze the transcriptome of HPV genotypes in invasive cervical carcinoma biopsies with multiple infections. A previous work, based on exfoliated cells (collected by cytobrush), reported the presence of viral mRNA by qRT-PCR from at least one genotype in coinfected pre-cancerous lesions and invasive cancer, however the study was restricted to *E6E7* transcripts, and a single invasive cervical cancer with multiple infection was analyzed [45].

This work is limited with respect to: (i) the strategies for genotype identification that did not rule out the presence of other non-*Alphapapillomavirus* genotypes; (ii) the impossibility of detecting a likely integration of viral genotypes that do not express their genomes; and (iii) the expression of genotypes undetected by reverse

Table 2
Counting of splicing junction reads for *E6E7* alternative transcripts.

Genotype	*E6E7* transcripts	Splicing junction	Samples						
			HPVco_1	HPVco_3	HPVco_4	HPVco_5	HPVco_7	HPVco_8	HPVco_9
HPV16	*E6*I*	226^409	611	0	1535	2	2088	394	675
	*E6*II*	226^526	53	0	445	0	218	58	28
	E6^E7	226^742	1	0	1	0	1	0	0
	E6/E7	–	7	0	41	0	299	20	144
HPV18	*E6*I*	233^416	0	–	0	4500	–	–	–
	E6^E7	233^791	0	–	0	0	–	–	–
	E6/E7	–	0	–	0	583	–	–	–
HPV31	*E6*I*	210^413	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–
HPV33	*E6*I*	231^509	–	–	–	0	–	–	–
	*E6*II*	231^785	–	–	–	0	–	–	–
	E6/E7	–	–	–	–	0	–	–	–
HPV35	*E6*I*	228^419	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–
HPV39	*E6*I*	231^420	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–
HPV45	*E6*I*	230^413	–	804	–	–	–	0	0
	E6/E7	–	–	126	–	–	–	0	0
HPV52	*E6*I*	224^502	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–
HPV58	*E6*I*	232^510	–	3	–	–	–	–	–
	E6/E7	–	–	0	–	–	–	–	–
HPV68	*E6*I*	232^415	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–

Genotype	*E6E7* transcripts	Splicing junction	Samples						
			HPVco_11	HPVco_12	HPVco_13	HPVco_14	HPVco_16	HPVco_18	HPVco_19
HPV16	*E6*I*	226^409	4	–	3104	1	2	859	5117
	*E6*II*	226^526	0	–	288	0	0	67	527
	E6^E7	226^742	0	–	0	0	0	2	4
	E6/E7	–	0	–	197	0	0	41	305
HPV18	*E6*I*	233^416	–	–	5	2003	–	0	0
	E6^E7	233^791	–	–	0	0	–	0	0
	E6/E7	–	–	–	0	422	–	0	0
HPV31	*E6*I*	210^413	–	253	0	–	0	0	–
	E6/E7	–	–	109	0	–	0	0	–
HPV33	*E6*I*	231^509	–	–	–	0	649	0	–
	*E6*II*	231^785	–	–	–	0	54	0	–
	E6/E7	–	–	–	–	0	128	0	–
HPV35	*E6*I*	228^419	–	0	–	–	–	–	–
	E6/E7	–	–	0	–	–	–	–	–
HPV39	*E6*I*	231^420	0	0	–	–	0	0	–
	E6/E7	–	33	0	–	–	0	0	–
HPV45	*E6*I*	230^413	–	–	–	0	–	–	–
	E6/E7	–	–	–	–	0	–	–	–
HPV52	*E6*I*	224^502	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–
HPV58	*E6*I*	232^510	–	–	–	–	–	–	–
	E6/E7	–	–	–	–	–	–	–	–
HPV68	*E6*I*	232^415	–	–	–	–	0	–	–
	E6/E7	–	–	–	–	–	0	–	–

hybridization and vice-versa due to a low amount of DNA of specific viral genotypes.

5. Conclusion

In conclusion, the present study indicated that, in invasive cervical carcinomas infected by multiple HPV genotypes, a single HPV was preferentially expressed. This supported the hypothesis that, in spite of the presence of multiple genotypes, a single HPV genotype may be associated with cancer development.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.05.009.

Availability of data and materials

Data generated in this work (RNA-seq reads - fastq files - of each samples and fasta files of chimeric transcripts) were deposited in Gene Expression Omnibus (GEO - http://www.ncbi.nlm.nih.gov/geo) with access number GSE144293.

Ethics approval and consent to participate

This work was approved by the 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.

Table 3
Integration analyses of multiple HPV genotypes in invasive cervical samples by identification of chimeric transcripts using RNA-seq.

Samples	HPV type	Chimeric transcripts				
		HPV breakpoint (ORF)	Human breakpoint	Locus (Gene)	Coverage [a]	PCR/sanger [b]
HPVco_1	16	880 (*E1*)	50,999,332	6p12.3	76	+
HPVco_3	45	1210 (*E1*)[c]	246,377,386[c]	1q44 (*SMYD3*)	17	−
HPVco_4	16	880 (*E1*)	153,418,067	1q21.3 (*S100A7A*)	1650	+
HPVco_5	18[f]	724 (*E7*)	89,951,604	4q22.1 (*MMRN1*)	2662	+[d]
		1095 (*E1*)	89,946,311	4q22.1 (*MMRN1*)	1966	+[d]
HPVco_7	16	4214 (*E5/L2*)[e]	17,377,488	11p15.1	38	−
HPVco_8	16	1561 (*E1*)	42,726,986	22q13.2	250	+
	45	929 (*E1*)	246,381,380	1q44	3	−
HPVco_9	16[f]	1056 (*E1*)	14,952,133	20p12.1 (*MACROD2*)	17	+
		2775: 2776 (*E1*)	14,952,115: 14,952,116	20p12.1 (*MACROD2*)	10	+
HPVco_11	39	940: 942 (*E1*)	52,641,201: 52,641,203	6p12.2	78	+
HPVco_12	31	871: 872 (*E1*)	100,146,971: 100,146,972	3q12.1 (*CMSS1*)	437	+[d]
HPVco_13	ND	ND	ND	ND	ND	ND
HPVco_14	18	929 (*E1*)	44,129,352	1p34.1 (*KLF17*)	4270	+
HPVco_16	ND	ND	ND	ND	ND	ND
HPVco_18	16[f]	1486: 1488 (*E1*)	20,631,287: 20,631,289	21q21.1	247	+
		880 (*E1*)	20,592,134	21q21.1	211	+
HPVco_19	16	3199: 3197 (*E2*)	17,509,493: 17,509,495	7p21.1	44	+

[a] Chimeric transcript coverage was obtained from the number of reads covering the integration junction.
[b] Validation of chimeric transcripts by RT-PCR and Sanger sequencing. Electropherograms shown at Additional File 6.
[c] The exact HPV-human integration border was not found due a 6 nucleotides gap between them.
[d] Chimeric transcript confirmed only by RT-PCR, the sanger sequence was not obtained with a satisfied quality.
[e] Disruption at intergenic region between genes *E5* and *L2* from HPV16.
[f] Two independents chimeric transcripts. ND – Not determined.

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, 573806/2008–0 and 304339/2018-0), 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

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

Declaration of Competing Interest

The authors declare that they do not have conflicting interests.

Acknowledgements

This work is part of A.C.B.´s Ph.D thesis carried out at the Genetics Post-Graduate Program (PGGEN) of Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil. The authors are grateful to Hector N. Seuanez for a critical revision of the manuscript.

References

[1] J.M. Walboomers, M.V. Jacobs, M.M. Manos, F.X. Bosch, J.A. Kummer, K.V. Shah, P.J. Snijders, J. Peto, C.J. Meijer, N. Muñoz, Human papillomavirus is a necessary cause of invasive cervical cancer worldwide, J. Pathol. 189 (1999) 12–19, https://doi.org/10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F.
[2] 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.
[3] Z.-M. Zheng, C.C. Baker, Papillomavirus genome structure, expression, and post-transcriptional regulation, Front. Biosci. J. Virtual Libr. 11 (2006) 2286–2302.
[4] M. Ajiro, Z.-M. Zheng, Oncogenes and RNA splicing of human tumor viruses, Emerg. Microbes Infect. 3 (2014) e63, , https://doi.org/10.1038/emi.2014.62.
[5] 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] K.S. Cuschieri, H.A. Cubie, M.W. Whitley, A.L. Seagar, M.J. Arends, C. Moore, G. Gilkisson, E. McGoogan, Multiple high risk HPV infections are common in cervical neoplasia and young women in a cervical screening population, J. Clin. Pathol. 57 (2004) 68–72.
[11] H. Trottier, S. Mahmud, M.C. Costa, J.P. Sobrinho, E. Duarte-Franco, T.E. Rohan, A. Ferenczy, L.L. Villa, E.L. Franco, Human papillomavirus infections with multiple types and risk of cervical neoplasia, Cancer Epidemiol. Biomark. Prev. 15 (2006) 1274–1280, https://doi.org/10.1158/1055-9965.EPI-06-0129.
[12] R. Senapati, B. Nayak, S.K. Kar, B. Dwibedi, HPV genotypes co-infections associated with cervical carcinoma: special focus on phylogenetically related and non-vaccine targeted genotypes, PLoS One 12 (2017) e0187844, , https://doi.org/10.1371/journal.pone.0187844.
[13] L.D. Brot, B. Pellegrini, S.T. Moretti, D.M. Carraro, F.A. Soares, R.M. Rocha, G. Baiocchi, I.W. da Cunha, V.P. de Andrade, Infections with multiple high-risk HPV types are associated with high-grade and persistent low-grade intraepithelial lesions of the cervix, Cancer Cytopathol. 125 (2017) 138–143, https://doi.org/10.1002/cncy.21789.
[14] L.M. de Almeida, L.F.L. Martins, V.B. Pontes, F.M. Corrêa, R.C. Montenegro, L.C. Pinto, B.M. Soares, J.P.C.B. Vidal, S.P. Félix, N. Bertoni, M. Szklo,

M.A.M. Moreira, Human papillomavirus genotype distribution among cervical Cancer patients prior to Brazilian national HPV immunization program, J. Environ. Public Health 2017 (2017) 1645074, https://doi.org/10.1155/2017/1645074.
[15] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, Bioinforma. Oxf. Engl. 27 (2011) 863–864, https://doi.org/10.1093/bioinformatics/btr026.
[16] A.C. Brant, A.N. Menezes, S.P. Felix, L.M. de Almeida, M. Sammeth, M.A.M. Moreira, Characterization of HPV integration, viral gene expression and E6E7 alternative transcripts by RNA-Seq: a descriptive study in invasive cervical cancer, Genomics (2018), https://doi.org/10.1016/j.ygeno.2018.12.008.
[17] 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.
[18] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration, Brief. Bioinform. 14 (2013) 178–192, https://doi.org/10.1093/bib/bbs017.
[19] 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.
[20] R Core Team, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, (2008).
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] M.A. Ozbun, Human papillomavirus type 31b infection of human keratinocytes and the onset of early transcription, J. Virol. 76 (2002) 11291–11300, https://doi.org/10.1128/jvi.76.22.11291-11300.2002.
[26] 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.
[27] P.J. Snijders, A.J. van den Brule, H.F. Schrijnemakers, P.M. Raaphorst, C.J. Meijer, J.M. Walboomers, Human papillomavirus type 33 in a tonsillar carcinoma generates its putative E7 mRNA via two E6* transcript species which are terminated at different early region poly(A) sites, J. Virol. 66 (1992) 3172–3178.
[28] T. Mesplède, D. Gagnon, F. Bergeron-Labrecque, I. Azar, H. Sénéchal, F. Coutlée, J. Archambault, p53 degradation activity, expression, and subcellular localization of E6 proteins from 29 human papillomavirus genotypes, J. Virol. 86 (2012) 94–107, https://doi.org/10.1128/JVI.00751-11.
[29] G. Halec, M. Schmitt, B. Dondog, E. Sharkhuu, N. Wentzensen, T. Gheit, M. Tommasino, F. Kommoss, F.X. Bosch, S. Franceschi, G. Clifford, L. Gissmann, M. Pawlita, Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer, Int. J. Cancer 132 (2013) 63–71, https://doi.org/10.1002/ijc.27605.
[30] Y. Li, X. Wang, T. Ni, F. Wang, W. Lu, J. Zhu, X. Xie, Z.-M. Zheng, Human papillomavirus type 58 genome variations and RNA expression in cervical lesions, J. Virol. 87 (2013) 9313–9322, https://doi.org/10.1128/JVI.01154-13.
[31] 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.
[32] 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.
[33] L.S.A. de Resende, S.H. Rabelo-Santos, L.O. Sarian, R.R. Figueiredo Alves, A.A. Ribeiro, L.C. Zeferino, S. Derchain, A portrait of single and multiple HPV type infections in Brazilian women of different age strata with squamous or glandular cervical lesions, BMC Infect. Dis. 14 (2014) 214, https://doi.org/10.1186/1471-2334-14-214.
[34] M.L. Nogueira Dias Genta, T.R. Martins, R.V. Mendoza Lopez, J.C. Sadalla, J.P.M. de Carvalho, E.C. Baracat, J.E. Levi, J.P. Carvalho, Multiple HPV genotype infection impact on invasive cervical cancer presentation and survival, PLoS One 12 (2017), https://doi.org/10.1371/journal.pone.0182854 e0182854.
[35] L. Ponce-Benavente, P. Rejas-Pinelo, M.A. Aguilar-Luis, C. Palomares-Reyes, L. Becerra-Goicochea, L. Pinillos-Vilca, W. Silva-Caso, L.E. Costa, P. Weilg, J. Alvitrez-Arana, J. Bazán-Mayra, J. Del Valle-Mendoza, Frequency and coinfection between genotypes of human papillomavirus in a population of asymptomatic women in northern Peru, BMC Res. Notes. 11 (2018) 530, https://doi.org/10.1186/s13104-018-3644-7.
[36] V.N. Pimenoff, S. Tous, Y. Benavente, L. Alemany, W. Quint, F.X. Bosch, I.G. Bravo, S. de Sanjosé, Distinct geographic clustering of oncogenic human papillomaviruses multiple infections in cervical cancers: results from a worldwide cross-sectional study, Int. J. Cancer (2018), https://doi.org/10.1002/ijc.31964.
[37] M. Li, X. Du, M. Lu, W. Zhang, Z. Sun, L. Li, M. Ye, W. Fan, S. Jiang, A. Liu, M. Wang, Y. Meng, Y. Li, Prevalence characteristics of single and multiple HPV infections in women with cervical cancer and precancerous lesions in Beijing, China, J. Med. Virol. 91 (2019) 473–481, https://doi.org/10.1002/jmv.25331.
[38] P. Wu, H. Xiong, M. Yang, L. Li, P. Wu, C. Lazare, C. Cao, P. Gao, Y. Meng, W. Zhi, S. Lin, J. Hu, J. Wei, D. Ma, J. Liu, P. Yin, H. Xing, Co-infections of HPV16/18 with other high-risk HPV types and the risk of cervical carcinogenesis: a large population-based study, Gynecol. Oncol. 155 (2019) 436–443, https://doi.org/10.1016/j.ygyno.2019.10.003.
[39] M.A. Oyervides-Muñoz, A.A. Pérez-Maya, C.N. Sánchez-Domínguez, A. Berlanga-Garza, M. Antonio-Macedo, L.D. Valdéz-Chapa, R.M. Cerda-Flores, V. Trevino, H.A. Barrera-Saldaña, M.L. Garza-Rodríguez, Multiple HPV infections and viral load Association in Persistent Cervical Lesions in Mexican women, Viruses 12 (2020) 380, https://doi.org/10.3390/v12040380.
[40] 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.
[41] J. Ham, N. Dostatni, J.M. Gauthier, M. Yaniv, The papillomavirus E2 protein: a factor with many talents, Trends Biochem. Sci. 16 (1991) 440–444.
[42] S. Jeon, B.L. Allen-Hoffmann, P.F. Lambert, Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells, J. Virol. 69 (1995) 2989–2997.
[43] S. Jeon, P.F. Lambert, Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 1654–1658.
[44] Z. Wu, T.-Y. Li, M.-Y. Jiang, L. Yu, J. Zhao, H. Wang, X. Zhang, W. Chen, Y.-L. Qiao, Human papillomavirus (HPV) 16/18 E6 oncoprotein expression in infections with single and multiple genotypes, Cancer Prev. Res. (Phila.) (2019), https://doi.org/10.1158/1940-6207.CAPR-18-0343.
[45] E. Andersson, C. Kärrberg, T. Rådberg, L. Blomqvist, B.-M. Zetterqvist, W. Ryd, M. Lindh, P. Horal, Type-dependent E6/E7 mRNA expression of single and multiple high-risk human papillomavirus infections in cervical neoplasia, J. Clin. Virol. 54 (2012) 61–65, https://doi.org/10.1016/j.jcv.2012.01.012.