OPEN ACCESS Trends in Bioinformatics

ISSN 1994-7941
DOI: 10.3923/tb.2021.1.12

Research Article

Transcriptional Patterns of High-Risk Human Papillomavirus Types 16, 18, 45, 68b Genes

[1,2]Arkom Chaiwongkot, [4]Masahide Seki, [4]Yutaka Suzuki, [2,3]Teerasit Techawiwattanaboon and [1,2]Parvapan Bhattarakosol

[1]Applied Medical Virology Research Unit, Faculty of Medicine, Chulalongkorn University, 1873 Rama 4 Road, Pathumwan, Bangkok, 10330, Thailand
[2]Department of Microbiology, Faculty of Medicine, Chulalongkorn University, 1873 Rama 4 Road, Pathumwan, Bangkok, 10330, Thailand
[3]Chula Vaccine Research Center (Chula VRC), Center of Excellence in Vaccine Research and Development, Chulalongkorn University, Pathumwan, Bangkok 10330, Thailand
[4]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

Abstract

Background and Objective: Persistent high-risk human papillomavirus (HR-HPV) infections together with their oncogene expressions are necessary for cervical cancer development. We aimed to detect transcription patterns and the major viral oncogene transcripts expressed in five cervical cancer cell lines containing integration form of four HR-HPV types 16, 18, 45 and 68b. **Materials and Methods:** RNA extracted from five cervical cancer cell lines (CaSki, SiHa, HeLa, MS751 and ME180) were used for RNA sequencing analysis on an Illumina HiSeq2500. The HISAT2, Cufflinks, Cuffmerge and Cuffdiff were used to analyze the data. The HPV transcripts were quantified as fragments per kilobase per million (FPKM). Real-time PCR was performed to validate and differential viral gene expressions. **Results:** The major and common transcripts obtained from four HR-HPV types 16, 18, 45 and 68b were E6*I with splicing event occur within the E6 gene fused with full-length E7 ORF, 226^409 (HPV16), 233^416 (HPV18), 230^412 (HPV45) and 129^311(HPV68b) and were associated with full-length E7 oncogene expression. Real-time PCR revealed that full-length HR-HPV E7 oncogene was highly expressed in all five cell lines. **Conclusion:** Splicing patterns that occurred within the E6 gene resulting in the E6*I transcript facilitates E7 oncogene translation. Various transcripts obtained from four HR-HPV types are useful for further functional study of truncated viral proteins expressed from different HPV types.

Key words: High-risk human papillomavirus, transcription patterns, RNA sequencing, cervical cancer, splicing pattern, E6*1, truncated E6

Citation: Chaiwongkot, A., M. Seki, Y. Suzuki, T. Techawiwattanaboon and P. Bhattarakosol, 2021. Transcriptional patterns of high-risk human papillomavirus types 16, 18, 45, 68b genes. Trends Bioinform., 14: 1-12.

Corresponding Author: Arkom Chaiwongkot, Applied Medical Virology Research Unit, Department of Microbiology, Faculty of Medicine, Chulalongkorn University, 1873 Rama 4 Road, Pathumwan, Bangkok, 10330, Thailand

Copyright: © 2021 Arkom Chaiwongkot *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The long term persistent infection and upregulated expression of high-risk human papillomaviruses (HR-HPVs) E6 and E7 oncoproteins leads to chromosomal instability and cervical cancer progression. These E6 and E7 oncoproteins disrupt the normal function of host proteins necessary for normal cell cycle regulation for example, E6 causes degradation of p53 and E7 inactivates retinoblastoma tumor suppressor protein (pRb) leading to cell cycle progression[1,2]. Epidemiological reports showed that HPV16 is mostly found in cervical cancer cases accounting for 50% worldwide, followed by HPV18 (~20%) and other HR-HPV types such as HPV45, 31, 33, 58, 52 depends on regions[3,4].

HPV16, 18 and 31 transcription patterns have been widely studied and revealed that viral oncogene transcripts E6 and E7 are expressed from the early promoter as bicistronic or polycistronic transcripts with two or more open reading frames (ORFs) located on the same mRNA with splicing variants found in E6 region in some transcript types[5-13]. It has been reported that the transcript encoding the full length of E6 and E7 mRNA translated only the first E6 ORF that located close to promoter[14,15] due to the distance between two ORFs is short, consequently, only the first ORF is translated[16]. It was reported that monocistronic mRNA containing E7 ORF expressed from a promoter located within the E6 gene such as p542 for HPV16 is required to efficiently translate E7 protein[17,18]. Splicing event within the E6 intronic region is the alternative mechanism to help express E7 protein and E6 disruption by splicing event is associated with cervical carcinogenesis[14,19,20]. Different splicing patterns within the E6 region of HPV16 have been identified with one Splicing Donor (SD) site at position 226 and three splicing acceptor (SA) sites at position 409, 526 and 742, they are named as E6*I (SD226^SA409), E6*II (SD226^SA526) and E6*X(E6*^E7*) (SD226^SA742)[12]. Two splicing patterns within the E6 gene of HPV18 have been identified including E6*I (SD233^SA416) and E6*^E7* (SD233^SA791)[10,11,21]. There is no splicing event within the E6 region of low risk HPVs[22].

Most of the previously published papers focused on HPV transcription patterns; however, there was limited quantitative information of major transcripts that are associated with E6 and E7 oncogene expression. A recent study using RNAseq assay reported quantitative results of different HPV16 transcripts[19]. Other HR-HPVs that are necessary for cervical carcinogenesis should be concerned to understand the transcription patterns among HR-HPV types.

Thus, we aimed to employ RNAseq analysis to identify major transcripts that are associated with E6 and E7 oncogene expressions of not only HPV16 and HPV18 but also HPV45 and HPV68b in five cervical cancer cell lines. The major transcripts that were associated with E6* and E7 oncogene expressions obtained from four HR-HPV types 16, 18, 45 and 68b with integration form were identified in the present study.

MATERIALS AND METHODS

Study area: The study was carried out at Department of Microbiology, Faculty of Medicine, Chulalongkorn University, Thailand and Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Japan. This research project was conducted from April, 2017-September, 2018.

Cervical cancer cell lines: Five cervical cancer cell lines containing different HR-HPV types were used in the present study including CaSki (containing approximately 600 copies of integrated HPV16 per cell, (CRL-1550 Lot No. 3794357)), SiHa (containing approximately 1-2 copies of integrated HPV16 per cell (HTB-35 Lot No. 4031219)), HeLa (containing integrated HPV18), MS751 (containing integrated HPV45 (HTB-34 Lot No. 58078726)) and ME180 (containing integrated HPV68b (HTB-33 Lot 57758662)). All cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (HyClone, South Logan, Utah, USA) with 10% fetal bovine serum in a CO_2 incubator at 37°C. This study has been approved (COA No. 677/2015) by the Institutional Review Board of the Faculty of Medicine, Chulalongkorn University.

RNA preparation for RNA sequencing: Total RNA was extracted from cervical cancer cell lines using an RNeasy® Mini Kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. Approximately 10,000,000 cells were used. The RNA was eluted twice in a total volume of 100 µL of RNase-free H_2O. RNA was quantified by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and samples with RIN value >8.0 were used for RNAseq analysis. Illumina TrueSeq Stranded mRNA Sample Preparation Kit was used. Samples were sequenced on an Illumina HiSeq2500 (Illumina, Inc., San Diego, CA, USA) using a paired-end run.

Data analysis: The Hisat2 v2.1.0 aligner was used to map the raw data reads onto the viral reference sequences as

Table 1: Primers for validation of HPV transcripts by real-time PCR

HPV types	Target gene	Forward (5'-3')	Reverse (5'-3')
HPV16	E7 full length	CAGCTCAGAGGAGGAGGATG	GCCCATTAACAGGTCTTCCA
	E6 full length	CATTATTGTTATAGTTTGTATGG	TGGAATCTTTGCTTTTTGTCC
	E6*I409	ACTGCGACGTGAGGTGTATTAAC	TGGAATCTTTGCTTTTTGTCC
	E6*II526	ACTGCGACGTGAGATCATCAAGAACACG	GTCGAGTCTCCTCCTCCTAC
	E6*X742	ACTGCGACGTGAGTGTGACTCTACG	GGTTTCTGAGAACAGATGGGGC
	E6*IV2709	ACTGCGACGTGAGGACGTGGTCCAG	TCTAGGCGCATGTGTTTCCAAT
	E6*III3356	ACTGCGACGTGAGCAGCAACGAAG	GGTCGCTGGATAGTCGTCTG
HPV18	E7 full length	TATGCATGGACCTAAGGC	GTGTGACGTTGTGGTTCGGCTCG
	E6 full length	TCTGTGTATGGAGACACATTGGA	CACTGGCCTCTATAGTGCCC
	E6*416	CTTACAGAGGTGCCTGCGG	CACTGGCCTCTATAGTGCCC
HPV45	E7 full length	CGAGTCAGAGGAGGAAAACG	CACAAAGGACAAGGTGCTCA
	E6 full length	TCTGTATATGGAGAGACACTGG	CTCGGTACTGTCCAGCTATGC
	E6*412	AACGCACAGAGGTGCCTGCGGTGCC	CTCGGTACTGTCCAGCTATGC
	E6*640	AACGCACAGAGAATGAATTAGATC	TTTTGTGACGCTGTGGTTCGGCTCG
HPV68b	E7 full length	ACCCGACCATGCAGTTAATC	TCTAGCTTCCGCAGGTTCTC
	E6 full length	GGGACGGGGTACCATTAGCTGC	CATATAACTTTGTATTAGTTATGG
	E6*311	CGGACAGAGGTGCATGTGTTGCCTG	CCTCTCGTTTACTGGTCCAGC
	E6*406	CGGACAGAGGACAGTGTCGCCAC	GCATGGACATAACTCTAACAC
House keeping genes	GAPDH	GCACCGTCAAGGCTGAGAAC	ATGGTGGTGAAGACGCCAGT

follows, HPV16 (NC_001526.2), HPV18 (NC_001357.1), HPV68b (FR751039.1) and HPV45 (X74479.1). Following the alignment, the viral reads were assembled into transcripts using Cufflinks v. 2.2.1. The transcripts were merged into full spliced transcripts (Cuffmerge) and quantified viral isoforms (Cuffdiff) as Fragments per Kilobase per Million (FPKM). The analyzed data were visualized, using the Integrated Genomic Viewer (IGV) to identify the expressed transcripts and splicing junction of each HPV type.

Validation of HPV transcripts by real-time polymerase chain reaction (real time-PCR): Primers were designed to detect the viral transcripts in five cell lines by real-time RT-PCR. cDNA was synthesized using Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA). Real-time PCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA). Primer sequences were shown in Table 1. The PCR conditions were: initial denaturing at 98°C for 30 sec, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. GAPDH was used as a housekeeping gene and delta-delta Ct was calculated to investigate differences in gene expression.

RESULTS

Analysis of expressed transcripts: The reads were mapped onto the viral reference sequences to identify the viral transcripts that were expressed in each cervical cancer cell lines (CaSki (HPV16), SiHa (HPV16), HeLa (HPV18), MS751 (HPV45) and ME180 (HPV68b)). Transcripts containing truncated E6 and full-length E7 genes were highly detected in all four HR-HPV types 16, 18, 45 and 68b. Full-length E6, full-length E1 and truncated E1 transcripts were also detected but in less amount (Fig. 1a-e). CUFF.1.1 (CaSki, Fig. 1a), CUFF.1.2 (SiHa, Fig. 1b), CUFF.1.1 (HeLa, Fig. 1c), CUFF.3.1 (MS751, Fig. 1d) and CUFF.3.2 (ME180, Fig. 1e) were highly expressed as shown by the coverage and FPKM of transcripts. The FPKM values of viral transcripts were shown in supplementary data S1. Transcript containing full-length E2 and truncated E2 ORF were found in HPV16 positive CaSki cell line (Fig. 1a). The amplification of papillomavirus oncogene transcripts (APOTs) and sanger sequencing assays were performed and revealed that the HPV16 E2/E4 gene was retained in the CaSki cell line (supplementary data S2).

The splicing junctions of all four HPV types were found within E6 and E1 regions. For the E6 region, one Splicing Donor (SD) at the 5 end and different Splicing Acceptor (SA) positions at the 3 end were found as follow; three splicing junctions were found in HPV16 positive cervical cancer cell lines, CaSki and SiHa, SD226^SA409(E6*I), SD226^SA526(E6*II) and SD226^SA742 (E6*X). Two splicing junctions were found in HPV18 (HeLa), SD233^SA416 (E6*I), SD233^SA635, HPV45 (MS751); SD230^SA412 (E6*I), SD230^SA640 and HPV68b (ME180); SD129^SA311 (E6*I), SD129^SA406.

Splicing junctions within E1 region found in CaSki and SiHa were SD880^SA3358, SD880^SA3361, SD880^SA3391, SD880^SA1726, SD880^SA2405, SD880^SA2582(E1C), SD880^SA2709 (E2*), SD880^SA3020, SD880^SA3078, SD880^SA3329, SD577^SA6810, SD898^SA1725, SD1302^SA2709, SD1302^SA3358 (E2C), SD1760^SA3391, SD1263^SA3391, SD2309^SA3461 and the other forms were SD96^SA1063, SD226^ SA2709 (E6*IV), SD226^SA3329, SD226^SA3358 (E6*III), SD226^SA3361, SD226^SA3391 and

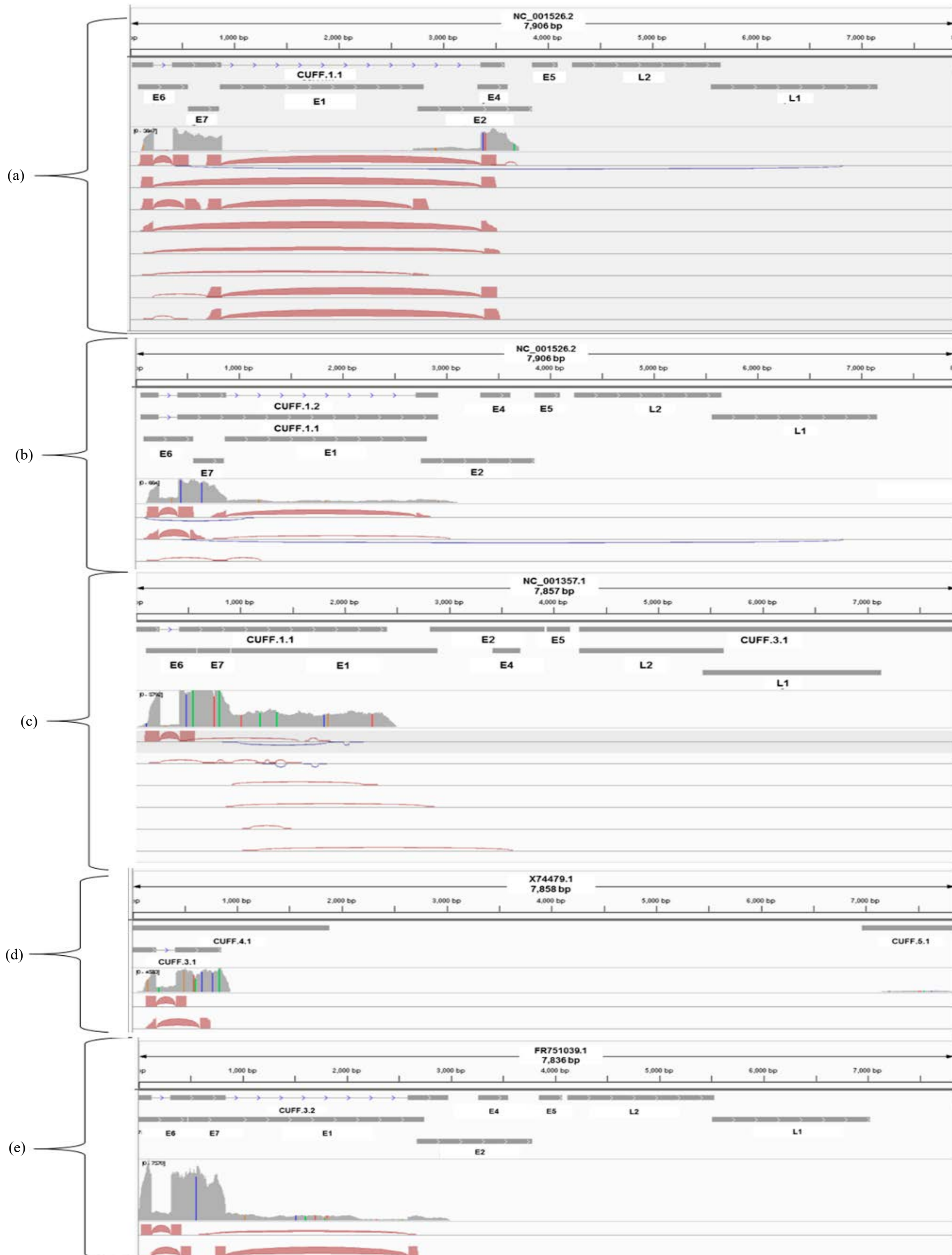

Fig. 1(a-e): Read depth maps and splicing events visualized by IGV of four HR-HPV types 16, 18, 45 and 68b transcripts obtained from five different cervical cancer cell lines

(a) CaSki (HPV16), (b) SiHa (HPV16), (c) HeLa (HPV18), (d) MS751 (HPV45) and (e) ME180 (HPV68b). The fragments per kilobase per million (FPKM) of CUFF and other viral transcripts were shown in supplementary S1 excel

S1 excel: Fragments per kilobase per million (FPKM) of CUFF and various HR-HPV transcripts

Tracking_id	Class_code	Nearest_ref_id	Gene_id	Gene_short_name	tss_id	Locus	Length	Coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi	FPKM_status
CaSki												
CUFF.1.1	-	-	CUFF.1	-	-	NC_001526.2:25-3588	905	1995.28	664626	630633	698620	OK
gene0	-	-	CUFF.1	E6	-	NC_001526.2:82-559	477	0	0	0	819.678	OK
gene1	-	-	CUFF.1	E7	-	NC_001526.2:561-858	297	0	0	0	2885.15	OK
gene2	-	-	CUFF.1	E1	-	NC_001526.2:864-2813	1949	162.464	54116.7	49891.2	58342.2	OK
gene3	-	-	CUFF.1	E2	-	NC_001526.2:2754-3852	1098	796.817	265420	249748	281092	OK
gene4	-	-	CUFF.1	E4	-	NC_001526.2:3331-3619	288	0	0	0	3237.83	OK
gene5	-	-	CUFF.1	E5	-	NC_001526.2:3848-4100	252	0	0	0	793.096	OK
gene6	-	-	CUFF.2	L2	-	NC_001526.2:4235-5657	1422	0.504699	168.115	0	459.3	OK
gene7	-	-	CUFF.2	L1	-	NC_001526.2:5559-7155	1596	0.880522	293.302	0	637.551	OK
CUFF.1	-	-	CUFF.1	E1,E2,E4,E5,E6,E7	-	NC_001526.2:25-4100	-	-	984163	946204	1.02E+06	OK
CUFF.2	-	-	CUFF.2	L1,L2	-	NC_001526.2:4235-7155	-	-	461.417	6.54258	916.292	OK
SiHa												
CUFF.1.1	-	-	CUFF.1	-	-	NC_001526.2:54-2921	2685	74.3923	162393	146412	178375	OK
CUFF.1.2	-	-	CUFF.1	-	-	NC_001526.2:54-2921	858	233.887	510559	454547	566571	OK
gene0	-	-	CUFF.1	E6	-	NC_001526.2:82-559	477	173.334	378377	311933	444821	OK
gene1	-	-	CUFF.1	E7	-	NC_001526.2:561-858	297	0	0	0	18907.5	OK
gene2	-	-	CUFF.1	E1	-	NC_001526.2:864-2813	1949	0	0	0	763.339	OK
gene3	-	-	CUFF.1	E2	-	NC_001526.2:2754-3852	1098	0	0	0	1514.47	OK
gene4	-	-	CUFF.1	E4	-	NC_001526.2:3331-3619	288	0	0	0	4547.77	OK
gene5	-	-	CUFF.1	E5	-	NC_001526.2:3848-4100	252	0	0	0	5197.45	OK
gene6	-	-	CUFF.2	L2	-	NC_001526.2:4235-5657	1422	0	0	0	0	OK
gene7	-	-	CUFF.2	L1	-	NC_001526.2:5559-7155	1596	0	0	0	0	OK
CUFF.1	-	-	CUFF.1	E1,E2,E4,E5,E6,E7	-	NC_001526.2:54-4100	-	-	1.05E+06	962702	1.14E+06	OK
CUFF.2	-	-	CUFF.2	L1,L2	-	NC_001526.2:4235-7155	-	-	0	0	0	OK
HeLa												
CUFF.1.1	-	-	CUFF.1	-	-	NC_001357.1:0-2404	2222	2981.31	440110	420894	459327	OK
gene0	-	-	CUFF.2	E6	-	NC_001357.1:104-581	477	0	0	0	362.57	OK
gene1	-	-	CUFF.2	E7	-	NC_001357.1:589-907	318	0	0	0	1003.45	OK
gene2	-	-	CUFF.2	E1	-	NC_001357.1:913-2887	1974	0	0	0	50.8667	OK
gene3	-	-	CUFF.2	E2	-	NC_001357.1:2816-3914	1098	0	0	0	102.363	OK
gene4	-	-	CUFF.2	E4	-	NC_001357.1:3417-3684	267	0	0	0	331.738	OK
gene5	-	-	gene5	E5	-	NC_001357.1:3935-4157	222	0	0	0	0	OK
CUFF.3.1	-	-	CUFF.3	-	-	NC_001357.1:4243-7857	3614	17.0838	2521.96	2103.66	2940.27	OK
gene6	-	-	CUFF.3	L2	-	NC_001357.1:4243-5632	1389	0	0	0	63.7681	OK
gene7	-	-	CUFF.3	L1	-	NC_001357.1:5429-7136	1707	13.725	2026.14	1387.29	2664.99	OK
CUFF.1	-	-	CUFF.1	-	-	NC_001357.1:0-2404	-	-	440110	420894	459327	OK
CUFF.2	-	-	CUFF.2	E1,E2,E4,E6,E7	-	NC_001357.1:104-3914	-	-	0	0	0	OK
gene5	-	-	gene5	E5	-	NC_001357.1:3935-4157	-	-	0	0	0	OK
CUFF.3	-	-	CUFF.3	L1,L2	-	NC_001357.1:4243-7857	-	-	4548.1	3872.41	5223.78	OK

S1 excel: Continue

Tracking_id	Class_code	Nearest_ref_id	Gene_id	Gene_short_name	tss_id	Locus	Length	Coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi	FPKM_status
MS751												
NC_001357.1_gene0	-	-	NC_001357.1_gene0	NC_001357.1_E6	-	NC_001357.1:104-581	477	0	0	0	0	OK
NC_001357.1_gene1	-	-	NC_001357.1_gene1	NC_001357.1_E7	-	NC_001357.1:589-907	318	0	0	0	0	OK
NC_001357.1_gene5	-	-	NC_001357.1_gene5	NC_001357.1_E5	-	NC_001357.1:3935-4157	222	0	0	0	0	OK
NC_001357.1_gene2	-	-	CUFF.1	NC_001357.1_E1	-	NC_001357.1:913-2887	1974	0	0	0	0	OK
NC_001357.1_gene3	-	-	CUFF.1	NC_001357.1_E2	-	NC_001357.1:2816-3914	1098	0	0	0	0	OK
NC_001357.1_gene4	-	-	CUFF.1	NC_001357.1_E4	-	NC_001357.1:3417-3684	267	0	0	0	0	OK
NC_001357.1_gene6	-	-	CUFF.2	NC_001357.1_L2	-	NC_001357.1:4243-5632	1389	0	0	0	0	OK
NC_001357.1_gene7	-	-	CUFF.2	NC_001357.1_L1	-	NC_001357.1:5429-7136	1707	0	0	0	0	OK
CUFF.3.1	-	-	CUFF.3	-	-	X74479.1:1-854	671	2428.7	891922	847225	936618	OK
CUFF.4.1	-	-	CUFF.4	-	-	X74479.1:1-1878	1877	599.41	220128	208872	231385	OK
CUFF.5.1	-	-	CUFF.5	-	-	X74479.1:6962-7858	896	237.328	87156.8	80998	93315.6	OK
NC_001357.1_gene0	-	-	NC_001357.1_gene0	NC_001357.1_E6	-	NC_001357.1:104-581	-	-	0	0	0	OK
NC_001357.1_gene1	-	-	NC_001357.1_gene1	NC_001357.1_E7	-	NC_001357.1:589-907	-	-	0	0	0	OK
NC_001357.1_gene5	-	-	NC_001357.1_gene5	NC_001357.1_E5	-	NC_001357.1:3935-4157	-	-	0	0	0	OK
CUFF.1	-	-	CUFF.1	NC_001357.1_E1, NC_001357.1_E2, NC_001357.1_E4	-	NC_001357.1:913-3914	-	-	0	0	0	OK
CUFF.2	-	-	CUFF.2	NC_001357.1_L1, NC_001357.1_L2	-	NC_001357.1:4243-7136	-	-	0	0	0	OK
CUFF.3	-	-	CUFF.3	-	-	X74479.1:1-854	-	-	891922	847219	936624	OK
CUFF.4	-	-	CUFF.4	-	-	X74479.1:1-1878	-	-	220128	208872	231385	OK
CUFF.5	-	-	CUFF.5	-	-	X74479.1:6962-7858	-	-	87156.8	80998	93315.6	OK
ME180												
NC_001357.1_gene5	-	-	NC_001357.1_gene5	NC_001357.1_E5	-	NC_001357.1:3935-4157	222	0	0	0	0	OK
gene0	-	-		NC_001357.1_E6	-	NC_001357.1:104-581	477	0	0	0	0	OK
NC_001357.1_gene1	-	-	NC_001357.1_gene1	NC_001357.1_E7	-	NC_001357.1:589-907	318	0	0	0	0	OK
NC_001357.1_gene6	-	-	CUFF.1	NC_001357.1_L2	-	NC_001357.1:4243-5632	1389	0	0	0	0	OK
NC_001357.1_gene7	-	-	CUFF.1	NC_001357.1_L1	-	NC_001357.1:5429-7136	1707	0	0	0	0	OK
NC_001357.1_gene2	-	-	CUFF.2	NC_001357.1_E1	-	NC_001357.1:913-2887	1974	0	0	0	0	OK
NC_001357.1_gene3	-	-	CUFF.2	NC_001357.1_E2	-	NC_001357.1:2816-3914	1098	0	0	0	0	OK
NC_001357.1_gene4	-	-	CUFF.2	NC_001357.1_E4	-	NC_001357.1:3417-3684	267	0	0	0	0	OK
FR751039.1_gene0	-	-	CUFF.3	FR751039.1_E6	-	FR751039.1:0-477	477	0	0	0	321.367	OK
CUFF.3.2	-	-	CUFF.3	-	-	FR751039.1:0-2973	1044	3678.55	776556	742111	811001	OK
FR751039.1_gene1	-	-	CUFF.3	FR751039.1_E7	-	FR751039.1:484-817	333	0	0	0	701.942	OK
FR751039.1_gene2	-	-	CUFF.3	FR751039.1_E1	-	FR751039.1:823-2746	1923	619.222	130720	124221	137220	OK
FR751039.1_gene3	-	-	CUFF.3	FR751039.1_E2	-	FR751039.1:2672-3785	1113	0	0	0	93.9538	OK
FR751039.1_gene4	-	-	CUFF.3	FR751039.1_E4	-	FR751039.1:3267-3552	285	0	0	0	296.286	OK
FR751039.1_gene5	-	-	FR751039.1_gene5	FR751039.1_E5	-	FR751039.1:3849-4071	222	0	0	0	0	OK
FR751039.1_gene6	-	-	CUFF.4	FR751039.1_L2	-	FR751039.1:4117-5527	1410	2.0071	423.706	169.09	678.322	OK
FR751039.1_gene7	-	-	CUFF.4	FR751039.1_L1	-	FR751039.1:5507-7025	1518	2.30104	485.759	226.688	744.831	OK

S1 excel: Continue

Tracking_id	Class_code	Nearest_ref_id	Gene_id	Gene_short_name	tss_id	Locus	Length	Coverage	FPKM	FPKM_conf_lo	FPKM_conf_hi	FPKM_status
NC_001357.1_gene5	-	-	NC_001357.1_gene5	NC_001357.1_E5	-	NC_001357.1:3935-4157	-	-	0	0	0	OK
	-	-		NC_001357.1_E6	-	NC_001357.1:104-581	-	-	0	0	0	OK
NC_001357.1_gene1	-	-	NC_001357.1_gene1	NC_001357.1_E7	-	NC_001357.1:589-907	-	-	0	0	0	OK
CUFF.1	-	-	CUFF.1	NC_001357.1_L1, NC_001357.1_L2	-	NC_001357.1:4243-7136	-	-	0	0	0	OK
CUFF.2	-	-	CUFF.2	NC_001357.1_E1, NC_001357.1_E2, NC_001357.1_E4	-	NC_001357.1:913-3914	-	-	0	0	0	OK
CUFF.3	-	-	CUFF.3	FR751039.1_E1, FR751039.1_E2, FR751039.1_E4, FR751039.1_E6, FR751039.1_E7	-	FR751039.1:0-3785	-	-	907276	872208	942344	OK
FR751039.1_gene5	-	-	FR751039.1_gene5	FR751039.1_E5	-	FR751039.1:3849-4071	-	-	0	0	0	OK
CUFF.4	-	-	CUFF.4	FR751039.1_L1, FR751039.1_L2	-	FR751039.1:4117-7025	-	-	909.466	544.728	1274.2	OK

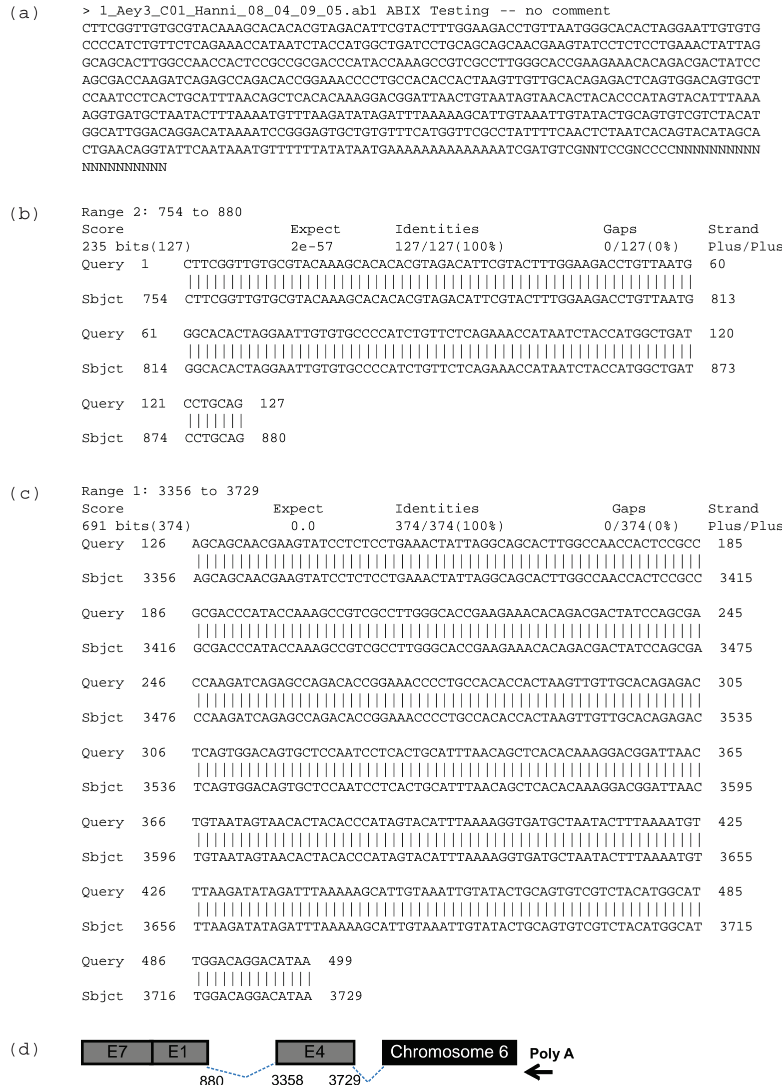

Fig. S2: HPV16 sequence transcript (S2A) obtained from CaSki cell line detected by APOT assay and direct sequencing was presented by BLAST analysis (S2B and S2C) and proposed transcript (S2D)

SD579^SA6809. Splicing junctions within E1 region in HPV18 (HeLa) were SD929^SA2779, SD977^SA1836, SD1342^SA1436, SD1987^SA2047 and one splicing event within E7 region, SD599^SA619. HPV68b (ME180) were SD839^SA2586, SD683^SA2586, SD839^SA2586 and no E1 splicing junctions were found in HPV45 (MS751). Transcripts found in four HR-HPV types 16, 18, 45 and 68b in five cervical cancer cell lines were expressed as biscistronic and polycistronic as shown in Fig. 2a-e.

Major transcripts of HR-HPV types 16, 18, 45 and 68b: According to coverage and FPKM values visualized by IGV software, major transcripts found in CaSki (HPV16), SiHa (HPV16), HeLa (HPV18), MS751 (HPV45) and ME180 (HPV68b) were SD226^SA409-SD880^SA3358 (Fig. 2a), SD226^SA409-SD880^SA2709 (Fig. 2b), SD233^SA416 (Fig. 2c), SD230^SA412 (Fig. 2d) and SD129^SA331-SD839^SA2586 (Fig. 2e), respectively. All major transcripts of four HR-HPVs contain disruption within the E6 gene.

Validation of HPV oncogene expressions by real-time PCR: HPV Transcripts Were Validated To Confirm the RNAseq results, primers specific for a full length and spliced HPV transcript types were designed as shown in Table 1. It has been shown that full-length E7 transcript of all 4 HR-HPV types was highly expressed. A spliced transcript was the second most expressed, however, full-length E6 transcript was less expressed (Fig. 3).

Fig. 2(a-e): Map of HPV16, 18, 45 and 68b transcripts detected in five cervical cancer cell lines
(a) CaSki (HPV16), (b) SiHa (HPV16), (c) HeLa (HPV18), (d) MS751 (HPV45) and (e) ME180 (HPV68b)

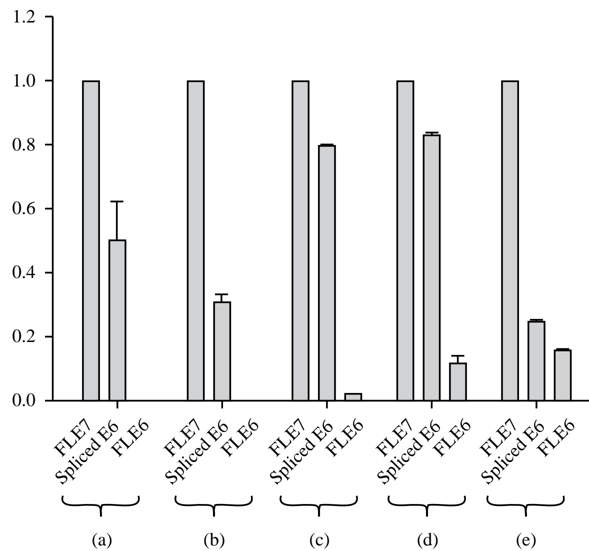

Fig. 3: Validation of HPV16, 18, 45 and 68b oncogene expressions (full-length E7 (FLE7), spliced E6 (E6*1) and FLE6) detected in five cell lines by real-time PCR

DISCUSSION

Four HR-HPV types 16, 18, 45 and 68 mRNA obtained from five different cervical cancer cell lines were used for RNAseq analysis to identify major transcripts necessary for the expression of viral oncogenes. Expression of viral oncogenes E6 and E7 are detected as bicistronic and polycistronic, however, it has been reported that a bicistronic transcript encoding the full-length E6 and E7 ORFs translates only E6 protein[14,15]. The mechanism is that the spacing between the E6 ORF and E7 ORF within this full-length transcript is only two nucleotides, thus, the scanning ribosome does not have enough time to discard its releasing factors and to efficient reinitiation of E7 protein translation. To efficiently translate the E7 gene, a monocistronic transcript containing the full-length E7 ORF transcribed from a promoter located within the E6 region should be expressed[17,18]. Deep RNA sequencing was performed to identify and quantitate the major transcripts that were expressed in five cervical cancer cell lines containing four different HR-HPV types included HPV16, 18, 45 and 68b. The RNAseq results of four HR-HPV types in the present study revealed less mRNA transcripts encoding the full-length E6 and monocistronic full-length E7 transcripts (Fig. 1-2). Thus, these two transcript types were not required for increased E6 and E7 protein expression. The major transcripts that were detected in all four HR-HPV types were transcripts with disruption of the E6 intronic region where a splicing event occurred that consistent with previously published papers that E7 oncoprotein was efficiently expressed from mRNA with splicing event occurs within E6 gene[14,19,20,23]. The explanation is that splicing of E6 intron creates a frameshift, resulting in the spliced mRNA which has enough nucleotide space for the translation termination of truncated E6 protein and re-initiation of full-length E7 translation[14]. However, it has been reported that the distance between the splicing acceptor at the E6 region and the start codon of the E7 gene must be optimal to efficiently translate E7 protein[15,23]. Quantitative analysis as shown by coverage and FPKM values of major transcripts revealed that transcript types 226^409 (HPV16), 233^416 (HPV18), 230^412 (HPV45) and 129^311(HPV68b) with full-length E7 ORF were mostly expressed (Fig. 2 a-e) and these transcripts were named E6*I[14]. The HPV16, for example, the distance between the SA409 to E7 start codon at nucleotide position 562 is 153 bp that is the optimal space for strongly translate E7 gene when compared to SA526 to E7 start codon 562 (distance 36 bp). The transcript with splicing acceptor within the E7 gene and unspliced within the E1 gene were proposed to encode for E1 protein[24-26]. It has been reported that splicing event within the E6 gene was found only in HR-HPVs but not Low risk HPVs[22]. The level of E6*I transcript found in CaSki and SiHa were not largely different as shown by RNAseq (FPKM 664,626 and 510,559 in CaSki and SiHa, respectively) and real-time PCR (Fig. 3), although there is the difference in copy number of integrated HPV16 within each cell line, that may be due to high methylation (>90%) and closed chromatin were found in the Long Control Region (LCR) of HPV16 in CaSki cell lines (600 copies per cell), however, no methylation was found in HPV16 LCR in SiHa cell line (1-2 copies per cell)[27-29].

It has been shown by APOTs assay that the HPV16 E2/E4 gene was retained in the CaSki cell line and splicing acceptor was also found in the E2/E4 gene at position 3358 (supplementary data S2). In accordance with RNAseq results, in which the major transcript of HPV16 found in the CaSki cell line was SD226/SA409-SD880/SA3358. The importance of SA3358 for HPV16 E6 and E7 translation was reported[30]. When comparing HPV16 transcripts found in cervical cancer cell lines, CaSki and SiHa in the present study with previous report transcripts of the W12 cell line (W12 cell lines derived from LSIL cells), it was found that transcript encoding the full-length E6 and E7 ORFs with splicing donor at 880 was found only in cervical cancer cell lines[7,12]. This transcript type was previously reported to be expressed only in CIN3 and Squamous Cell Carcinoma (SCC)[19]. Transcript type E6*I was mostly detected in cervical cancer[9,19,22]. In the present study, HPV16 oncogene transcripts E6*I containing truncated E6 and full-length E7 were present in high levels compared to E6*II and E6*X in both CaSki and SiHa cervical cancer cell lines. E6*I and E7 transcripts were also highly expressed in HR-HPV types 18, 45 and 68b. One study showed that E7 oncoprotein of HPV16 and HPV18 was highly expressed in CaSki and HeLa, respectively, that correlated with the major transcript for E7 gene expression (E6*I) detected in the present study[31].

The cervical cancer cell line MS751 revealed transcripts of truncated E6, E7 and part of E1 ORFs that consistent with the previous report on the characterization of HPV45 DNA in MS751 cell line in which E6, E7 and part of E1 ORFs were retained after viral integration into host chromosome[32]. Apart from E6*I, E7 and E1 transcripts, both HPV18 and 45 (Both HPVs are in the same group, Alpha 7 papillomavirus) expressed non-coding region, however, their significance is not well understood.

In summary, full-length HPV E7 ORF was found in most expressed transcripts, followed by truncated E6 (E6*1). Full-length E6 ORF was less detected. E2 transcript was also found in CaSki cell lines. HR-HPV16, 18, 45, 68b expressed

common major transcripts with splicing donor and acceptor sites that occurred within the E6 gene (E6*I) that facilitate E7 oncogene translation. Thus, major transcripts encoded for truncated E6*I isoform and E7 proteins expressed from HR-HPV types may be involved in carcinogenesis[33] could be used as biomarkers for diagnostics and prognostics of HPV related diseases. The role of truncated proteins that are transcribed from various HR-HPV transcripts in carcinogenesis is limited, further functional experiments of mentioned proteins is of interest. The transcripts might be used as the target for cervical cancer therapy which may help to improve the treatment. Although, the present study did not performed RNAseq in clinical samples due to very low amount of RNA obtained from cervical scrapes, the data presented in the present study would be useful for functional analysis of the role of truncated HR-HPV proteins in cervical cancer cell lines.

CONCLUSION

RNA sequencing analysis of four HR-HPV types including 16, 18, 45 and 68b obtained from five cervical cancer cell lines revealed various transcript types, however, common major transcripts with splicing donor and acceptor sites that occurred within the E6 gene (E6*I) were found in all four HR-HPV types. The major transcript could be used as biomarkers for diagnostics and prognostics of HPV related diseases. The important of truncated E6 protein (E6*I) in carcinogenesis need to be further investigated.

SIGNIFICANCE STATEMENT

This study discovers the common major transcripts as shown in Fig. 1a-e named as CUFF that were highly expressed from HR-HPV types 16, 18, 45 and 68b as presented by FPKM values (supplementary data S1). The transcript data can be beneficial for researchers to design primers for cloning of various viral genes especially truncated forms such as E6*I, E6*II for further study the role and the involvement of these viral proteins expressed from different HR-HPV types in carcinogenesis and viral-host protein interactions. Viral gene knock down experiments using cervical cancer cell lines as a model could be applied. This study will help the researcher to uncover the critical role of truncated HR-HPV E6 proteins that limited data were reported so far. Thus, a new molecular pathway and the specific role of truncated E6 proteins driven carcinogenesis may be arrived at.

ACKNOWLEDGMENTS

This work was supported by Chulalongkorn University Office of International Affairs Scholarship for Short-term Research, Ratchadapiseksompoch Fund, Faculty of Medicine, Chulalongkorn University, grant number RA59/013 and was partly supported by the TRF Grant for a new researcher from Thailand Research Fund Grant (Grant number TRG6080008). We thank Vanida Mungmee for her kindly help in cell culture preparation, Megumi Kombu for her help with the RNAseq library construction and Yuuta Kuze for his help with the data analyses. We thank the Suzuki lab at the University of Tokyo for providing facilities necessary for the RNAseq experiment.

REFERENCES

1. Munger, K., A. Baldwin, K.M. Edwards, H. Hayakawa and C.L. Nguyen *et al*., 2004. Mechanisms of human papillomavirus-induced oncogenesis. J. Virol., 78: 11451-11460.
2. Münger, K., J.R. Basile, S. Duensing, A. Eichten, S.L. Gonzalez, M. Grace and V.L. Zacny, 2001. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. Oncogene, 20: 7888-7898.
3. Clifford, G.M., J.S. Smith, M. Plumier, N. Munoz and S. Franceschi, 2003. Human papillomavirus type in invasive cervical cancer worldwide: A meta-analysis. Br. J. Cancer, 88: 63-73.
4. Munoz, N., F.X. Bosch, S. de Sanjose, R. Herrero and X. Castellsague *et al*., 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N. Engl. J. Med., 348: 518-527.
5. Ozbun, M.A. and C. Meyers, 1998. Human papillomavirus type 31b E1 and E2 transcript expression correlates with vegetative viral genome amplification. Virology, 248: 218-230.
6. Ozbun, M.A., 2002. Human papillomavirus type 31b infection of human keratinocytes and the onset of early transcription. J. Virol., 76: 11291-11300.
7. Milligan, S.G., T. Veerapraditsin, B. Ahamet, S. Mole and S.V. Graham, 2007. Analysis of novel human papillomavirus type 16 late mRNAs in differentiated W12 cervical epithelial cells. Virology, 360: 172-181.
8. Stubenrauch, F. and L.A. Laimins, 1999. Human papillomavirus life cycle: Active and latent phases. Semin. Cancer Biol., 9: 379-386.
9. Schmitt, M., V. Dalstein, T. Waterboer, C. Clavel, L. Gissmann and M. Pawlita, 2011. The HPV16 transcriptome in cervical lesions of different grades. Mol. Cell. Probes, 25: 260-265.

10. Wang, X., C. Meyers, H.K. Wang, L.T. Chow and Z.M. Zheng, 2011. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. J. Virol., 85: 8080-8092.
11. Ajiro, M., S. Tang, J. Doorbar and Z.M. Zheng, 2016. Serine/arginine-rich splicing factor 3 and heterogeneous nuclear ribonucleoprotein A1 regulate alternative RNA splicing and gene expression of human papillomavirus 18 through two functionally distinguishable cis elements. J. Virol., 90: 9138-9152.
12. Graham, S.V., 2010. Human papillomavirus: Gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. Future Microbiol., 5: 1493-1506.
13. Graham, S.V. and A.A.A. Faizo, 2017. Control of human papillomavirus gene expression by alternative splicing. Virus Res., 231: 83-95.
14. Tang, S., M. Tao, J.P. McCoy Jr. and Z.M. Zheng, 2006. 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: 4249-4263.
15. Smotkin, D., H. Prokoph and F.O. Wettstein, 1989. Oncogenic and nononcogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. J. Virol., 63: 1441-1447.
16. Al-Allaf, F.A., Z. Abduljaleel, M. Athar, M.M. Taher and W. Khan *et al*., 2019. Modifying inter-cistronic sequence significantly enhances IRES dependent second gene expression in bicistronic vector: Construction of optimised cassette for gene therapy of familial hypercholesterolemia. Non-Coding RNA Res., 4: 1-14.
17. Braunstein, T.H., B.S. Madsen, B. Gavnholt, M.W. Rosenstierne, C.K. Johnsen and B. Norrild, 1999. Identification of a new promoter in the early region of the human papillomavirus type 16 genome. J. Gen. Virol., 80: 3241-3250.
18. Glahder, J.A., C.N. Hansen, J. Vinther, B.S. Madsen and B. Norrild, 2003. A promoter within the E6 ORF of human papillomavirus type 16 contributes to the expression of the E7 oncoprotein from a monocistronic mRNA. J. Gen. Virol., 84: 3429-3441.
19. Chen, J., Y. Xue, M. Poidinger, T. Lim and S.H. Chew *et al*., 2014. Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements during carcinogenic progression. Virology, 462-463: 14-24.
20. Lin, K., X. Lu, J. Chen, R. Zou, L. Zhang and X. Xue, 2015. E6-associated transcription patterns in human papilloma virus 16-positive cervical tissues. Oncol. Lett., 9: 478-482.
21. Toots, M., A. Männik, G. Kivi, M. Ustav, E. Ustav and M. Ustav, 2014. The transcription map of human papillomavirus type 18 during genome replication in U2OS cells. PLoS ONE, Vol. 9. 10.1371/journal.pone.0116151.
22. Zheng, Z.M., 2010. Viral oncogenes, noncoding RNAs and RNA splicing in human tumor viruses. Int. J. Biol. Sci., 6: 730-755.
23. Zheng, Z.M., M. Tao, K. Yamanegi, S. Bodaghi and W. Xiao, 2004. 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: 1091-1108.
24. Wu, C., N. Kajitani and S. Schwartz, 2017. Splicing and polyadenylation of human papillomavirus type 16 mRNAs. Int. J. Mol. Sci., Vol. 18. 10.3390/ijms18020366.
25. Schwartz, S., 2013. Papillomavirus transcripts and posttranscriptional regulation. Virology, 445: 187-196.
26. Johansson, C. and S. Schwartz, 2013. Regulation of human papillomavirus gene expression by splicing and polyadenylation. Nat. Rev. Microbiol., 11: 239-251.
27. Chaiwongkot, A., S. Vinokurova, C. Pientong, T. Ekalaksananan and B. Kongyingyoes *et al*., 2013. Differential methylation of E2 binding sites in episomal and integrated HPV 16 genomes in preinvasive and invasive cervical lesions. Int. J. Cancer, 132: 2087-2094.
28. Arce, J.D.C., E. Gockel-Krzikalla and F. Rosl, 2012. Silencing of multi-copy HPV16 by viral self-methylation and chromatin occlusion: A model for epigenetic virus-host interaction. Hum. Mol. Genet., 21: 1693-1705.
29. Stu¨nkel, W. and H.U. Bernard, 1999. The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. J. Virol., 73: 1918-1930.
30. Li, X., C. Johansson, C.C. Palacios, A. Mossberg, S. Dhanjal, M. Bergvall and S. Schwartz, 2013. Eight nucleotide substitutions inhibit splicing to HPV-16 3'-splice site SA3358 and reduce the efficiency by which HPV-16 increases the life span of primary human keratinocytes. PLoS ONE, Vol. 8. 10.1371/journal.pone.0072776.
31. Seedorf, K., T. Oltersdorf, G. Krämmer and W. Röwekamp, 1987. Identification of early proteins of the human papilloma viruses type 16 (HPV 16) and type 18 (HPV 18) in cervical carcinoma cells. EMBO J., 6: 139-144.
32. Geisbill, J., U. Osmers and M. Durst, 1997. Detection and characterization of human papillomavirus type 45 DNA in the cervical carcinoma cell line MS751. J. Gen. Virol., 78: 655-658.
33. Paget-Bailly, P., K. Meznad, D. Bruyère, J. Perrard and M. Herfs *et al*., 2019. Comparative RNA sequencing reveals that HPV16 E6 abrogates the effect of E6*I on ROS metabolism. Sci. Rep., Vol. 9. 10.1038/s41598-019-42393-6.