

**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-Graduação**



**INSTITUTO NACIONAL DE CÂNCER  
PÓS-GRADUAÇÃO EM CONCOLOGIA**

**Ayslan Castro Brant**

**Análise da integração genômica do Papilomavírus Humano e a  
expressão de seus genes em câncer cervical**

**Orientador: Dr. Miguel Angelo Martins Moreira**

**Rio de Janeiro**

**Fevereiro 2015**

**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-Graduação**



**INSTITUTO NACIONAL DE CÂNCER  
PÓS-GRADUAÇÃO EM ONCOLOGIA**

**Ayslan Castro Brant**

**Análise da integração genômica do Papilomavírus Humano e a  
expressão de seus genes em câncer cervical**

Dissertação de mestrado apresentada  
ao Instituto Nacional de Câncer José  
de Alencar Gomes da Silva como parte  
dos requisitos para obtenção do grau  
de Mestre em Oncologia

**Orientador:** Dr. Miguel Angelo Martins Moreira

**Rio de Janeiro**

**Fevereiro 2015**

B816a Brant, Ayslan Castro.

Análise da integração genômica do Papilomavírus Humano e a expressão de seus genes em câncer cervical. / Ayslan Castro Brant. – Instituto Nacional de Câncer José Alencar Gomes da Silva, 2015.

118 f. il.: color.

Orientador: Dr. Miguel Angelo Martins Moreira.

Dissertação (Mestrado em Oncologia)– Instituto Nacional de Câncer José Alencar Gomes da Silva, 2015.

1. Genômica. 2. Papilomavírus Humano. 3. Neoplasias do Colo do Útero. I. Moreira, Miguel Angelo Martins (orient.). II. Instituto Nacional de Câncer José Alencar Gomes da Silva. III. Título.

616.99414

**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-Graduação**



**INSTITUTO NACIONAL DE CÂNCER  
PÓS-GRADUAÇÃO EM ONCOLOGIA**

**Ayslan Castro Brant**

**Análise da integração genômica do Papilomavírus Humano e a  
expressão de seus genes em câncer cervical**

**Orientador:** Dr. Miguel Angelo Martins Moreira

**Aprovada em:**

**Examinadores:**

Dr<sup>a</sup>. Claudia Esther Alicia Rocio Hassan – Presidente

Dr<sup>a</sup>. Silvia Maria Baeta Cavalcanti

Dr. Marc Andre Michael Thorsten Sammeth

Dr. Marcelo Alves Soares – Suplente I

Dr. Francisco Prosdocimi de Castro Santos – Suplente II

**Rio de Janeiro**

**Fevereiro 2015**

## Análise da integração genômica do Papilomavírus Humano e a expressão de seus genes em câncer cervical

Ayslan Castro Brant

### RESUMO

O câncer de colo de útero é o terceiro tipo de neoplasia com maior incidência mundial em mulheres, apresentando o Vírus do Papiloma Humano (HPV) como fator necessário mas não suficiente para o seu desenvolvimento. A imortalização e o poder de transformação das células neoplásicas se devem aos oncogenes virais *E6* e *E7*, que geram oncoproteínas capazes de mediar estímulos mitogênicos e antiapoptóticos através de sua interação com proteínas reguladoras do ciclo celular. Esses genes são expressos como um pré-RNA *E6/E7* bicistônico, e por meio da existência de diferentes sítios de *splicing*, duas ou mais isoformas de mRNA *E6/E7* são produzidos. O genoma do HPV é mantido na forma episomal no estágio inicial das lesões displásicas de baixo grau, porém, nas células pré-cancerosas avançadas e na maioria dos carcinomas, ele está integrado no DNA cromossômico. Essa integração causa o aumento da expressão dos oncogenes *E6* e *E7*, promovendo vantagens seletiva para o crescimento tumoral. O objetivo do presente trabalho foi analisar o padrão de integração do genoma viral no genoma do hospedeiro, a expressão dos genes do HPV e os produtos de *splicing* do RNA *E6/E7* em tumores cervicais. Para esse fim extraímos o RNA de 25 biópsias (14 infectadas com HPV16, 10 com HPV18 e uma co-infectada pelo HPV16 e HPV18). A partir do RNA das amostras, foi realizado o sequenciamento de RNA (RNA-Seq) na plataforma *Hi-Seq 2500* (Illumina). Através da busca por *contigs* quiméricos pela montagem “de novo” do transcriptoma das células, encontramos integração em 18 amostras. Na maioria das amostras o rompimento do genoma viral ocorreu nos genes *E1* ou *E2*. Cerca de 52,6% das integrações ocorreram em regiões intergênicas do genoma celular, o restante ocorreu dentro dos genes *MCL1*, *PIGR*, *PXMP4*, *TP63*, *MMP13*, *SLC16A14*, *HRH1*, *RAD51B* e *NRXN1*. Nas amostras analisadas, encontramos três padrões de expressão do genoma viral: amostras com maior expressão dos genes *E6/E7*; equivalentes expressões de *E6/E7* e de *E2/E4*; e amostras que apresentaram expressão de *E2/E4* e *E5* maiores do que de *E6/E7*. Para o HPV 16 detectamos 5 transcritos alternativos *E6/E7* (*E6\*I*, *E6\*II*, *E6\*X*, *E6\*V* e *E6\*VI*) mais o transcrito *E6+E7* que não sofre *splicing*. Já para o HPV 18 detectamos os transcritos *E6\*I* e o *E6+E7*. O nível de expressão em todas as amostras infectadas pelo HPV16 foi *E6\*I* > *E6\*II* > *E6+E7*, e em algumas amostras encontramos baixas expressões dos transcritos *E6\*X*, *E6\*V* e *E6\*VI*. Todas as amostras infectadas com HPV18 apresentaram o transcrito *E6\*I* mais expresso do que *E6+E7*.

Human Papillomavirus genomic integration and gene expression at cervical cancer

Ayslan Castro Brant

ABSTRACT

Cervical cancer is the third most incident type of cancer in women worldwide, and the presence of Human Papillomavirus (HPV) is necessary but not sufficient for its development. The immortalization and transformation of neoplastic cells are carried out by the viral oncogenes *E6* and *E7*, which will form proteins that can interfere with specific proteins responsible for cell cycle regulation. Those genes are expressed as a bicistronic *E6/E7* pre-mRNA, producing two or more different alternative *E6/E7* mRNAs due to alternative splicing. At initial phase of cervical infection, the HPV genome is maintained as an episomal DNA, but in tumors, the viral genome is integrated into the nuclear DNA of host cells. Once integrated, *E6* and *E7* viral oncogenes are overexpressed, what will improve selective advantages to tumoral development. Our goal is to analyze the HPV genome integration pattern at host genome, the HPV gene expression and the *E6/E7* splicing products in cervical tumors. To this purpose, we extracted total RNA from 25 cervical cancers samples (14 infected with HPV16, 10 with HPV 18 and 1 co-infected with HPV16 and HPV18). The RNA sequencing (RNA-Seq) at *Hi-Seq* 2500 (Illumina) was done with the RNA extracted from each sample. To determine the HPV integration, we made a *de novo* construction of cellular transcriptome and a search for chimeric contigs. We have found integration in 18 samples, and in most samples, the HPV genome break was observed at *E1* or *E2* genes. We found that 52,6% of integrations occurred at intergenic regions of cellular genome, and the remaining integrations occurred inside genes like *MCL1*, *PIGR*, *PXMP4*, *TP63*, *MMP13*, *SLC16A14*, *HRH1*, *RAD51B* and *NRXN1*. In the analyzed samples, we found three viral genome expression pattern: samples with higher *E6/E7* expression; samples with equivalent *E6/E7* and *E2/E4* expression; and samples with *E2/E4* and *E5* expression higher than *E6/E7*. To HPV 16 we found five different types of *E6/E7* alternative transcripts (*E6\**I**, *E6\**II**, *E6\**X**, *E6\**V** and *E6\**VI**), plus the *E6+E7* transcript. In respect to HPV 18 we detected the *E6\**I** and *E6+E7* transcripts. For all samples infected by HPV 16, the expression level were *E6\**I** > *E6\**II** > *E6+E7*, and the transcripts *E6\**X**, *E6\**V** and *E6\**VI** were found at low levels in a few samples. All samples infected with HPV18, the transcript *E6\**I** is expressed higher than *E6+E7*.

Este trabalho foi realizado na Divisão de Genética da Coordenação de Pesquisa do Instituto Nacional de Câncer José de Alencar Gomes da Silva, sob orientação do Dr. Miguel Angelo Martins Moreira, e contou com apoio financeiro da FAPERJ, CNPq, INCT-Câncer, Organização Panamericana de Saúde e INCA/Ministério da Saúde.

## **Agradecimentos**

Ao meu orientador Miguel por ter me dado a oportunidade de desenvolver este projeto, pela sua paciência, ensinamentos, compreensão e disponibilidade para sempre tirar dúvidas e debater qualquer questão.

Ao Albert por ter atuado como um co-orientador, sendo peça fundamental no meu aprendizado de bioinformática e no desenvolvimento deste trabalho.

A Shay pela amizade, conselhos e ajuda durante as etapas de bancada deste projeto.

Ao Sérgio e João pelas trocas de conhecimento, dicas e opiniões a respeito deste trabalho.

A Carolina Furtado pela ajuda com a montagem das bibliotecas e na corrida no Illumina.

A Tati Pastor e Mayra pela amizade, discussões, dicas e troca de conhecimentos.

Ao meu amigo Leonardo Paulista pela ajuda com a criação e edição das imagens deste trabalho.

Ao demais companheiro de laboratório, pesquisadores e funcionários do departamento de genética do INCA.

A Mônica, pela sua paciência, companheirismo, palavras de apoio e por ter me dado forças durante os momentos mais difíceis no desenvolvimento deste trabalho.

Aos meus pais pelo apoio, dedicação e por continuarem acreditando na minha capacidade, fazendo de tudo para que minha formação seja possível.

A pós-graduação do INCA.

Aos membros da banca avaliadora.

As agências de fomento: INCT para Controle do Câncer, OPAS (Organização Panamericana de Saúde), CNPq, FAPERJ e INCA/Ministério da Saúde.



## Lista de abreviaturas e siglas

µg	Micrograma
µl	Mililitro
3'ss	Sítio de <i>splicing</i> 3'
5'ss	Sítio de <i>splicing</i> 5'
ADN	Adenocarcinoma
BCL	Arquivos de densidade
BNT	Banco Nacional de Tumores
CCE	Carcinoma de Células Escamosas
cDNA	DNA complementar
CEP	Comitê de Ética em Pesquisa
CIN	Neoplasia intraepitelial cervical
cm	Centímetros
cm <sup>3</sup>	Centímetros cúbicos
DNA	Ácido Desoxirribonucleico
<i>E</i>	Genes de expressão precoce do HPV
ESE	Ativador exonico de <i>splicing</i>
ESS	Silenciador exonico de <i>splicing</i>
hnRNP	Ribonucleoproteínas heterogenias nucleares
HPV	Papilomavírus Humano
INCA	Instituto Nacional de Câncer
ISE	Ativador intronico de <i>splicing</i>
Kb	Quilo pares de bases
<i>L</i>	Genes de expressão tardia do HPV
LCR	Longa região de controle do HPV

LJI	Leituras de junção de integração
LJS	Leituras de junção de <i>splicing</i>
LOC	Arquivos de localização
M	Molaridade
ml	Mililitro
MLU	Média de leituras <i>unspliced</i>
mm <sup>3</sup>	Milímetros cúbicos
nM	Nanomolar
°C	Graus Celsius
ORF	Regiões abertas de leituras
P105	Promotor precoce HPV 18
P670	Promotor tardio HPV 16
P97	Promotor precoce HPV 16
pb	Pares de bases
PCR	Reação em cadeia da polimerase
pM	Picomolares
Poli-A	Região repetitiva de adeninas (poliadenilação)
Poli-T	Região repetitiva de timinas
Poly-AE	Poliadenilação precoce
Poly-AL	Poliadenilação tardia
Q	Valores de qualidade das bases
R1	Sequências filtradas da primeira leitura pelo <i>PRINSEQ</i>
R2	Sequências filtradas da segunda leitura pelo <i>PRINSEQ</i>
RJI	Região de junção de integração
RJS	Região de junção de <i>splicing</i>
RNA	Ácido ribonucleico

RNA <sub>m</sub>	RNA mensageiro
RNA <sub>r</sub>	RNA ribossômico
RNA-seq	Sequenciamento de RNA
SA	Sítio de <i>splicing</i> acceptor
SD	Sítio de <i>Splicing</i> doador
snRNP	Pequenas ribonucleoproteínas
SR	Proteína rica em serine-arginine
TCLE	Termo de Consentimento Livre e Esclarecido

## Lista de figuras

Figura 1.1: Estimativa dos tipos de câncer feminino mais comum na população brasileira realizada pelo Instituto Nacional do Câncer (INCA), para o ano de 2014....	2
Figura 1.2: Representação estrutural do genoma do HPV.....	4
Figura 1.3: Ciclo reprodutivo do HPV e progressão para o câncer cervical. ....	8
Figura 1.4: Progressão da infecção cervical provocada pelo papilomavírus humano (HPV) para câncer invasivo.....	10
Figura 1.5: Representação da integração do genoma do HPV 16 no genoma da célula hospedeira.. ....	12
Figura 1.6: Componente reguladores do splicing alternativo.. ....	15
Figura 1.7: Sequências e fatores reguladores do <i>splicing</i> .....	16
Figura 1.8: <i>Splicing</i> do íntron 1 do pré-RNA bicistrônico <i>E6/E7</i> do HPV 16 ou HPV 18 é essencial para a produção da oncoproteína E7.. ....	18
Figura 1.9: <i>Splicing</i> alternativo pré-RNA <i>E6/E7</i> no HPV18.. ....	19
Figura 1.10: <i>Splicing</i> alternativo pré-RNA <i>E6/E7</i> no HPV16.. ....	19
Figura 4.1: Representação de leituras que validam uma junção de integração (LJI).. ....	31
Figura 4.2: Possíveis transcritos alternativos <i>E6/E7</i> para o HPV 16.....	33
Figura 4.3: Possíveis transcritos alternativos <i>E6/E7</i> para o HPV 18.....	34
Figura 4.4: Representação de leituras <i>unspliced</i> . ....	35

## Lista de tabelas

Tabela 4.1: Lista das 25 amostras de cânceres cervicais utilizadas no estudo.....	24
Tabela 5.1: Resultados da corrida RNA-Seq 50 x 1 para as amostras HPV16_13 e HPV16_14.....	36
Tabela 5.2: Resultados da corrida RNA-Seq 100 x 1.....	37
Tabela 5.3: Resultados da corrida RNA-Seq 100 x 2.....	37
Tabela 5.4: Número total de leituras obtidas para cada amostra.....	39
Tabela 5.5: Análise da integração em amostras infectadas com HPV 16.....	41
Tabela 5.6: Análise da integração em amostras infectadas com HPV 18.....	42
Tabela 5.7: Análise da integração em amostras co-infectadas com HPV 16 e HPV 18.....	43
Tabela 5.8: Mapeamento das leituras contra a sequência de referência do HPV 16.....	46
Tabela 5.9: Mapeamento das leituras contra a sequência de referência do HPV 18.....	47
Tabela 5.10: Mapeamento das leituras da amostra co-infectada contra as sequências de referência do HPV 16 e do HPV 18.....	47
Tabela 5.11: Média de cobertura da expressão do genoma viral em tumores infectados por HPV 16.....	49
Tabela 5.12: Média de cobertura da expressão do genoma viral em tumores infectados por HPV 18.....	49
Tabela 5.13: Média de cobertura da expressão do genoma viral no tumor co-infectado.....	50
Tabela 5.14: Análise da proporção dos diferentes transcritos alternativos <i>E6/E7</i> em amostras de tumores cervicais infectadas com HPV16.....	52
Tabela 5.15: Análise da proporção dos diferentes transcritos alternativos <i>E6/E7</i> em amostras de tumores cervicais infectadas com HPV18.....	53

Tabela 5.16: Análise da proporção dos diferentes transcritos alternativos <i>E6/E7</i> na amostra de tumor cervicais co-infectadas pelo HPV 16 e HPV 18.....	54
--	----

## Sumário

<b>1- Introdução .....</b>	<b>1</b>
1.1- Epidemiologia do câncer de colo de útero .....	1
1.2- HPV e Câncer .....	2
1.3- Estrutura e organização do genoma do HPV .....	3
1.3.1- Função dos produtos gênicos.....	5
1.4- O ciclo reprodutivo do HPV .....	7
1.4.1- Integração do genoma viral .....	10
1.5- Expressão do genoma do HPV .....	13
1.5.1- <i>Splicing</i> alternativo .....	14
1.5.2- Transcritos alternativos <i>E6/E7</i> .....	17
<b>2- Justificativa .....</b>	<b>21</b>
<b>3- Objetivos .....</b>	<b>22</b>
3.1- Objetivo geral .....	22
3.2- Objetivos específicos .....	22
<b>4- Material e métodos .....</b>	<b>23</b>
4.1- População de estudo .....	23
4.2- As biópsias.....	25
4.3- Extração de DNA e RNA.....	25
4.4- Integridade e concentração do RNA .....	25
4.5- Sequenciamento do RNA mensageiro .....	25
4.5.1- Preparação da biblioteca de RNA .....	26
4.5.2- Hibridização e clusterização na <i>flow cell</i> .....	27
4.5.3- Sequenciamento por síntese .....	28
4.5.4- Processamento dos dados .....	28
4.5.4.1- Geração de arquivos <i>fastq</i> e <i>demultiplexing</i> das amostras.....	28

4.5.4.2- Eliminação das leituras de baixa qualidade .....	29
4.5.4.3- Análise da integração viral no genoma hospedeiro .....	30
4.5.4.4- Análise de expressão do HPV .....	31
<b>5- Resultados .....</b>	<b>36</b>
5.1- Corridas no sequenciador <i>HiSeq</i> 2500 (Illumina).....	36
5.2- Quantidade de leituras obtidas por amostra .....	37
5.3- Integração viral no genoma hospedeiro .....	40
5.4- Expressão do genoma do HPV .....	45
5.4.1- Mapeamento das leituras contra o genoma do HPV .....	45
5.4.2- Cobertura da expressão dos genes virais .....	48
5.4.3- Transcritos alternativos E6/E7 .....	50
<b>6- Discussão.....</b>	<b>55</b>
6.1- Integração do genoma viral.....	55
6.2- Padrão de expressão gênica viral e integração do genoma .....	61
6.3- Transcritos alternativos <i>E6/E7</i> .....	63
<b>7- Conclusões .....</b>	<b>67</b>
<b>8- Referências bibliográficas .....</b>	<b>69</b>
<b>Anexo 4.1 .....</b>	<b>76</b>
<b>Anexo 4.2.....</b>	<b>77</b>
<b>Anexo 4.3.....</b>	<b>81</b>
<b>Anexo 4.4.....</b>	<b>86</b>
<b>Anexo 4.5.....</b>	<b>89</b>
<b>Anexo 5.1 .....</b>	<b>90</b>
<b>Anexo 5.2.....</b>	<b>91</b>
<b>Anexo 5.3.....</b>	<b>92</b>
<b>Anexo 5.4.....</b>	<b>106</b>
<b>Anexo 5.5.....</b>	<b>116</b>



## **1- Introdução**

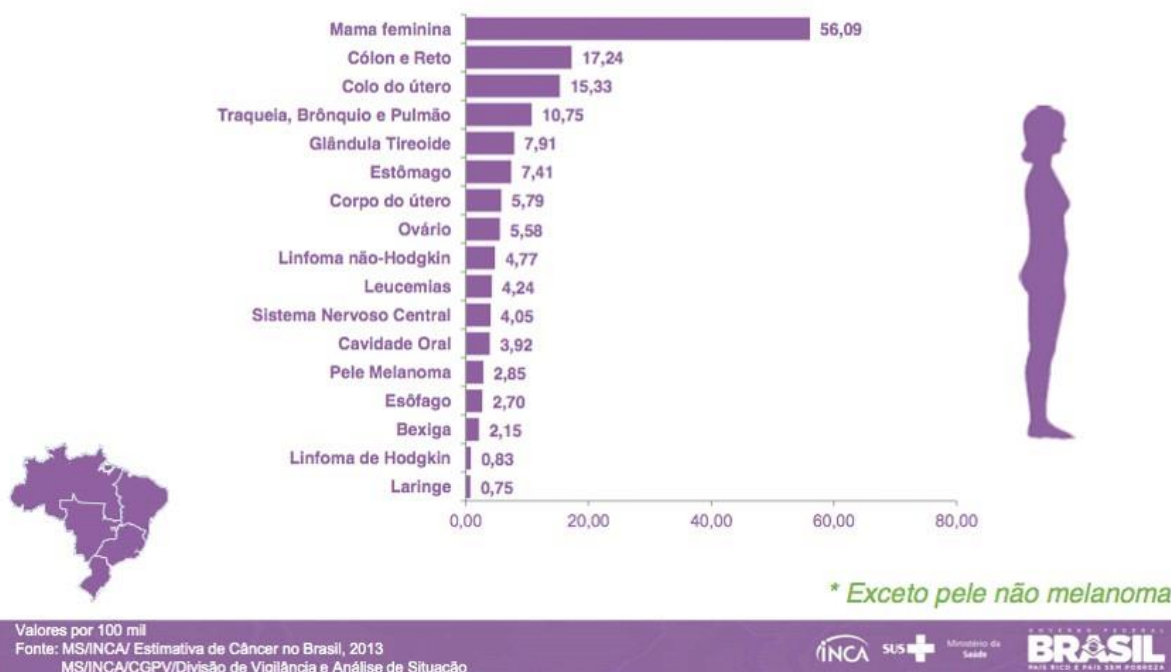
### **1.1- Epidemiologia do câncer de colo de útero**

O número mundial de casos de câncer de colo de útero aumentou de 378 mil casos no ano de 1980 para 454 mil casos em 2010, apresentando uma taxa de crescimento anual de 0,6%. A taxa de mortalidade desse câncer vem diminuindo ao longo dos anos, porém em 2010 ele foi responsável pela morte de cerca de 200 mil mulheres no mundo (FOROUZANFAR *et al.*, 2011). Para o ano de 2012 foi estimada a ocorrência de 527 mil novos casos, passando a ser o quarto tipo de câncer mais comum entre as mulheres (INCA, 2014). Mais de 85% dos casos de câncer cervical ocorrem em países subdesenvolvidos, sendo que a maioria das mortes decorrentes do câncer de colo uterino ocorrem nesses países (SHULZHENKO *et al.*, 2014).

No Brasil, o câncer de colo de útero é o terceiro tipo de câncer mais incidente entre as mulheres, sendo superadas apenas pelo câncer de mama e pelo câncer de colo e reto (figura 1.1). Para o ano de 2014, foram estimados 15.590 novos casos do câncer de colo de útero, com uma incidência de 15,33 novos casos a cada 100 mil mulheres (INCA, 2014).



**Taxas brutas de incidência das localizações primárias\* estimadas para 2014, em mulheres, Brasil.**



**Figura 1.1:** Estimativa dos tipos de câncer feminino mais comum na população brasileira realizada pelo Instituto Nacional do Câncer (INCA), para o ano de 2014. Esses dados não incluem o câncer de pele não melanoma. Obtidos de **INCA, 2014**.

## 1.2- HPV e Câncer

O Papilomavírus Humano (HPV) é o principal fator de risco para o desenvolvimento de lesões intraepiteliais de alto grau e do câncer do colo do útero, sendo considerado um fator necessário, mas não suficiente, para ativar o desenvolvimento tumoral. Além do HPV, fatores do hospedeiro como a imunidade, características genéticas e o comportamento sexual também são alguns dos fatores de risco associados ao desenvolvimento do câncer (BODILY; LAIMINS, 2011; NAKAGAWA *et al.*, 2000).

Com base em estudos epidemiológicos do Condiloma Acuminata, Zur Hausen propôs pela primeira vez uma associação entre HPV e o câncer de colo uterino. O Condiloma Acuminata é uma doença transmitida através de contato sexual no qual apresenta como característica verrugas genitais, que podem se tornar malignas,

onde está presente o HPV (ZUR HAUSEN, 1976). Posteriormente, em 1978 partículas virais de HPV foram encontradas em lesões na cérvix uterina por DELLA TORRE *et al.*, 1978 e LAVERTY *et al.*, 1978. Cerca de 99% das neoplasias intraepiteliais cervicais contém sequências virais (WALBOOMERS *et al.*, 1999).

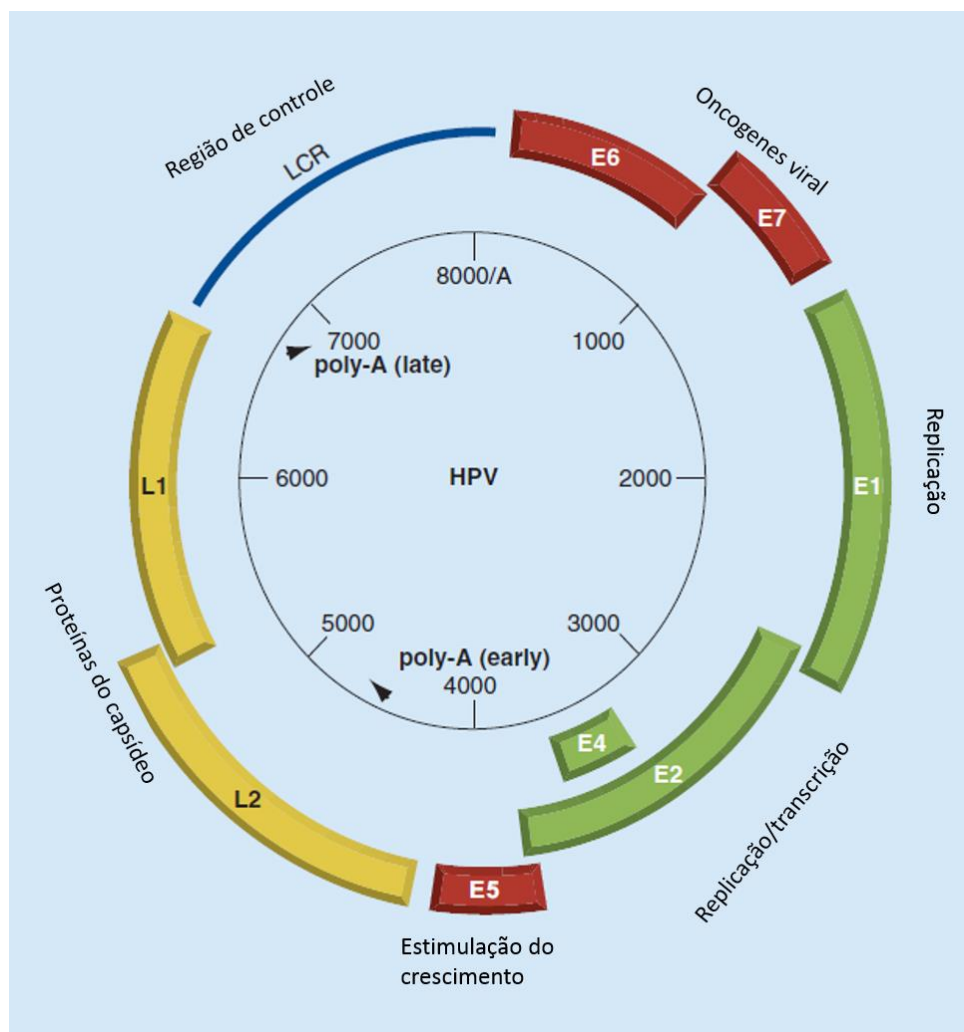
Já foram identificados mais de 100 tipos de HPV, dos quais aproximadamente 40 infectam o trato genital (DE VILLIERS *et al.*, 2004). Os HPVs genitais podem ser subdivididos em HPVs de baixo risco, que são encontrados principalmente em verrugas genitais, e HPVs de alto risco, que são frequentemente associados com câncer cervical invasivo (JACOBS *et al.*, 1995). Estudos epidemiológicos feitos em 2003 identificaram 15 tipos de HPV como sendo de alto risco para o desenvolvimento de câncer cervical (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 e 82) e 12 de baixo risco (6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81 e CP6108). Além desses, 3 tipos de HPV foram considerados provavelmente de alto risco (26, 53, 66) e outros 3 foram classificados como HPV de risco indeterminado (34, 57 e 83) (MUÑOZ *et al.*, 2003). Estudos de meta-análise sugerem que o HPV16 e HPV18 são os que apresentam maior prevalência em mulheres com lesões cervicais de alto grau e câncer cervical em diferentes regiões do mundo (DE SANJOSÉ *et al.*, 2007).

### **1.3- Estrutura e organização do genoma do HPV**

O papilomavírus humano é um vírus pertencente à família papillomaviridae, que apresenta tropismo por células escamosas epiteliais. Seu genoma é formado por uma molécula de DNA fita dupla circular com aproximadamente 8 Kb envolvida por proteínas do capsídeo viral que tem forma icosaédrica, é um vírus não-envelopado, com diâmetro entre 52 – 55 nm (MCLAUGHLIN-DRUBIN; MEYERS; MUNGER, 2012).

O genoma do HPV pode ser dividido principalmente em duas regiões: região codificante e região não codificante. Na região codificante encontram-se os genes de expressão precoce (E – *early*) e os genes de expressão tardia (L – *late*). A região não codificante é uma região de controle da transcrição e duplicação do genoma viral, sendo denominada longa região de controle (LCR – *long control region*) (NGUYEN; RAMÍREZ-FORT; RADY, 2014).

A região codificante dos genes *E* ocupa aproximadamente 50% do genoma do HPV e apresenta seis regiões abertas de leituras (ORF – *open read frame*) correspondentes aos genes *E1*, *E2*, *E4*, *E5*, *E6* e *E7*. Esses genes codificam proteínas que são necessárias para a replicação viral (*E1*), para o controle de transcrição do DNA viral (*E2*) e que interferem no ciclo celular da célula hospedeira (*E4*, *E5*, *E6* e *E7*). A região codificante dos genes *L* ocupam aproximadamente 40% do genoma viral e apresentam ORF para os genes *L1* e *L2*, que irão codificar proteínas estruturais do vírion (figura 1.2). A LCR representa 10% do genoma viral e não codifica nenhuma proteína funcional. Porém, nessa região encontram-se o sítio de origem de replicação e múltiplos sítios para ligação de fatores de transcrição, que são importantes para determinar o início da transcrição do genoma viral pela RNA polimerase II (BERNARD, 2002; MORSHED *et al.*, 2014; ZHENG; BAKER, 2006).



**Figura 1.2:** Representação estrutural do genoma do HPV. O genoma do HPV é formado por uma molécula de DNA circular fita dupla com aproximadamente 8 Kb. Esse genoma apresenta três regiões

principais: região de controle da transcrição (LCR), genes de expressão precoces (*E*) e genes de expressão tardia (*L*). Cada um dos genes do HPV apresenta funções importantes para o ciclo reprodutivo viral. O genoma do HPV ainda apresenta duas regiões de poli-adenilação, uma para os transcritos dos genes precoces - poly-A (early) e outra para os genes de expressão tardia - poly-A (late). Adaptada de **LIE; KRISTENSEN, 2008**.

O genoma do HPV apresenta dois promotores principais, um promotor precoce que é responsável pelo controle da expressão de quase todos os genes *E* e, um promotor tardio, que irá controlar a expressão dos genes *L*. Para o HPV 16 esses dois promotores são respectivamente o P97 e o P670. O promotor P97 está localizado *upstream* a ORF do gene *E6* (dentro da LCR) (SMOTKIN; WETTSTEIN, 1986), já o promotor P670 está localizado dentro da ORF do gene *E7* (GRASSMANN *et al.*, 1996). O promotor P105 do HPV 18 é equivalente ao promotor P97 no HPV 16 (ZHENG; BAKER, 2006). O término da transcrição dos transcritos precoces e tardios são determinados respectivamente pelos sítios de poli-adenilação precoces (poly-AE), localizado dentro da ORF de *L2*, e pelo sítio de poli-adenilação tardio (poly-AL), localizado dentro da LCR (SCHWARTZ, 2008; WANG *et al.*, 2011).

Devido ao fato do genoma do HPV apresentar apenas dois promotores e dois sítios de poliadenilação, todos os oito genes virais (*E1*, *E2*, *E4*, *E5*, *E6*, *E7*, *L1*, *L2*) são transcritos a partir da mesma fita molde de DNA, gerando transcritos policistrônicos ou bicistrônicos, referentes aos genes *E* e aos genes *L*. Esses transcritos são processados por diversos eventos de *splicing* alternativo gerando, assim, mais de oito produtos gênicos (JOHANSSON; SCHWARTZ, 2013).

### **1.3.1- Função dos produtos gênicos**

A proteína E1 é fundamental para que ocorra a replicação do DNA viral. Ela é a única enzima codificada pelo HPV que apresenta atividade de DNA helicase. Uma vez que essa proteína se liga a origem de replicação viral irá recrutar a maquinaria de replicação da célula promovendo a replicação do DNA viral (IARC, 2007). Atua, ainda, como um agente repressivo na transcrição e inibe a replicação do DNA, mantendo constante o número de cópias de DNA epissomal dentro das células (BEUTNER; TYRING, 1997; GNANAMONY; PEEDICAYIL; ABRAHAM, 2007).

A proteína E2 apresenta três funções principais no ciclo reprodutivo do HPV. A primeira função é regular a expressão de outros genes virais, atuando como um repressor ou ativador transcricional. Essa capacidade de regulação se deve a ligação da proteína E2 em sítios específicos localizado na LCR. A expressão dos genes *E6* e *E7* é modulada por essa função de E2 (IARC, 2007). Já a segunda função da proteína E2 é promover a replicação do DNA viral, uma vez que ela recruta a proteína E1 no sítio de origem de replicação (BEUTNER; TYRING, 1997), enquanto a terceira principal função da proteína E2 é transferir o genoma viral para as células filhas durante a divisão celular do hospedeiro (IARC, 2007).

A proteína E4 é a mais abundante das proteínas expressas pelo HPV, sendo que sua função no ciclo reprodutivo do HPV e na geração do câncer cervical ainda não está bem definido. Aparentemente a proteína E4 apresenta várias funções no ciclo reprodutivo viral, apresentando um importante papel na liberação e na transmissão das partículas virais (DOORBAR, 2013). Estudos em diferentes tipos de HPV indicam que E4 apresenta a capacidade de destruir o complexo de queratina celular facilitando, assim, a liberação e/ou a transmissão das partículas virais (BRYAN; BROWN, 2000). Também foi observado que essa proteína tem a capacidade de inibir a proliferação celular na fase G2 do ciclo celular (DAVY; DOORBAR, 2007).

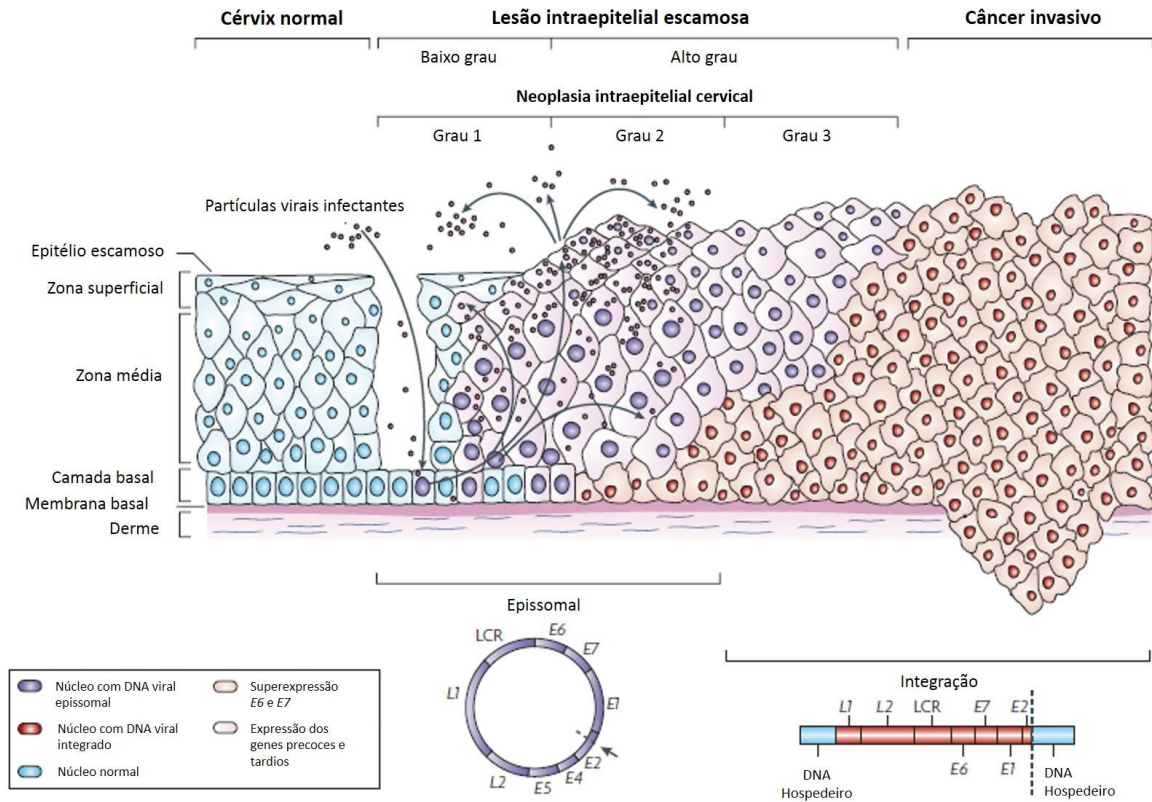
A proteína E5 está sendo considerada o terceiro oncogenes viral, porém a sua função também não está bem estabelecida. Há indícios de que essa proteína contribui quantitativamente na fase produtiva do ciclo reprodutivo do HPV (IARC, 2007). Além disso, evita que as células infectadas sejam reconhecidas pelo sistema imunológico (MORSLED *et al.*, 2014).

As oncoproteínas E6 e E7 apresentam um papel fundamental na transformação maligna das células infectadas pelo HPV, provocando uma diminuição na regulação do controle do ciclo celular e da maturação celular (GNANAMONY; PEEDICAYIL; ABRAHAM, 2007). A proteína E6 se liga a proteína p53 levando a sua degradação proteolítica, gerando como consequência uma replicação descontrolada das células infectadas (BEUTNER; TYRING, 1997). A proteína E7 se liga à proteína pRb e promove a sua degradação. Com isso, ocorre a perda do controle do ciclo celular, o que favorece a replicação do DNA viral, uma vez que a célula é mantida na fase S do ciclo (IARC, 2007).

As proteínas L1 e L2 são proteínas estruturais do capsídeo viral, onde a L1 é conhecida como proteína maior e a L2 como proteína menor do capsídeo (IARC, 2007). A ORF do gene *L1* é a região mais conservada do genoma do HPV, com isso, sua sequência nucleotídica é utilizada para a identificação de diferentes tipos e subtipos de HPV. Se a sequência nucleotídica de *L1* apresentar diferença maior do que 10% em comparação com outro HPV, conclui-se que esses vírus são de tipos diferentes. Entretanto, se a diferença for entre 2% e 10%, os HPVs são de diferentes subtipos; enquanto que, se a diferença de *L1* for menor do que 1% os vírus são de variantes diferentes (IARC, 2007).

#### **1.4- O ciclo reprodutivo do HPV**

Durante o ciclo reprodutivo normal do HPV (figura 1.3), na fase produtiva da infecção os vírus infectam as células proliferativas da camada basal do epitélio estratificado escamoso através de uma micro-lesão. Dentro da célula o material genético viral é liberado do capsídeo e transportado para o interior do núcleo, onde permanece na forma epissomal. Nas células basais, utilizando dos mecanismos de replicação celular, o promotor precoce é ativado permitindo a expressão dos genes *E*, porém, as expressões dos oncogenes *E6* e *E7* acabam sendo reprimidas pela proteína E2 (STEGER; CORBACH, 1997). Nessa etapa, ocorre um aumento do número de cópias do DNA epissomal para aproximadamente 50-100 genomas por células (SHULZHENKO *et al.*, 2014).



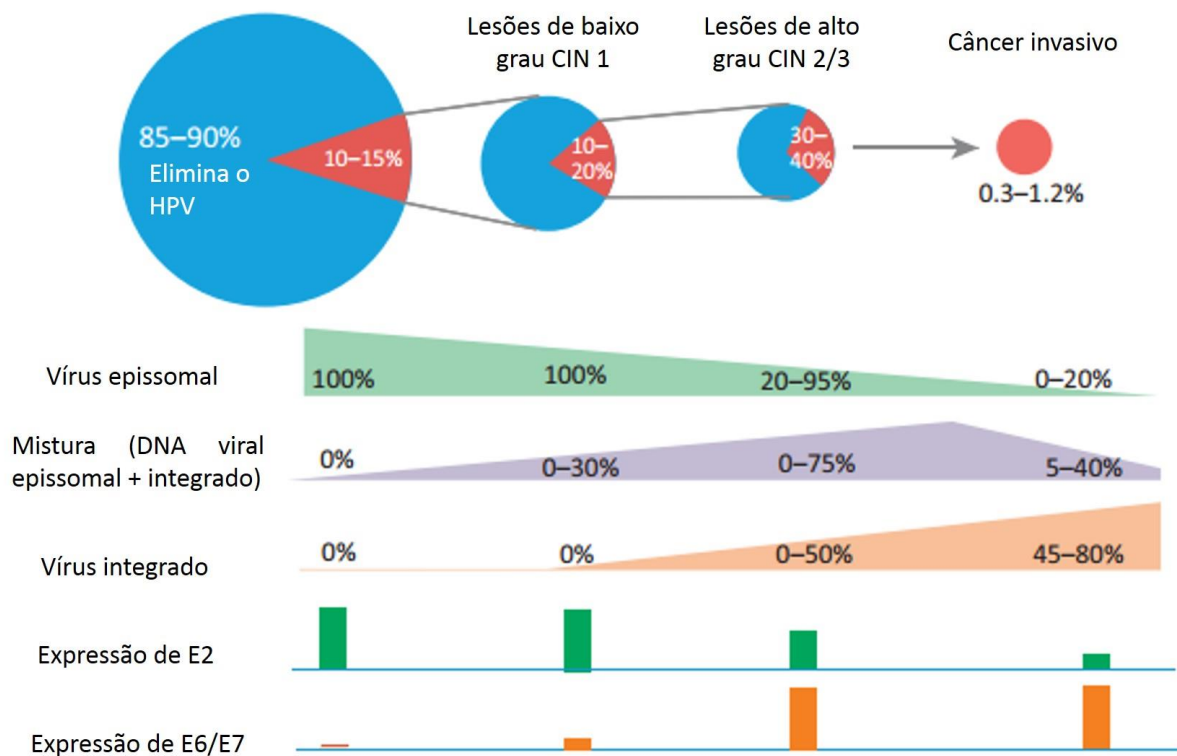
**Figura 1.3:** Ciclo reprodutivo do HPV e progressão para o câncer cervical. Através de uma micro-lesão no epitélio o papilomavírus humano (HPV) infectam as células proliferativas da camada basal do epitélio. Após a infecção, os genes precoces (*E*) são expressos e o DNA viral se replica a partir do DNA epissomal (núcleo roxo). A medida que as células infectadas migram para a zona média e para a zona superficial do epitélio, o genoma viral se replica de forma mais intensa e, os genes tardios *L1* e *L2* são expressos. Após a replicação do DNA epissomal e da síntese das proteínas do capsídeo viral *L1* e *L2*, novas partículas virais são montadas e liberadas na superfície do epitélio, podendo assim infectar novas células da camada basal. Neoplasias intraepiteliais de baixo grau (CIN 1) são observados durante a fase produtiva da infecção pelo HPV. A progressão das lesões intraepiteliais escamosas para neoplasia intraepitelial cervical de alto grau (CIN 2/3) e para câncer invasivo está associada com a integração do genoma do HPV no genoma da célula hospedeira (núcleo vermelho). Geralmente a integração provoca uma perda para proteína E2 devido a ruptura que ocorre no gene *E2* e, como consequência, é observado uma super-expressão dos oncogenes virais *E6* e *E7*. Adaptada de **WOODMAN; COLLINS; YOUNG, 2007**.

A medida que as células basais se diferenciam, a divisão celular cessa e as células migram para a camada suprabasal do epitélio. É observado que neste ponto, além do promotor tardio do HPV ser ativado, ocorre um aumento da expressão dos genes *E6* e *E7*. Como consequência da expressão desses oncogenes, as células infectadas voltam a sintetizar seu DNA e o processo de apoptose é inibido, dando início a formação de um processo displásico. Aproveitando a reativação da



multiplicação celular, o genoma viral se multiplica em centenas ou até mesmo milhares de cópias e, as proteínas L1 e L2 do capsídeo são sintetizadas. Com o DNA viral multiplicado e as proteínas do capsídeo sintetizadas, o vírion é formado e novas partículas virais são liberadas na camada superior do epitélio, ficando disponível para infectar novas células da camada basal (figura 1.3) (CHENG *et al.*, 1995; HUMMEL; HUDSON; LAIMINS, 1992).

Durante a fase produtiva da lesão, a infecção é observada através da presença de mudanças pré-malignas na histologia do epitélio cervical. Nessas infecções iniciais as alterações formam displasias epiteliais médias, que são classificadas como neoplasia intraepitelial cervical de baixo grau (CIN1 – low-grade cervical intraepithelial neoplasia) (MARTIN; O’LEARY, 2011; WOODMAN; COLLINS; YOUNG, 2007). Na maioria dos casos essa infecção é eliminada pelo sistema imunológico feminino. Aproximadamente de 10 a 15% dos casos, a infecção pelo HPV se torna persistente, isso se deve, provavelmente, ao mecanismo viral de inibição e escape do sistema imunológico, além da deficiência na resposta imunológica da mulher (BODILY; LAIMINS, 2011; NAKAGAWA *et al.*, 2000). A infecção persistente é conhecida como fase de transformação da infecção, no qual é observado um aumento da expressão dos genes *E6* e *E7*. O aumento das proteínas *E6* e *E7* irá bloquear a regulação negativa do ciclo celular através da degradação das proteínas p53 e pRb (SHULZHENKO *et al.*, 2014). Nessa etapa as lesões epiteliais aparecem como displasia moderada (CIN2) ou lesões epiteliais de alto grau (CIN3) (MARTIN; O’LEARY, 2011). A progressão das lesões CIN2 e CIN3 para o câncer invasivo está relacionada com a integração do genoma viral no genoma do hospedeiro (Figura 1.4).



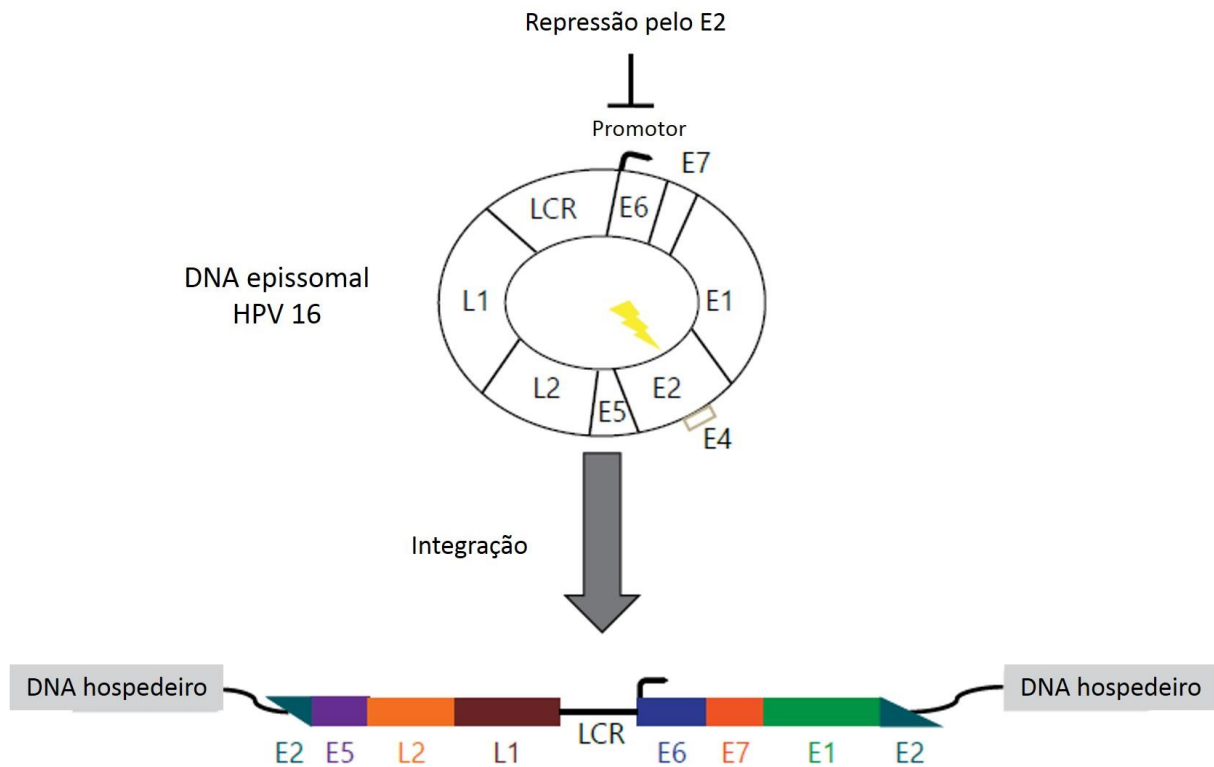
**Figura 1.4:** Progressão da infecção cervical provocada pelo papilomavírus humano (HPV) para câncer invasivo. Do total de mulheres que apresentam infecção pelo HPV, a grande maioria tem a infecção eliminada pelo próprio sistema imunológico, apenas uma pequena parcela progride para a formação de neoplasias intraepiteliais cervicais (CIN 1 e CIN 2/3) ou para o câncer invasivo. A progressão da infecção está relacionada com a integração do genoma viral no genoma do hospedeiro, podendo o genoma viral ser encontrado em diferentes estados físicos [apenas epissomal, epissomal + integrado (mistura) ou apenas integrado]. Nas diferentes etapas da lesão cervical, a integração ainda pode afetar de diferentes formas a expressão do gene *E2* e dos oncogenes *E6* e *E7*. Adaptada de **SHULZHENKO et al., 2014**.

#### 1.4.1- Integração do genoma viral

Alguns estudos vem demonstrando que a integração do DNA do HPV no genoma do hospedeiro é uma etapa importante para o desenvolvimento do câncer cervical (figura 1.3), porém o mecanismo no qual leva a integração é pouco conhecido (JEON; ALLEN-HOFFMANN; LAMBERT, 1995). A mudança do estado físico do genoma do HPV de epissomal para integrado está fortemente associado com a severidade das lesões neoplásicas. Lesões CIN1 apresentam o genoma do HPV exclusivamente na forma epissomal, enquanto que mais de 75% dos casos de CIN2 e CIN3 apresentam uma mistura do genoma do HPV, apresentando tanto a forma epissomal quanto a integrada (figura 1.4). No estágio de câncer invasivo,

raramente é observado a presença apenas do genoma episossomal, sendo a forma integrada a mais comum (presente de 45 a 80% dos casos), porém em diversos casos o genoma viral pode ser encontrado nos dois estados (CRICCA *et al.*, 2007; TONON *et al.*, 2001). A frequência com que a integração viral é encontrada em carcinomas provocados pelo HPV 16 e pelo HPV 18 é bastante alta, sendo aproximadamente 80% e 100% respectivamente (PETT; COLEMAN, 2007). Como pode ser observado, há diversos estudos para entender a relação da integração com a progressão tumoral, porém esses resultados ainda são bem diferenciados.

A integração não é uma etapa normal do ciclo reprodutivo do vírus HPV de alto risco, pois ela causa a quebra de genes virais que são essenciais para a síntese de uma partícula viral infectante (PETT; COLEMAN, 2007). Normalmente a quebra do genoma do HPV ocorre *downstream* aos genes *E6* e *E7*, sendo com maior frequência na região dos genes *E1* ou *E2* (WOODMAN; COLLINS; YOUNG, 2007) (figura 1.5). Essa quebra resulta na perda da função do gene repressor viral *E2*, o que leva a uma remodelação do promotor *E6/E7* (HAM *et al.*, 1991). Esse processo irá aumentar a estabilidade dos RNAm que codificam os oncogenes *E6* e *E7*, promovendo vantagens seletiva para a proliferação celular (JEON; ALLEN-HOFFMANN; LAMBERT, 1995; JEON; LAMBERT, 1995). O aumento da expressão de *E6* e *E7* irá levar perda da função de importantes proteínas celulares como a p53 e a pRb (MÜNGER *et al.*, 2004). Eventualmente o HPV integrado pode permanecer de forma silenciosa em um pequeno número de células por um longo período sem desencadear o crescimento seletivo dessas células (SHULZHENKO *et al.*, 2014).



**Figura 1.5:** Representação da integração do genoma do HPV 16 no genoma da célula hospedeira. Nas lesões intraepiteliais de baixo grau o genoma viral é normalmente encontrado na forma episossomal, no qual a proteína gerada pela expressão do gene *E2* irá reprimir a expressão dos oncogenes *E6* e *E7*. Principalmente durante a carcinogênese, o DNA viral é integrado no genoma do hospedeiro e, para que isso ocorra, é necessário que haja uma quebra no genoma viral, que normalmente ocorre no gene *E2* (seta amarela). Com a falta da proteína *E2*, após a integração e com a modulação do promotor dos genes *E* localizados na longa região de controle (LCR), os genes *E6* e *E7* são super-expressos, contribuindo para a progressão tumoral. Adaptada de **NGUYEN; RAMÍREZ-FORT; RADY, 2014**.

Tem sido especulado que importantes genes celulares são afetados pela integração do genoma viral, e essa interferência também contribui para o aumento do risco da progressão das lesões neoplásicas induzidas pelo HPV (DÜRST *et al.*, 1987). Mutagênese insercional provocada pela integração de sequências virais pode ativar oncogenes celulares ou inativar genes supressores tumorais das células hospedeiras (PETER *et al.*, 2010). A integração do genoma viral pode ocorrer em qualquer local do genoma, mas envolve frequentemente sítios frágeis, pontos de translocações conhecidos e regiões de transcrição ativas (KRAUS *et al.*, 2008). Estudos afirmam que a integração ocorre preferencialmente dentro ou perto de genes potencialmente tumorais, como o gene *MYC*; além disso, identificaram

múltiplos genes que são afetados pela integração do HPV 18, e dentre estes encontram-se genes supressores tumorais, proto-oncogenes e genes envolvidos na manutenção da integridade do DNA (FERBER *et al.*, 2003; VINOKUROVA *et al.*, 2005).

### 1.5- Expressão do genoma do HPV

Na literatura pouco se conhece a respeito da expressão gênica do HPV e conseqüentemente quais os tipos de transcritos irão levar a síntese de quais proteínas. A maioria dos estudos foram realizados com o HPV 16, tendo então seu mecanismo de expressão gênica melhor entendido. Apesar do HPV 18 ser o segundo tipo de HPV com maior importância para o desenvolvimento do câncer, muito pouco se sabe a respeito do seu mecanismo de expressão gênica. A organização do genoma desse tipo de HPV e o mapa dos seus transcritos começou a ser compreendida melhor por WANG *et al.*, 2011.

O genoma do HPV 16 contém um promotor precoce p97 e um promotor tardio p670. Para que ocorra a expressão de cada gene do HPV no momento exato do ciclo reprodutivo viral, é necessário que exista um mecanismo de controle do processamento do RNA viral. Sendo assim, na etapa inicial do ciclo reprodutivo viral, no qual apenas o promotor precoce é utilizado, é importante que apenas o sinal poli-adenilação precoce (poly-AE) e os sítios de *splicing* dos genes precoces sejam utilizados evitando, assim, a expressão dos genes tardios. Esse controle sugere que o sinal poli-adenilação tardio (poly-AL) e os sítios de *splicing* dos genes tardios sejam reprimidos durante a fase que está correndo a divisão celular. Por outro lado, quando os genes tardios precisam ser expressos, durante a fase de liberação das partículas virais, é necessário que o poly-AE e os sítios de *splicing* dos genes precoces sejam inibidos (SCHWARTZ, 2008).

Células com alta taxa proliferativa infectadas por HPV 16 possuem uma expressão constante dos oncogenes *E6* e *E7*, dessa forma, a expressão desses genes é considerada um pré-requisito para que ocorra a proliferação celular (DOORBAR, 2005). Nessas células o ciclo replicativo do HPV 16 está estagnado na fase inicial, apresentando uma expressão contínua dos genes precoces e repressão da expressão dos genes tardios *L1* e *L2*, dessa forma as células de câncer cervical nunca expressam as proteínas *L1* e *L2* (SCHWARTZ, 2008).

Estudos mostraram que enquanto o genoma do HPV 16 estiver na forma episossomal, independentemente se for em lesões de alto grau ou em câncer invasivo, irá ocorrer a expressão dos genes *E1* e *E2*, mesmo que seja em baixos níveis (HEBNER; LAIMINS, 2006). Também foi observado que o RNAm *E4* em HPV 16 é o mais abundante, isso se deve pois o sítio de *splicing* (SA3358) que é utilizado para gerar esse RNAm pode ser utilizado tanto nos pré-RNAm precoces quanto nos tardios (RUSH; ZHAO; SCHWARTZ, 2005).

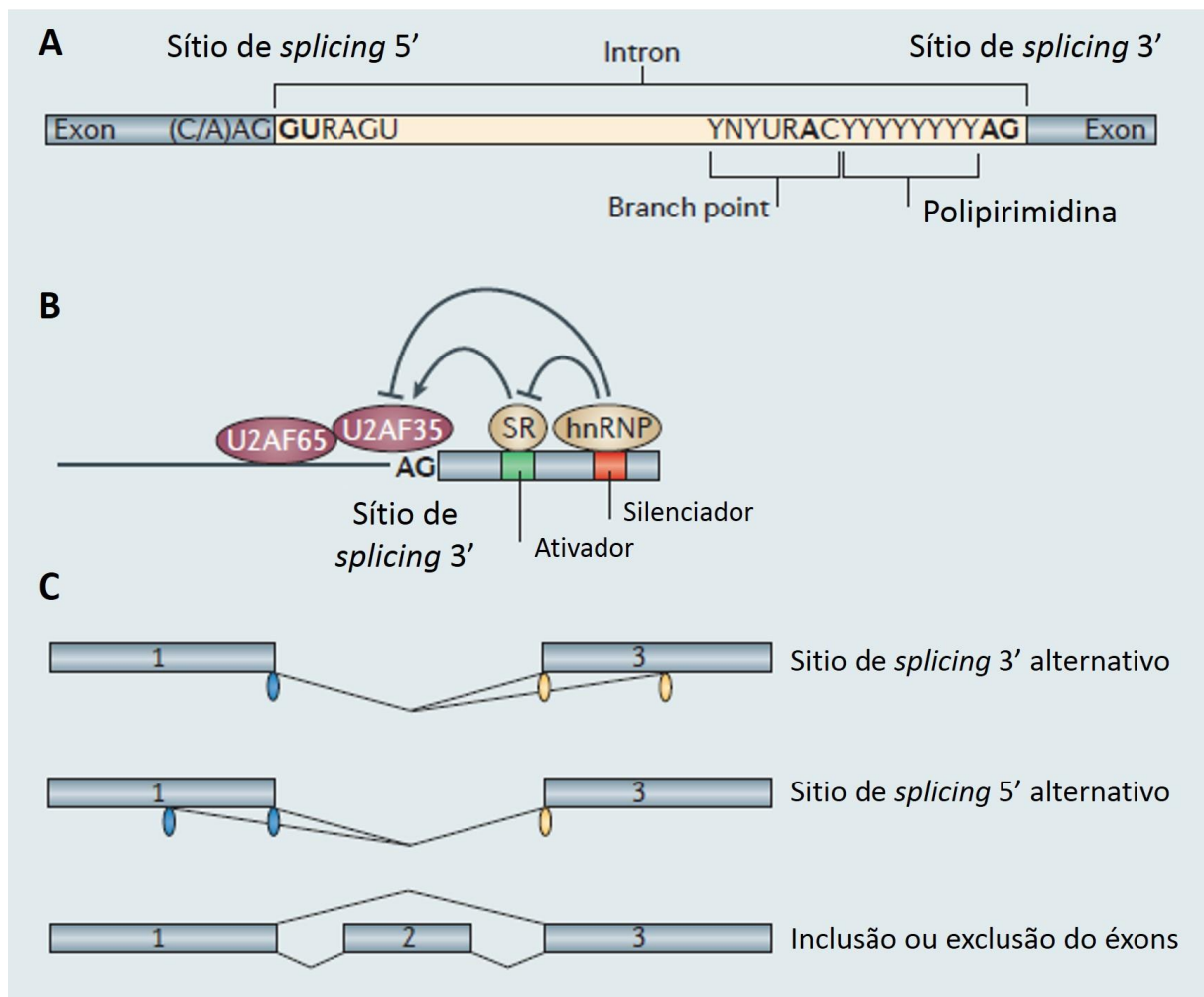
O processo de *splicing* do RNAm é um importante mecanismo de regulação da expressão gênica a nível pós-transcricional, sendo fundamental durante o ciclo reprodutivo do HPV. Pelo fato do genoma do HPV normalmente apresentar apenas dois promotores (um precoce e outro tardio) e dois sítios de poliadenilação (poly-AE e poly-AL), os transcritos formados geralmente são policistrônicos, apresentando múltiplos éxons e íntrons. Então, para codificar as diversas proteínas a partir desse RNAm policistrônico, o HPV necessita do mecanismo de *splicing* alternativo (ZHENG; BAKER, 2006).

### **1.5.1- *Splicing* alternativo**

O RNAm sofre *splicing* através de um complexo de proteínas chamadas de spliceossomo, no qual irão eliminar íntrons não codificantes do pré-RNAm e combinar os éxons formando, assim, o RNAm codificante. Para a formação do spliceossomo é necessário a combinação de cinco pequenas ribonucleoproteínas - snRNP (U1, U2, U4, U5 e U6) com diversos outros fatores de *splicing* (ZHENG; BAKER, 2006).

As zonas de transições do éxon-íntron são determinadas pela presença do sítio de *splicing* 5' (5'ss ou sítio doador – SD) e do sítio de *splicing* 3' (3'ss ou sítio acceptor – SA). Um íntron sempre começa com o dinucleotídeo GU e termina com AG, além disso, na extremidade do sítio de *splicing* 3' apresenta uma região chamada de *Branch point* que possui uma adenina (A) conservada, seguida de uma sequência de polipirimidina (figura 1.6 A). Durante o processamento do pré-RNAm, o 5'ss é complementar ao RNA splicessomal U1 (U1snRNA) da U1 snRNP presente no spliceossomo e, o 3'ss contendo o *Branch point* e a sequência polipirimidina, interage com os componentes E2snRNP e U2AF65-U2AF35 (figura 1.6 B). Esses

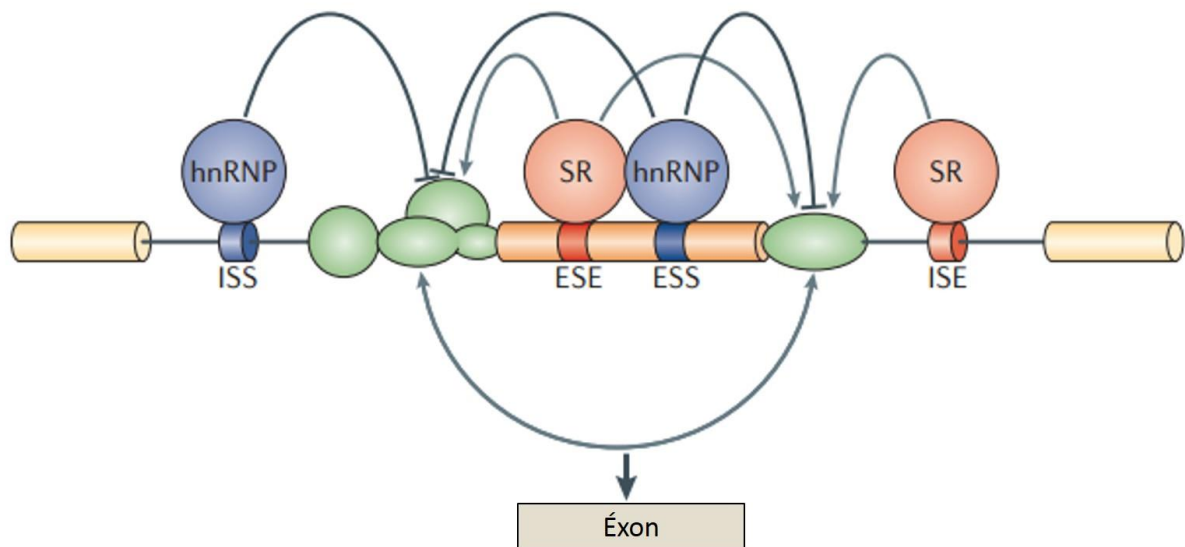
reconhecimentos dos sítios de *splicing* pelos componentes do spliceossomo são necessários, mas não suficiente para determinar uma transição éxon-íntron e assim ocorrer o *splicing* (JOHANSSON; SCHWARTZ, 2013).



**Figura 1.6:** Componente reguladores do splicing alternativo. Na zona de transição éxon-íntron apresenta o sítio de *splicing* 5' (5'ss) e o sítio de splicing 3' (3'ss). Um íntron sempre é iniciado com o dinucleotídeo GU e termina com AG, além disso, na região 3'ss do íntron há uma sequência chamada de *Branch point* (que apresenta uma adenina conservada) e uma sequência de polipirimidina (A). Durante o processo do *splicing*, a região 3'ss é reconhecida pelos componentes U2AF65-U2AF35 do spliceossomo. O reconhecimento e o recrutamento das proteínas do spliceossomos nos sítos de splicing são regulados por proteínas reguladoras (SR e hnRNP) que se ligam às regiões ativadoras ou silenciadoras do RNAm (B). O *splicing* alternativo pode ocorrer quando diferentes sítios de splicing 3' ou 5' são recrutados, ou ainda quando há a exclusão ou inclusão de um éxons, formando assim diferentes RNAm a partir do mesmo pré-RNAm (C). Adaptada de JOHANSSON; SCHWARTZ, 2013.

Em regiões adjacentes aos sítios de *splicing* existem sequências conservadas ativadoras e silenciadoras no qual se ligam determinados proteínas reguladoras,

tornando o reconhecimento desses sítios mais eficazes, estimulando ou inibindo o seu uso (CHEN; MANLEY, 2009). Essas proteínas que se ligam nos ativadores e silenciadores pertencem respectivamente as famílias das proteínas SR (do inglês serine-arginine-rich protein) e das ribonucleoproteínas heterogenias nucleares (hnRNP) (figura 1.6 B) (JOHANSSON; SCHWARTZ, 2013; LONG; CACERES, 2009). Essas sequências ativadoras e silenciadoras de *splicing* existem tanto em éxons quanto em íntrons, sendo chamadas de ativador exônico de *splicing* (ESE - *exonic splicing enhancer*); silenciador exônico de *splicing* (ESS - *exonic splicing silencer*); ativador intrônico de *splicing* (ISE - *intronic splicing enhancer*); silenciador intrônico de *splicing* (ISS - *intronic splicing silencer*) (figura 1.7) (KORNBLIHTT *et al.*, 2013; LADD; COOPER, 2002; ZHENG, 2004). Essas proteínas reguladoras que se ligam nas regiões reguladoras do *splicing*, ou recrutam as proteínas do spliceossomo para os sítios de *splicing*, ou interferem negativamente com esse recrutamento. Sendo assim, a eficiência do *splicing* é determinada pelo número e combinações dos fatores ativadores e silenciadores, além da sua combinação com os elementos formadores do spliceossomo (JOHANSSON; SCHWARTZ, 2013).



**Figura 1.7:** Sequências e fatores reguladores do *splicing*. O *splicing* é regulado por regiões de sequências reguladoras presentes no pré-RNA, no qual podem ser ativadoras ou silenciadoras, localizadas tanto nos íntrons (ativador intrônico de *splicing* - ISE; silenciador intrônico de *splicing* - ISS) quanto nos éxons (ativador exônico de *splicing* - ESE; silenciador exônico de *splicing* - ESS). Nessas sequências se ligam proteínas reguladoras pertencentes as famílias SR (serine-arginine-rich protein) e hnRNP (ribonucleoproteínas heterogenias nucleares). Essas proteínas quando ligadas nos



sítios reguladores podem recrutar ou impedir a ligação das proteínas que forma o spliceossomo (representados em verde) nos sítios de *splicing* 5' e 3'. Adaptada de **KORNBLIHTT et al., 2013**.

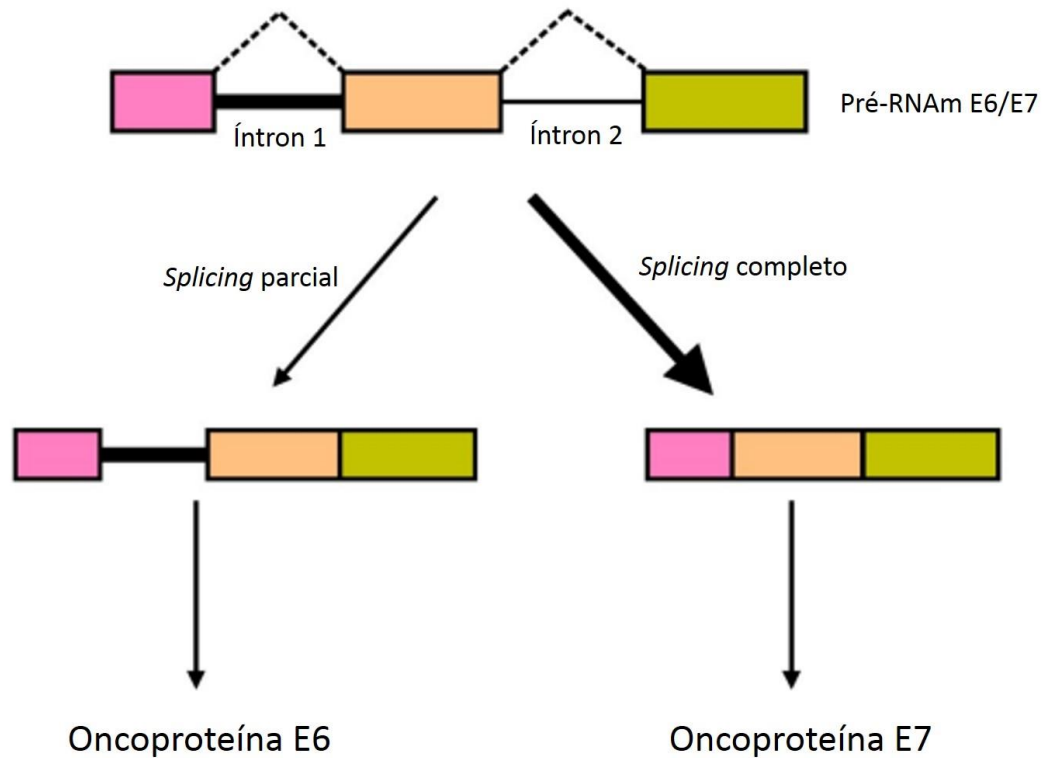
Nas células do humano, diversos RNAm sofrem o *splicing* alternativo no qual um éxon do pré-RNAm pode ser combinado de várias maneiras, criando diferentes RNAm que codificam diferentes proteínas (figura 1.6 C). Pelo fato do genoma viral ser pequeno e apresentar genes que possuem sobreposição de suas ORFs, os vírus também dependem do *splicing* alternativo para aumentarem o repertório de proteínas que eles podem codificar. Além disso, o *splicing* alternativo é utilizado para controlar temporalmente a expressão gênica viral durante seu ciclo reprodutivo. (JOHANSSON; SCHWARTZ, 2013; ZHENG, 2004).

### **1.5.2- Transcritos alternativos E6/E7**

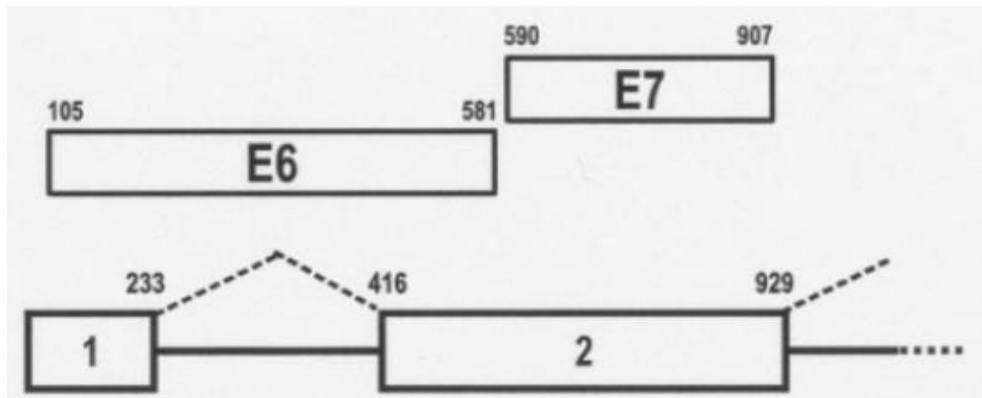
A transcrição dos genes *E6* e *E7* ocorre de forma diferenciada nos HPVs de alto risco e de baixo risco. Para o HPV de alto risco, os oncogenes *E6* e *E7* são transcritos utilizando o mesmo promotor e o mesmo sítio de poliadenilação, gerando um pré-RNAm bicistrônico *E6/E7*. Esse pré-RNAm apresenta três éxons e dois íntrons, sendo o íntron 1 localizado dentro da ORF do *E6*. No HPV de baixo risco, os genes *E6* e *E7* são expressos individualmente, necessitando de dois promotores distintos para formar o RNAm *E6* e o RNAm *E7*. Além disso, o RNAm *E6* do HPV de baixo risco não apresenta íntron. Sendo assim, *splicing* na ORF de *E6* ocorre predominante no pré-RNAm do HPV de alto risco, não sendo observados no HPV de baixo risco. O pré-RNAm bicistrônico *E6/E7* no HPV 16 e HPV18 quando sofre *splicing* completo, eliminando os dois íntrons, codifica a proteína *E7*, porém se sofrer *splicing* parcial, eliminando apenas o íntron 2, codifica a proteína *E6* (figura 1.8) (ZHENG, 2010).

O íntron 1 do pré-RNAm *E6/E7* no HPV 18 possui apenas um sítio de *splicing* 5' (SD233) e um sítio de *splicing* 3' (SA416) (figura 1.9), já no HPV 16 apresenta três 5'ss (SD191, SD221 e SD226) e três 3'ss (SA409, SA526 e SA742) (figura 1.10 A) (AJIRO et al., 2012; ZHENG; BAKER, 2006). Na literatura foram identificados cinco transcritos alternativos *E6/E7* originados do *splicing* alternativo do íntron 1 do pré-RNAm *E6/E7* no HPV 16: *E6\*1* (formado a partir da utilização dos SD226 e SA409,

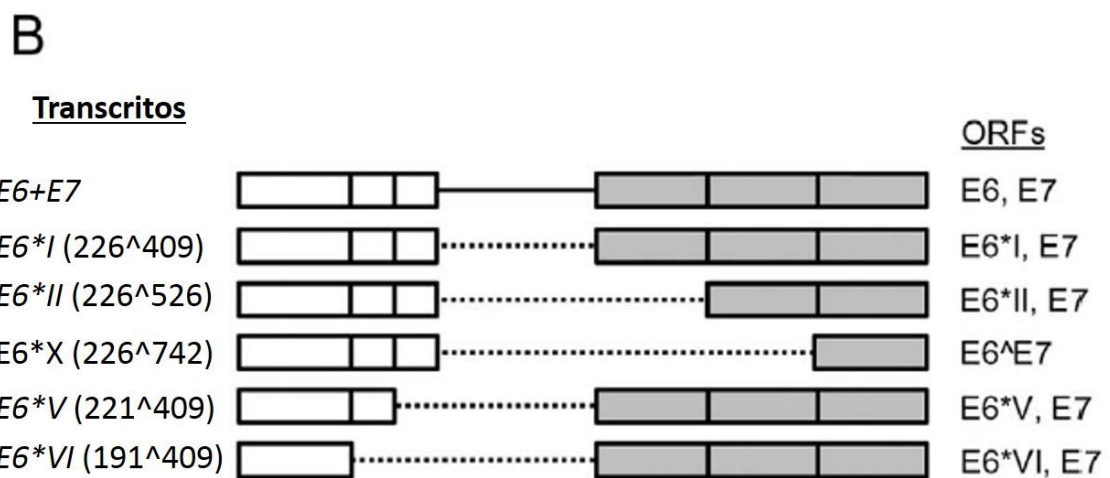
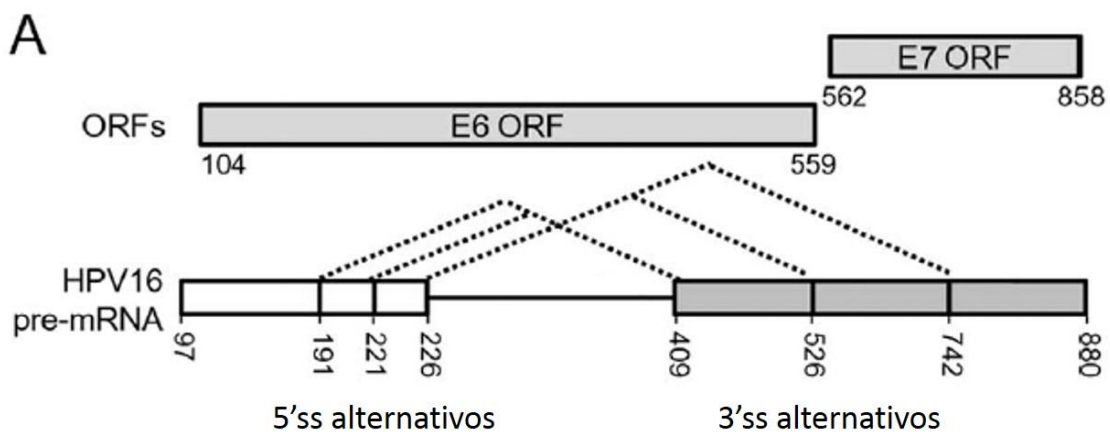
formando a junção de *splicing* 226^409), *E6\*II* (226^526), *E6\*X* (226^742), *E6\*V* (221^409) e *E6\*VI* (191^409) (figura 1.10 B) (AJIRO *et al.*, 2012; TANG *et al.*, 2006). O *splicing* do íntron 1 no HPV 18 forma apenas o transcrito alternativo *E6\*I* (233^416) (ZHENG *et al.*, 2004).



**Figura 1.8:** *Splicing* do íntron 1 do pré-RNA bicistrônico *E6/E7* do HPV 16 ou HPV 18 é essencial para a produção da oncoproteína *E7*. O pré-RNA *E6/E7* apresentam três éxons (caixas coloridas) e dois íntrons (linhas). A maioria dos RNAs *E6/E7* em câncer cervical e linhagens celulares sofrem *splicing* completo, eliminando os dois íntrons, permitindo assim a produção da oncoproteína *E7*. Uma menor quantidade desses pré-RNAs sofre *splicing* parcial, retendo o íntron 1 e mantendo a ORF de *E6* intacta, codificando assim a oncoproteína *E6*. Adaptada de ZHENG, 2010.



**Figura 1.9:** *Splicing* alternativo pré-RNA*m* *E6/E7* no HPV18. O íntron 1 do pré-RNA*m* *E6/E7* do HPV 18 apresentam apenas dois sítios de *splicing* alternativos, um sítio de *splicing* 5' (233) e o sítio de *splicing* 3' (416). Esse pré-RNA*m* sofre *splicing* alternativo formando a junção de *splicing* 233^416, gerando o transcrito alternativo *E6\**I**, que é importante para codificar a oncoproteína *E7*. Adaptada de **TANG et al., 2006**.



**Figura 1.10:** *Splicing* alternativo pré-RNA*m* *E6/E7* no HPV16. O íntron 1 do pré-RNA*m* *E6/E7* do HPV 16 apresenta três sítios de *splicing* 5' (5'ss) – 191, 221, 226 – e, três sítios de *splicing* 3' (3'ss) – 409,

526 e 742 **(A)**. A retenção desse íntron gera o transcrito *E6+E7*, que apresenta a ORF dos genes *E6* e *E7* intactas, fato que beneficiará a expressão do gene *E6*. O *splicing* alternativo do íntron 1 gera cinco transcritos alternativos *E6/E7*: *E6\*I*, *E6\*II*, *E6\*X*, *E6\*V* e *E6\*VI*, que apresentam respectivamente as junções de *splicing* 226^409, 226^526, 226^742, 221^409 e 191^409. A retirada do íntron 1 pelo processo de *splicing* causa a ruptura da ORF de *E6*, gerando ORF parciais *E6\**. Essa ruptura é fundamental para que ocorra a expressão do oncogene *E7* **(B)**. Adaptada de **AJIRO et al., 2012**.

A remoção do íntron 1 do pré-RNA bicistrônico *E6/E7* pelo mecanismo de *splicing* leva a ruptura da ORF do *E6*, impedindo que a proteína *E6* seja sintetizada por completo, sintetizando apenas um polipeptídeo truncado *E6\** (PIM; BANKS, 1999). A perda dessa ORF de *E6* é importante para que haja a expressão de *E7*, exceto para o transcrito *E6\*X* do HPV16, cujo a ORF de *E7* também é afetada pela remoção do íntron 1. O gene *E6* será expresso apenas pelo transcrito *E6+E7*, originado do pré-RNA bicistrônico *E6/E7* em que houve a retenção do íntron 1. O mecanismo no qual leva a retenção desse íntron no transcrito *E6+E7* ainda não é conhecido (TANG et al., 2006). Sabendo que o espaço físico entre o término da tradução de *E6* e o início da tradução de *E7* no transcrito *E6+E7* é de apenas 2 nucleotídeos - o que dificulta a polimerase traduzir *E7* logo após *E6* ser traduzido -, tem sido proposta uma teoria no qual diz que no HPV 16 o *splicing* do íntron 1 beneficia a expressão de *E7*. Sendo assim, quando o íntron 1 é retirado se cria um códon de terminação imediatamente após a junção de *splicing*, formando um espaçamento suficiente para que o ribossomo termine a síntese do polipeptídeo *E6\** e inicie a tradução do gene *E7* (KOZAK, 1987; SEDMAN et al., 1991).

## 2- Justificativa

Muito pouco se conhece na literatura a respeito da integração do genoma viral no genoma do hospedeiro em casos de câncer cervical. Alguns estudos demonstram uma tendência da integração ocorrer em importantes sítios gênicos como por exemplo proto-oncogenes e genes supressores tumorais (FERBER *et al.*, 2003; PETER *et al.*, 2010), porém ainda são poucos os indícios para afirmar que esses sítios descritos são, de fato, sítios preferencias para a ocorrência da integração do genoma do HPV. A análise feita a partir do RNAm permite a identificação de transcritos quiméricos, sintetizados a partir de sequências gênicas viral e humanas, indicando assim os sítios de integração no genoma hospedeiro e o gene viral rompido para a integração.

Não se compreende bem o efeito da integração na expressão dos diversos genes do HPV, sendo pouco entendido quais genes passam a ser mais ou menos expressos quando ocorre essa integração.

Existem dados sobre os transcritos dos oncogenes virais *E6* e *E7* provenientes do genoma viral integrado principalmente em linhagens celulares e em lesões precursoras associados aos HPV16 e HPV18 (Tang *et al.*, 2006). Pouco se conhece a respeito dos produtos alternativos *E6/E7* originados a partir do mecanismo de *splicing* alternativo, não sabendo ao certo quantas isoformas desses transcritos existem para cada tipo viral e se há alguma relação com o desenvolvimento tumoral.

Dessa forma, análise das sequências de RNAm através do sequenciamento de RNA (RNA-seq) de amostras de cânceres cervicais associados aos HPVs 16 e 18, pode permitir responder a essas questões.

### **3- Objetivos**

#### **3.1- Objetivo geral**

Analisar o padrão de integração do genoma viral no genoma do hospedeiro e a expressão dos genes do HPV em tumores cervicais.

#### **3.2- Objetivos específicos**

- Identificar os genes humanos e virais afetados pela integração do HPV.
- Quantificar a expressão dos genes do HPV nos tumores de diferentes tipos histológicos.
- Analisar a existência e a expressão das diferentes isoformas dos transcritos alternativos *E6/E7* em diferentes tumores.

## **4- Material e métodos**

### **4.1- População de estudo**

O grupo amostral estudado foi composto por 25 biópsias de tumores cervicais de pacientes previamente diagnosticadas com câncer de colo de útero, matriculadas no Hospital do Câncer II do Instituto Nacional de Câncer José de Alencar Gomes da Silva (INCA) entre Julho de 2011 e Março de 2014 (Registro CEP nº 156/10; anexo 4.1). Todas as pacientes incluídas no estudo assinaram o Termo de Consentimento Livre e Esclarecido (TCLE; anexo 4,2). Todas as biópsias foram obtidas antes do início do tratamento das pacientes.

A seleção dessas amostras foi realizada com base no tipo viral e na classificação histológica tumoral (Tabela 4.1). O tipo viral presente nas biópsias e a classificação histológica tumoral foram determinados previamente. Dessa forma, foram selecionadas 14 amostras identificadas com infecção por HPV 16, das quais 8 amostras foram classificadas histologicamente como Carcinoma de Células Escamosas (CCE) e 6 como Adenocarcinoma (ADN). Outras 10 amostras apresentavam infecção por HPV18, nas quais 4 eram do tipo histológico CCE e 6 do tipo ADN. Uma amostra foi identificada com co-infecção do HPV 16 e do HPV 18, sendo do tipo histológico CCE.

**Tabela 4.1:** Lista das 25 amostras de cânceres cervicais utilizadas no estudo. Em 14 foram determinados a presença de HPV 16, em 10 de HPV 18 e uma amostra foi observada a co-infecção por HPV 16 e HPV 18. Os tumores ainda foram classificados de acordo com seu tipo histológico, Adenocarcinoma (ADN) ou Carcinoma de Células Escamosas (CCE).

<b>Amostras de tumores cervicais</b>		
	<b>Amostra</b>	<b>Histologia</b>
HPV 16	HPV16_1	ADN
	HPV16_2	ADN
	HPV16_3	CCE
	HPV16_4	CCE
	HPV16_5	CCE
	HPV16_6	CCE
	HPV16_7	CCE
	HPV16_8	CCE
	HPV16_9	CCE
	HPV16_10	ADN
	HPV16_11	ADN
	HPV16_12	ADN
	HPV16_13	ADN
	HPV16_14	CCE
HPV 18	HPV18_1	CCE
	HPV18_2	ADN
	HPV18_3	ADN
	HPV18_4	ADN
	HPV18_5	CCE
	HPV18_6	CCE
	HPV18_7	ADN
	HPV18_8	ADN
	HPV18_9	CCE
	HPV18_10	ADN
HPV 16 HPV 18	HPV16/18_1	CCE



## **4.2- As biópsias**

As biópsias, com dimensões aproximadas de 1 cm<sup>3</sup>, foram armazenadas em criotubos contendo 1 ml de RNA Later (*Life-Technologies*) por 24 horas. Após isso, o RNA Later foi retirado e as biópsias divididas em diversos fragmentos menores, sendo armazenadas separadamente em freezer a -80°C no Banco Nacional de Tumores (BNT) do INCA. No BNT, essas amostras foram cadastradas e identificadas com códigos de barras.

## **4.3- Extração de DNA e RNA**

Um fragmento de cada amostra tumoral de aproximadamente 2 mm<sup>3</sup> foi retirado do freezer -80°C, descongelado e fragmentado com a ajuda de um bisturi descartável. Posteriormente o DNA e o RNA das amostras foram extraídos com o *Kit Qiagen Allprep DNA/RNA mini kit* (QIAGEN). O DNA de cada amostra foi armazenado no freezer -20°C, já o RNA foi armazenado no freezer -80°C.

## **4.4- Integridade e concentração do RNA**

Após a extração, com o objetivo de observar a integridade do RNA, 3 µl do RNA total extraído foi submetido à eletroforese em gel de agarose 1% (Invitrogen) diluída em 100 ml de tampão NaOH 1X (NaOH 10mM; H<sub>3</sub>BO<sub>3</sub> 42,5mM) e corado com 1 µl do corante de corrida “*blue green*” (LGC Biotecnologia). A eletroforese foi realizada em cubas horizontais contendo tampão de corrida NaOH 1X e submetidas a uma voltagem de 5 volts/cm durante 30 minutos. O RNA foi considerado íntegro quando possível observar as bandas equivalentes ao RNA ribossômico (RNAr) 28S e 18S. Para isso, o gel foi visualizado em um transiluminador de luz ultravioleta. A concentração e o grau de pureza do RNA foram analisados com o uso do espectrofotômetro NanoDrop (Thermo Scientific).

## **4.5- Sequenciamento do RNA mensageiro**

Utilizando o RNA total das amostras de câncer cervical foi feito o sequenciamento de RNA (RNA-Seq) através da plataforma *HiSeq™ 2500* (Illumina).

Para isso, foram realizadas as seguintes etapas: Preparação das bibliotecas de RNA; Hibridização e clusterização na *flow cell*; Sequenciamento por síntese; Processamento dos dados.

#### 4.5.1- Preparação da biblioteca de RNA

O RNA total extraído e quantificado foi utilizado para a construção das bibliotecas de RNA com o Kit *TruSeq RNA Sample Prep* da Illumina, seguindo as recomendações do fabricante.

A primeira etapa da construção das bibliotecas foi a diluição das amostras em água ultra pura (Invitrogen), de tal forma que cada uma tivesse uma quantidade entre 0,1 a 4 µg de RNA total em um volume final de 50 µl. A quantidade de RNA utilizado para cada amostra variou de acordo com a concentração do RNA extraído. Após a diluição inicial, o RNAm contendo cauda de poliadenilação (poli-A) foi capturado por oligonucleotídeos poli-T ligados às partículas magnéticas contidas no kit.

Uma vez separado, o RNAm passou por uma etapa de eluição a 94°C por 3 minutos. Esta incubação, devido a presença de cátions divalentes, permitiu a fragmentação aleatória do RNAm em pequenos fragmentos, entre 200 pb e 300 pb. Posteriormente, foi feita a construção da primeira fita do cDNA através da enzima transcriptase reversa na presença de primers randômicos. A construção da segunda fita do cDNA foi feita utilizando as enzimas DNA polimerase I e RNase H.

Com o cDNA sintetizado, uma única molécula de adenina (A) foi adicionada nas extremidades 3'. Essa etapa é importante para impedir que um cDNA se ligue a outro e, é importante durante o processo de ligação dos adaptadores que apresentam uma timina (T) na extremidade (para a ligação com os fragmentos de cDNA) e uma sequência identificadora (índice). Dessa forma, através da sequência nucleotídica presente no índice foi possível identificar as diferentes amostras corridas na mesma linha da *flow cell* (corrida multiplex). A ligação dos adaptadores à molécula de cDNA ocorreu através da enzima DNA ligase.

Uma etapa de enriquecimento por meio da reação em cadeia da polimerase (PCR) foi realizada com o objetivo aumentar a quantidade de fragmentos de cDNA com adaptadores nas duas extremidades. Esta etapa é fundamental para a

hibridização dos fragmentos na *flow cell*, uma vez que os adaptadores contêm sequências complementares aos oligonucleotídeos presentes fixados a esta *flow cell*. Essa reação foi realizada utilizando um conjunto de primers específicos que se anelam ao final dos adaptadores.

A última etapa da preparação das bibliotecas foi a quantificação por PCR em tempo real utilizando o kit *Library Quant Kit Illumina (Kappa Biosystems)*. Dessa forma, por meio de uma quantificação relativa, foi estimada a concentração das bibliotecas através da construção de uma curva padrão com amostras de concentrações conhecidas.

#### **4.5.2- Hibridização e clusterização na *flow cell***

Foram realizadas três corridas de RNA-seq, duas de leitura única (*single read*) e uma de leitura pareada (*pair end*):

- (1) Corrida de leituras de 52 pares de base (52 pb), em três linhas. Realizada apenas com as amostras HPV16\_13 e HPV16\_14. As duas bibliotecas foram misturadas de maneira equimolar (2 nM), desnaturadas com 0,1 M de NaOH, diluídas a 8 pM, e clusterizadas nas três linhas da *flowcell*.
- (2) Corrida de leituras de 100 pares de base (100 pb), com duas linhas. Uma linha contendo as 12 bibliotecas das amostras infectadas com HPV 16 (HPV16\_1 a HPV16\_12), e a segunda linha contendo as 10 bibliotecas das amostras com HPV 18 (HPV18\_1 a HPV18\_10) e a amostra co-infectada com o HPV 16 e o HPV 18 (HPV16/18\_1), num total de 11 bibliotecas. Para cada uma das linhas, as bibliotecas foram misturadas de maneira equimolar (2 nM), desnaturadas com 0,1 M de NaOH, diluídas a 12 pM, e clusterizadas em cada linha da *flowcell*.
- (3) Similar a segunda corrida, mas com leituras *pair end* de 101 pb x 101 pb.

A preparação da *flow cell* ocorreu no equipamento cBot (Illumina) utilizando o *TruSeq® SR Cluster Kit v3* (Illumina) para as amostras da primeira corrida, e o *TruSeq® PE Cluster Kit v3* (Illumina) para as amostras da segunda e da terceira corrida. Durante esse procedimento, ocorreu a hibridização das bibliotecas na superfície da *flow cell*. Uma vez que os fragmentos já estavam hibridizados, ocorreu

uma série de reações de PCR *in loco* (amplificação por ponte) formando, assim, os agrupamentos de DNA molde (*clusters*). Após a reação de amplificação por ponte, os fragmentos amplificados foram linearizados e ocorreu a hibridização com *primers* de sequenciamento, em regiões específicas dos adaptadores.

#### **4.5.3- Sequenciamento por síntese**

As bibliotecas clusterizadas na *flow cell* foram sequenciadas utilizando o kit *TruSeq SBS Kit v3* (50 ou 200 ciclos), de acordo com as recomendações do fabricante. Para a primeira e segunda corrida, além da leitura dos 52 pb e 100 pb respectivamente, também foram lidas 6 pares de bases referente ao índice. Na terceira corrida (*pair end*), a primeira leitura de 101 pares de bases foi seguida de mais 6 pares de bases referentes ao índice, e 101 pares de bases da segunda leitura.

#### **4.5.4- Processamento dos dados**

As sequências nucleotídicas do RNA<sub>m</sub> (leituras), obtidas nas corridas *single read* e *pair end* no Illumina, foram processadas e analisadas com ferramentas de bioinformática. O processamento dos dados envolveu: geração dos arquivos *fastq* e *demultiplexing* das amostras (com a identificação das amostras através dos índices), eliminação das leituras de baixa qualidade, análise da integração do genoma viral no hospedeiro e, análise da expressão do genoma do HPV.

##### **4.5.4.1- Geração de arquivos *fastq* e *demultiplexing* das amostras**

Os arquivos de intensidade (*basecalls* - BCL) e localização (LOC) gerados no sequenciador Illumina foram utilizados para a criação de arquivos *fastq* através do programa *Casava 1.8* (Illumina). Arquivos *fastq* são arquivos texto que apresentam sequências nucleotídicas seguidas por valores de qualidade para cada uma das bases. Durante esta etapa, os resultados das diferentes amostras da corrida multiplex, foram separadas uma das outras (*demultiplexing*) através da identificação das sequências dos índices (*barcodes*).

A geração dos arquivos *fastq* e *demultiplexing* foram realizados através de um *script configureBclToFastq.pl* juntamente com um arquivo *samplesheet.csv* referente a cada corrida. Uma *Sample Sheet* é um arquivo tabular, que para cada linha da *flow cell* correlaciona as amostras com as sequências dos índices adicionadas a elas.

Os critérios opcionais adicionados ao *script* foram a adição de uma base de *mismatch* com o objetivo de evitar perda de leituras devido a erros ocorridos na região do índice. Sendo assim, no *script* foi permitido que as leituras tivessem *mismatch* de 1 base nessa região. Além disso, também foi utilizado o critério opcional – *fastq-cluster-count 0*, no qual determina que todos os clusters formados para cada amostra gerem apenas um arquivo *fastq*.

#### 4.5.4.2- Eliminação das leituras de baixa qualidade

Após a geração das leituras no formato *fastq*, estas foram filtradas de acordo com o valor de qualidade pelo programa *PRINSEQ* (SCHMIEDER; EDWARDS, 2011). O filtro de qualidade para as corridas realizadas foi feito com o *script prinseq-lite.pl*, seguindo as recomendações presentes no manual do programa. Sendo assim, foram selecionadas as leituras que apresentavam um tamanho mínimo de 30 bases e valores de qualidade (Q) maiores do que 20. Esse valor de qualidade expressa a probabilidade de erro para cada base, ou seja, para Q = 20, cada base apresenta uma probabilidade de erro igual a 0,01 (ILLUMINA, 2011). Outros critérios utilizados para o filtro foram a eliminação de sequências com mais de 60% de bases ambíguas por posição e, a eliminação das leituras que apresentam caudas poli-A/T/N 3' e 5' menores do que 5 bases. O controle de qualidade das leituras geradas é uma etapa fundamental para garantir que os dados utilizados não sejam de baixa qualidade, nem sequências artificiais ou ainda sequências contaminadas.

Como *output*, para cada amostra da corrida *single read* é gerado um arquivo *fastq*, no qual estão presentes todas as leituras que passaram pelo filtro de qualidade do *PRINSEQ*. Para amostras da corrida *pair end*, são formados quatro arquivos *fastq*: o primeiro para as sequências filtradas da primeira leitura (R1); o segundo para as sequências filtradas da segunda leitura (R2), que são pareadas com a R1; o terceiro e o quarto são os arquivos R1 e R2 que não são pareados (*Singletons*). Como as amostras que analisamos no modo *single read* da primeira corrida foram corridas em três linhas, seus arquivos gerados pelo *PRINSEQ* foram agrupados através do comando UNIX *cat*.

#### 4.5.4.3- Análise da integração viral no genoma hospedeiro

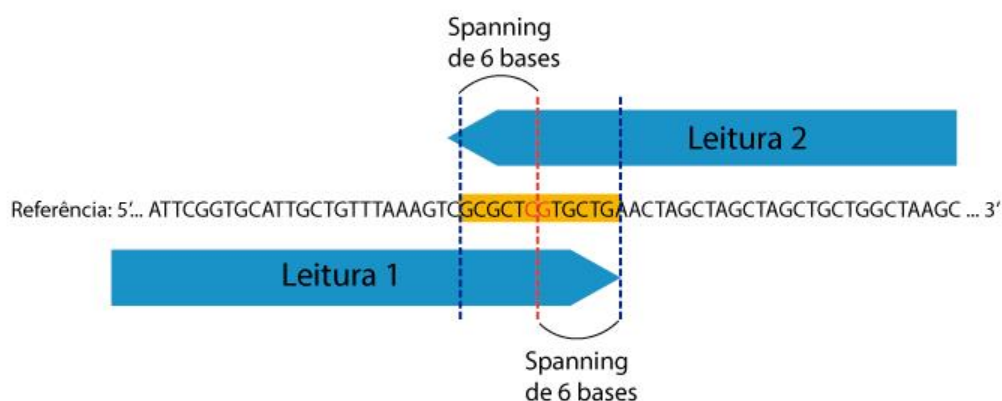
Para a busca por integração do genoma viral no genoma do hospedeiro foram realizadas montagens “*de novo*” do transcriptoma das amostras tumorais através do programa *Trinity* (GRABHERR *et al.*, 2011). Os opcionais utilizados no *script* foram selecionados de acordo com o manual do programa. Para as amostras HPV16\_13 e HPV16\_14 foram utilizadas as leituras obtidas pelo filtro de qualidade do *PRINSEQ*, já para as demais amostras foram utilizadas apenas os arquivos R1 e R2 geradas pelo *PRINSEQ* da terceira corrida – não sendo utilizados os arquivos *singletons* nem os arquivos da segunda corrida. O tamanho mínimo estabelecido para a construção dos *contigs* foi de 300 pb.

A integração foi observada a partir da busca de *contigs* quiméricos, no qual apresentam sequências do genoma do HPV juntamente com sequências do genoma humano. Para a seleção desses *contigs* quiméricos, perante o conjunto de todos os *contigs* construídos para cada amostra, foi utilizado a ferramenta *Blast*. Primeiramente, utilizando o *Blast* na plataforma *stingray@galaxy* (WAGNER *et al.*, 2014) buscamos os *contigs* que apresentavam sequências do HPV infectante. Desses, selecionamos aqueles que apresentavam *E-value* menor que  $1 \times 10^{-6}$  e identidade maior ou igual a 98%. Esses *contigs* selecionados foram separados dos demais por meio de um *script extractsequence.pl* (obtido no fórum <http://seqanswers.com/forums/showthread.php?t=9498> - Anexo 4.3). Uma vez que os *contigs* que apresentam sequências do HPV foram separados foi feita uma comparação contra o genoma humano no *Blast online* (NCBI), visando a identificação dos *contigs* quiméricos. Através da comparação da sequência nucleotídicas dos *contigs*, com a sequência do genoma do HPV infectante e do genoma humano, foi possível observar em qual ponto do gene viral houve a quebra e em qual gene humano ocorreu a integração. As amostras, no qual não foram encontradas *contigs* quiméricos, não tiveram a integração viral confirmada.

Após a identificação dos *contigs* quiméricos, suas sequências nucleotídicas foram utilizadas como referências para a construção de mapas de cobertura das leituras. Esses mapeamentos foram realizados através do programa *BowTie 2* (LANGMEAD; SALZBERG, 2012), utilizando parâmetros padrões segundo o manual do programa. Esta é uma etapa fundamental para estipular o valor de cobertura das regiões de junção de integração HPV/Humano. Essa região de junção de integração

(RJI) é composta por 12 nucleotídeos, sendo 6 bases *upstream* e outras 6 bases *downstream* do sítio de integração (figura 4.1).

Com os mapas construídos para cada um dos *contigs*, obtivemos o valor de cobertura das RJI através da contabilização do número de leituras de junção de integração (LJI). Determinada leitura é considerada uma LJI quando ela apresenta pelo menos 6 bases atravessando (*spanning*) o sítio de integração (figura 4.1). A contabilização das LJI foi feita através da ferramenta *view* do programa *SAMtools* (LI *et al.*, 2009). Com essa ferramenta foi possível selecionarmos todas as leituras que tivessem pelo menos uma base cobrindo a RJI. Através de um programa de editor de texto contabilizamos o número de leituras que apresentavam a sequência nucleotídica das 12 bases da RJI, obtendo assim o número de leituras com *spanning* mínimo de 6 bases que cobrem essa junção.



**Figura 4.1:** Representação de leituras que validam uma junção de integração (LJI). O sítio onde ocorreu a integração na sequência de referência é indicado pela linha tracejada em vermelho, entre os nucleotídeos da junção de integração (nucleotídeos em vermelho). A integração é indicada quando as leituras, representadas pelas setas azuis, apresentam pelo menos 6 bases atravessando o sítio de integração, que se encontra na região de junção de integração - (RJI; indicada pela área em amarelo) limitada pelas linhas tracejadas azuis. A sequência nucleotídica demonstrada na ilustração é apenas representativa, e não corresponde à uma sequência real.

#### 4.5.4.4- Análise de expressão do HPV

A análise da expressão dos transcritos dos genes do HPV contido nas amostras tumorais foi realizada através do mapeamento das leituras contra a sequência de referência do genoma do HPV infectante. Essas sequências foram

obtidas no banco de dados do *GenBank* (<http://www.ncbi.nlm.nih.gov/genbank>) sob o número de acesso K02718.1 para o HPV 16 e AY262282.1 para o HPV 18.

O mapeamento foi realizado através do programa *BowTie 2*. Como *input*, além da sequência de referência, foi utilizado todos os arquivos *fastq* obtidos pelo *PRINSEQ* de cada amostra. Por meio do programa *GATK* (MCKENNA *et al.*, 2010), para cada mapa construído, obtivemos o valor de cobertura por base. Esses valores foram utilizados para calcular a cobertura média de cada gene viral expresso, no qual foi feito através da média aritmética dos valores de cobertura obtidos a cada base do gene analisado.

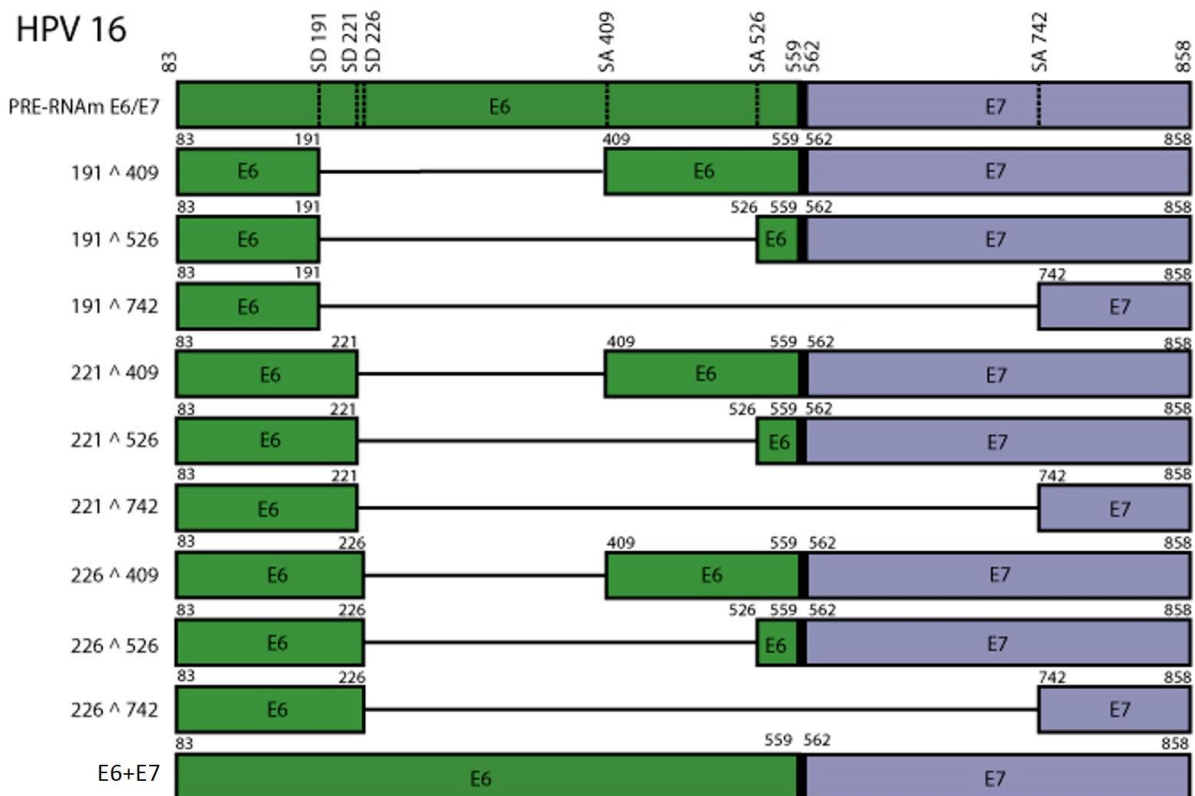
Para a identificação dos diferentes transcritos alternativos *E6/E7* expressos em cada amostra de tumor cervical, construímos sequências de referências específicas para cada um desses possíveis transcritos do HPV 16 e do HPV 18. Com o programa *BowTie 2*, mapeamos as leituras obtidas de cada amostra contra essas referências construídas.

As construções das sequências dos transcritos alternativos *E6/E7* foram realizadas utilizando as sequências dos genes *E6* e *E7* obtidos da referência do HPV 16 (K02718.1) e HPV 18 (AY262282.1), com base nos sítios de *splicing* doadores (SD) e aceptores (SA) previamente conhecidos (AJIRO *et al.*, 2012; DOORBAR *et al.*, 1990; SHERMAN *et al.*, 1992; TANG *et al.*, 2006).

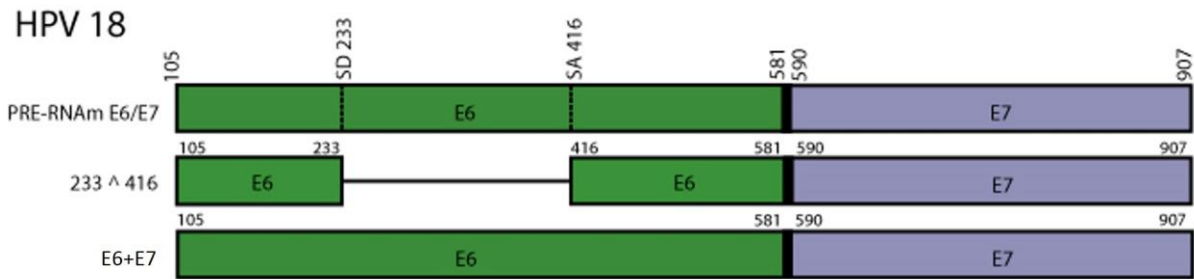
O pré-RNA *E6/E7* do HPV 16 apresenta seis sítios de *splicing*, sendo 3 sítios doadores (referentes aos nucleotídeos nas posições genômica 191, 221, 226) e 3 sítios aceptores (nucleotídeos nas posições genômica 409, 526, 742) (AJIRO *et al.*, 2012). Visto isso, é possível a formação de 9 tipos de transcritos alternativos *E6/E7*, fora o transcrito *E6+E7*. A construção das sequências nucleotídicas desses transcritos alternativos foram realizadas através da eliminação dos nucleotídeos presentes entre o sítio de *splicing* doador e o sítio de *splicing* acceptor, formando transcritos com diferentes junções de *splicing* e tamanhos distintos (Figura 4.2): 191<sup>^</sup>409 (559 pb), 191<sup>^</sup>526 (442 pb), 191<sup>^</sup>742 (256 pb), 221<sup>^</sup>409 (589 pb), 221<sup>^</sup>526 (472 pb), 221<sup>^</sup>742 (256 pb), 226<sup>^</sup>409 (594 pb), 226<sup>^</sup>526 (477 pb) e 226<sup>^</sup>742 (261 pb). O transcrito *E6+E7*, apresenta 776 pb de comprimento. As sequências construídas, em formato *fasta*, dos possíveis transcritos podem ser vistas no anexo 4.4.



No HPV 18 o pré-RNA *E6/E7* apresenta apenas dois sítios de *splicing*, um sítio doador (SD) localizado no nucleotídeo na posição genômica 233, e um sítio aceptor (SA) localizado na posição 416 (TANG *et al.*, 2006). Com isso, se forma apenas um tipo de transcrito alternativo *E6/E7*, e o transcrito *E6+E7*. A sequência nucleotídica utilizada como referência, assim como no HPV 16, foi construída através da eliminação dos nucleotídeos presentes entre o SD e o SA. Assim, foi construído a sequência do transcrito com junção de *splicing* 233^416, apresentando 621 pb de comprimento (Figura 4.3). O transcrito *E6+E7* de HPV 18, apresenta 803 pb de comprimento. As sequências construídas, em *fasta*, para os transcritos alternativo de HPV 18 estão demonstradas no anexo 4.5.



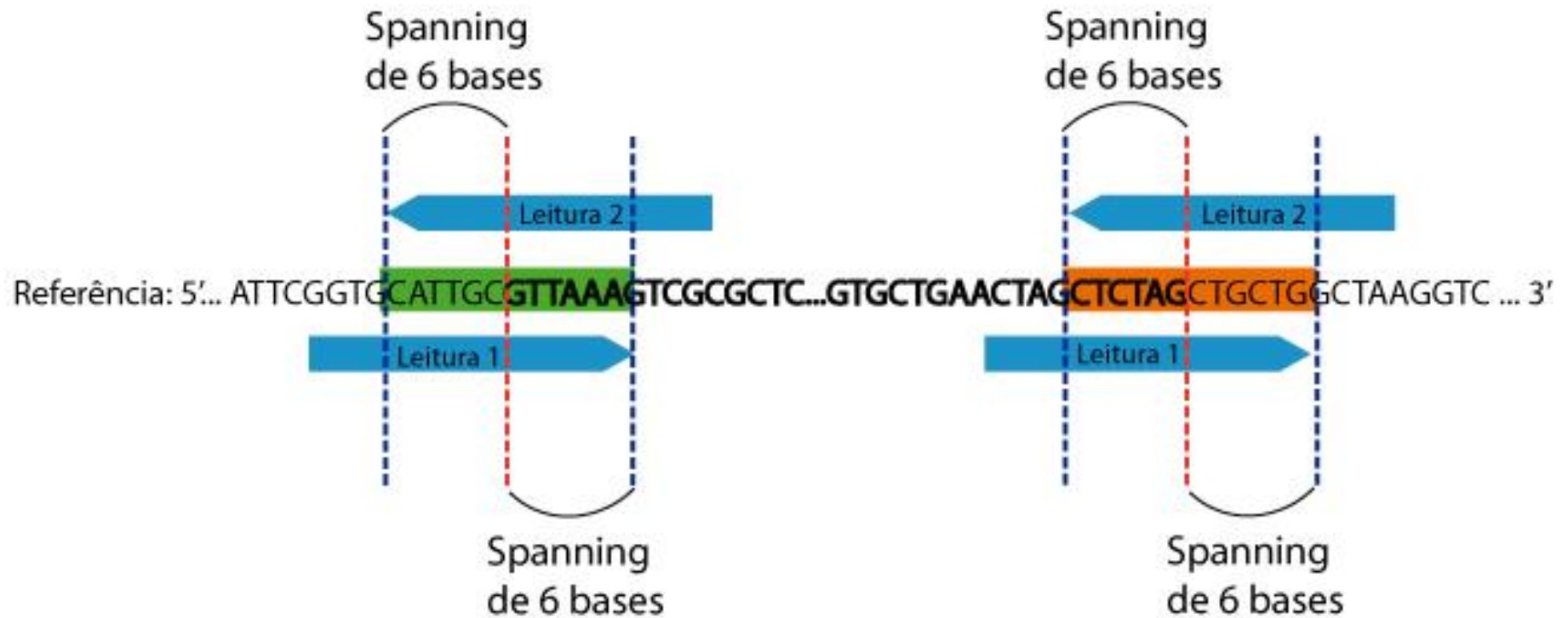
**Figura 4.2:** Possíveis transcritos alternativos *E6/E7* para o HPV 16. O pré-RNA *E6/E7* do HPV 16 apresentam 3 sítios de *splicing* doadores (SD) e 3 sítios de *splicing* aceptores (SA). Durante o processamento desse pré-RNA pelo mecanismo de *splicing* é possível a formação de 9 transcritos alternativos *E6/E7*, que apresentam as seguintes junções de *splicing*: 191^409, 191^526, 191^742, 221^409, 221^526, 221^742, 226^409, 226^526 e 226^742. Ainda é possível a formação de um transcrito que não sofre o *splicing*: *E6+E7*. As linhas contínuas entre os SD e SA representam as sequências nucleotídica que foram eliminadas para a construção das sequências de referências dos transcritos alternativos.



**Figura 4.3:** Possíveis transcritos alternativos *E6/E7* para o HPV 18. O pré-RNA *E6/E7* do HPV 18 apresenta apenas um sítio de *splicing* doador (SD) e um sítio de *splicing* aceptor (SA). Durante o processamento desse transcrito pelo mecanismo de *splicing*, é possível formar apenas o transcrito alternativo com junção de *splicing* 233<sup>^</sup>416 e o transcrito *E6+E7*, que não sofre o *splicing*. A linha contínua entre os SD e SA representa a sequência nucleotídica que foi eliminada para a construção da sequência de referência do transcrito alternativo.

O valor de cobertura da expressão para cada um dos transcritos alternativos *E6/E7* que sofrem *splicing* foi obtido através da contabilização do número de leituras de junção de *splicing* (LJS) que cobrem a região de junção de *splicing* (RJS). A RJS é composta por 12 nucleotídeos, 6 bases *upstream* e 6 bases *downstream* do sítio de junção de *splicing*. Para uma leitura ser considerada LJS é necessário apresentar *spanning* de pelo menos 6 bases da junção de *splicing*. A contagem das LJS foi realizada através da ferramenta *view* do programa *SAMtools* e da busca das leituras que apresentavam as 12 bases da RJS por um programa de edição de texto.

Já para os transcritos alternativos *E6+E7*, a cobertura da expressão foi obtida através da média de leituras *unspliced* (MLU), que cobrem as extremidades 5' e 3' da região *unspliced* (figura 4.4). A região *unspliced* é composta por uma sequência de nucleotídeos conservadas apenas nos transcritos *E6+E7*, sendo dos nucleotídeos 227 a 408 para o HPV 16 e, dos nucleotídeos 234 a 405 para o HPV 18. Leituras *unspliced* são leituras que apresentam um *spanning* de pelo menos 6 bases nas extremidades 5' ou na extremidade 3' da região *unspliced*. A contagem dessas leituras *unspliced* nas duas extremidades da região *unspliced* foi feita com a ferramenta *view* (*SAMtools*) e pela busca da sequência de 12 nucleotídeos, 6 *upstream* e 6 *downstream* da primeira e da última base da região *unspliced* (figura 4.4).



**Figura 4.4:** Representação de leituras *unspliced*. O transcrito alternativo *E6+E7* apresenta uma sequência de nucleotídeos, que está presente apenas nesse transcrito, chamada de região *unspliced* (bases em negrito). Uma leitura é considerada uma leitura *unspliced* quando ela apresenta um *spanning* de pelo menos 6 bases da extremidade 5' (caixa verde) ou a extremidade 3' (caixa laranja) da região *unspliced*. O valor de cobertura da expressão desse transcrito alternativo foi obtido através da média de leituras *unspliced* (MLU) para as duas extremidades. (A sequência nucleotídica demonstrada na ilustração é apenas representativa, não correspondendo a sequência real)

## 5- Resultados

### 5.1- Corridas no sequenciador *HiSeq 2500* (Illumina)

A clusterização das *flow cells* das três corridas realizadas no *HiSeq 2500* geraram grandes quantidades de *clusters* em todas as linhas. Na primeira corrida, as três linhas utilizadas para as amostras HPV16\_13 e HPV16\_14 geraram um total de  $684,34 \times 10^6$  clusters, apresentando uma densidade média de aproximadamente 825 mil clusters por milímetro quadrado (K/mm<sup>2</sup>). As três linhas apresentaram mais de 88% das bases com qualidade maior ou igual a 30 ( $Q \geq 30$ ). Na segunda e na terceira corridas, as linhas em que foram clusterizadas as amostras HPV16\_1 a HPV16\_12 geraram um total de  $325,55 \times 10^6$  clusters, com uma densidade média de aproximadamente 588 K/mm<sup>2</sup> e mais de 87% das bases apresentando  $Q \geq 30$ . Já nas linhas utilizadas para as amostras infectadas com HPV 18 e para a amostra da co-infecção foi gerado um total de  $400,02 \times 10^6$  clusters, com densidade média de aproximadamente 723 K/mm<sup>2</sup> e mais de 86% das bases com  $Q \geq 30$ .

Os valores do número de clusters, densidade e porcentagem de bases que apresentaram  $Q \geq 30$  específicos para cada linha e para cada corrida, estão representados nas tabelas 5.1 para as amostras HPV16\_13 e HPV16\_14, e para as demais amostras nas tabelas 5.2 e 5.3.

**Tabela 5.1:** Resultados da corrida RNA-Seq 50 x 1 para as amostras HPV16\_13 e HPV16\_14. Para cada linha da *flow cell* foi obtido a densidade de *clusters*, em milhares por milímetro quadrado (K/mm<sup>2</sup>), +/- o desvio padrão; número de *clusters*, em milhões ( $\times 10^6$ ); porcentagem de bases que apresentam qualidade maior ou igual a 30 ( $\%Q \geq 30$ ).

Corrida 50 X 1			
Linha	Densidade (K/mm <sup>2</sup> )	Clusters (x10 <sup>6</sup> )	%Q $\geq$ 30
1	1038 $\pm$ 97	286,86	88,9
2	627 $\pm$ 100	173,25	95,5
3	811 $\pm$ 116	224,23	92,3

**Tabela 5.2:** Resultados da corrida RNA-Seq 100 x 1. As 12 amostras, HPV16\_1 a HPV16\_12, foram corridas na linha 1 e as 10 amostras infectadas por HPV 18 e a co-infectadas ocuparam a linha 2. Para cada linha da *flow cell* foi obtido a densidade de *clusters*, em milhares por milímetro quadrado (K/mm<sup>2</sup>), +/- o desvio padrão; número de *clusters*, em milhões (x 10<sup>6</sup>); porcentagem de bases que apresentam qualidade maior ou igual a 30 (%Q ≥ 30).

<b>Corrida 100 X 1</b>			
<b>Linha</b>	<b>Densidade (K/mm<sup>2</sup>)</b>	<b>Clusters (x10<sup>6</sup>)</b>	<b>%Q ≥ 30</b>
1	598 ± 94	165,42	95,1
2	730 ± 110	201,69	93,8

**Tabela 5.3:** Resultados da corrida RNA-Seq 100 x 2. As 12 amostras, HPV16\_1 a HPV16\_12, foram corridas na linha 1 e as 10 amostras infectadas por HPV 18 e a co-infectadas ocuparam a linha 2. Para cada linha da *flow cell* foi obtido a densidade de *clusters*, em milhares por milímetro quadrado (K/mm<sup>2</sup>), +/- o desvio padrão; número de *clusters*, em milhões (x 10<sup>6</sup>); porcentagem de bases que apresentam qualidade maior ou igual a 30 (%Q ≥ 30).

<b>Corrida 100 X 2</b>				
<b>Linha</b>	<b>Densidade (K/mm<sup>2</sup>)</b>	<b>Clusters (x10<sup>6</sup>)</b>	<b>%Q ≥ 30</b>	
			<b>Leitura 1</b>	<b>Leitura 2</b>
1	579 ± 93	160,13	94,0	87,4
2	717 ± 98	198,33	92,4	86,0

## 5.2- Quantidade de leituras obtidas por amostra

O RNA-seq gerou para cada amostra infectada com HPV 16 uma média de 42 milhões de leituras, já para as amostras infectadas com HPV 18, foram obtidos uma média de 48 milhões de leituras e, a amostra que apresentou co-infecção gerou cerca de 61 milhões de leituras. Através da utilização do filtro de qualidade do *PRINSEQ* foram eliminados em média 0,82% e 0,7% das leituras obtidas pelas amostras infectadas com HPV 16 e com HPV 18, respectivamente. Já para a amostra co-infectada ocorreu uma perda de 0,66% das leituras.

Para cada amostra, os valores exatos da quantidade de leituras obtidas, a porcentagem de leituras que foram eliminadas pelo filtro de qualidade e a quantidade de leituras selecionadas para as análises, estão demonstrados na tabela 5.4.

**Tabela 5.4:** Número total de leituras obtidas para cada amostra. Foi contabilizado o número total de leituras obtidos pela corrida RNA-seq e o número de leituras selecionadas após passarem pelos diversos critérios estabelecidos pelo filtro de qualidade do *PRINSEQ*. As perdas de leituras estão demonstradas em porcentagem.

<b>Número de leituras por amostra</b>				
	<b>Amostra</b>	<b>RNA-Seq</b>	<b>Prinseq</b>	<b>Perda (%)</b>
HPV 16	HPV16_1	18.631.868	18.488.000	0,77
	HPV16_2	12.629.125	12.453.980	1,39
	HPV16_3	44.762.827	44.377.920	0,86
	HPV16_4	64.709.467	64.382.672	0,51
	HPV16_5	45.533.103	45.318.652	0,47
	HPV16_6	49.609.704	49.363.964	0,50
	HPV16_7	28.962.581	28.606.205	1,23
	HPV16_8	51.201.516	50.579.704	1,21
	HPV16_9	36.347.928	36.060.303	0,79
	HPV16_10	38.762.868	38.465.798	0,77
	HPV16_11	44.985.757	44.689.039	0,66
	HPV16_12	20.518.364	20.273.124	1,20
	HPV16_13	76.763.607	76.425.396	0,44
	HPV16_14	61.069.079	60.659.618	0,67
HPV 18	HPV18_1	60.700.998	60.277.430	0,70
	HPV18_2	59.385.231	59.005.639	0,64
	HPV18_3	27.830.185	27.593.828	0,85
	HPV18_4	53.759.449	53.327.505	0,80
	HPV18_5	50.174.368	49.804.500	0,74
	HPV18_6	35.462.183	35.213.704	0,70
	HPV18_7	71.387.041	70.926.211	0,65
	HPV18_8	46.434.304	46.131.225	0,65
	HPV18_9	42.990.945	42.698.997	0,68
	HPV18_10	41.733.681	41.461.932	0,65
HPV16 HPV 18	HPV16/18_1	61.628.645	61.222.630	0,66

### 5.3- Integração viral no genoma hospedeiro

A análise da integração do genoma do HPV no genoma do hospedeiro, através da busca por *contigs* quiméricos construídos pela montagem “*de novo*”, mostrou integração em 18 das 25 amostras analisadas. Dentre as 14 amostras de tumores cervicais infectadas com HPV 16, a integração foi encontrada em 8 amostras (tabela 5.5), já para as 10 amostras infectadas com HPV 18, foi encontrado integração em 9 (tabela 5.6). Também foi encontrado integração para o HPV 16 na amostra que apresentou co-infecção (tabela 5.7).

Analisando as sequências nucleotídicas na região de junção de integração obtidas para cada *contig* foi possível identificar em qual região do genoma do HPV houve a ruptura (anexo 5.1) e qual a região do cromossomo da célula tumoral ocorreu a integração (anexo 5.2).



**Tabela 5.5:** Análise da integração em amostras infectadas com HPV 16. A integração foi identificada através da presença de *contigs* quiméricos formados a partir de uma montagem “*de novo*” do transcriptoma das amostras. A análise da sequência nucleotídica dos *contigs* quiméricos pela ferramenta *blast* permitiu a identificação do gene viral e da posição nucleotídica no genoma viral em que ocorreu a quebra (posição quebra), além do gene humano e a posição nucleotídica que ocorreu a integração. O cromossomo no qual está localizado o gene humano em que ocorreu a integração e a função desse gene também foram identificadas. A cobertura dos *contigs* formados foram obtidos através do número de leituras de junção de integração (LJI).

Análise da integração em amostras infectadas com HPV 16								
Amostra	Integração	Gene HPV	Posição quebra	Gene Humano	Cromossomo	Posição integração	Função gene	LJI
HPV16_1	Encontrada	<i>E1</i>	880	<i>MCL1</i>	1q21	436 (R)	Proteína anti-apoptótica	310
HPV16_2	Encontrada	<i>E6</i>	526	<i>PIGR</i>	1q31	206950244 (G)	Receptor imunoglobulina	32
HPV16_3	Encontrada	<i>E1</i>	880	INT	4q31.1	139583009 (G)	-	334
HPV16_4	Encontrada	<i>E1</i>	880	<i>PXMP4</i>	20q11.22	237 (R)	Proteína 4 da membrana do peroxissomo	245
HPV16_5	Encontrada	<i>E1</i>	880	INT	13q22.1	73145553 (G)	-	179
		<i>E2</i>	3721	INT	Xq21.31	89010655 (G)	-	37
HPV16_6	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV16_7	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV16_8	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV16_9	Encontrada	<i>E1</i>	880	INT	11q22.2	102768316 (G)	-	606
HPV16_10	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV16_11	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV16_12	Encontrada	INT ( <i>E5-L2</i> )	4193	INT	9p11.2	41480558 (G)	-	2
HPV16_13	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV16_14	Encontrada	<i>E2</i>	2847	INT	2q31.2	178048625 (G)	-	496

ND – não determinado; INT – região intergênica; (R) – em relação a posição no RNAm; (G) – em relação a posição no genoma.

**Tabela 5.6:** Análise da integração em amostras infectadas com HPV 18. A integração foi identificada através da presença de *contigs* quiméricos formados a partir de uma montagem “*de novo*” do transcriptoma das amostras. A análise da sequência nucleotídica dos *contigs* quiméricos pela ferramenta *blast* permitiu a identificação do gene viral e da posição nucleotídica no genoma viral em que ocorreu a quebra (posição quebra), além do gene humano e a posição nucleotídica que ocorreu a integração. O cromossomo no qual é localizado o gene humano em que ocorreu a integração e a função desse gene também foram identificadas. A cobertura dos *contigs* formados foram obtidos através do número de leituras de junção de integração (LJI).

Análise da integração em amostras infectadas com HPV 18								
Amostra	Integração	Gene HPV	Posição quebra	Gene Humano	Cromossomo	Posição integração	Função gene	LJI
HPV18_1	Encontrada	<i>E1</i>	929	INT	15q21	58225819 (G)	-	172
HPV18_2	Encontrada	<i>E1</i>	929	<i>TP63</i>	3q28	1367 (R)	Supressor tumoral	216
HPV18_3	Não Encontrada	ND	ND	ND	ND	ND	ND	ND
HPV18_4	Encontrada	<i>E7</i>	724	<i>MMP13</i>	11q22.3	660 (R)	Quebra da matriz extracelular	187
HPV18_5	Encontrada	<i>E1</i>	929	<i>SLC16A14</i>	2q36.3	1839 (R)	ND	912
HPV18_6	Encontrada	<i>E2</i>	3053	INT	20q13.12	45123711 (G)	-	308
HPV18_7	Encontrada	<i>E1</i>	929	INT	14q23.2	61926194 (G)	-	685
		<i>E2</i>	3854			61925836 (G)		681
HPV18_8	Encontrada	<i>E1</i>	929	<i>HRH1</i>	3p25	2128 (R)	Receptor Histamina H1	61
HPV18_9	Encontrada	<i>E1</i>	929	<i>RAD51B</i>	14q23	1111 (R)	Reparo de DNA	26
		<i>L2</i>	5117			1548 (R)		25
HPV18_10	Encontrada	<i>E1</i>	929	<i>NRXN1</i>	2p16.3	1013422 (G)	Adesão celular	74
		<i>E7</i>	724			1739 (R)		164

ND – não determinado; INT – região intergênica; (R) – em relação a posição no RNAm; (G) – em relação a posição no genoma.

**Tabela 5.7:** Análise da integração em amostras co-infectadas com HPV 16 e HPV 18. A integração foi identificada através da presença de *contigs* quiméricos formados a partir de uma montagem “*de novo*” do transcriptoma das amostras. A análise da sequência nucleotídica dos *contigs* quiméricos pela ferramenta *blast* permitiu a identificação do gene viral e da posição nucleotídica no genoma viral em que ocorreu a quebra (posição qubra), além do gene humano e a posição nucleotídica que ocorreu a integração. O cromossomo no qual está localizado o gene humano em que ocorreu a integração e a função desse gene também foram determinados. A cobertura dos *contigs* formados foram obtidos através do número de leituras de junção de integração (LJI).

Análise da integração em amostras co-infectadas com HPV 16 e HPV18									
Amostra	HPV	Integração	Gene HPV	Posição quebra	Gene Humano	Cromossomo	Posição integração	Função gene	LJI
HPV16/18_1	HPV16	Encontrada	<i>E1</i>	880	INT	5q33.2	155777884 (G)	-	14
	HPV18	Não Encontrada	ND	ND	ND	ND	ND	ND	ND

ND – não determinado; INT – região intergênica; (G) – em relação a posição no genoma.

Observamos que para as amostras integradas que apresentavam HPV 16 a quebra no genoma viral ocorreu com maior frequência (5 de 8 integrações) no gene *E1*, porém, quebras em outros genes virais também foram observados (tabela 5.5). Na amostra HPV16\_5 foram encontradas duas integrações independentes, no qual uma quebra ocorreu no gene *E1* e a outra quebra no gene *E2*. Também foi observado quebra no gene *E2* na integração da amostra HPV16\_14. Quebra no gene *E6* foi observada na amostra HPV16\_2, e na amostra HPV16\_12 ocorreu quebra na região intergênica entre os genes *E5* e *L2*.

Nas amostras infectadas com HPV 18 as rupturas também ocorreram com maior frequência no gene *E1*, afetando 7 das 9 integrações (tabela 5.6). Outros genes virais que foram rompidos são o gene *E7* na amostra HPV18\_4 e o gene *E2* na amostra HPV18\_6. Três amostras apresentaram duas integrações independentes: (1) na amostra HPV18\_7 as rupturas ocorreram nos genes *E1* e *E2*; (2) na amostra HPV18\_9 romperam nos genes *E1* e *L2*; e (3) na amostra HPV18\_10 os genes quebrados foram o *E1* e *E7*.

Para a amostra co-infectada só foi encontrado integração para o genoma do HPV 16 (tabela 5.7). Nessa integração identificamos que a quebra ocorreu no gene *E1*.

Nas amostras que apresentam infecções pelo HPV 16 foi observado que, das 8 que continham DNA viral integrado, em 5 essa integração ocorreu em regiões intergênicas dos cromossomos. Nas três amostras restantes, a integração ocorreu em genes específicos (tabela 5.5). Para as 9 amostras que apresentavam DNA genômico do HPV 18 integrado, em 3 a integração ocorreu em regiões intergênicas e em 6 ocorreram em genes específicos (tabela 5.6). Os cromossomos no qual ocorreram essas integrações variaram entre as amostras.

Nas amostras HPV16\_3, HPV16\_9, HPV16\_12 e HPV16\_14, as integrações ocorreram em regiões intergênicas dos cromossomos 4, 11, 9 e 2 respectivamente. A amostra HPV16\_5, como apresenta duas integrações independentes, foi observado que uma integração ocorreu na região intergênica do cromossomo 13 e outra na região intergênica do cromossomo X. Na amostra HPV16\_1, encontramos integração no gene *myeloid cell leukemia 1 (MCL1)*, localizado no cromossomo 1. Na amostra HPV16\_2, a integração aconteceu no gene *polymeric immunoglobulin receptor (PIGR)*, também localizado no cromossomo 1. E para a amostra HPV16\_4,

a integração foi identificada no gene *peroxisomal membrane protein 4 (PXMP4)*, no cromossomo 20.

A integração das amostras HPV18\_1 e HPV18\_6 ocorreu na região intergênica dos cromossomos 15 e 20 respectivamente. Na amostra HPV18\_7, suas duas integrações ocorreram na região intergênica do cromossomo 14. As amostras HPV18\_2, HPV18\_4, HPV18\_5 e HPV18\_8, apresentaram integrações respectivamente nos genes *tumor protein p63 (TP63)*, *matrix metalloproteinase 13 (MMP13)*, *solute carrier family 16 (SLC16A14)* e *histamine receptor H1 (HRH1)*, localizados nos cromossomos 3, 11, 2 e 3. As duas integrações na amostra HPV18\_9 ocorreram em regiões próximas no gene *RAD51 paralog B (RAD51B)*, localizado no cromossomo 14. Esse mesmo fenômeno foi observado para a amostra HPV18\_10, tendo suas duas integrações ocorrendo no gene *neurexin 1 (NRXN1)* no cromossomo 2. O genoma do HPV 16 que foi integrado na amostra co-infectada, a integração ocorreu na região intergênica do cromossomo 5 (tabela 5.7).

Os *contigs* quiméricos formados das amostras infectadas com HPV16 obtiveram uma cobertura média de 249 LJI (tabela 5.5). O *contig* formado para a amostra HPV16\_14 foi o que apresentou a maior cobertura, com 496 leituras, e o *contig* que apresentou a menor cobertura foi da amostra HPV16\_14, com 2 leituras.

A cobertura média obtida pelos *contigs* quiméricos formados para as amostras infectadas pelo HPV 18 foi de 292,5 LJI (tabela 5.6). Com 912 leituras, o *contig* da amostra HPV18\_5 foi o que obteve a maior cobertura, já os *contigs* da amostra HPV18\_9 foram os que apresentaram a menor cobertura, com 26 e 25 leituras. A amostra co-infectada formou um *contig* quimérico para o genoma do HPV 16 com cobertura de 14 LJI (tabela 5.7).

## **5.4- Expressão do genoma do HPV**

### **5.4.1- Mapeamento das leituras contra o genoma do HPV**

Do número médio de aproximadamente 42 milhões de leituras das amostras infectadas com o HPV 16, por volta de 13 mil (0,031%) foram mapeadas contra o genoma de referência desse tipo viral. Já para a média das leituras obtidas para as amostras infectadas com HPV 18, das aproximadamente 48 milhões de leituras,

cerca de 9 mil (0,018%) mapearam contra o genoma de referência desse HPV infectante. Para a amostra co-infectada, das aproximadamente 61 milhões de leituras obtidas, apenas 15 leituras mapearam contra o genoma de referência do HPV 18 e mais de 9 mil leituras (0,014%) mapearam contra o genoma de referência do HPV 16.

A quantidade de leituras mapeadas para cada amostra pode ser observada na tabela 5.8 para as amostras infectadas com HPV 16, na tabela 5.9 para as amostras infectadas com HPV 18 e na tabela 5.10 para a amostra com co-infecção.

**Tabela 5.8:** Mapeamento das leituras contra a sequência de referência do HPV 16. Para cada amostra foi obtida a quantidade total de leituras selecionadas pelo filtro de qualidade (total de leituras) e a quantidade de leituras que foram mapeadas contra o genoma de referência do HPV 16 (Leituras mapeadas). A partir disso, foi obtido a porcentagem de leituras mapeadas contra a referência (% mapeada).

<b>Mapeamento das amostras infectadas com HPV 16 contra o genoma de referência do HPV 16</b>			
<b>Amostra</b>	<b>Total de leituras</b>	<b>Leituras mapeadas</b>	<b>% Mapeada</b>
HPV16_1	18.488.000	15.147	0,0813
HPV16_2	12.453.980	1.323	0,0105
HPV16_3	44.377.920	4.496	0,0100
HPV16_4	64.382.672	7.752	0,0120
HPV16_5	45.318.652	8.313	0,0183
HPV16_6	49.363.964	4.495	0,0091
HPV16_7	28.606.205	35.625	0,1230
HPV16_8	50.579.704	25	0,00004
HPV16_9	36.060.303	5.403	0,0149
HPV16_10	38.465.798	15.284	0,0394
HPV16_11	44.689.039	17	0,00003
HPV16_12	20.273.124	7.810	0,0381
HPV16_13	76.425.396	68.655	0,0894
HPV16_14	60.659.618	8.838	0,0145

**Tabela 5.9:** Mapeamento das leituras contra a sequência de referência do HPV 18. Para cada amostra foi obtida a quantidade total de leituras selecionadas pelo filtro de qualidade (Total de leituras) e a quantidade de leituras que foram mapeadas contra o genoma de referência do HPV 18 (Leituras mapeadas). A partir disso, foi obtido a porcentagem de leituras mapeadas contra a referência (% Mapeada).

<b>Mapeamento das amostras infectadas com HPV 18 contra o genoma de referência do HPV 18</b>			
<b>Amostra</b>	<b>Total de leituras</b>	<b>Leituras mapeadas</b>	<b>% Mapeada</b>
HPV18_1	60.277.430	19.445	0,0320
HPV18_2	59.005.639	3.332	0,0056
HPV18_3	27.593.828	2.560	0,0092
HPV18_4	53.327.505	1.898	0,0035
HPV18_5	49.804.500	13.765	0,0274
HPV18_6	35.213.704	5.022	0,0142
HPV18_7	70.926.211	29.614	0,0415
HPV18_8	46.131.225	10.591	0,0228
HPV18_9	42.698.997	2.678	0,0062
HPV18_10	41.461.932	6.759	0,0162

**Tabela 5.10:** Mapeamento das leituras da amostra co-infectada contra as sequências de referência do HPV 16 e do HPV 18. Foi obtida a quantidade total de leituras selecionadas pelo filtro de qualidade (Total de leituras) e a quantidade de leituras que foram mapeadas contra os genomas de referências (Leituras mapeadas). A partir disso, foi obtido a porcentagem de leituras mapeadas para cada referência (% Mapeada).

<b>Mapeamento da amostra co-infectada contra o genoma de referência do HPV 16 e HPV 18</b>				
<b>Amostra</b>	<b>Total de leituras</b>	<b>Referência</b>	<b>Leituras mapeadas</b>	<b>% Mapeada</b>
HPV16/18_1	61.222.630	HPV 16	9.918	0,0161
		HPV 18	15	0,00002

#### 5.4.2- Cobertura da expressão dos genes virais

A cobertura média da expressão dos genes virais diferiu entre as amostras de tumores cervicais. Das 14 amostras infectadas com HPV 16, 10 apresentaram maior cobertura média por base para os genes *E6/E7*, três amostras apresentaram os genes *E2/E4* com a maior cobertura média e apenas uma amostra apresentou maior cobertura para o gene *E5* (tabela 5.11). Todas as amostras infectadas com HPV 18 apresentaram maior cobertura média para os genes *E6/E7* (tabela 5.12). Na amostra co-infectada, o gene do HPV 16 que obteve a maior cobertura média foi o *E2/E4*, já o gene *E6/E7* do HPV 18 foi o único que apresentou cobertura para esse genótipo viral, com uma baixa cobertura média por base de 0,37 vezes (tabela 5.13).

Analisando a expressão do genoma total, observamos que as amostras infectadas por HPV 16 e classificadas como adenocarcinoma (ADN) apresentaram, em média, uma cobertura por base de 57,67 vezes, já as classificadas como carcinoma de células escamosas (CCE) apresentaram cobertura por base de 33,92 vezes para cada amostra. As amostras ADN infectadas com HPV 18 apresentaram em média uma cobertura de 51,02 vezes e, as de CCE apresentaram cobertura de 54,59 vezes. Na amostra co-infectada, a cobertura média por base da expressão do genoma do HPV 16 foi de 47,04 vezes e do genoma do HPV 18 foi de 0,03 vezes. Gráficos de cobertura por base da expressão do genoma viral para cada amostra estão apresentados nos anexos 5.3, 5.4 e 5.5.



**Tabela 5.11:** Média de cobertura da expressão do genoma viral em tumores infectados por HPV 16. A análise da expressão do genoma viral das amostras infectadas pelo HPV 16 foi estimado através da média de cobertura por base obtido para o genoma viral e para a sequência de cada gene ou região genômica.

<b>Média de cobertura da expressão dos genes viral de tumores cervicais infectados por HPV 16</b>								
	<b>Amostra</b>	<b>Genoma</b>	<b>E1</b>	<b>E2/E4</b>	<b>E5</b>	<b>E6/E7</b>	<b>L1</b>	<b>L2</b>
<b>ADN</b>	HPV16_1	49,78	20,97	57,06	10,53	156,51	18,83	0,42
	HPV16_2	2,95	0,97	0	0	17,38	0	0
	HPV16_10	75,94	47,68	176,97	33,9	99,81	3,49	2,26
	HPV16_11	0,12	0	0	0	0,94	0,12	0
	HPV16_12	44,84	12,86	118,54	16,15	21,4	10,21	2,3
	HPV16_13	172,89	38,68	534,96	767,36	195,62	2	2,25
<b>CCE</b>	HPV16_3	11,29	5,81	0	0	66,8	0	0
	HPV16_4	22,34	15,65	0	0	110,35	0,12	0
	HPV16_5	23,4	4,76	27,4	1,24	100,52	0,24	0,14
	HPV16_6	24,94	10,74	58,81	27,2	28,05	0,93	0,63
	HPV16_7	161,43	258,64	0	0	305,22	0,12	0
	HPV16_8	0,11	0,05	0,16	0,41	0,64	0	0
	HPV16_9	12,64	2,33	0,51	0	72,38	0	0
	HPV16_14	15,27	23,5	15,59	0	48,19	0,57	0,07

**Tabela 5.12:** Média de cobertura da expressão do genoma viral em tumores infectados por HPV 18. A análise da expressão do genoma viral das amostras infectadas pelo HPV 18 foi estimado através da média de cobertura por base obtido para o genoma viral e para a sequência de cada gene ou região genômica.

<b>Média de cobertura da expressão dos genes viral de tumores cervicais infectados por HPV 18</b>								
	<b>Amostra</b>	<b>Genoma</b>	<b>E1</b>	<b>E2/E4</b>	<b>E5</b>	<b>E6/E7</b>	<b>L1</b>	<b>L2</b>
<b>ADN</b>	HPV18_2	15,05	2,08	1,24	0	73,69	0	0
	HPV18_3	14,52	6,87	19,79	10,65	33,87	0,34	0,36
	HPV18_4	8,06	0,78	0,27	0,38	39,34	0	0
	HPV18_7	169,82	134,11	0	262,41	438,29	1,04	1,56
	HPV18_8	61,41	1,83	129,19	98,45	154,58	0,27	0,71
	HPV18_10	37,27	1,44	46,13	40,33	117,45	2,18	0,57
<b>CCE</b>	HPV18_1	113,25	27,53	212,78	202,68	221,65	0,2	0,79
	HPV18_5	64,43	3,59	0,36	0	308,34	0	0
	HPV18_6	22,52	0,75	0	0	112,73	0	0
	HPV18_9	18,17	7,04	17,57	19,68	21,34	7,2	7,33

**Tabela 5.13:** Média de cobertura da expressão do genoma viral no tumor co-infectado. A análise da expressão do genoma do HPV 16 e HPV 18 das amostras co-infectadas foi estimada através da média de cobertura por base obtida para o genoma viral e para a sequência de cada gene ou região genômica.

Média de cobertura da expressão dos genes viral de tumores cervicais co-infectados por HPV 16 e HPV18								
Amostra	Vírus	Genoma	E1	E2/E4	E5	E6/E7	L1	L2
HPV16/18_1	HPV 16	47,04	1,84	99,74	87,37	70,22	3,05	1,31
	HPV 18	0,03	0	0	0	0,37	0	0

### 5.4.3- Transcritos alternativos E6/E7

Do total de transcritos *E6/E7* expressos, foi analisado o padrão de *splicing* alternativo das amostras de tumores cervicais e calculado a proporção da expressão de cada transcrito alternativo *E6/E7* encontrado.

Observamos que em todas as amostras infectadas pelo HPV 16, o transcrito alternativo *E6/E7* mais expresso foi o *E6\*I* (226<sup>409</sup>), apresentando uma média de 610 LJS por amostra, o que equivale aproximadamente a 89% dos transcritos *E6/E7* expressos. Exceto para as amostras HPV16\_8 e HPV16\_11, que expressaram apenas o transcrito *E6\*I*, o segundo transcrito com maior expressão foi o *E6\*II*, apresentando uma média aproximada de 57 LJS, cerca de 7% do total dos transcritos *E6/E7*. Já o terceiro transcrito mais expresso foi o *E6+E7*, apresentando cerca de 22 MLU por amostra, e representando cerca de 3% dos transcritos *E6/E7*. Os transcritos alternativos *E6\*X* e *E6\*VI* foram encontrados em 5 e 2 amostras respectivamente, sendo responsáveis por uma proporção muito baixa na expressão dos transcritos *E6/E7* totais. O transcrito *E6\*V* e os transcritos que apresentam junções de *splicing* 191<sup>526</sup>, 191<sup>742</sup>, 221<sup>526</sup> e 221<sup>742</sup> não foram encontrados em nenhuma amostra tumoral infectada pelo HPV 16. A quantidade de LJS, MLU e a proporção dos transcritos alternativos *E6/E7* expressos para cada amostra infectada pelo HPV 16 pode ser observada na tabela 5.14.

Todas as amostras infectadas pelo HPV 18 apresentaram maior expressão do transcrito *E6\*I*, obtendo uma média de aproximadamente 748 LJS por amostra, o que representa, em média, 93% da expressão dos transcritos *E6/E7* expressos. O transcrito *E6+E7* apresenta uma média de aproximadamente 44 MLU por amostra,

equivalente a cerca de 6% dos transcritos E6/E7 total. Os valores obtidos para cada amostra contendo HPV 18 encontra-se na tabela 5.15.

A proporção da expressão para os transcritos alternativos *E6/E7* do HPV 16 na amostra com co-infecção foi de 90,09% do transcrito *E6\*I*, 6,47% do transcrito *E6\*II* e 3,04% do transcrito *E6+E7*. Os transcritos *E6\*X* e *E6\*V* foram expressos em baixas proporções, equivalendo respectivamente a 0,12% e 0,25% dos transcritos *E6/E7* expressos. O transcrito *E6\*VI* e os transcritos que apresentam as junções de *splicing* 191<sup>^</sup>526, 191<sup>^</sup>742, 221<sup>^</sup>526 e 221<sup>^</sup>742 não foram encontrados. A expressão do genoma do HPV 18 nessa amostra obteve apenas 1 LJS para o transcrito *E6\*I*, obtendo nenhuma MLU para o transcrito alternativo *E6+E7*.

**Tabela 5.14:** Análise da proporção dos diferentes transcritos alternativos *E6/E7* em amostras de tumores cervicais infectadas com HPV16. Para cada amostra foi contabilizado o número de leituras de junção de *splicing* (LJS) obtido para os transcritos alternativos *E6/E7* que sofrem *splicing* e, para o transcrito *E6+E7* foi estimada a média de leituras *unspliced* (MLU). Foi calculada a proporção, em porcentagem (%), da expressão de cada um dos transcritos alternativos expressos pelas amostras.

Proporção de transcritos alternativos <i>E6/E7</i> em tumores cervicais infectados por HPV16 (em %)																					
Amostra	191^409 ( <i>E6*VI</i> )		191^526		191^742		221^409 ( <i>E6*V</i> )		221^526		221^742		226^409 ( <i>E6*I</i> )		226^526 ( <i>E6*II</i> )		226^742 ( <i>E6*X</i> )		<i>E6+E7</i>		
	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	MLU	%	
ADN	HPV16_1	1	0,06	0	0	0	0	0	0	0	0	0	0	1296	88,89	110	7,55	1	0,06	48,5	3,32
	HPV16_2	0	0	0	0	0	0	0	0	0	0	0	0	159	89,83	14	7,90	0	0	4	2,25
	HPV16_10	0	0	0	0	0	0	0	0	0	0	0	0	567	79,24	114	15,93	2	0,27	32,5	4,54
	HPV16_11	0	0	0	0	0	0	0	0	0	0	0	0	2	100	0	0	0	0	0	0
	HPV16_12	0	0	0	0	0	0	0	0	0	0	0	0	159	86,41	18	9,78	0	0	7	3,80
	HPV16_13	0	0	0	0	0	0	0	0	0	0	0	0	1062	83,72	126	9,93	0	0	80,5	6,34
CCE	HPV16_3	0	0	0	0	0	0	0	0	0	0	0	0	512	91,92	24	4,30	0	0	21	3,77
	HPV16_4	0	0	0	0	0	0	0	0	0	0	0	0	844	88,74	67	7,04	0	0	40	4,20
	HPV16_5	1	0,09	0	0	0	0	0	0	0	0	0	0	947	88,33	99	9,23	2	0,18	23	2,14
	HPV16_6	0	0	0	0	0	0	0	0	0	0	0	0	155	82,66	18	9,6	1	0,53	13,5	7,20
	HPV16_7	0	0	0	0	0	0	0	0	0	0	0	0	1914	92,82	123	5,96	0	0	25	1,21
	HPV16_8	0	0	0	0	0	0	0	0	0	0	0	0	3	100	0	0	0	0	0	0
	HPV16_9	0	0	0	0	0	0	0	0	0	0	0	0	622	90,53	51	7,42	4	0,58	10	1,45
	HPV16_14	0	0	0	0	0	0	0	0	0	0	0	0	303	84,40	44	12,25	0	0	12	3,34

ADN – Adenocarcinoma; CCE – Carcinomas de Células Escamosas

**Tabela 5.15:** Análise da proporção dos diferentes transcritos alternativos *E6/E7* em amostras de tumores cervicais infectadas com HPV18. Para cada amostra foi contabilizado o número de leituras de junção de *splicing* (LJS) obtido para o transcrito *E6\*I* e, para o transcrito *E6+E7* foi estimada a média de leituras *unspliced* (MLU). Foi calculada a proporção, em porcentagem (%), da expressão de cada um dos transcritos alternativos expressos pelas amostras.

<b>Proporção de transcritos alternativos <i>E6/E7</i> em tumores cervicais infectados por HPV18 (em %)</b>					
Amostra		233 <sup>416</sup> ( <i>E6*I</i> )		<i>E6+E7</i>	
		LJS	%	MLU	%
ADN	HPV18_2	331	94,30	20	5,69
	HPV18_3	178	94,42	10,5	5,57
	HPV18_4	198	97,29	5,5	2,70
	HPV18_7	2589	96,62	90,5	3,37
	HPV18_8	458	86,49	71,5	13,50
	HPV18_10	400	90,29	43	9,70
CCE	HPV18_1	1165	95,96	49	4,03
	HPV18_5	1520	93,22	110,5	6,77
	HPV18_6	553	92,78	43	7,21
	HPV18_9	93	97,89	2	2,10

ADN – Adenocarcinoma; CCE – Carcinomas de Células Escamosas

**Tabela 5.16:** Análise da proporção dos diferentes transcritos alternativos *E6/E7* na amostra de tumor cervicais co-infectadas pelo HPV 16 e HPV 18. Foi contabilizado o número de leituras de junção de *splicing* (LJS) obtido para os transcritos alternativos *E6/E7* que sofrem *splicing* e, para o transcrito *E6+E7* foi estimada a média de leituras *unspliced* (MLU). Ainda foi calculada a proporção, em porcentagem (%), da expressão de cada um dos transcritos alternativos expressos pelas amostras.

Amostra		Proporção de transcritos alternativos <i>E6/E7</i> em tumores cervicais co-infectados por HPV16 e HPV18 (em %)																							
		HPV 16																		HPV18					
		191^409 ( <i>E6*VI</i> )		191^526		191^742		221^409 ( <i>E6*V</i> )		221^526		221^742		226^409 ( <i>E6*I</i> )		226^526 ( <i>E6*II</i> )		226^742 ( <i>E6*X</i> )		<i>E6+E7</i>		233^416 ( <i>E6*I</i> )		<i>E6+E7</i>	
LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	LJS	%	MLU	%	LJS	%	MLU	%
CCE	HPV16/18_1	0	0	0	0	0	0	2	0,25	0	0	0	0	696	90,09	50	6,47	1	0,12	23,5	3,04	1	100	0	0

CCE – Carcinomas de Células Escamosas

## 6- Discussão

### 6.1- Integração do genoma viral

O evento principal para a progressão de uma neoplasia intraepitelial cervical (CIN) para o câncer cervical é a integração do genoma do HPV no genoma do hospedeiro. Acredita-se que a principal razão para a progressão tumoral após a integração do HPV, é a expressão descontrolada dos oncogenes viral, o que acaba gerando uma pressão seletiva favorável as células neoplásicas. Além disso, se especula que genes celulares importantes são afetados por essa integração, contribuindo ainda mais para a geração do câncer de colo uterino (WENTZENSEN; VINOKUROVA; VON KNEBEL DOEBERITZ, 2004).

Diversas técnicas vêm sendo utilizadas para analisar o sítio de integração do genoma do HPV. Uma dessas técnicas é a hibridização *In situ*, no qual através de sondas específicas se pode ter uma estimativa da distribuição das integrações no genoma do hospedeiro (MINCHEVA; GISSMANN; ZUR HAUSEN, 1987). Os protocolos mais utilizados para identificar o sítio de integração são baseados na técnica de PCR. Três dos principais protocolos com base na PCR são: (1) amplificação dos transcritos oncogênicos do HPV (APOT) (KLAES *et al.*, 1999), (2) amplificação da sequência integrada do HPV franqueando com sequencias celulares (DIPS) (LUFT *et al.*, 2001) e, (3) amplificação da região de fusão utilizando *primers* específicos e enzimas de restrição (PCR de restrição) (THORLAND *et al.*, 2000). Em 2013, foi publicado um trabalho analisando o sítio de integração do genoma viral em 51 amostras de câncer cervical infectadas com HPV 16 através do sequenciamento de nova geração na plataforma *HiSeq* 2000 (Illumina). Nesse trabalho foi realizado o sequenciamento do DNA das amostras tumorais seguido da identificação de potenciais sítios de junção HPV-célula (XU *et al.*, 2013). Para identificarmos os sítios de integração do genoma viral no genoma do hospedeiro, realizamos o sequenciamento de RNA (RNA-Seq) de 25 amostras de tumores cervicais. A partir das sequências obtidas, identificamos os sítios de integração através da busca por transcritos quiméricos apresentando sequências HPV-Humano. Metodologia de identificação de integração semelhante a que utilizamos, foi realizada por OJESINA *et al.*, 2014.

No câncer o genoma viral na forma integrada é o mais comum, sendo encontrado em 45% a 80% dos tumores, dependendo do tipo viral infectante. Em diversas análises já foram observados que no câncer e em lesões intraepiteliais de alto grau o genoma viral pode ser encontrado comitadamente na forma epissomal e integrado (CRICCA *et al.*, 2007; TONON *et al.*, 2001). A frequência da integração nos HPVs de alto risco é bastante alta, sendo encontrada em aproximadamente 80% das neoplasias infectadas com HPV 16 e em praticamente todas infectadas com o HPV 18 (PETT; COLEMAN, 2007; WOODMAN; COLLINS; YOUNG, 2007). Uma análise de 83 amostras infectadas com HPV 16, incluindo CIN 2/3 e câncer cervical, encontrou 46,9% das amostras com DNA epissomal, 37,4% apresentaram a forma integrado e 15,7% apresentaram DNA integrado e epissomal (HÄFNER *et al.*, 2008). Como se pode observar a proporção de amostras de neoplasias intraepiteliais cervical e câncer cervical que apresentam DNA viral integrado e/ou epissomal variam nos diferentes trabalhos já publicados. Em nossa análise encontramos 57,1% (8 de 14 amostras) das amostras de câncer cervical infectados com HPV 16 e 90% (9 de 10 amostras) das amostras infectadas com HPV 18 apresentando DNA viral integrado. Também encontramos integração do genoma do HPV 16 na amostra co-infectada pelo HPV 16 e pelo HPV 18. A metodologia que utilizamos nessa análise, pelo fato de ser uma análise da sequência do RNA das células tumorais, nos permitiu identificar apenas as integrações, não sendo possível identificar as amostras que apresentam apenas a forma epissomal ou ainda as amostras que apresentam o DNA viral em ambas as formas. Além disso, por se tratar da análise do transcriptoma dos tumores, identificamos apenas a integração nos sítios genômicos que foram expressos, não identificando integrações que tenham ocorrido dentro de íntrons e em outras regiões genômicas não codificantes.

Como foi revisado por WOODMAN, COLLINS e YOUNG em 2007, a quebra do genoma do HPV para que ocorra a integração, acontece preferencialmente nos genes *E1* ou *E2*. Das amostras de tumor cervical infectadas com HPV 16 e/ou HPV 18 observamos que 13 apresentaram o rompimento do genoma viral no gene *E1*, já rompimento no gene *E2* foram observados em 4 amostras, sendo assim, conforme descrito na literatura, a maioria dos rompimentos para a integração de fato ocorreram em *E1* e *E2*. Três amostras apresentaram rompimento em genes virais diferentes de *E1* e *E2*, dentre elas, a amostra HPV16\_2 sofreu a ruptura em *E6*, amostra HPV16\_12 rompeu na região intergênica entre os genes *E5* e *L2*, e a



amostra HPV18\_4 com ruptura em *E7*. Em destaque, encontramos quatro amostras que apresentaram duas integrações independentes, no qual os genes rompidos para cada uma das integrações foram diferentes: HPV16\_5 e HPV18\_7 com rupturas em *E1* e *E2*; HPV18\_9 rupturas em *E1* e *L2*; e HPV18\_10 com rupturas em *E1* e *E7* (tabelas 5.5, 5.6 e 5.7; anexo 5.1).

Através da metodologia que utilizamos, observamos que todas as rupturas no gene *E1* ocorreram no mesmo nucleotídeo: 880 ou 929 no genoma do HPV 16 e HPV 18, respectivamente. A quebra nesses nucleotídeos já foi observada anteriormente por SCHMITZ *et al.*, 2012. Segundo os autores, pelo fato dos nucleotídeos 880, no HPV 16, e 929, no HPV 18, se tratarem de sítios de *splicing* doadores existentes em *E1*, as integrações encontradas nesse ponto são originadas a partir do mecanismo de *splicing* entre a ORF de *E1* com o éxon de algum gene conhecido. Sendo assim, uma vez que nossas análises foram feitas com base em sequências de RNA, acreditamos que as indicações de rompimento do genoma viral em *E1*, - nas posições 880 (HPV 16) ou 929 (HPV18) -, obtidos pelos *contigs* quiméricos, na verdade podem não ser exatamente o ponto em que ocorreu esse rompimento. Dessa forma, o rompimento no genoma viral pode estar localizado no próprio sítio doador de *splicing* indicado ou em qualquer outro ponto *downstream* a esse sítio. Fato semelhante pode ter ocorrido na integração da amostra HPV16\_2, no qual a análise do *contig* quimérico nos indicou quebra do genoma viral no sítio de *splicing* 526. Nesse caso, como o nucleotídeo 526 é um sítio de *splicing* acceptor, acreditamos que o transcrito formado também é resultado de um *splicing*, além disso, o sítio em que ocorreu a quebra no genoma viral pode ter ocorrido nesse SA ou em qualquer outro ponto *upstream* a ele. Nas demais amostras, as integrações que encontramos a quebra no genoma do HPV não ocorreram em sítios de *splicing* conhecidos, portanto o ponto de quebra no genoma viral indicado pela metodologia que utilizamos, de fato deve ser o local onde ocorreu o rompimento. Para identificarmos precisamente o sítio de quebra das amostras no qual os transcritos quiméricos foram originados de um produto de *splicing*, precisamos fazer a análise do DNA extraídos desses tumores, buscando assim sequências de DNA quiméricas HPV/humano.

Conforme observado por KRAUS *et al.* em 2008, a integração do genoma viral pode ocorrer em qualquer lugar no genoma humano, mas elas tendem a ocorrer com maior frequência em determinados sítios do genoma hospedeiro. Já foram

mapeados mais de 250 sítios de integração no genoma humano. Esses sítios foram encontrados em todos os cromossomos humanos, porém a maioria deles eram sítios frágeis, pontos de quebra de translocação e regiões de transcrição ativas. Sítios frágeis são regiões cromossômicas propensas a sofrerem quebras, facilitando assim a integração de um DNA exógeno (SUTHERLAND; RICHARDS, 1999). Estudos observaram a presença de integrações em diversas bandas citogenéticas nos diversos cromossomos, porém as encontradas com maior frequência são 4q13.3, 8q24.21, 13q22.1 e 17q21.2. Dentre estes, o sítio mais predominante está localizado em 8q24.21, perto do *locus* do gene *MYC* (SCHMITZ *et al.*, 2012). Nas 18 amostras integradas que analisamos, encontramos uma ampla distribuição das integrações nos cromossomos, sendo assim não observamos uma preferência das integrações por determinada região cromossômica (tabelas 5.5, 5.6 e 5.7).

A posição nucleotídica exata em que acontece a integração varia entre as amostras, mesmo elas ocorrendo na mesma região cromossômica, também não há uma sequência de nucleotídeos características que esteja associada a integração em determinada região. Em diversos casos a integração do HPV ocorre dentro ou perto de genes conhecidos, porém o mais frequente é que ela ocorra em regiões intergênicas (WENTZENSEN; VINOKUROVA; VON KNEBEL DOEBERITZ, 2004). Em um trabalho publicado por KRAUS *et al.* em 2008, foi observado que frequentemente a integração ocorre dentro de regiões de transcrições ativas do genoma do hospedeiro, além disso, esses autores encontraram 38% das integrações dentro de genes conhecidos. Alguns desses genes afetados pela integração são genes importantes para o desenvolvimento tumoral, como os genes *MYC*, *TP63*, *NR4A2*, *APM-1*, *FANCC*, *TNFAIP2* e *hTERT* (WENTZENSEN; VINOKUROVA; VON KNEBEL DOEBERITZ, 2004). Em nosso trabalho, 52,6% dos sítios de integração que foram encontrados estão localizados em regiões intergênicas, os restantes das integrações foram encontrados dentro de 9 genes diferentes: *MCL1*, *PIGR*, *PXMP4*, *TP63*, *MMP13*, *SLC16A14*, *HRH1*, *RAD51B* e *NRXN1* (tabelas 5.5, 5.6 e 5.7).

Encontramos na amostra HPV16\_1 a integração na região 1q21 dentro do gene *MCL1*, que é responsável pela síntese de uma proteína anti-apoptótica membro da família BCL-2. Integração na região 1q21 também foram observadas em outros trabalhos, porém as integrações foram associadas aos genes *HS8015* (THORLAND *et al.*, 2000) e *RPS27* (WENTZENSEN *et al.*, 2002). Segundo

THORLAND *et al.*, 2000 essa região 1q21 é um sítio frágil. A integração da amostra HPV16\_4 ocorreu na região 20q11.22, dentro do gene *PXMP4* que codifica uma proteína da membrana do peroxissomo. A região 20q11.22 já foi descrita com um ponto de integração do genoma viral, porém o gene afetado por essa integração foi o *CBFA2T2* (XU *et al.*, 2013). A região 3q28 segundo Wentzensen (2004) se trata de um sítio frágil, e integrações nesse sítio já foram descritas afetando os genes *TP73* e *TPRG1*, ambos pertencentes a família dos fatores de transcrição de p53 (DALL *et al.*, 2008; NAMBARU *et al.*, 2009). Encontramos integração em 3q28 na amostra HPV18\_2, ocorrendo dentro do gene *TP63*, que também faz parte da família de p53. Como os genes contidos na região 3q28 já foram diretamente relacionados com o desenvolvimento de diferentes tipos de tumores, integração nessa região pode ser de grande importância no desenvolvimento do câncer cervical. Na amostra HPV18\_5 a integração ocorreu no sítio 2q36.3 dentro do gene *SLC16A14*, integração nessa mesma região foi descrita por XU *et al.* em 2013, porém segundo os autores o gene afetado pela integração foi o *PID1*. Já a amostra HPV18\_8 encontramos integração em 3p25 no gene *HRH1*, responsável pela codificação do receptor da histamina. Integração em 3p25 também já foi descrita em uma região intergênica, próxima ao protóncogene *RAF1* (DÜRST *et al.*, 1987). Através das nossas análises ainda encontramos integrações no gene *PIGR*, que sintetiza um receptor da imunoglobulina, e no gene *MMP13* que codifica uma proteína da família metaloproteinase responsável pela quebra da matriz extracelular durante o desenvolvimento embrionário. A integração nos genes *PIGR* e *MMP13* foram encontrados respectivamente na amostra HPV16\_2 na região 1q31 e na amostra HPV18\_4 na região 11q22.3.

Segundo VINOKUROVA *et al.*, 2005, em quase todas amostras tumorais por eles analisadas foram encontrados apenas uma integração por amostra. Em nossas análises, observamos quatro amostras tumorais com duas integrações independentes: HPV16\_5, HPV18\_7, HPV18\_9 e HPV18\_10. As duas integrações independentes na amostra HPV16\_5 ocorreram em regiões intergênicas localizadas nos sítios 13q22.1 e Xq21.31 (tabela 5.5). Integração na região 13q22.1 também foi observada por XU *et al.*, 2013, no qual a integração ocorreu dentro ou na região intergênica próximo do gene *KLF5*. Nas amostras HPV18\_7, HPV18\_9 e HPV18\_10 observamos que as duas integrações independentes sofridas por elas ocorreram próximas, em apenas uma região cromossômica. Segundo WENTZENSEN;

VINOKUROVA; VON KNEBEL DOEBERITZ em 2004, diversas integrações na mesma região cromossômica em um tumor é um evento raro, porém esse acúmulo de integrações na mesma região foi encontrado no *locus* do gene *MYC*. Na amostra HPV18\_7 as duas integrações ocorreram na região intergênica da banda 14q23.2, integração simples nessa mesma região foi observada por XU *et al.*, 2013, porém ocorrendo dentro do gene *HIF1A*. As duas integrações na amostra HPV18\_9 ocorreram na região 14q23, no gene *RAD51B*. Já foi descrito na literatura recorrência de integração do genoma do HPV no gene *RAD51B* em diferentes tumores infectados com HPV 16, HPV 18 e HPV 52 (OJESINA *et al.*, 2014). Esse gene atua no reparo de DNA e, já foi associado ao desenvolvimento tumoral de diversos tipos de cânceres (THOMAS *et al.*, 2009). A amostra HPV18\_10 apresentou as duas integrações ocorrendo na região cromossômica 2p16.3, no gene *NRXN1*, responsável por gerar uma importante proteína que promove a adesão celular no sistema nervoso de vertebrados.

Para as amostras HPV16\_3, HPV16\_9, HPV16\_12, HPV16\_14, HPV18\_1 e HPV18\_6 a integração ocorreu na região intergênica das bandas cromossômicas 4q31.1, 11q22.2, 9p11.2, 2q31.2, 15q21 e 20q13.12 respectivamente (tabelas 5.5, 5.6 e 5.7). Na amostra co-infectada HPV16/18\_1, a integração do HPV 16 também ocorreu em região intergênica, porém em 5q33.2. De acordo com o nosso conhecimento, integrações do genoma do HPV nessas regiões cromossômicas ainda não foram descritas na literatura.

A cobertura das regiões em que ocorreram as integrações foi obtida através do número de leituras (LJI) que cobrem a região de junção de integração nos *contigs* formados para os transcritos quiméricos. Esse valor de cobertura é importante para determinar se a integração encontrada de fato existe ou se trata de um artefato formado pela metodologia empregada. De acordo com a metodologia para análise de integração através de RNA-Seq utilizada por OJESINA *et al.*, 2014, um sítio de integração só foi definido se ele apresentar pelo menos seis leituras cobrindo a região de junção de integração. Em nossa análise, consideramos uma LJI apenas as leituras que tivessem um *spanning* de pelo menos de 6 bases, aumentando assim a chance dessa leitura estar cobrindo a região de integração e não uma região aleatória do genoma. Sendo assim, apenas a integração encontrada para a amostra HPV16\_12 obteve menos de seis LJI cobrindo a região da integração, apresentando cobertura de apenas duas LJI. Além da hipótese de que a integração dessa amostra

pode ser um artefato, acreditamos que a baixa cobertura da região de integração do seu *contig* formado, também pode estar ligada a baixa expressão dessa região, uma vez que essa integração ocorreu em uma região intergênica.

Para confirmarmos os sítios de integrações encontrados em nossas análises, precisamos validar os resultados encontrados utilizando metodologias moleculares como, por exemplo, APOT, DIPS, PCR de restrição e PCR convencional a partir do DNA tumoral.

## **6.2- Padrão de expressão gênica viral e integração do genoma**

A expressão do genoma do HPV nas amostras tumorais foi estimada através da média de cobertura das bases pertencentes a cada gene. Sendo assim, quanto maior for a expressão de determinado gene, maior será a sua média de cobertura. Como não fizemos uma padronização da expressão do genoma do HPV com a expressão de algum gene constitutivo celular, não foi possível comparar a expressão do HPV entre os diferentes tumores, foi possível apenas fazer uma comparação relativa entre as expressões de cada gene viral dentro do mesmo tumor. Ainda, esse padrão de expressão gênica viral está associado a presença de cópias episomais e integradas e, principalmente, está relacionada à ação inibitória promovida pela proteína E2.

O processo de como ocorre a expressão do genoma do HPV ainda não é bem conhecido, porém se sabe que a expressão desse genoma ocorre de forma diferenciada dependendo do estágio da infecção e do estado físico do genoma, se ele se encontra na forma episomal ou integrada (NGUYEN; RAMÍREZ-FORT; RADY, 2014; WOODMAN; COLLINS; YOUNG, 2007). Na análise das diferentes amostras de câncer cervical, encontramos três padrões na expressão do genoma do HPV. O padrão em que se encontra a maioria das amostras (10 amostras), observamos maior expressão dos oncogenes *E6/E7*. O segundo padrão mais frequente (8 amostras), também observamos uma maior expressão dos oncogenes *E6/E7* porém comitadamente é observado alta expressão dos genes *E2/E4*. O terceiro padrão (5 amostras) observamos amostras em que apresentam uma maior expressão dos genes *E2/E4* e *E5* em relação a expressão dos oncogenes *E6/E7*. Ainda observamos que duas amostras, juntamente com o genoma do HPV 18 na

amostra com co-infecção, apresentaram baixa expressão do genoma do HPV (tabelas 5.11, 5.12 e 5.13; anexos 5.3, 5.4 e 5.5).

Observamos que as amostras HPV16\_2, HPV16\_3, HPV16\_4, HPV16\_7, HPV16\_9, HPV16\_14, HPV18\_2, HPV18\_4, HPV18\_5 e HPV18\_6, apresentaram um padrão de expressão no qual os oncogenes *E6/E7* estão mais expressos em relação aos demais genes virais. Exceto para as amostras HPV16\_2, HPV16\_7 e HPV18\_4, a quebra do genoma viral para a integração ocorreu nos genes *E1* ou *E2*. Como a quebra em *E2* impede a formação da proteína *E2* e a quebra em *E1* impede a transcrição de *E2* pelo promotor precoce, os genes *E6/E7* são super-expressos uma vez que não há a ação de *E2* interferindo na transcrição desses oncogenes (JEON; ALLEN-HOFFMANN; LAMBERT, 1995; JEON; LAMBERT, 1995). As amostras HPV16\_2 e HPV18\_4 identificamos o rompimento para a integração dentro dos genes *E6* e *E7* respectivamente, porém mesmo assim elas apresentaram uma alta expressão dos oncogenes *E6/E7*. Acreditamos que nessa amostra essa alta expressão se deve a presença de outras cópias do genoma viral (integrado ou não ao genoma da célula tumoral). Para a amostra HPV16\_7 a metodologia utilizada não mostrou presença de integração, mas não há como afirmar que não existam cópias do genoma viral integrado. Análise da integração através de outras metodologias moleculares pode ser capaz de nos responder essas questões.

Alta expressão dos genes *E6/E7* seguido de altas expressões dos genes *E2*, *E4* e *E5* foi o padrão encontrado para as amostras HPV16\_1, HPV16\_5, HPV18\_1, HPV18\_3, HPV18\_7, HPV18\_8, HPV18\_9 e HPV18\_10. Exceto para as amostras HPV18\_3, a ruptura do genoma viral para a integração, ocorreu nos genes *E1* e/ou *E2*, sendo esse o provável motivo para a alta expressão dos oncogenes *E6/E7*. Na amostra HPV18\_3 não encontramos integração do genoma viral. As altas expressões dos genes *E2*, *E4* e *E5* nos leva a supor que provavelmente esses tumores cervicais possuem o genoma viral tanto na forma episomal quanto na forma integrada, conforme já relatado por CRICCA *et al.*, 2007 e TONON *et al.*, 2001. Em destaque, a amostra HPV18\_9 além da alta expressão dos oncogenes *E6/E7* e dos genes *E2*, *E4* e *E5* apresentou expressão dos genes tardios *L1* e *L2*. Segundo SCHWARTZ (2008), em HPV16 os genes tardios em células de câncer cervical nunca se expressam. Como a amostra HPV18\_9 apresenta infecção por HPV 18, acreditamos que provavelmente essa hipótese de SCHWARTZ não pode ser seguida para os tumores com HPV 18.

As amostras HPV16\_6, HPV16\_10, HPV16\_12, HPV16\_13 e a co-infecção HPV16/18\_1 (genoma do HPV 16) apresentaram uma maior expressão dos genes *E2/E4* ou *E5*. Exceto para as amostras HPV16\_12 e o genoma do HPV 16 da amostra HPV16/18\_1, não foram encontradas integrações para as demais amostras, o que provavelmente explica a menor expressão dos oncogenes virais *E6/E7*. Na amostra HPV16\_12 a quebra para a integração ocorreu na região intergênica entre os genes *E5* e *L2*, com isso a ruptura nessa região não interfere na expressão do gene *E2* a partir do promotor precoce, permitindo assim que a proteína *E2* atue reprimindo a super-expressão dos oncogenes *E6/E7*. Como a amostra co-infectada (HPV16/18\_1) não apresentou uma alta expressão dos oncogenes *E6/E7* de HPV 16, mas sim dos genes *E2/E4* e *E5*, acreditamos que essa amostra possui genomas virais na forma epissomal e integrada.

Baixa expressão do genoma do HPV foram observados para as amostras HPV16\_8, HPV16\_11 e para o genoma do HPV 18 na amostra co-infectada HPV16/18\_1. Apesar da baixa expressão desses genomas, nessas três amostras os oncogenes *E6/E7* foram os que apresentaram a maior expressão. Acreditamos que essa baixa expressão está relacionada ao baixo número de cópias do genoma viral presente nas células, uma vez que quanto maior o número de cópias do genoma viral maior é a expressão dos genes virais (SHULZHENKO *et al.*, 2014). Outra possibilidade para explicar essa baixa expressão, é que o fragmento da biópsia analisada poderia ser composto por células tumorais e células normais, que não apresentam o HPV. Há alguns indícios de que determinados genótipos virais presente em uma co-infecção tendem a dominar na expressão dos seus genes, inibindo assim a expressão dos demais vírus presentes na infecção (ANDERSSON *et al.*, 2012). Dessa forma, com relação a amostra HPV16/18\_1, como ela apresentou alta expressão do genoma do HPV 16 e baixa expressão do HPV 18, acreditamos que pode estar ocorrendo uma competição na expressão gênica desses dois tipos virais.

### **6.3- Transcritos alternativos *E6/E7***

Como o íntron 1 do pré-RNA *E6/E7* do HPV 16 possui três sítios de *splicing* doadores (SD191, SD221 e SD226) e três sítios de *splicing* aceptores (SA409, SA526 e SA742) (AJIRO *et al.*, 2012), potencialmente o processamento pelo *splicing*

alternativo é capaz de formar 9 tipos de transcritos alternativos  $E6/E7$ , sem contar o transcrito  $E6+E7$ , que é formado quando não ocorre *splicing* do íntron 1. Na literatura já foram descritos a existência de 5 desses transcritos:  $E6^*I$  ( $226^{409}$ ),  $E6^*II$  ( $226^{526}$ ),  $E6^*X$  ( $226^{742}$ ),  $E6^*V$  ( $221^{409}$ ) e  $E6^*VI$  ( $191^{409}$ ), sendo esses dois últimos encontrados apenas por AJIRO (2012). Os outros quatro transcritos alternativos, que apresentam as junções de *splicing*  $161^{526}$ ,  $191^{742}$ ,  $221^{526}$  e  $221^{742}$ , são transcritos hipotéticos. De acordo com os nossos conhecimentos nenhum trabalho na literatura descreveu a hipótese da formação desses quatro transcritos nem observaram a sua existência. Já o íntron 1 do pré-RNA  $E6/E7$  do HPV 18 apresenta apenas um sítio doador e um sítio aceptor, formando os transcritos alternativos  $E6^*I$  ( $233^{416}$ ) e  $E6+E7$  (ZHENG *et al.*, 2004).

A expressão dos diferentes transcritos  $E6/E7$  nas amostras tumorais foram estimados através do número de leituras de junção de *splicing* (LJS) para os transcritos originados por *splicing* alternativo e pela média de leituras *unspliced* (MLU) para o transcrito alternativo que não sofre *splicing* ( $E6+E7$ ). Da mesma forma que na análise da expressão do genoma viral, não fizemos a quantificação relativa da expressão desses transcritos alternativos com a expressão de algum gene constitutivo das células, logo não podemos comparar os níveis de cada transcrito entre as amostras, apenas determinar quais tipos de transcritos são os mais e os menos expressos em determinada amostra tumoral.

Observamos que em todas as amostras infectadas pelo HPV 16, o transcrito  $E6^*I$  foi o tipo de transcrito alternativo  $E6/E7$  que obteve a maior expressão, sendo responsável, em média, por 89% dos transcritos  $E6/E7$ . Vários trabalhos já demonstraram que *splicing* do íntron 1 do pré-RNA  $E6/E7$  é altamente eficiente, e a maioria dos transcritos em tecidos e linhagens celulares de câncer cervical é o  $E6^*I$  (SHERMAN *et al.*, 1992; TANG *et al.*, 2006; ZHENG *et al.*, 2004). Ainda observamos que o segundo e o terceiro transcrito  $E6/E7$  com maior expressão são os transcritos  $E6^*II$  e  $E6+E7$  respectivamente, exceto nas amostras HPV16\_8 e HPV16\_11 que expressaram apenas o  $E6^*I$ . Segundo FORSLUND (1997) o nível da expressão do transcrito  $E6^*II$  em relação ao transcrito  $E6^*I$  é reduzido, e SCHMITT e PAWLITA (2011) verificaram que em linhagens celulares a expressão de  $E6^*II$  é dez vezes menor que de  $E6^*I$ . Os nossos resultados estão de acordo com esse achado, uma vez que nas amostras analisadas encontramos, em média, 610 LJS para o transcrito  $E6^*I$  e 57 LJS para o transcrito  $E6^*II$ , sendo  $E6^*I$  10,7 vezes mais expresso que o



*E6\*II*. Os transcritos *E6\*X* e *E6\*VI* não foram encontradas em todas as amostras analisadas, e quando encontradas apresentava uma baixa expressão. Já o transcrito *E6\*V* e os transcritos hipotéticos com junção de *splicing* 191<sup>^</sup>526, 191<sup>^</sup>742, 221<sup>^</sup>526 e 221<sup>^</sup>742 não foram encontrados em nenhuma amostra, o que indica que não são realmente formados, ou então, eles são expressos em níveis tão baixos que a metodologia que utilizamos não foi capaz de detectá-los.

Nas amostras infectadas com HPV 18, o transcrito *E6\*I* também foi o transcrito mais expresso, representando em média 93% dos transcritos alternativos *E6/E7* total. Acreditamos que a maior expressão desse transcrito também seja devido à alta eficiência do *splicing* do íntron 1 do pré-RNA *E6/E7*. O genoma do HPV 16 da amostra co-infectada HPV16/18\_1, assim como nas demais amostras infectadas com HPV 16, os transcritos *E6/E7* mais expressos foram *E6\*I* > *E6\*II* > *E6+E7*, já os transcritos *E6\*X* e *E6\*V* apresentaram baixas expressões. O genoma do HPV 18 na amostra co-infectada apresentou apenas o transcrito *E6\*I*, mas essa exclusividade de *E6\*I* pode ser devido ao baixo nível da expressão do genoma viral.

Experimentos já mostraram que de fato *splicing* do íntron 1 do pré-RNA *E6/E7* é fundamental para que promover a tradução da oncoproteína E7, por outro lado, retenção desse íntron é necessário para formar a oncoproteína E6. Porém, o mecanismo pelo qual o pré-RNA não sofre *splicing* do íntron 1 para a formação da proteína E6 não é conhecido (TANG *et al.*, 2006). Existem indícios que o fator de crescimento epitelial (EGF) pode estar ligado a essa regulação, uma vez que EGF se liga ao transcrito *E6/E7* e impede que ele sofra *splicing*. Ainda acredita-se que HPVs oncogênicos conseguem promover a retenção do íntron 1 utilizando suas próprias proteínas precoces e tardias para interferir com o spliceossomo (AJIRO *et al.*, 2012; ROSENBERGER *et al.*, 2010).

Diversas outras questões ligadas ao processo de formação dos transcritos alternativos *E6/E7* e o efeito de seus produtos, principalmente os polipeptídeos *E6\**, não estão bem esclarecidos. Alguns autores acreditam que no caso do HPV 16, o controle de qual sítio de *splicing* alternativo será escolhido para a eliminação do íntron 1 está relacionado com a presença das sequências ativadoras e silenciadoras de *splicing* (ESSE, ESS, ISE e ISS) (SCHWARTZ, 2008; ZHENG, 2010). Também existem evidências de que os sítios de *splicing* doadores e aceptores escolhidos para a retirada do íntron 1 é determinado pelo princípio da proximidade, onde o

menor fragmento do íntron será eliminado, ou seja, os SD e SA mais próximos uns dos outros serão os escolhidos (AJIRO *et al.*, 2012; REED; MANIATIS, 1986). Essa teoria está de acordo com os nossos achados, pois de fato o transcrito *E6\*1* no HPV 16 é o transcrito alternativo mais abundante, e para a sua formação é necessário a utilização dos SD e SA mais próximos. Já é conhecido que os transcritos *E6/E7* após a perda do íntron 1 podem gerar um ou diversos transcritos *E6\**, que são importantes para traduzir a oncoproteína E7 e também traduzem polipeptídeos *E6\**. Na literatura não há uma definição a respeito da função desses polipeptídeos, se eles são degradados ou se apresentam alguma função. Estudos realizados em 2014 demonstraram que o polipeptídeo *E6\*1* do HPV 16 foi capaz de reduzir o tamanho de tumores estimulados em camundongos. Os autores acreditam que essa redução se deve a capacidade do polipeptídeo *E6\*1* de se ligar a proteína E6 e impedir que ela atue em p53 e em outras proteínas (FILIPPOVA *et al.*, 2014). Se os demais polipeptídeos *E6\** são traduzidos e apresentam uma função semelhante ao *E6\*1*, ainda não é conhecido.

## 7- Conclusões

- Foi possível a identificação da integração do genoma viral no genoma do hospedeiro através da busca de *contigs* quiméricos apresentando sequências do HPV e do genoma humano. Por essa metodologia observamos que a integração do genoma viral em amostras de câncer cervical foi frequente, acometendo 57% das amostras infectadas pelo HPV 16 e 90% das infectadas pelo HPV 18. Já para amostra co-infectada observamos integração apenas do HPV 16.
- O rompimento do genoma viral para ocorrer a integração aconteceu mais frequentemente nos genes *E1* ou *E2*, o que acaba afetando a expressão da proteína *E2*.
- Observamos uma ampla distribuição das integrações em vários sítios cromossômicos, não identificamos a preferência de alguma região cromossômica para que ocorra a integração. Cerca de 52,6% dos sítios de integração que analisamos estavam localizados dentro de regiões intergênicas, os demais encontramos dentro dos genes *MCL1*, *PIGR*, *PXMP4*, *TP63*, *MMP13*, *SLC16A14*, *HRH1*, *RAD51B* e *NRXN1*.
- Tanto para o HPV 16 e HPV 18 encontramos três padrões na expressão de mRNAs do genoma do HPV: (1) Maior expressão dos oncogenes *E6/E7*; (2) Expressão equivalente dos oncogenes *E6/E7* e dos genes *E2/E4*; e (3) Maior expressão dos genes *E2/E4* e *E5*.
- Na amostra com múltipla infecção, observamos expressão preferencial do HPV16, sugerindo que esse vírus pode inibir a expressão do outro.
- Os transcritos alternativos *E6/E7* mais expressos nas amostras infectadas pelo HPV 16 foram  $E6^*I > E6^*II > E6+E7$ . O transcrito *E6^\*I* foi o transcrito alternativo mais comum nas amostras tumorais, representando em média 89% dos transcritos. Ainda observamos que a expressão do transcrito *E6^\*II* é aproximadamente 10 vezes menor do que o transcrito *E6^\*I*.
- Os transcritos alternativos *E6^\*X*, *E6^\*VI* e *E6^\*V* do HPV 16 aparentemente são transcritos raros, não sendo encontradas em todas amostras tumorais. Já os

transcritos hipotéticos com as junções de *splicing* 191<sup>^</sup>526, 191<sup>^</sup>742, 221<sup>^</sup>526 e 221<sup>^</sup>742 não foram encontrados em nenhuma amostra.

- O transcrito alternativo *E6/E7* mais expresso nas amostras infectadas com o HPV 18 foi o E6<sup>\*</sup>I, representando em média 93% desses transcritos.

## 8- Referências bibliográficas

AJIRO, M. *et al.* Intron definition and a branch site adenosine at nt 385 control RNA splicing of HPV16 E6\*1 and E7 expression. **PloS One**, v. 7, n. 10, p. e46412, 2012.

ANDERSSON, E. *et al.* Type-dependent E6/E7 mRNA expression of single and multiple high-risk human papillomavirus infections in cervical neoplasia. **Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology**, v. 54, n. 1, p. 61–65, maio 2012.

BERNARD, H.-U. Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches. **Antiviral Therapy**, v. 7, n. 4, p. 219–237, dez. 2002.

BEUTNER, K. R.; TYRING, S. Human papillomavirus and human disease. **The American Journal of Medicine**, v. 102, n. 5A, p. 9–15, 5 maio 1997.

BODILY, J.; LAIMINS, L. A. Persistence of human papillomavirus infection: keys to malignant progression. **Trends in Microbiology**, v. 19, n. 1, p. 33–39, jan. 2011.

BRYAN, J. T.; BROWN, D. R. Association of the human papillomavirus type 11 E1()E4 protein with cornified cell envelopes derived from infected genital epithelium. **Virology**, v. 277, n. 2, p. 262–269, 25 nov. 2000.

CHENG, S. *et al.* Differentiation-dependent up-regulation of the human papillomavirus E7 gene reactivates cellular DNA replication in suprabasal differentiated keratinocytes. **Genes & Development**, v. 9, n. 19, p. 2335–2349, 1 out. 1995.

CHEN, M.; MANLEY, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. **Nature Reviews. Molecular Cell Biology**, v. 10, n. 11, p. 741–754, nov. 2009.

CRICCA, M. *et al.* Viral DNA load, physical status and E2/E6 ratio as markers to grade HPV16 positive women for high-grade cervical lesions. **Gynecologic Oncology**, v. 106, n. 3, p. 549–557, set. 2007.

DALL, K. L. *et al.* Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. **Cancer Research**, v. 68, n. 20, p. 8249–8259, 15 out. 2008.

DAVY, C.; DOORBAR, J. G2/M cell cycle arrest in the life cycle of viruses. **Virology**, v. 368, n. 2, p. 219–226, 25 nov. 2007.

DELLA TORRE, G. *et al.* Viral particles in cervical condylomatous lesions. **Tumori**, v. 64, n. 5, p. 549–553, 31 out. 1978.

DE SANJOSÉ, S. *et al.* Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. **The Lancet infectious diseases**, v. 7, n. 7, p. 453–459, jul. 2007.

- DE VILLIERS, E.-M. *et al.* Classification of papillomaviruses. **Virology**, v. 324, n. 1, p. 17–27, 20 jun. 2004.
- DOORBAR, J. *et al.* Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. **Virology**, v. 178, n. 1, p. 254–262, set. 1990.
- DOORBAR, J. The papillomavirus life cycle. **Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology**, v. 32 Suppl 1, p. S7–15, mar. 2005.
- DOORBAR, J. The E4 protein; structure, function and patterns of expression. **Virology**, v. 445, n. 1-2, p. 80–98, out. 2013.
- DÜRST, M. *et al.* Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. **Proceedings of the National Academy of Sciences of the United States of America**, v. 84, n. 4, p. 1070–1074, fev. 1987.
- FERBER, M. J. *et al.* Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. **Oncogene**, v. 22, n. 46, p. 7233–7242, 16 out. 2003.
- FILIPPOVA, M. *et al.* The small splice variant of HPV16 E6, E6, reduces tumor formation in cervical carcinoma xenografts. **Virology**, v. 450-451, p. 153–164, fev. 2014.
- FOROUZANFAR, M. H. *et al.* Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. **Lancet**, v. 378, n. 9801, p. 1461–1484, 22 out. 2011.
- FORSLUND, O. *et al.* HPV 16 DNA and mRNA in cervical brush samples quantified by PCR and microwell hybridization. **Journal of Virological Methods**, v. 69, n. 1-2, p. 209–222, dez. 1997.
- GNANAMONY, M.; PEEDICAYIL, A.; ABRAHAM, P. An overview of human papillomaviruses and current vaccine strategies. **Indian Journal of Medical Microbiology**, v. 25, n. 1, p. 10–17, jan. 2007.
- GRABHERR, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature biotechnology**, v. 29, n. 7, p. 644–652, 15 maio 2011.
- GRASSMANN, K. *et al.* Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. **Journal of Virology**, v. 70, n. 4, p. 2339–2349, abr. 1996.
- HÄFNER, N. *et al.* Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. **Oncogene**, v. 27, n. 11, p. 1610–1617, 6 mar. 2008.
- HAM, J. *et al.* The papillomavirus E2 protein: a factor with many talents. **Trends in biochemical sciences**, v. 16, n. 11, p. 440–444, nov. 1991.

HEBNER, C. M.; LAIMINS, L. A. Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity. **Reviews in Medical Virology**, v. 16, n. 2, p. 83–97, abr. 2006.

HUMMEL, M.; HUDSON, J. B.; LAIMINS, L. A. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. **Journal of Virology**, v. 66, n. 10, p. 6070–6080, out. 1992.

IARC. **IARC monographs on the Evaluation of carcinogenic risks to humans - Volume 90 - Human Papillomaviruses** World Health Organization - International Agency for Research on Cancer, , 2007. Disponível em: <<http://monographs.iarc.fr/ENG/Monographs/vol90/>>

ILLUMINA. **CASAVA v1.8.2 User Guide**, dez. 2011.

INCA. **Estimativa | 2014 Incidência de Câncer no Brasil**, 2014. Disponível em: <<http://www.inca.gov.br/estimativa/2014>>

JACOBS, M. V. *et al.* Group-specific differentiation between high- and low-risk human papillomavirus genotypes by general primer-mediated PCR and two cocktails of oligonucleotide probes. **Journal of Clinical Microbiology**, v. 33, n. 4, p. 901–905, abr. 1995.

JEON, S.; ALLEN-HOFFMANN, B. L.; LAMBERT, P. F. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. **Journal of Virology**, v. 69, n. 5, p. 2989–2997, maio 1995.

JEON, S.; LAMBERT, P. F. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 5, p. 1654–1658, 28 fev. 1995.

JOHANSSON, C.; SCHWARTZ, S. Regulation of human papillomavirus gene expression by splicing and polyadenylation. **Nature Reviews. Microbiology**, v. 11, n. 4, p. 239–251, abr. 2013.

KLAES, R. *et al.* Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. **Cancer Research**, v. 59, n. 24, p. 6132–6136, 15 dez. 1999.

KORNBLIHTT, A. R. *et al.* Alternative splicing: a pivotal step between eukaryotic transcription and translation. **Nature Reviews. Molecular Cell Biology**, v. 14, n. 3, p. 153–165, mar. 2013.

KOZAK, M. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. **Molecular and Cellular Biology**, v. 7, n. 10, p. 3438–3445, out. 1987.

KRAUS, I. *et al.* The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. **Cancer research**, v. 68, n. 7, p. 2514–2522, 1 abr. 2008.

- LADD, A. N.; COOPER, T. A. Finding signals that regulate alternative splicing in the post-genomic era. **Genome Biology**, v. 3, n. 11, p. reviews0008.1–reviews0008.16, 2002.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, abr. 2012.
- LAVERTY, C. R. *et al.* Noncondylomatous wart virus infection of the postmenopausal cervix. **Pathology**, v. 10, n. 4, p. 373–378, out. 1978.
- LIE, A. K.; KRISTENSEN, G. Human papillomavirus E6/E7 mRNA testing as a predictive marker for cervical carcinoma. **Expert Review of Molecular Diagnostics**, v. 8, n. 4, p. 405–415, jul. 2008.
- LI, H. *et al.* The Sequence Alignment/Map format and SAMtools. **Bioinformatics (Oxford, England)**, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.
- LONG, J. C.; CACERES, J. F. The SR protein family of splicing factors: master regulators of gene expression. **The Biochemical Journal**, v. 417, n. 1, p. 15–27, 1 jan. 2009.
- LUFT, F. *et al.* Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. **International Journal of Cancer. Journal International Du Cancer**, v. 92, n. 1, p. 9–17, 1 abr. 2001.
- MARTIN, C. M.; O'LEARY, J. J. Histology of cervical intraepithelial neoplasia and the role of biomarkers. **Best Practice & Research. Clinical Obstetrics & Gynaecology**, v. 25, n. 5, p. 605–615, out. 2011.
- MCKENNA, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, v. 20, n. 9, p. 1297–1303, set. 2010.
- MCLAUGHLIN-DRUBIN, M. E.; MEYERS, J.; MUNGER, K. Cancer associated human papillomaviruses. **Current opinion in virology**, 1 jun. 2012.
- MINCHEVA, A.; GISSMANN, L.; ZUR HAUSEN, H. Chromosomal integration sites of human papillomavirus DNA in three cervical cancer cell lines mapped by in situ hybridization. **Medical Microbiology and Immunology**, v. 176, n. 5, p. 245–256, 1987.
- MORSHED, K. *et al.* Human Papillomavirus (HPV) - structure, epidemiology and pathogenesis. **Otolaryngologia Polska. The Polish Otolaryngology**, v. 68, n. 5, p. 213–219, out. 2014.
- MÜNGER, K. *et al.* Mechanisms of human papillomavirus-induced oncogenesis. **Journal of virology**, v. 78, n. 21, p. 11451–11460, nov. 2004.
- MUÑOZ, N. *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. **The New England journal of medicine**, v. 348, n. 6, p. 518–527, 6 fev. 2003.



NAKAGAWA, M. *et al.* Persistence of human papillomavirus type 16 infection is associated with lack of cytotoxic T lymphocyte response to the E6 antigens. **The Journal of Infectious Diseases**, v. 182, n. 2, p. 595–598, ago. 2000.

NAMBARU, L. *et al.* Prognostic significance of HPV physical status and integration sites in cervical cancer. **Asian Pacific journal of cancer prevention: APJCP**, v. 10, n. 3, p. 355–360, set. 2009.

NGUYEN, H. P.; RAMÍREZ-FORT, M. K.; RADY, P. L. The biology of human papillomaviruses. **Current Problems in Dermatology**, v. 45, p. 19–32, 2014.

OJESINA, A. I. *et al.* Landscape of genomic alterations in cervical carcinomas. **Nature**, v. 506, n. 7488, p. 371–375, 20 fev. 2014.

PETER, M. *et al.* Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. **The Journal of pathology**, v. 221, n. 3, p. 320–330, jul. 2010.

PETT, M.; COLEMAN, N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? **The Journal of pathology**, v. 212, n. 4, p. 356–367, ago. 2007.

PIM, D.; BANKS, L. HPV-18 E6\*1 protein modulates the E6-directed degradation of p53 by binding to full-length HPV-18 E6. **Oncogene**, v. 18, n. 52, p. 7403–7408, 9 dez. 1999.

REED, R.; MANIATIS, T. A role for exon sequences and splice-site proximity in splice-site selection. **Cell**, v. 46, n. 5, p. 681–690, 29 ago. 1986.

ROSENBERGER, S. *et al.* Alternative splicing of human papillomavirus type-16 E6/E6\* early mRNA is coupled to EGF signaling via Erk1/2 activation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 15, p. 7006–7011, 13 abr. 2010.

RUSH, M.; ZHAO, X.; SCHWARTZ, S. A splicing enhancer in the E4 coding region of human papillomavirus type 16 is required for early mRNA splicing and polyadenylation as well as inhibition of premature late gene expression. **Journal of Virology**, v. 79, n. 18, p. 12002–12015, set. 2005.

SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics (Oxford, England)**, v. 27, n. 6, p. 863–864, 15 mar. 2011.

SCHMITT, M.; PAWLITA, M. The HPV transcriptome in HPV16 positive cell lines. **Molecular and Cellular Probes**, v. 25, n. 2-3, p. 108–113, jun. 2011.

SCHMITZ, M. *et al.* Loss of gene function as a consequence of human papillomavirus DNA integration. **International Journal of Cancer. Journal International Du Cancer**, v. 131, n. 5, p. E593–602, 1 set. 2012.

SCHWARTZ, S. HPV-16 RNA processing. **Frontiers in Bioscience: A Journal and Virtual Library**, v. 13, p. 5880–5891, 2008.

SEDMAN, S. A. *et al.* The full-length E6 protein of human papillomavirus type 16 has transforming and trans-activating activities and cooperates with E7 to immortalize keratinocytes in culture. **Journal of Virology**, v. 65, n. 9, p. 4860–4866, set. 1991.

SHERMAN, L. *et al.* 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. **International Journal of Cancer. Journal International Du Cancer**, v. 50, n. 3, p. 356–364, 1 fev. 1992.

SHULZHENKO, N. *et al.* Ménage à trois: an evolutionary interplay between human papillomavirus, a tumor, and a woman. **Trends in Microbiology**, v. 22, n. 6, p. 345–353, jun. 2014.

SMOTKIN, D.; WETTSTEIN, F. O. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. **Proceedings of the National Academy of Sciences of the United States of America**, v. 83, n. 13, p. 4680–4684, jul. 1986.

STEGER, G.; CORBACH, S. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. **Journal of Virology**, v. 71, n. 1, p. 50–58, jan. 1997.

SUTHERLAND, G. R.; RICHARDS, R. I. Fragile sites-cytogenetic similarity with molecular diversity. **American Journal of Human Genetics**, v. 64, n. 2, p. 354–359, fev. 1999.

TANG, S. *et al.* The E7 oncoprotein is translated from spliced E6\*1 transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. **Journal of Virology**, v. 80, n. 9, p. 4249–4263, maio 2006.

THOMAS, G. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). **Nature Genetics**, v. 41, n. 5, p. 579–584, maio 2009.

THORLAND, E. C. *et al.* Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. **Cancer Research**, v. 60, n. 21, p. 5916–5921, 1 nov. 2000.

TONON, S. A. *et al.* Physical status of the E2 human papilloma virus 16 viral gene in cervical preneoplastic and neoplastic lesions. **Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology**, v. 21, n. 2, p. 129–134, maio 2001.

VINOKUROVA, S. *et al.* Clonal history of papillomavirus-induced dysplasia in the female lower genital tract. **Journal of the National Cancer Institute**, v. 97, n. 24, p. 1816–1821, 21 dez. 2005.

WAGNER, G. *et al.* STINGRAY: system for integrated genomic resources and analysis. **BMC research notes**, v. 7, p. 132, 2014.

- WALBOOMERS, J. M. *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. **The Journal of pathology**, v. 189, n. 1, p. 12–19, set. 1999.
- WANG, X. *et al.* Construction of a Full Transcription Map of Human Papillomavirus Type 18 during Productive Viral Infection. **Journal of Virology**, v. 85, n. 16, p. 8080–8092, 15 ago. 2011.
- WENTZENSEN, N. *et al.* Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions. **Oncogene**, v. 21, n. 3, p. 419–426, 17 jan. 2002.
- WENTZENSEN, N.; VINOKUROVA, S.; VON KNEBEL DOEBERITZ, M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. **Cancer Research**, v. 64, n. 11, p. 3878–3884, 1 jun. 2004.
- WOODMAN, C. B. J.; COLLINS, S. I.; YOUNG, L. S. The natural history of cervical HPV infection: unresolved issues. **Nature Reviews. Cancer**, v. 7, n. 1, p. 11–22, jan. 2007.
- XU, B. *et al.* Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. **PLoS ONE**, v. 8, n. 6, 18 jun. 2013.
- ZHENG, Z.-M. *et al.* 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. **Journal of Molecular Biology**, v. 337, n. 5, p. 1091–1108, 9 abr. 2004.
- ZHENG, Z.-M. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. **Journal of Biomedical Science**, v. 11, n. 3, p. 278–294, jun. 2004.
- ZHENG, Z.-M. Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. **International Journal of Biological Sciences**, v. 6, n. 7, p. 730–755, 2010.
- ZHENG, Z.-M.; BAKER, C. C. Papillomavirus genome structure, expression, and post-transcriptional regulation. **Frontiers in Bioscience: A Journal and Virtual Library**, v. 11, p. 2286–2302, 2006.
- ZUR HAUSEN, H. Condylomata acuminata and human genital cancer. **Cancer research**, v. 36, n. 2 pt 2, p. 794, fev. 1976.

## Anexo 4.1

### Registro no Comitê de Ética em Pesquisa do Instituto Nacional do Câncer



A(o): Dra. Liz Maria Almeida  
Pesquisador(a) Principal



Rio de Janeiro, 25 de fevereiro de 2011.

**Registro CEP nº 156/10** (Este nº. deve ser citado nas correspondências referentes a este estudo)

**Título do Estudo:** Estudo dos genótipos de HPV presentes em tumores do colo do útero em mulheres matriculadas no INCA

Prezado(a) Pesquisador(a),

Informo a composição do Comitê de Ética em Pesquisa do Instituto Nacional de Câncer que analisou e aprovou o estudo acima especificado.

*Adriana Scheliga  
Alessandra Oliveira Marins  
Ana Lúcia Amaral Eisenberg  
Ana Valéria Paranhos Miceli  
André Marcelo Machado Soares  
Carlos Henrique D Silva  
Célia Helena Fernandes Costa  
Cláudio Gustavo Stefanoff  
Cristiano Guedes Duque  
Danielle Nogueira Ramos  
Eliana Teixeira Maranhão  
Héilton Spindola Antunes  
Ivan da Costa Marques  
Maria de Fátima Batalha de Menezes  
Michele Quarti  
Paulo Alexandre Mora  
Rosane Vianna Jorge  
Urquilei dos Santos Pinheiro  
Walter Esteves Piñeiro  
Wilza Barbosa Felipe*

*Médica Oncologista Clínica HC I – Coordenadora  
Enfermeira do Serv. de Pesq. Clínica - CGTC  
Médica Patologista – DIPAT  
Psicóloga do Serviço de Psicologia - HC I  
Teólogo / Filósofo Extra INCA  
Médico HC IV – Coordenador Substituto  
Enfermeira da Oncologia – HC I  
Coordenador do Laboratório do BNT - CGTC  
Médico Oncologista - HCI  
Doutora em Epidemiologia e Bioestatística - CONPREV  
Fisioterapeuta do Serviço de Fisioterapia – HC I  
Dentista do Serviço de Pesquisa Clínica - CGTC  
Matemático Extra INCA  
Enfermeira da Divisão de Enfermagem – HC I  
Enfermeira - Extra INCA  
Médico Oncologista - HCII  
Doutora em Química Biológica da Divisão de Farmacologia - CGTC  
Representante de Usuários – CES/RJ  
Advogado Extra INCA  
Enfermeira da Comissão de Infecção Hospitalar - HC III*

Atenciosamente,

Dra. Adriana Scheliga  
Coordenadora do Comitê de Ética em Pesquisa  
CEP-INCA

## **Anexo 4.2**

### Termo de Consentimento Livre e Esclarecido

#### **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

**Projeto: Estudo dos genótipos de HPV presentes em tumores do colo do útero em mulheres matriculadas no INCA**

Nome do Voluntário: \_\_\_\_\_

Você está sendo convidada a participar de um estudo que tem por objetivo descrever as características biológicas do câncer do colo do útero e que envolve a coleta de material para a realização de uma biópsia (análise de pequenas amostras de tecido para o diagnóstico do tipo do tumor) e a coleta de uma pequena quantidade de sangue (5mL ou o equivalente a uma colher de chá).

O Câncer do colo do útero é o segundo tipo de tumor mais comum em mulheres. Sabe-se que sua origem está associada ao um vírus (HPV). Conhecer as características desse vírus e do tumor que ele originou é importante para compreender melhor a doença, o tratamento e as maneiras de prevenir o aparecimento desse tumor, como o desenvolvimento de vacinas.

Para que você possa decidir se quer participar ou não deste estudo, precisa conhecer seus benefícios, riscos e implicações.

#### **OBJETIVO DO ESTUDO**

Este estudo tem como objetivo identificar os diferentes tipos de HPV presentes nos tumores do colo do útero de pacientes atendidas no Instituto Nacional de Câncer e associar as características do tumor com as características do tipo de vírus presente.

#### **PROCEDIMENTOS DO ESTUDO**

Se você concordar em participar deste estudo será coletada uma amostra do tumor para a realização da biópsia (exame que diagnostica o tipo de tumor) ou uma amostra do mesmo no caso de você ser submetida a uma cirurgia, como parte do tratamento. A coleta de material para a biópsia será feita por uma médica(o) pesquisadora(or) participante desse estudo. Também serão coletados 5mL de sangue através de uma punção com agulha de uma veia de seu braço. Essas amostras serão levadas ao laboratório onde será isolado o DNA da biópsia do tumor e do sangue para ser submetido a procedimentos que permitirão identificar o vírus associado ao desenvolvimento do tumor. Você também responderá a um questionário com perguntas sobre

hábitos de vida, atividade sexual e uso de hormônios ao longo de sua vida. O DNA isolado das amostras de tumor e sangue será armazenado no Banco Nacional de Tumores após a utilização do mesmo para os procedimentos relacionados ao presente projeto. O material armazenado no BNT será guardado, sob a responsabilidade do INCA, para estudos futuros e você terá o direito de decidir sobre a participação em outros estudos que eventualmente venham a ser realizados a partir deste material. Se você concordar em participar deste projeto de pesquisa os pesquisadores participantes também consultarão seus registros médicos para obter dados que podem ser importantes para compreender o câncer do colo do útero.

### **MÉTODOS ALTERNATIVOS**

**Sua participação nesse estudo é totalmente voluntária. Você não é obrigada a participar do estudo. Se você não concordar em fornecer uma amostra de sangue, permitir a biópsia de seu tumor e responder ao questionário, você não participará desse estudo.**

### **RISCOS**

O seu tratamento será exatamente o mesmo caso você participe ou não deste estudo. E é importante que você saiba dos riscos implicados na coleta de uma amostra do tumor e na coleta do sangue. Esses dois procedimentos não fazem parte do tratamento usual para pacientes com esse tipo de tumor. A coleta da biópsia de seu tumor pode causar sangramento no local. Caso isso aconteça, esse sangramento será controlado pela médica(o) que realizará a coleta. Se você for submetida a uma cirurgia, como parte do tratamento, a coleta de uma amostra do tumor se dará a partir do material tumoral retirado durante a cirurgia e não haverá risco adicional. A coleta de sangue pode causar dor no local da punção, ou manchas rochas transitórias chamadas de equimoses.

### **BENEFÍCIOS**

Você não terá benefícios com os resultados dessa pesquisa. Entretanto os resultados gerados por ela poderão ajudar a compreender melhor o câncer do colo do útero e auxiliar na prevenção e no tratamento do mesmo, no futuro, para outras mulheres.

## **ACOMPANHAMENTO, ASSISTÊNCIA E RESPONSÁVEIS**

É importante que você saiba que o seu tratamento será o mesmo, caso aceite ou não participar desse projeto. É que você tem direito de saber qualquer informação gerada nessa pesquisa relacionada a sua doença. Essa pesquisa é coordenada pela Dr. Liz Maria de Almeida, e qualquer dúvida que tenha respeito pode ser sanada entrando em contato pelo telefone 21 3970-7409.

## **CARÁTER CONFIDENCIAL DOS REGISTROS**

Além da equipe de saúde que cuidará de você, seus registros médicos poderão ser consultados pelo Comitê de Ética em Pesquisa do Instituto Nacional de Câncer (CEP-INCA) e equipe de pesquisadores envolvidos nesse estudo. Seu nome não será revelado ainda que informações de seu registro médico sejam utilizadas para propósitos educativos ou de publicação, que ocorrerão independentemente dos resultados obtidos.

## **TRATAMENTO MÉDICO EM CASO DE DANOS**

Todo e qualquer dano decorrente do desenvolvimento deste projeto de pesquisa e que necessite de atendimento médico, ficará a cargo da instituição. Seu tratamento e acompanhamento médico independem de sua participação neste estudo.

## **CUSTOS**

Não haverá qualquer custo ou forma de pagamento para você pela sua participação nesse estudo.

## **BASES DA PARTICIPAÇÃO**

É importante que você saiba que a sua participação neste estudo é completamente voluntária e que você pode recusar-se a participar ou interromper sua participação a qualquer momento, sem penalidades ou perda de benefícios aos quais você tem direito. Caso você decida interromper sua participação no estudo, a equipe de pesquisadores deve ser comunicada e a coleta e o uso das amostras para os fins relativos ao estudo será imediatamente interrompida.





### Anexo 4.3

Script extractsequence.pl, obtido de:

<http://seqanswers.com/forums/showthread.php?t=9498>

```
#!/usr/bin/env perl
use strict;
use warnings;
use Bio::Perl;
use Getopt::Long;
use Data::Dumper;
use File::Basename;
use List::Util qw/min max/;
use Bio::SeqFeature::Generic;
my $settings={
contig=>[],
};
exit(main());
sub main{
die usage() if(@ARGV<1);
GetOptions($settings,('infile=s', 'start=s', 'nameOfOrganism=s', 'end=s',
'contig=s@', 'outfile=s', 'help'));
die usage() if($$settings{help});
my $file=$$settings{infile} || die "Missing infile\n".usage();
my $end||=$$settings{end};
my $outfile=$$settings{outfile} or die "Missing outfile\n".usage();
my $start=$$settings{start} || 1;
$start=1 if($start<1);
my $seq=extractSeqs($file,$start,$settings);
my $seqout=Bio::SeqIO->new(-file=>">$outfile");
$seqout->write_seq($_) for(@$seq);
print "Output file is in $outfile\n";
return 1;
}
sub extractSeqs{
```

```

my($file,$start,$settings)=@_;
my @return;
# get the correct contig Seq
my $targetSeq=[];
my $revcom=[];
if($$settings{contig}){
my %seqHash=seqHash($file);
for my $contigId(@{ $$settings{contig} }){
if($contigId=~/\^(.+)$/){
$contigId=$1;
push(@$revcom,1);
} else {push(@$revcom,0);}
die "Could not find contig $contigId in $file\n" if(!$seqHash{$contigId});
push(@$targetSeq,$seqHash{$contigId});
}
} else {
my $seqin=Bio::SeqIO->new(-file=>$file);
$targetSeq=[$seqin->next_seq];
$$settings{contig}=[map($_->id,@$targetSeq)];
}
# actually perform extraction
for(my $i=0;$i<@$targetSeq;$i++){
my $seq=$$targetSeq[$i];
my $rc=$$revcom[$i];
my $tmp=extractSeq($seq,$start,$rc,$settings);
push(@return,$tmp);
}
return \@return;
}
sub extractSeq{
my($targetSeq,$start,$revcom,$settings)=@_;
my $end=$$settings{end}||$targetSeq->length;
my @feat=$targetSeq->get_SeqFeatures();
$start=1 if($start<1);
$end=$targetSeq->length if($end>$targetSeq->length);

```

```

print "Extracting and highlighting features in contig ".$targetSeq->id." from
$start to $end\n";
my $truncSeq=$targetSeq->trunc($start,$end);
$truncSeq=$truncSeq->revcom if($revcom);
my $seqLength=$end-$start+1;
my $organism=($targetSeq->species)?
join(" ",$targetSeq->species->binomial):
join(" ","StrainXYZ");
$organism=$$settings{nameOfOrganism} if($$settings{nameOfOrganism});
# add in the source feature
$truncSeq->add_SeqFeature(
Bio::SeqFeature::Generic->new(
-start=>1,
-end=>$seqLength,
-primary=>'source',
-tag=>{
organism=>$organism,
},
)
);
# sort the features appropriately
@feat=sort {
return $b->location->start <=> $a->location->start if $revcom;
return $a->location->start <=> $b->location->start;
} @feat;
for my $feat(@feat){
my $subseqFeatStart=$feat->location->start-$start+1;
my $subseqFeatEnd=$feat->location->end-$start+1;
my $strand=$feat->strand;
if($revcom){
$subseqFeatStart=($end-$feat->location->start+1);
$subseqFeatEnd=($end-$feat->location->end+1);
($subseqFeatStart,$subseqFeatEnd)=$subseqFeatEnd,$subseqFeatStart);
$strand=$strand*-1;
} else {
}
}

```

```

    next if( ($subseqFeatStart<1 && $subseqFeatEnd<1) ||
($subseqFeatStart>$seqLength && $subseqFeatEnd>$seqLength) );
    my @gene=($feat->has_tag("gene"))?$feat->get_tag_values('gene'):();
    print join("_",$subseqFeatStart,$subseqFeatEnd,$strand,@gene)."\n";
    next if(ref($feat->location) ne "Bio::Location::Simple");
    $subseqFeatStart=1 if($subseqFeatStart<1);
    $subseqFeatEnd=$seqLength if($subseqFeatEnd>$seqLength);
    $feat->location->start($subseqFeatStart);
    $feat->location->end($subseqFeatEnd);
    $feat->strand($strand);
    $truncSeq->add_SeqFeature($feat);
}
#$truncSeq->seq(lc($truncSeq->seq)); # lc the sequence so that it can be
highlighted
#highlightFeatures($truncSeq,$settings);
return $truncSeq;
}
sub highlightFeatures{
my($seq,$settings)=@_;
my $highlight=$$settings{highlight_type}||"uc";
# TODO return false if not a rich seq
for my $feat($seq->get_SeqFeatures){
next if($feat->primary_tag eq 'source');
my($start,$end)=$feat->location->start,$feat->location->end);
if($highlight eq 'uc'){
my $firstPartOfTheSequence=($start>1)?$seq->subseq(1,$start-1):"";
my $lastPartOfTheSequence=($end<$seq->length)?$seq->subseq($end+1,$seq->length):"";
my $sequence=$firstPartOfTheSequence.uc($seq->subseq($start,$end)).$lastPartOfTheSequence;
$seq->seq($sequence);
}
}
}
# makes a hash of sequences, given a fasta file
sub seqHash{
my $filename=shift;

```

```

my($seqio,%seq);
if(-e $filename){
$seqio=Bio::SeqIO->new(-file=>$filename);
}
# sequence object
elsif(ref($filename)=~/seq/i) {
die ref($filename)." is not supported yet";
}
else{
die "Cannot determine the type of sequence for $filename";
}
while(my $seq=$seqio->next_seq){
my $id=$seq->id;
$seq{$id}=$seq;
# make a second key without genome info and zero padding
$id=~s/^\.+?_0*//;
$seq{$id}=$seq;
}
return %seq;
}
sub usage{
"Usage: perl $0 -i inputGenomicFile -s start -e end
-i input file: the file extension dictates the format
-o outfile
-s start coordinate. 1-based
-e end coordinate. 1-based
-c contig
put a negative sign in front of a contig to indicate revcom
you may have multiple -c args
-n name of organism (useful for genome browsers)
";
}

```

#### Anexo 4.4

Sequências de referências construídas para os possíveis transcritos alternativos E6/E7 do HPV 16. A região de junção de *splicing* (RJS) está destacada de amarelo e os nucleotídeos da junção de *splicing* em vermelho.

>191\_409

```
ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTATTA
ACTGTCAAAGCCACTGTGTCTCTGAAGAAAAGCAAAGACATCTGGACAAAAGCAAAGATTCCAT
AATATAAGGGGTCGGTGGACCGGTCGATGTATGTCTTGTGTCAGATCATCAAGAACACGTAG
AGAAACCCAGCTGTAATCATGCATGGAGATACACCTACATTGCATGAATATATGTTAGATTT
GCAACCAGAGACAACCTGATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAGGAGGAGG
ATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCCATTACAATATTGTAACC
TTTTGTTGCAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTTCG
TACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAAC
CATAA
```

>191\_526

```
ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATATCATCAAGAA
CACGTAGAGAAACCCAGCTGTAATCATGCATGGAGATACACCTACATTGCATGAATATATGT
TAGATTTGCAACCAGAGACAACCTGATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAG
GAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCCATTACAATAT
TGTAACCTTTTGTGCAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAG
ACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGTTCT
CAGAAACCATAA
```

>191\_742

```
ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATTGTGACTCTAC
GCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTTCGTACTTTGGAAGACCTGTTAATGG
GCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA
```

>221\_409

```
ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC
```

AAGCAACAGTTAC **TGCGACGTGTAT** TAACTGTCAAAAGCCACTGTGTCTGAAGAAAAGCAA  
AGACATCTGGACAAAAAGCAAAGATTCATAATATAAGGGTTCGGTGGACCGGTCGATGTAT  
GTCTTGTTCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCATGGAGATAC  
ACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATCTCTACTGTTATG  
AGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAA  
CCGGACAGAGCCCATTACAATATTGTAACCTTTTGTGCAAGTGTGACTCTACGCTTCGGTT  
GTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAG  
GAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA

>221\_526

ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC  
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC  
AAGCAACAGTTAC **TGCGACATCATC** AAGAACACGTAGAGAAACCCAGCTGTAATCATGCATG  
GAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATCTCTAC  
TGTTATGAGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACA  
AGCAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTGCAAGTGTGACTCTACGC  
TTCGGTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGC  
ACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA

>221\_742

ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC  
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC  
AAGCAACAGTTAC **TGCGACTGTGAC** TCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAG  
ACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCTGTTCT  
CAGAAACCATAA

>226\_409

ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC  
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC  
AAGCAACAGTTACTGCGA **CGTGAGGTGTAT** TAACTGTCAAAAGCCACTGTGTCTGAAGAAA  
AGCAAAGACATCTGGACAAAAAGCAAAGATTCATAATATAAGGGTTCGGTGGACCGGTCGA  
TGTATGTCTTGTTCAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCATGGA  
GATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATCTCTACTG  
TTATGAGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACAAG  
CAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTGCAAGTGTGACTCTACGCTT

CGGTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCAC  
ACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA

>226\_526

ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC  
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC  
AAGCAACAGTTACTGCGA **CGTGAGATCATC** AAGAACACGTAGAGAAACCCAGCTGTAATCAT  
GCATGGAGATACACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATC  
TCTACTGTTATGAGCAATTAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCT  
GGACAAGCAGAACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTGCAAGTGTGACTC  
TACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAA  
TGGGCACACTAGGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA

>226\_742

ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC  
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC  
AAGCAACAGTTACTGCGA **CGTGAGTGTGAC** TCTACGCTTCGGTTGTGCGTACAAAGCACACA  
CGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATCT  
GTTCTCAGAAACCATAA

>E6+E7

ATGCACCAAAGAGAACTGCAATGTTTCAGGACCCACAGGAGCGACCCAGAAAGTTAC  
CACAGTTATGCACAGAGCTGCAAACAACCTATACATGATATAATATTAGAATGTGTGTACTGC  
AAGCAACAGTTACTGCGACGTGAGGTATATGACTTTGCTTTTCGGGATTTATGCATAGTATA  
TAGAGATGGGAATCCATATGCTGTATGTGATAAATGTTTAAAGTTTTATTCTAAAATTAGTG  
AGTATAGACATTATTGTTATAGTTTGTATGGAACAACATTAGAACAGCAATACAACAAACCG  
TTGTGTGATTTGTTAATTAGGTGTATTAAGTGTCAAAGCCACTGTGTCCTGAAGAAAAGCA  
AAGACATCTGGACAAAAGCAAAGATTCCATAATATAAGGGGTCGGTGGACCGGTCGATGTA  
TGTCTTGTGCGAGATCATCAAGAACACGTAGAGAAACCCAGCTGTAATCATGCATGGAGATA  
CACCTACATTGCATGAATATATGTTAGATTTGCAACCAGAGACAACCTGATCTCTACTGTTAT  
GAGCAATTAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGA  
ACCGGACAGAGCCCATTACAATATTGTAACCTTTTGTGCAAGTGTGACTCTACGCTTCGGT  
TGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTA  
GGAATTGTGTGCCCCATCTGTTCTCAGAAACCATAA



## Anexo 4.5

Sequências de referências construídas para os possíveis transcritos alternativos E6/E7 do HPV 18. A região de junção de *splicing* (RJS) está destacada de amarelo e os nucleotídeos da junção de *splicing* em vermelho.

>233\_416

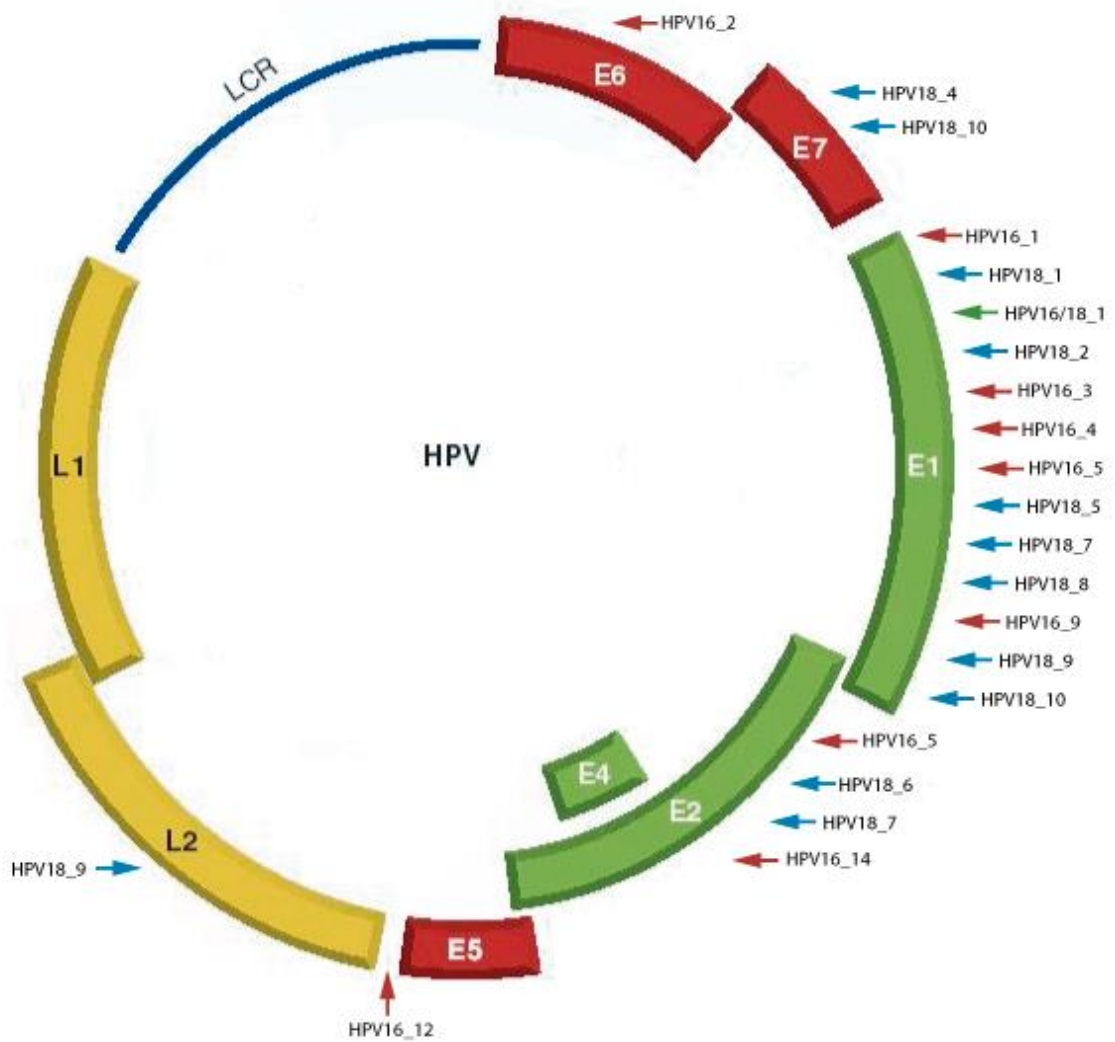
```
ATGGCGCGCTTTGAGGATCCAACACGGCGACCCTACAAGCTACCTGATCTGTGCACGG
AACTGAACACTTCACTGCAAGACATAGAAATAACCTGTGTATATTGCAAGACAGTATTGGAA
CTTACAGAGGTGCCTGCGGTGCCAGAAACCGTTGAATCCAGCAGAAAACTTAGACACCTTA
ATGAAAAACGACGATTTACAACATAGCTGGGCACTATAGAGGCCAGTGCCATTCGTGCTGC
AACCGAGCACGACAGGAACGACTCCAACGACGCAGAGAAACACAAGTATAATATTAAGTATG
CATGGACCTAAGGCAACATTGCAAGACATTGTATTGCATTTAGAGCCCCAAAATGAAATTC
GGTTGACCTTCTATGTCACGAGCAATTAAGCGACTCAGAGGAAGAAAACGATGAAATAGATG
GAGTTAATCATCAACATTTACCAGCCCGACGAGCCGAACCACAACGTCACACAATGTTGTGT
ATGTGTTGTAAGTGTGAAGCCAGAATTGAGCTAGTAGTAGAAAGCTCAGCAGACGACCTTCG
AGCATTCAGCAGCTGTTTCTGAACACCCTGTCCTTTGTGTGTCCGTGGTGTGCATCCCAGC
AGTAA
```

>E6+E7

```
ATGGCGCGCTTTGAGGATCCAACACGGCGACCCTACAAGCTACCTGATCTGTGCACGG
AACTGAACACTTCACTGCAAGACATAGAAATAACCTGTGTATATTGCAAGACAGTATTGGAA
CTTACAGAGGTATTTGAATTTGCATTTAAAGATTTATTTGTGGTGTATAGAGACAGTATACC
CCATGCTGCATGCCATAAATGTATAGATTTTTATTCTAGAATTAGAGAATTAAGACATTATT
CAGACTCTGTGTATGGAGACACATTGGAAAACTAACTAACACTGGGTTATACAATTTATTA
ATAAGGTGCCTGCGGTGCCAGAAACCGTTGAATCCAGCAGAAAACTTAGACACCTTAATGA
AAAACGACGATTTACAACATAGCTGGGCACTATAGAGGCCAGTGCCATTCGTGCTGCAACC
GAGCACGACAGGAACGACTCCAACGACGCAGAGAAACACAAGTATAATATTAAGTATGCATG
GACCTAAGGCAACATTGCAAGACATTGTATTGCATTTAGAGCCCCAAAATGAAATTCGGTT
GACCTTCTATGTCACGAGCAATTAAGCGACTCAGAGGAAGAAAACGATGAAATAGATGGAGT
TAATCATCAACATTTACCAGCCCGACGAGCCGAACCACAACGTCACACAATGTTGTGTATGT
GTTGTAAGTGTGAAGCCAGAATTGAGCTAGTAGTAGAAAGCTCAGCAGACGACCTTCGAGCA
TTCCAGCAGCTGTTTCTGAACACCCTGTCCTTTGTGTGTCCGTGGTGTGCATCCCAGCAGTA
A
```

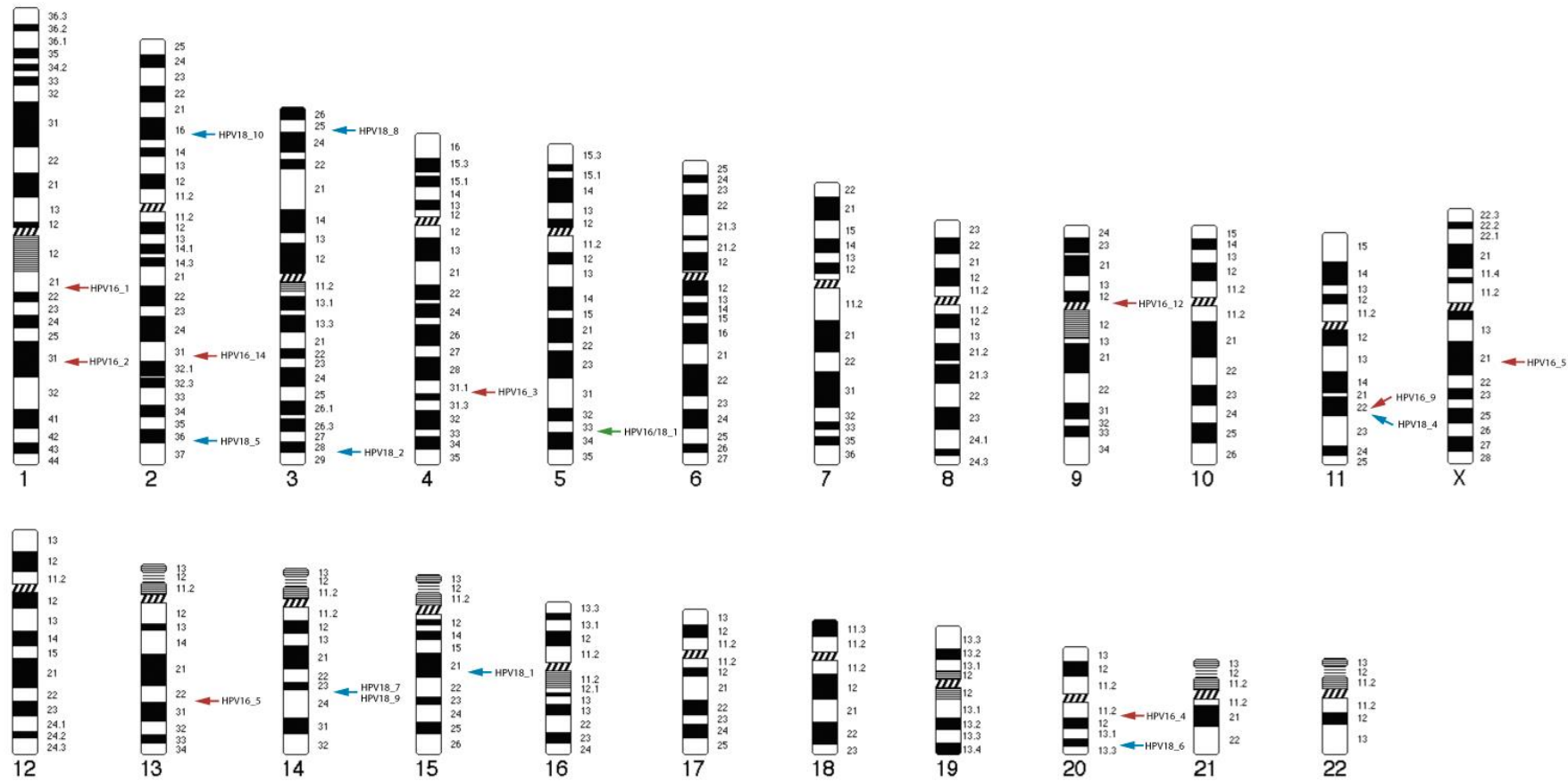
## Anexo 5.1

Esquema representativo do genoma do HPV demonstrando o gene em que ocorreu a quebra para cada amostra de tumor cervical. As setas vermelhas representam as amostras infectadas com HPV 16, setas azuis amostras com HPV 18 e a seta verde a amostra co-infectada.



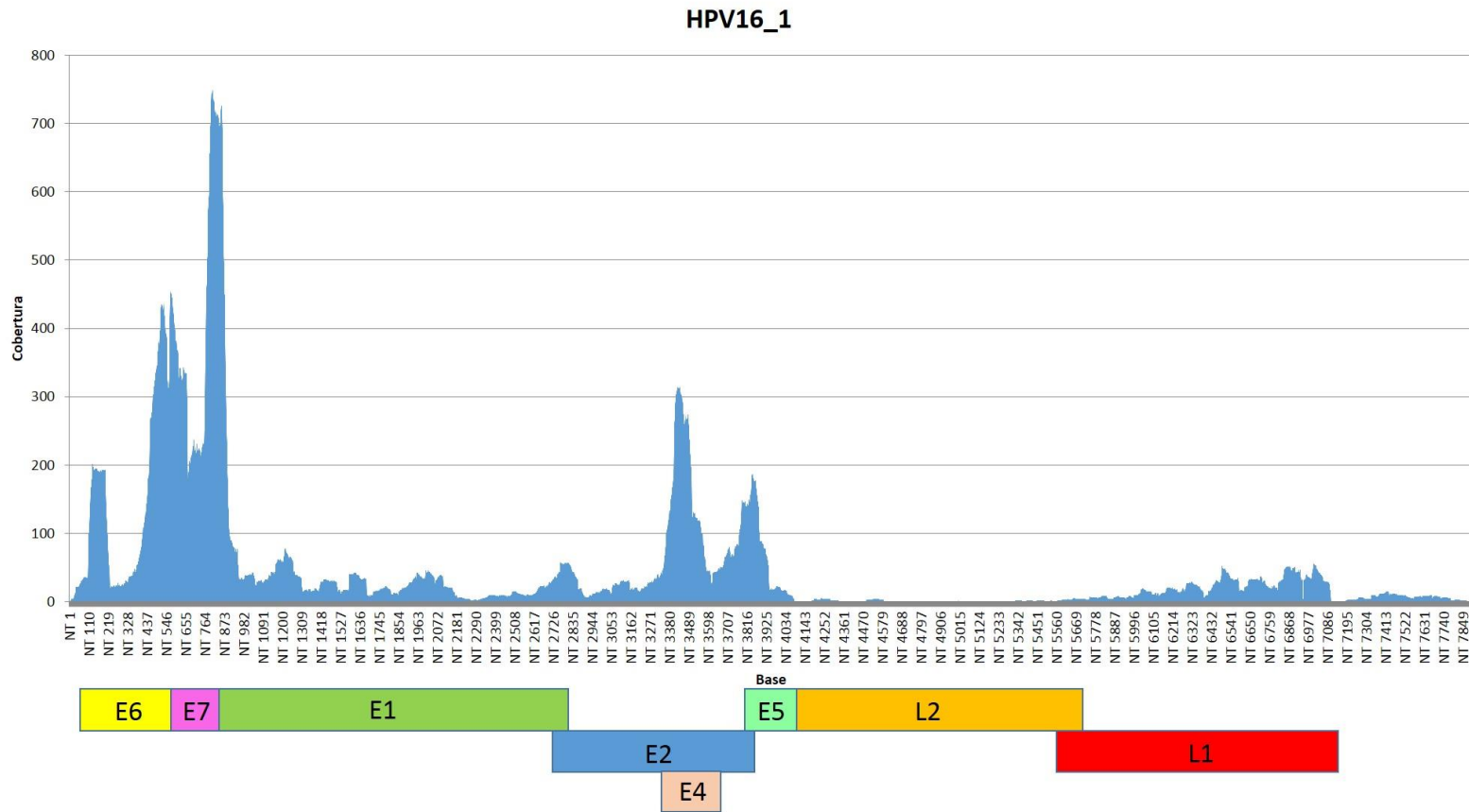
## Anexo 5.2

Representação no cariótipo humano feminino demonstrando o cromossomo e o braço em que ocorreu a integração do genoma do HPV para cada amostra de tumor cervical. As setas vermelhas representam as amostras infectadas com HPV 16, setas azuis amostras com HPV 18 e a seta verde a amostra co-infectada.

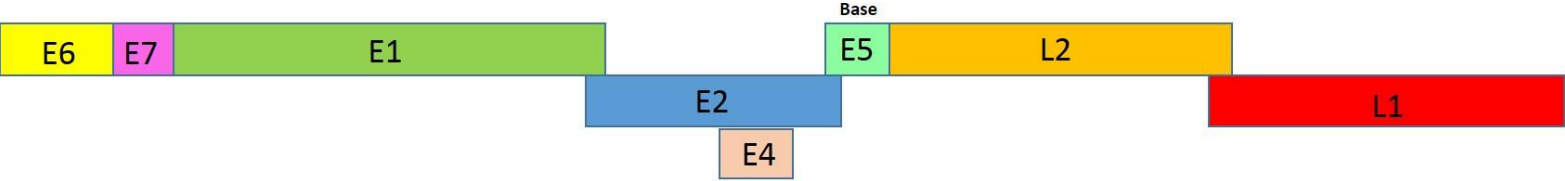
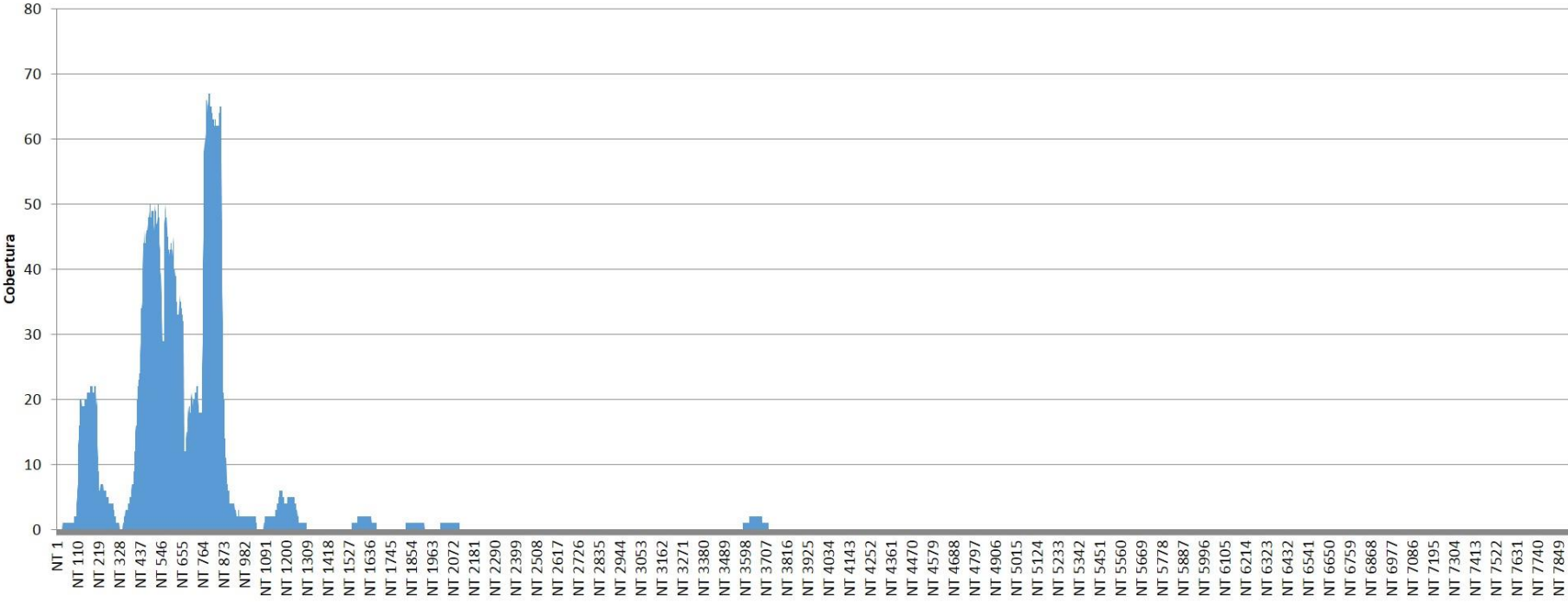


### Anexo 5.3

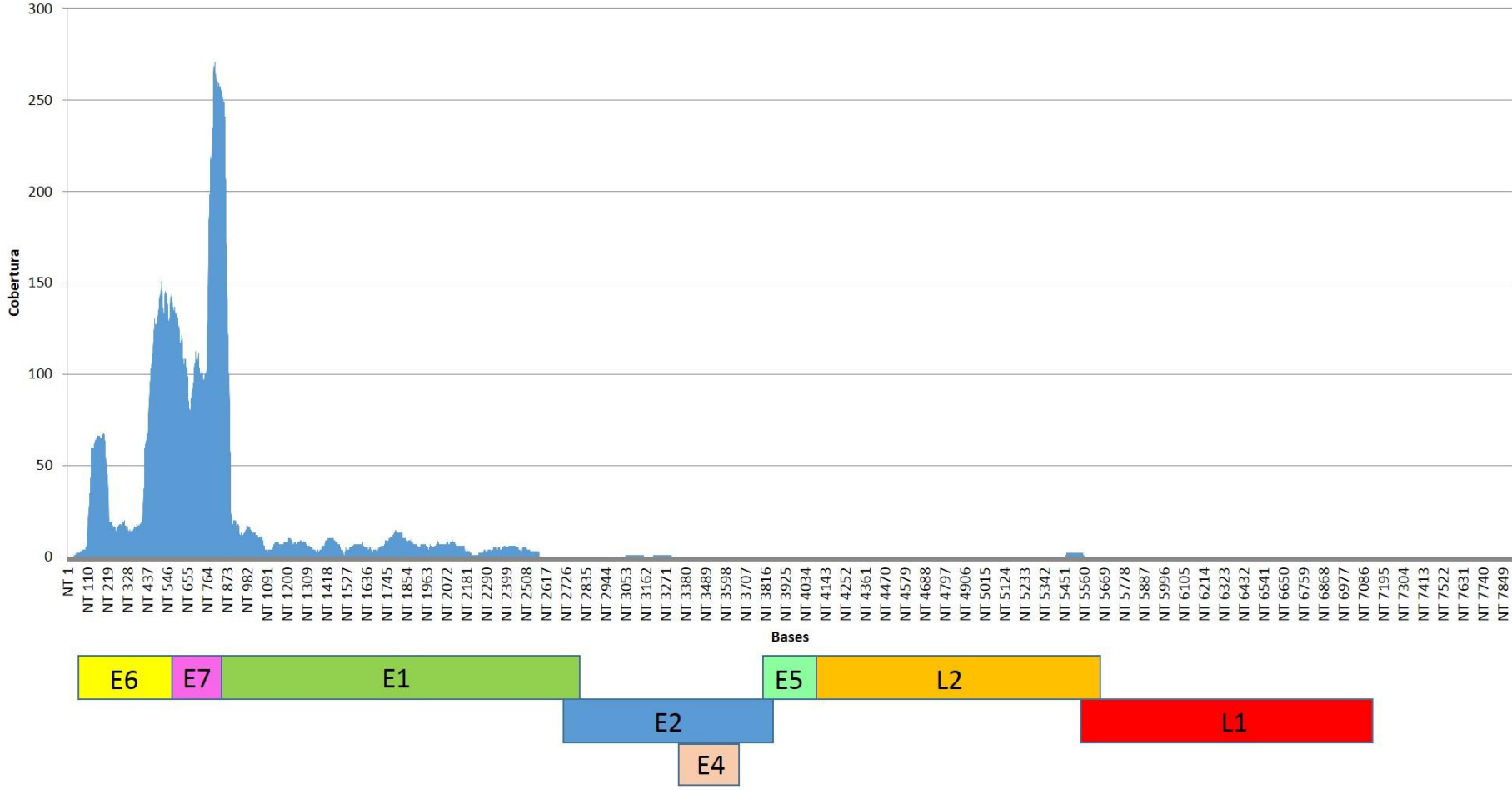
Gráficos de cobertura por base da expressão do genoma viral das amostras infectadas com HPV 16. As regiões nucleotídicas referentes a cada gene estão demarcadas abaixo do eixo x.



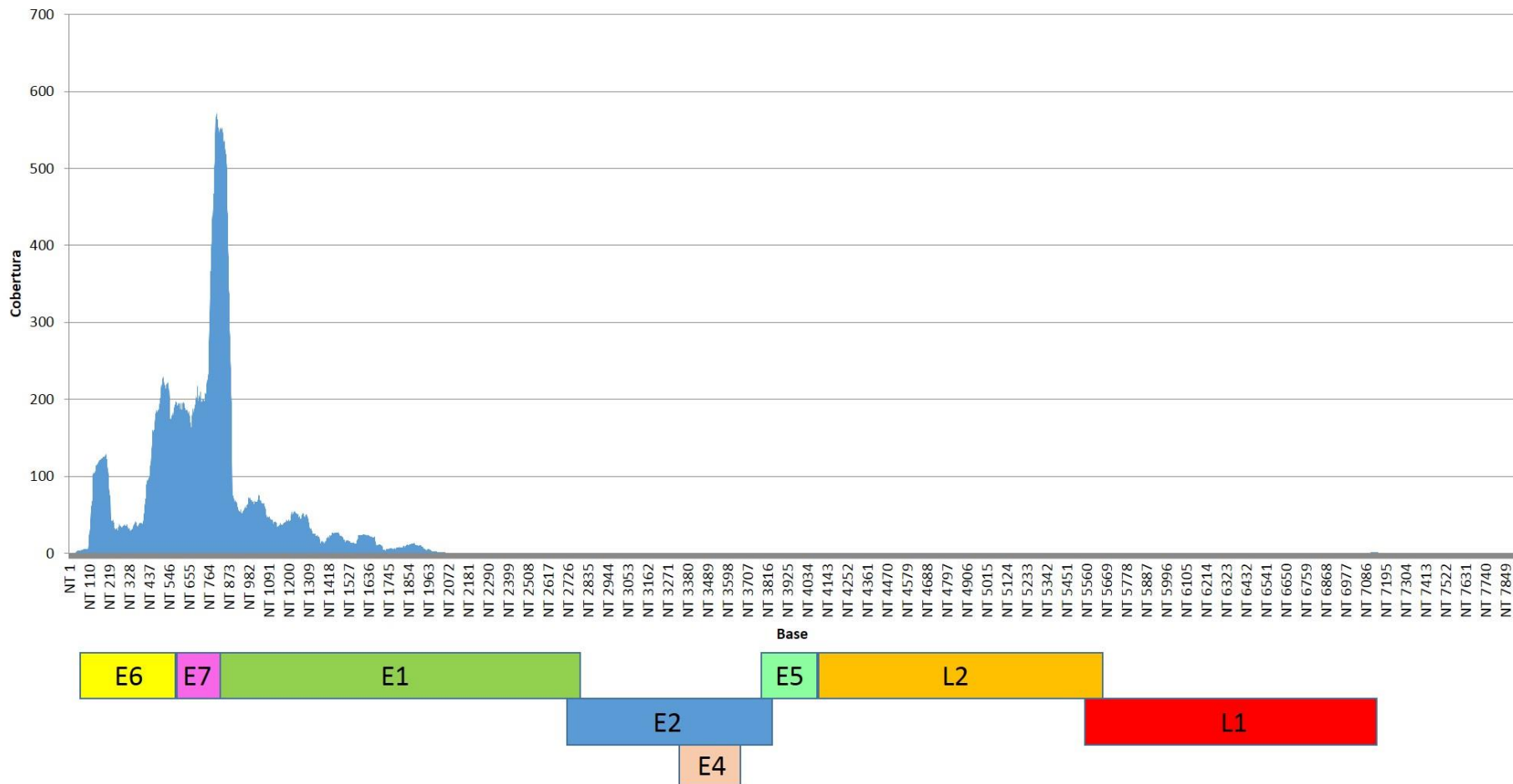
### HPV16\_2



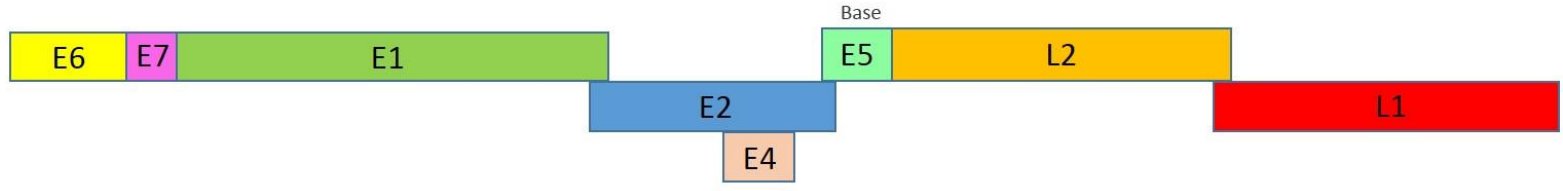
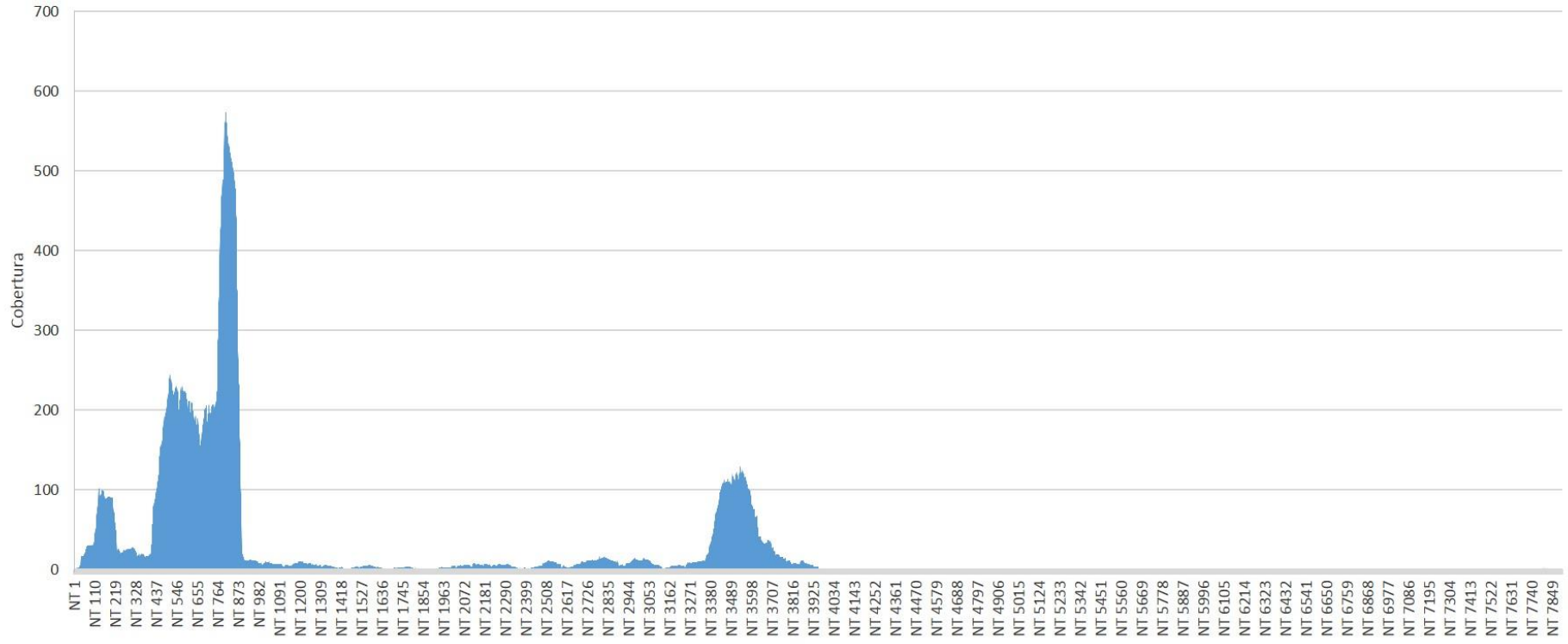
HPV16\_3



# HPV16\_4

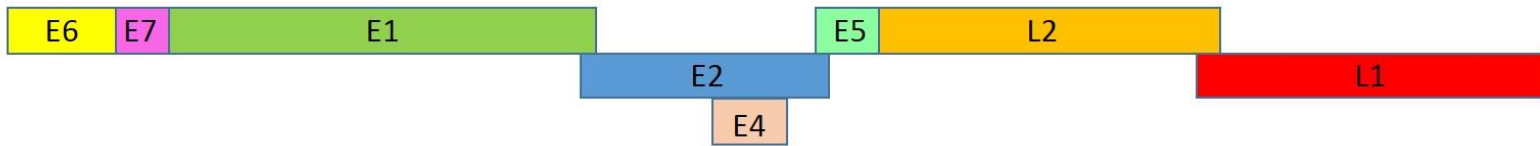
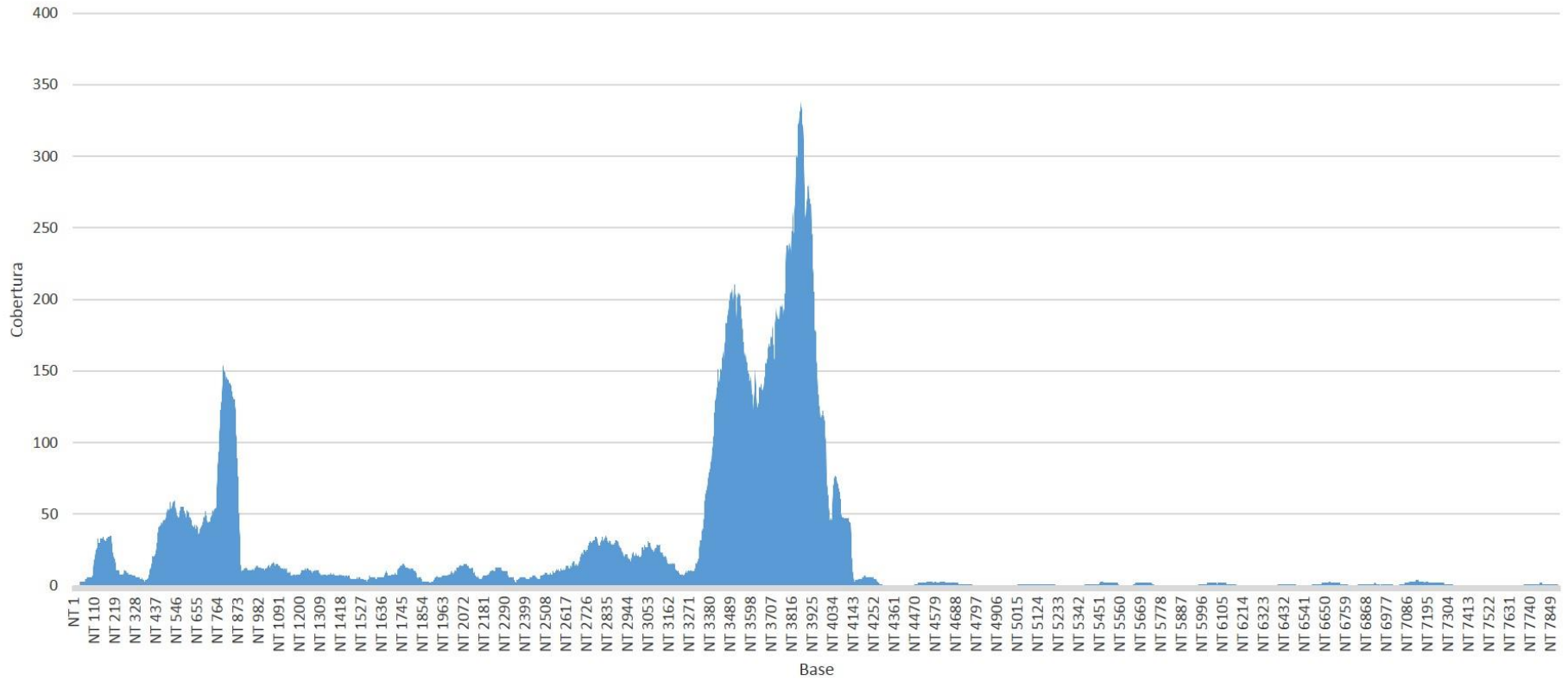


# HPV16\_5

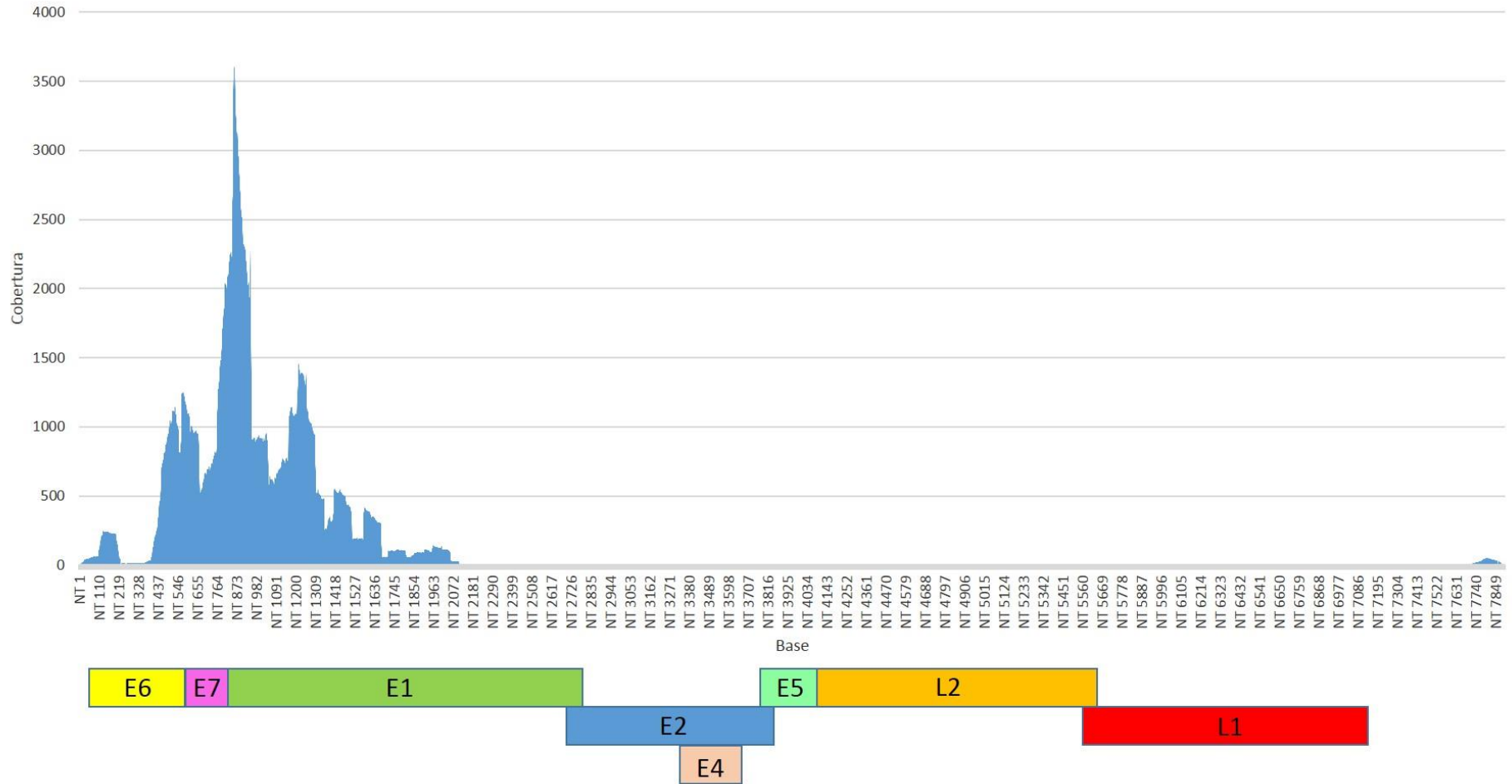




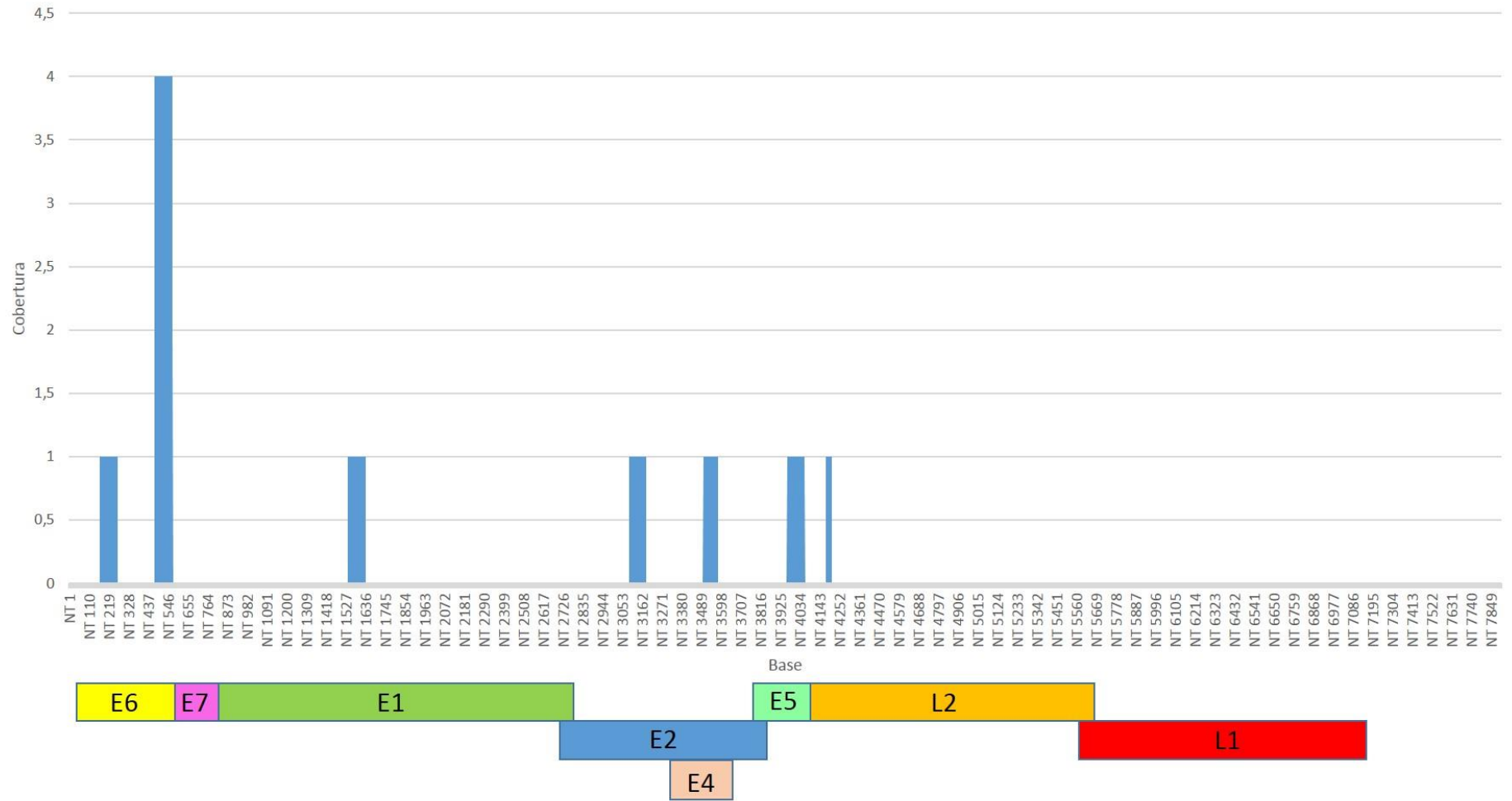
# HPV16\_6



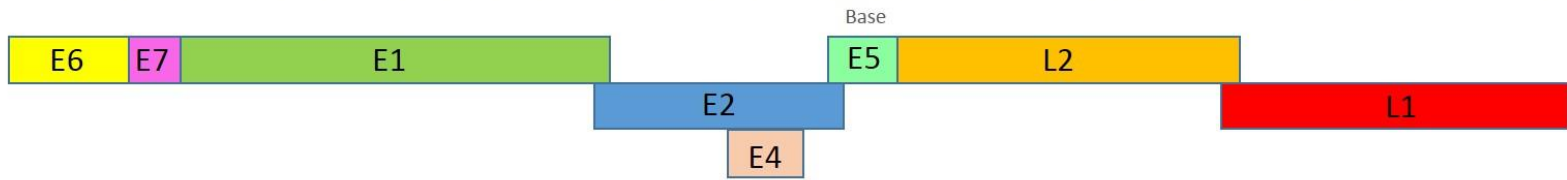
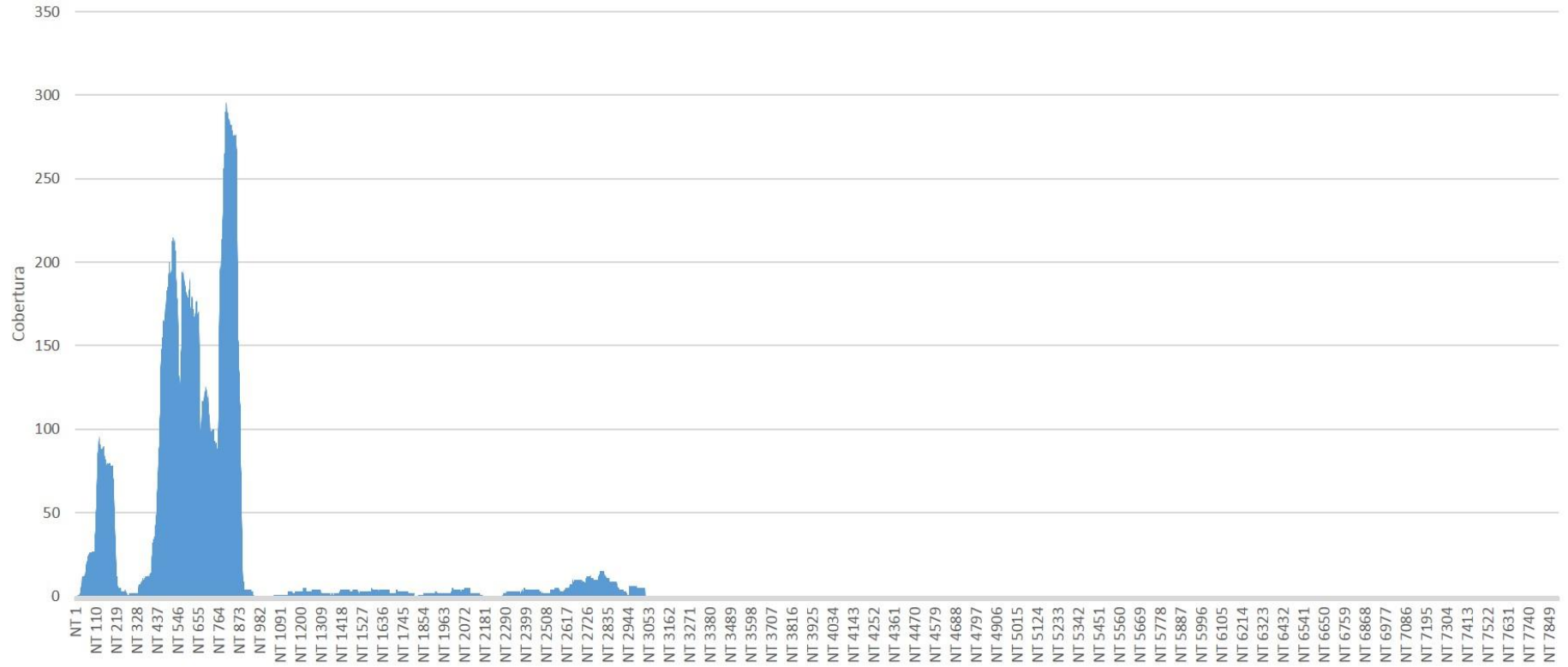
# HPV16\_7



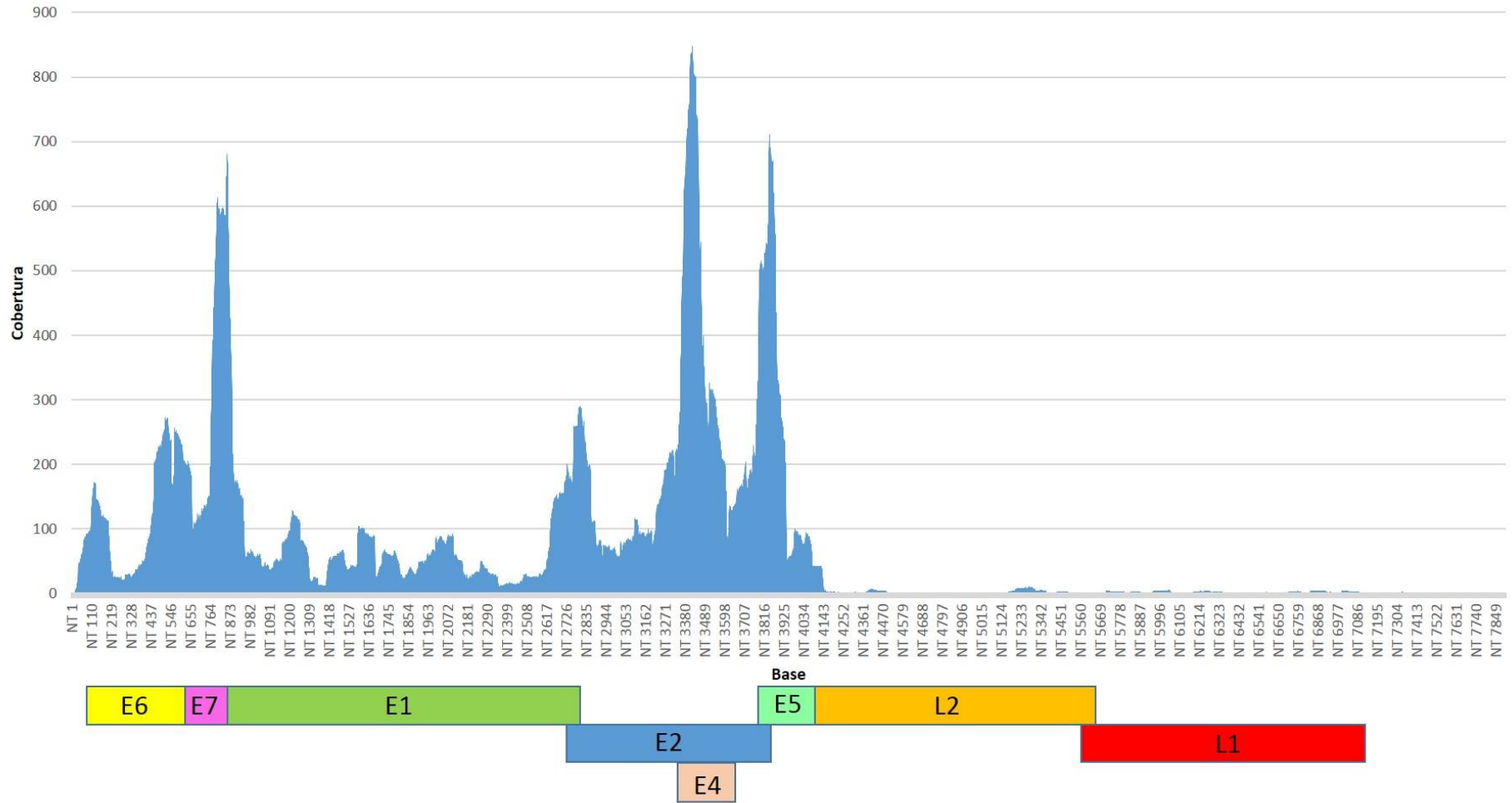
### HPV16\_8



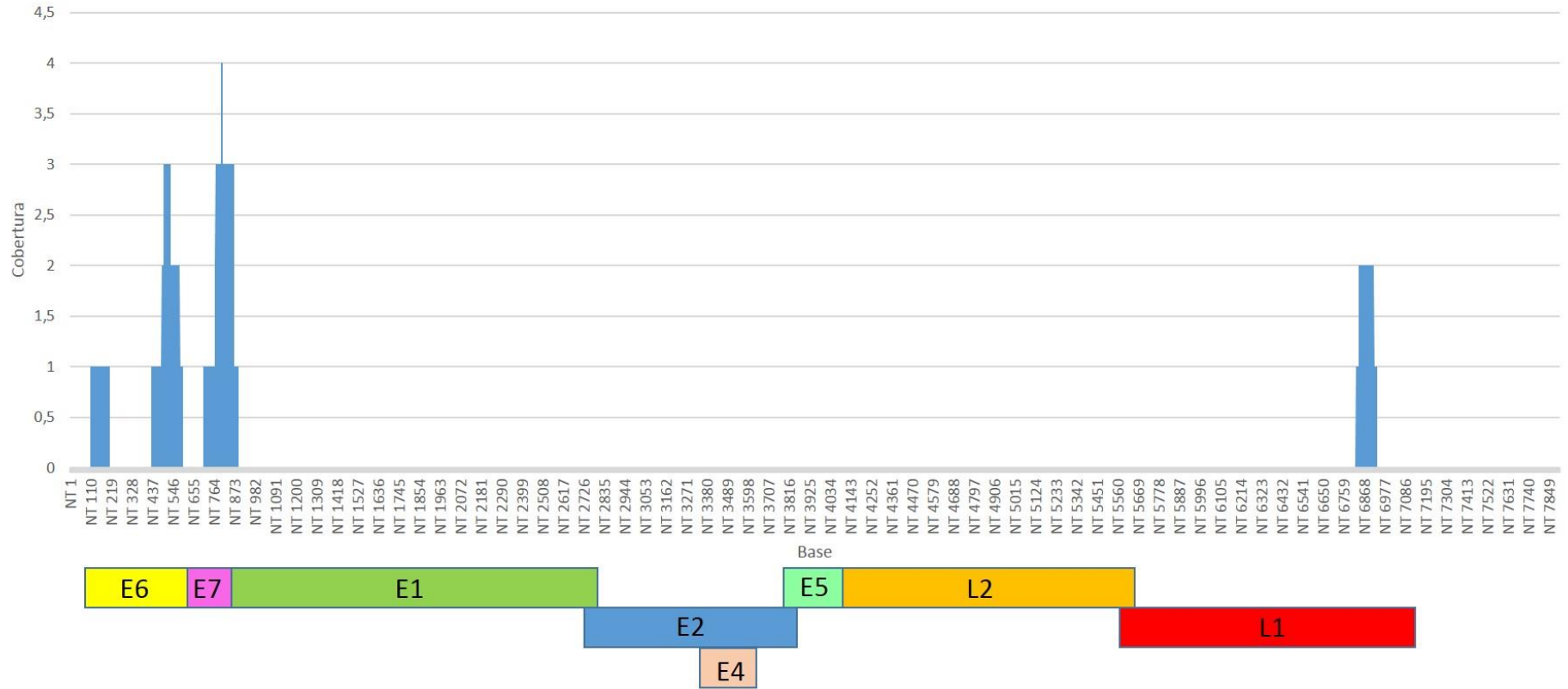
# HPV16\_9



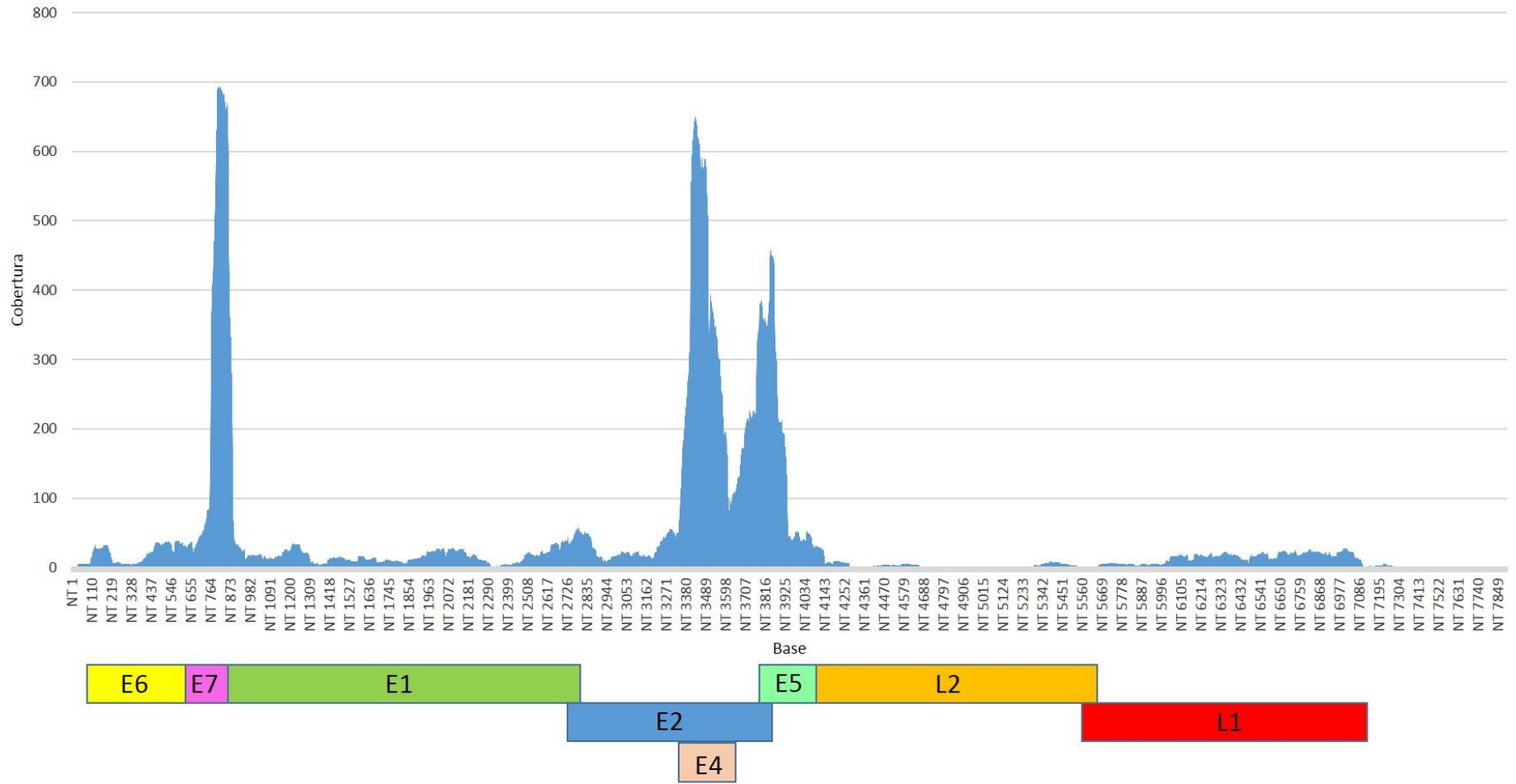
HPV16\_10



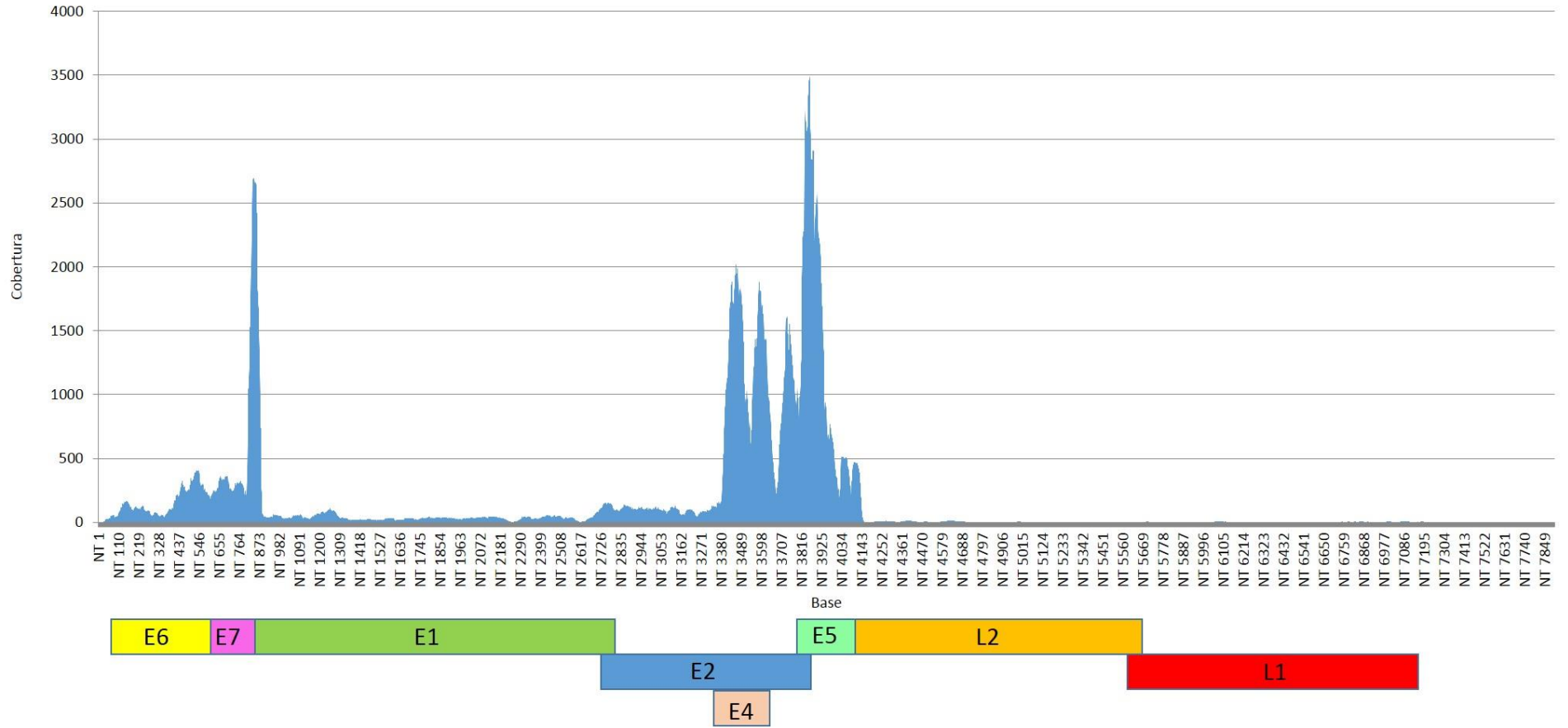
### HPV16\_11



HPV16\_12

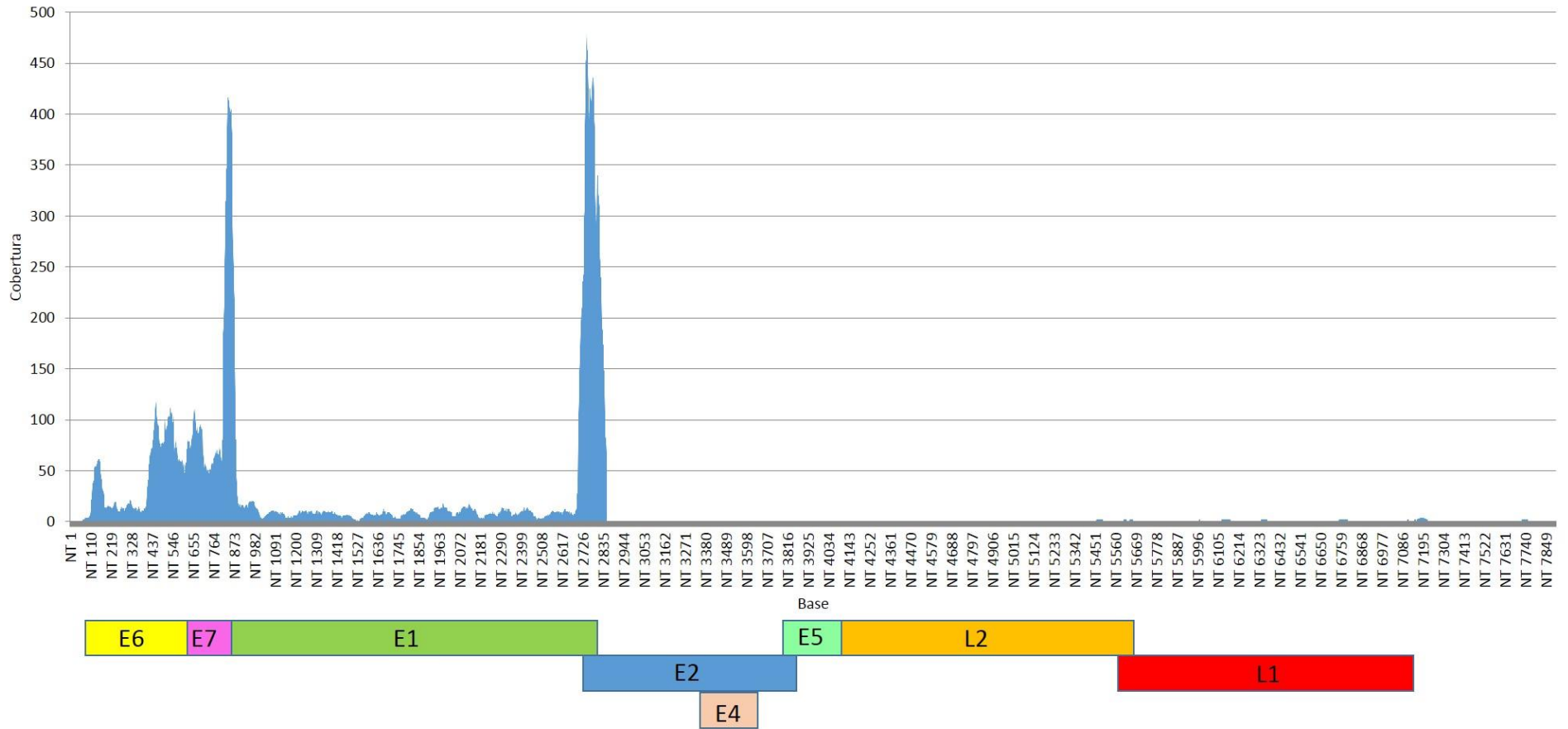


# HPV16\_13



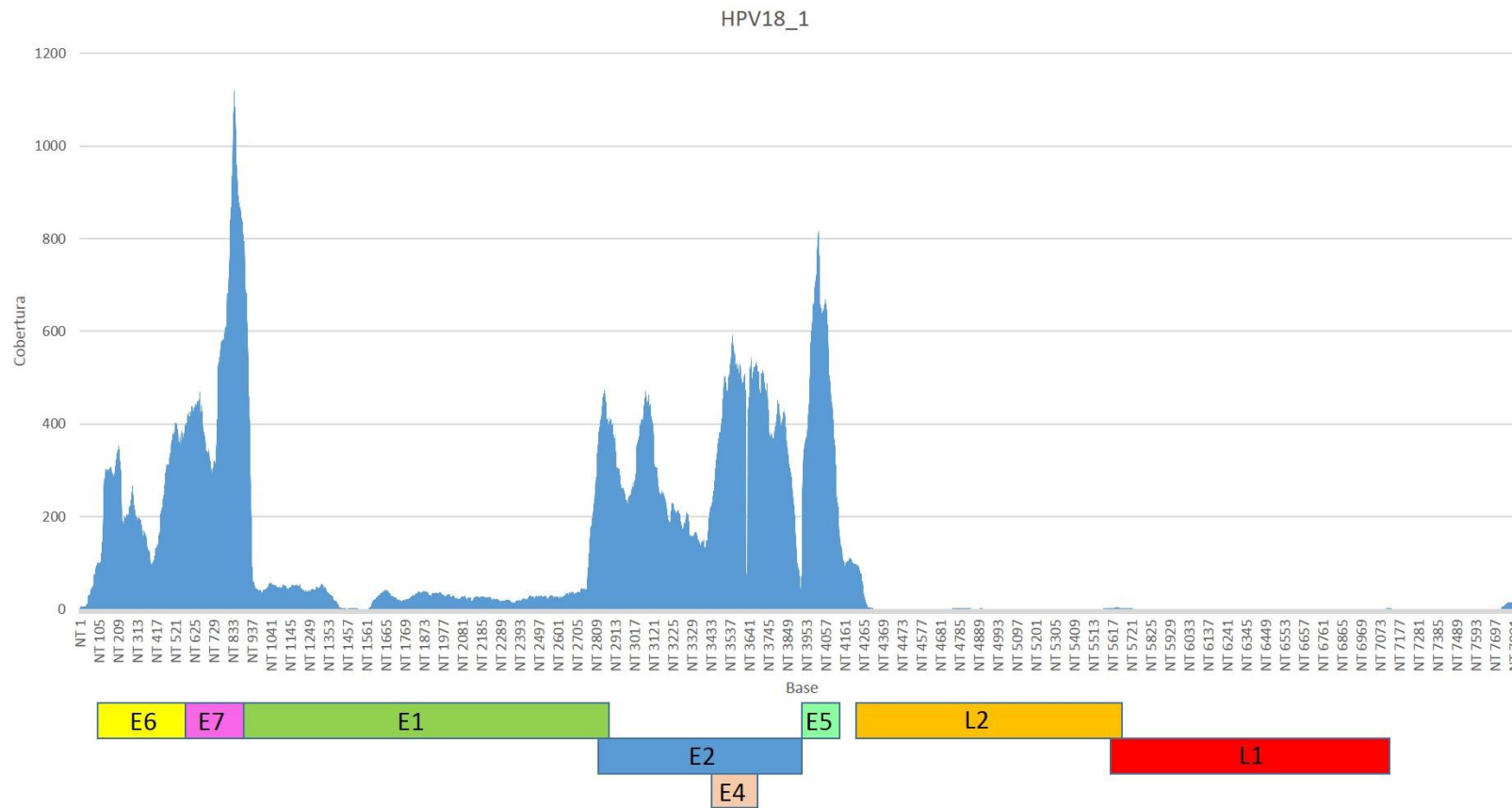


# HPV16\_14

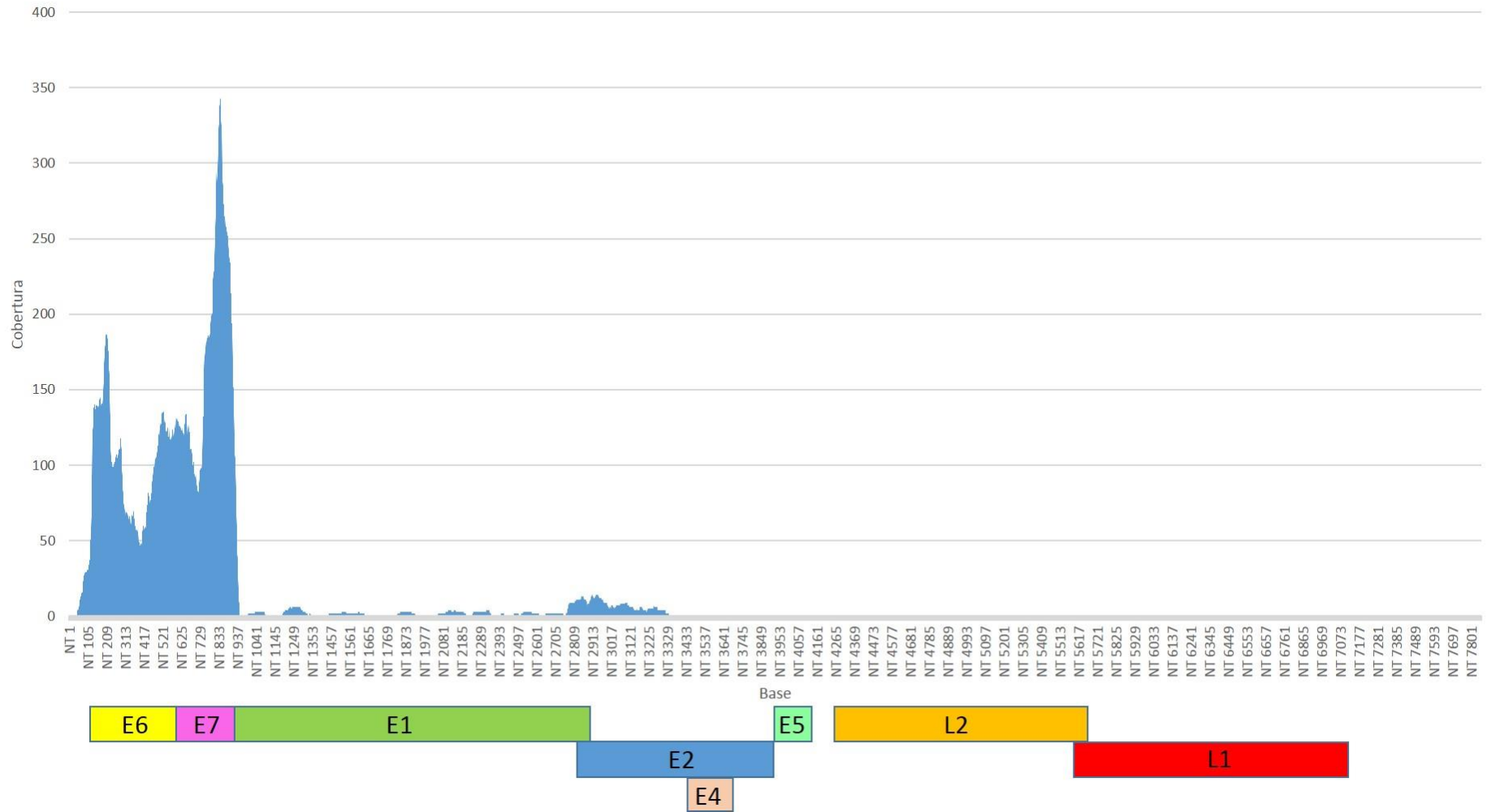


## Anexo 5.4

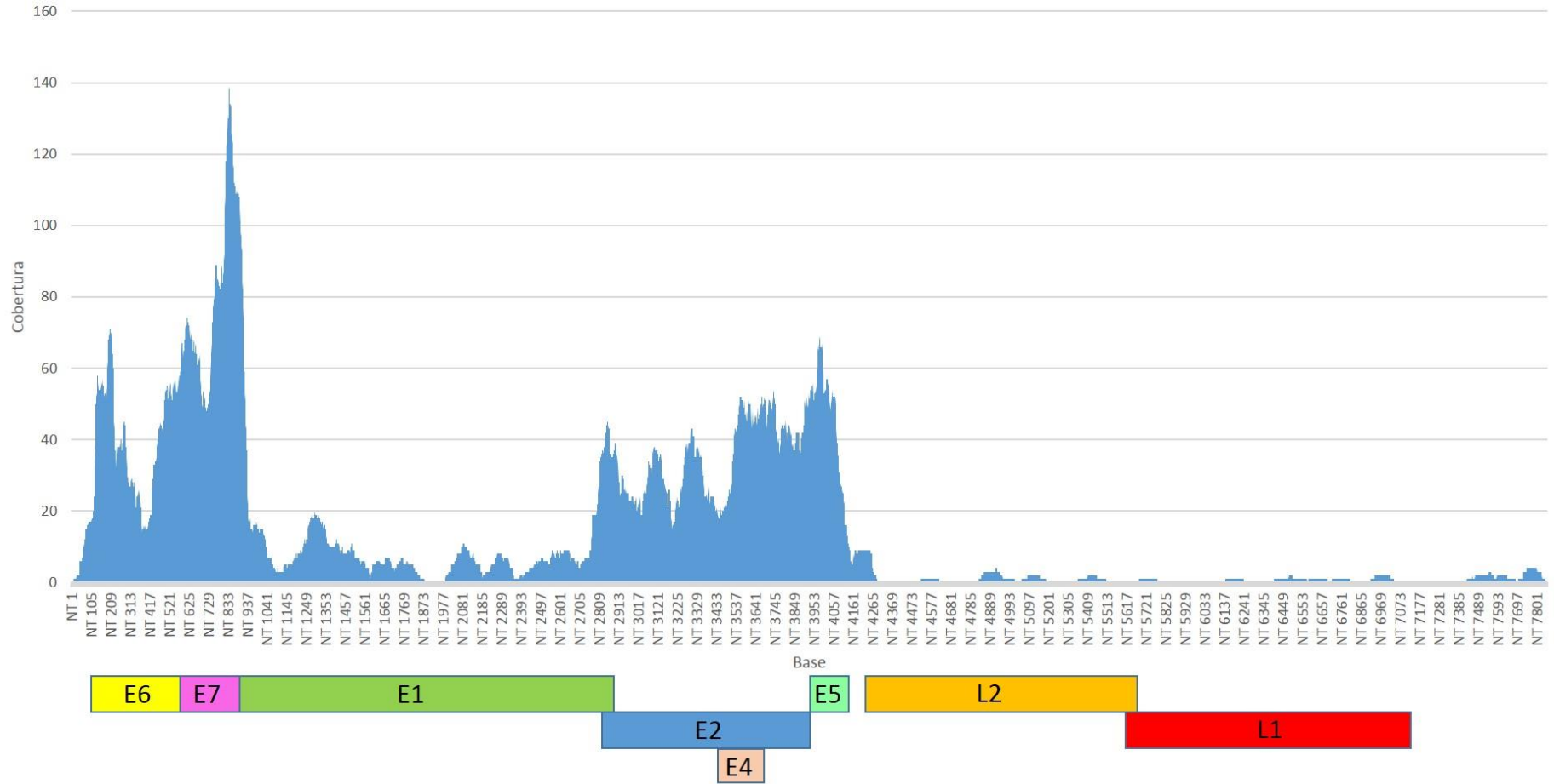
Gráficos de cobertura por base da expressão do genoma viral das amostras infectadas com HPV 18. As regiões nucleotídicas referentes a cada gene estão demarcadas abaixo do eixo x.



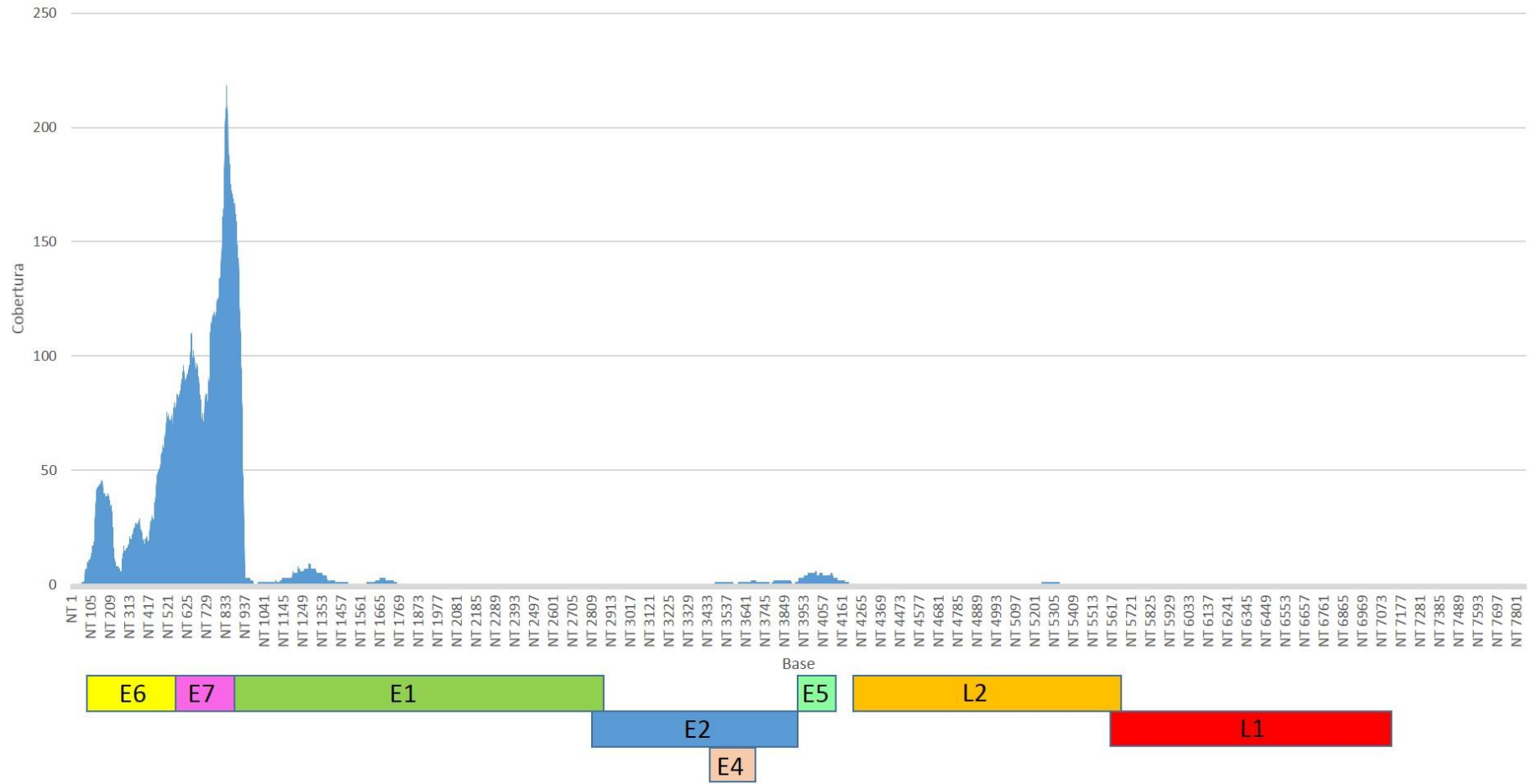
# HPV18\_2



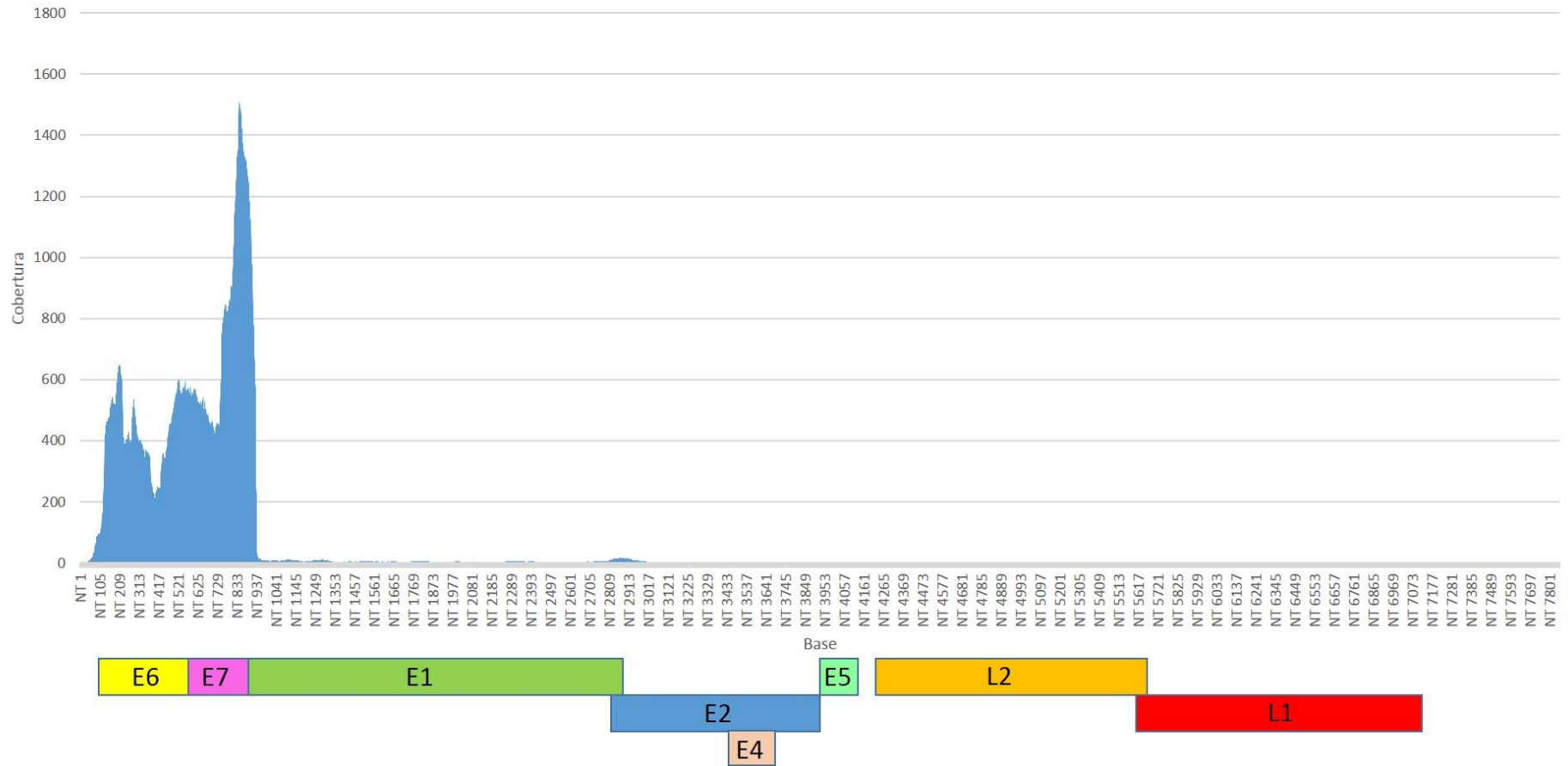
HPV18\_3



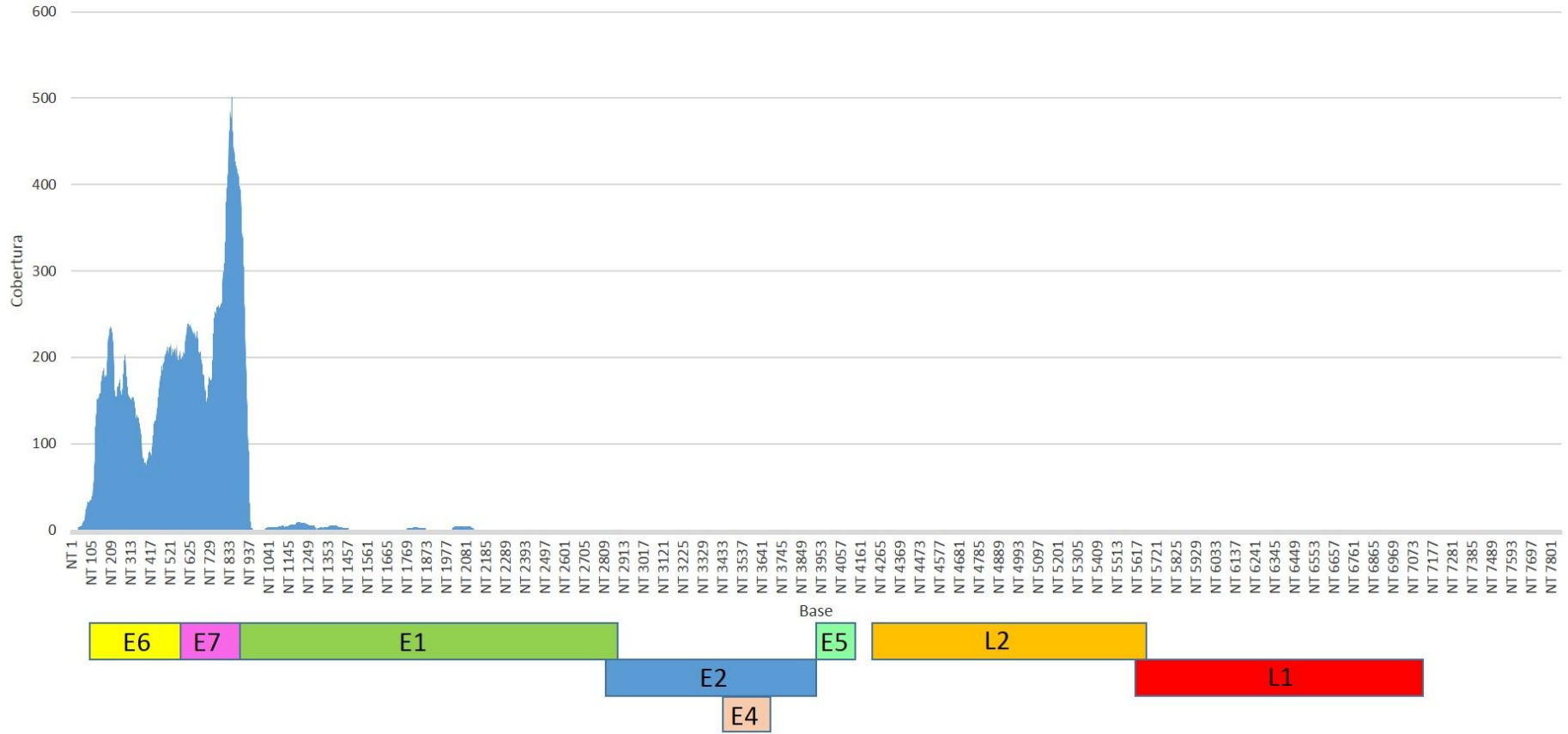
HPV18\_4

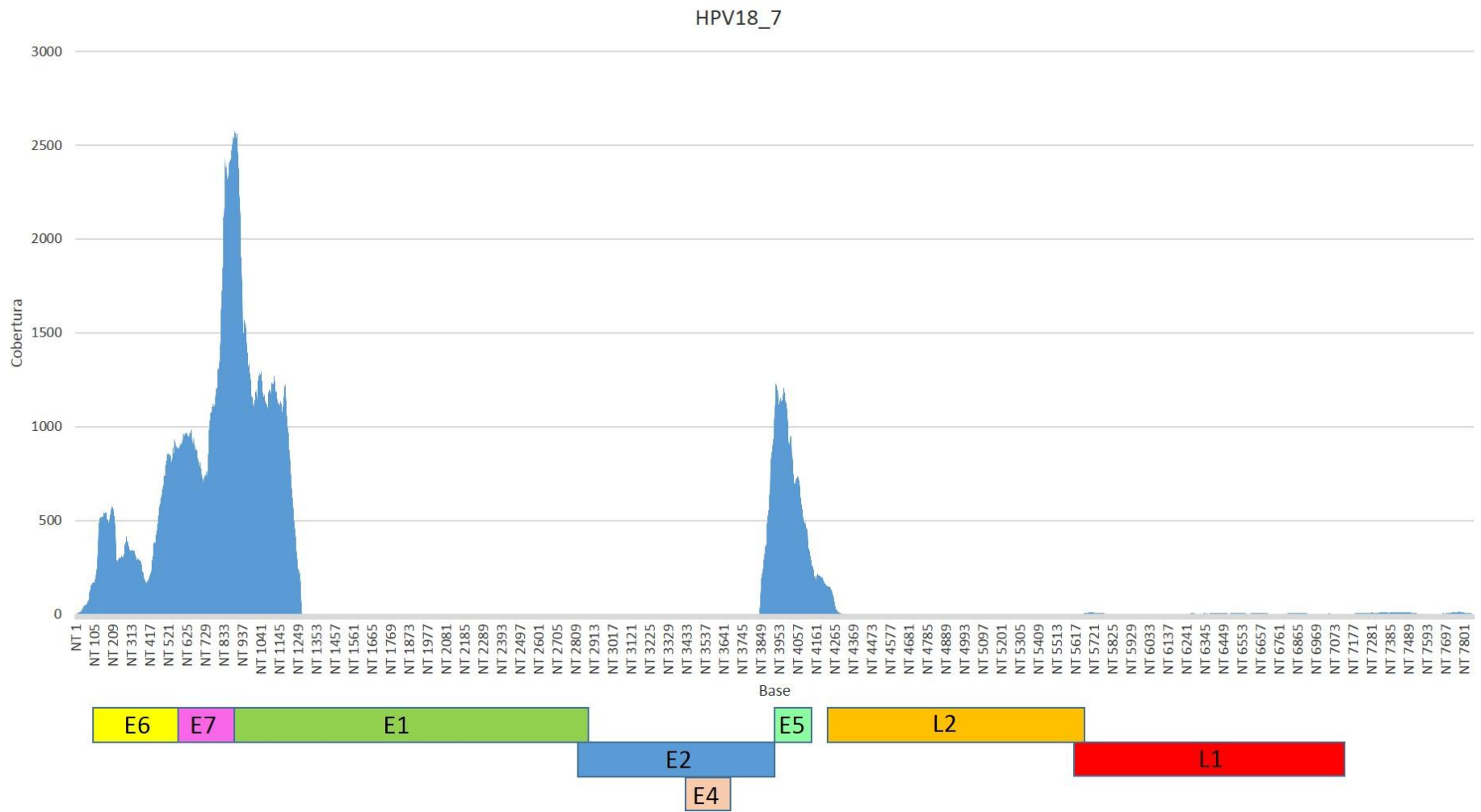


# HPV18\_5



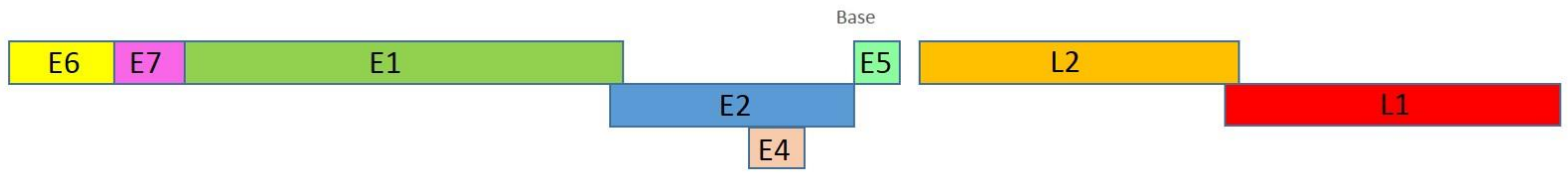
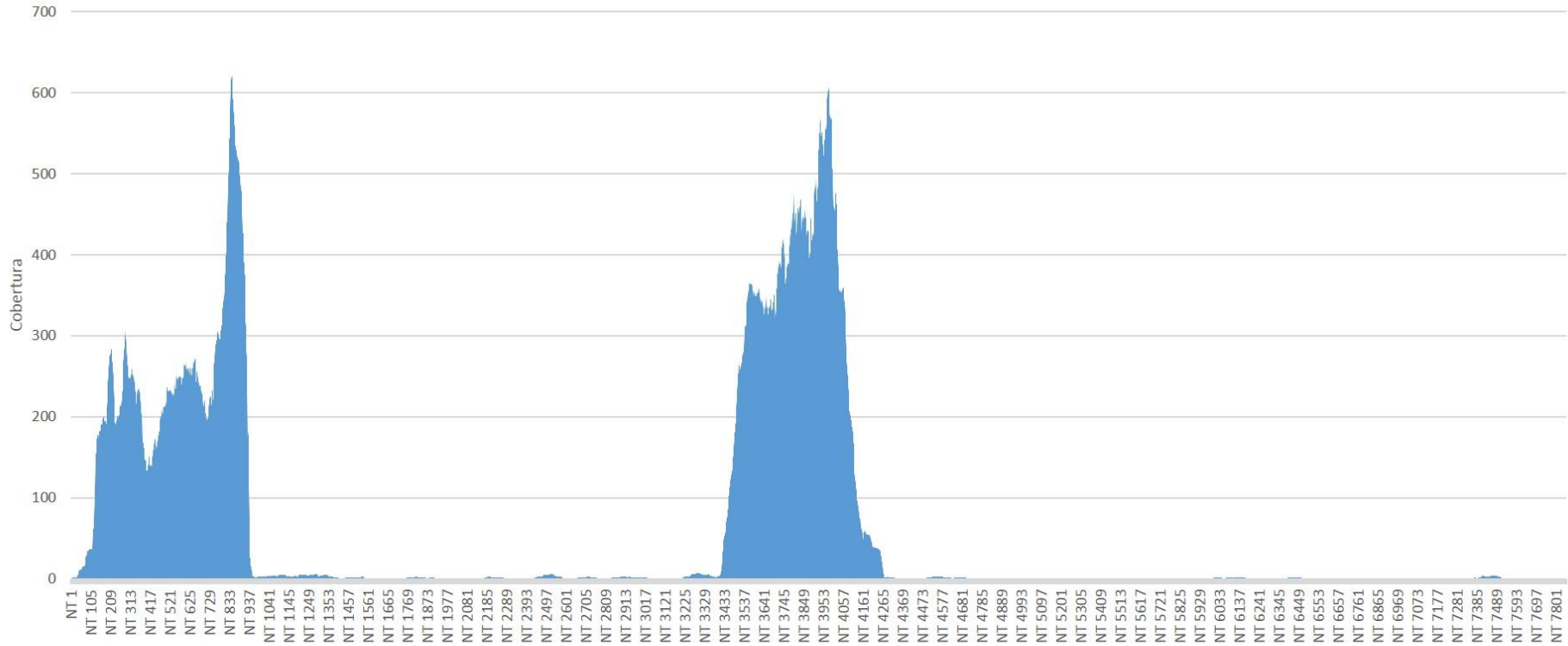
# HPV18\_6



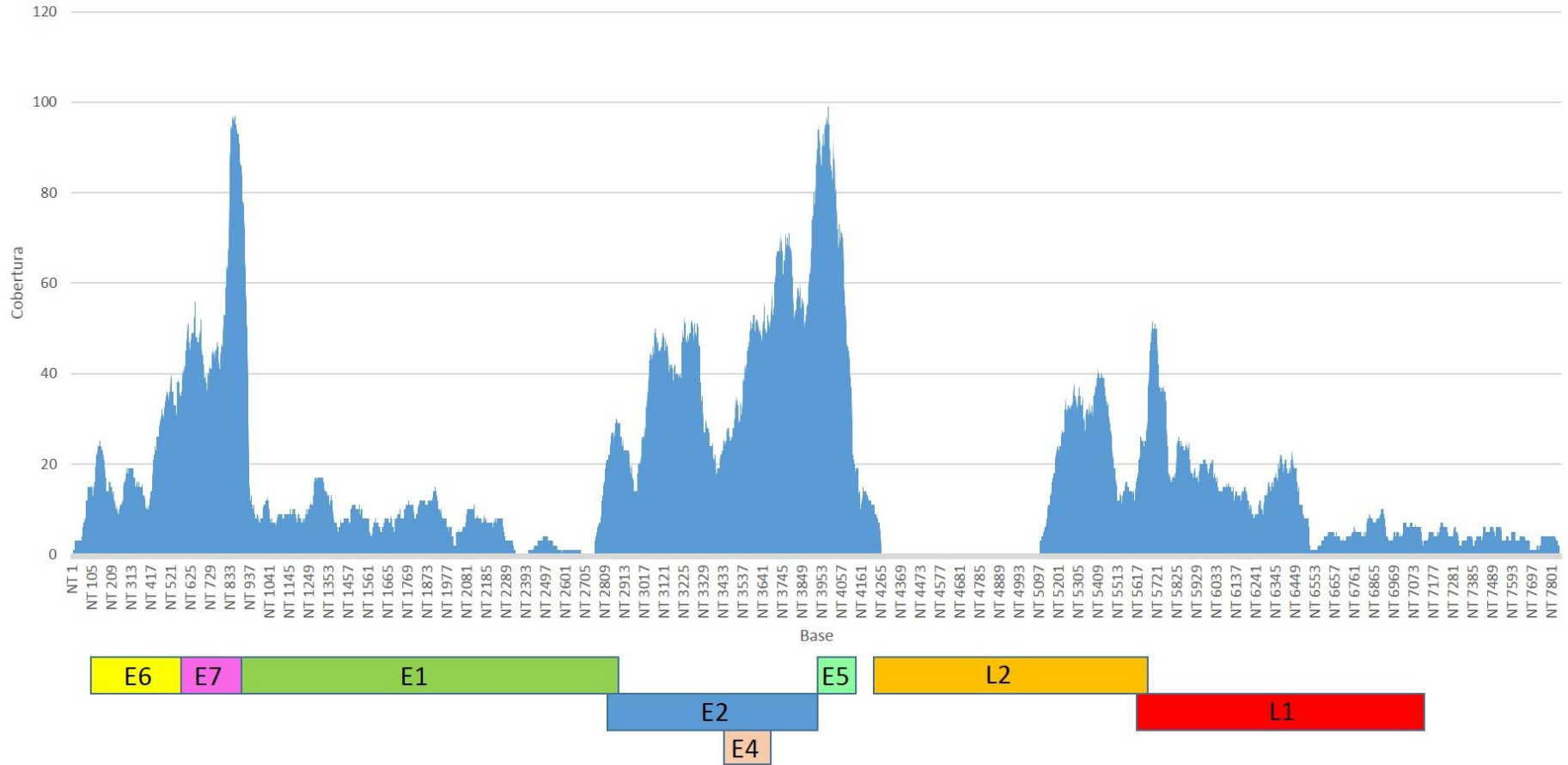




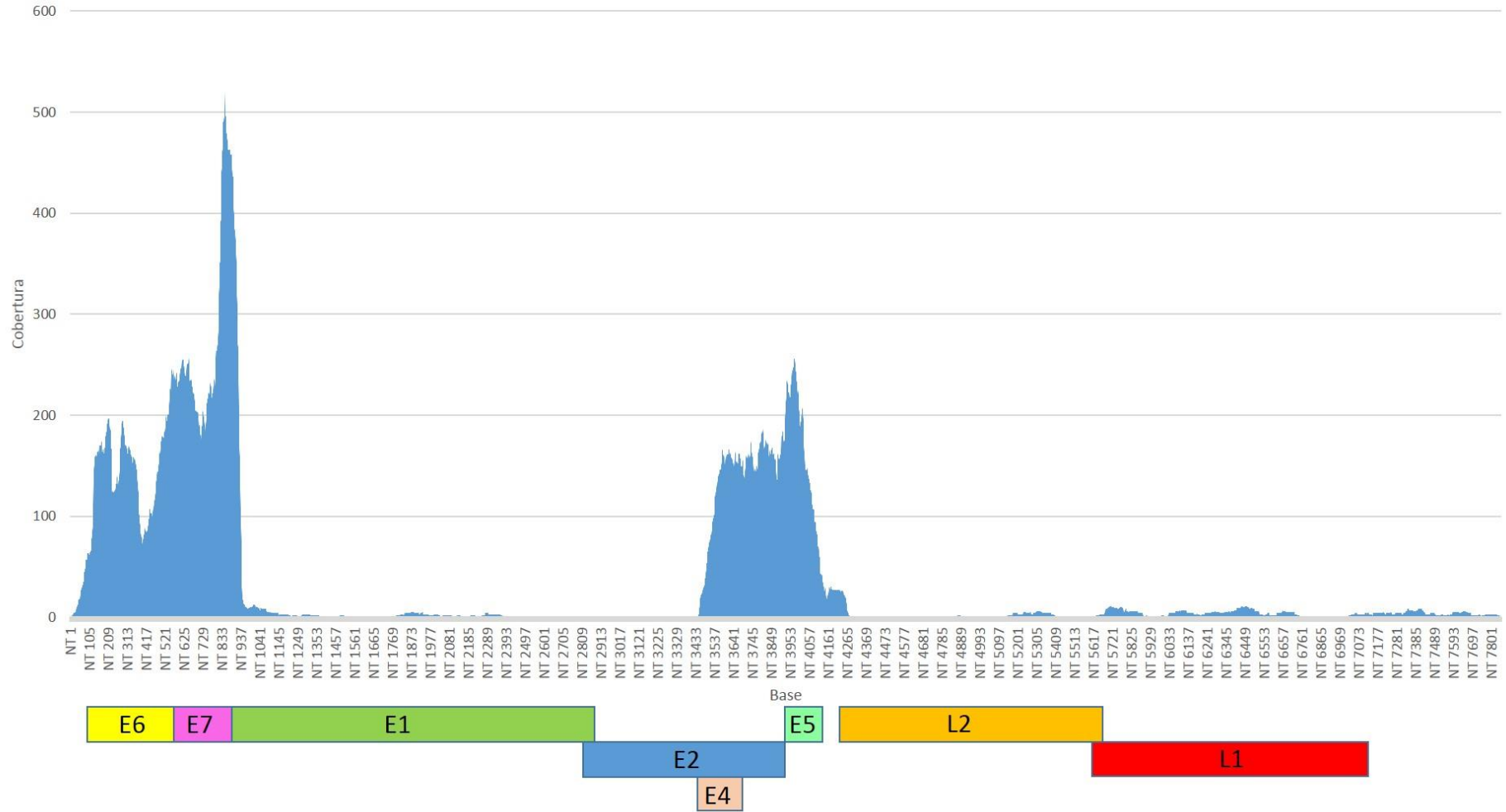
HPV18\_8



HPV18\_9

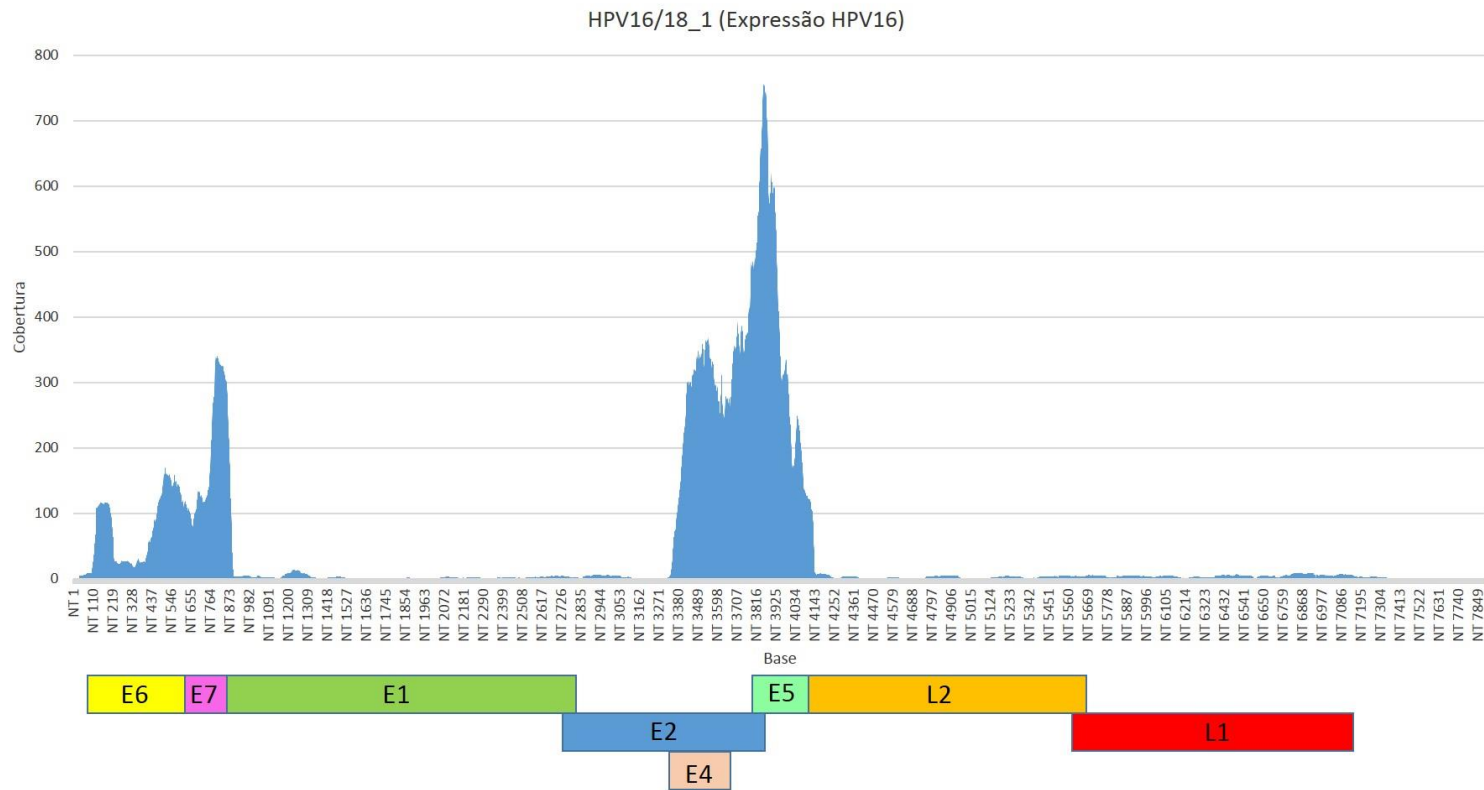


# HPV18\_10



## Anexo 5.5

Gráficos de cobertura por base da expressão do genoma viral da amostra co-infectada pelo HPV 16 e HPV18. As regiões nucleotídicas referentes a cada gene estão demarcadas abaixo do eixo x.



### HPV16/18\_1 (Expressão HPV18)

