UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

Programa de Pós Graduação em Ciências Biológicas (Bioquímica)

Ricardo Cesar Cintra

**Integração do DNA de Papilomavírus Humano no
genoma de células derivadas de esfregaços genitais,
anais e orais masculinos**

SÃO PAULO, 2016

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

Cintra, Ricardo Cesar

C574i Integração do DNA de Papilomavirus Humanos no genoma de
células derivadas de esfregaços genitais, anais e orais masculinos
/ Ricardo Cesar Cintra. -- São Paulo, 2016.
85p.

Dissertação (mestrado) – Instituto de Química da Universidade
de São Paulo. Departamento de Bioquímica.
Orientador : Villa, Luisa Lina

1. Disrupção gênica 2. Vírus oncogênicos 3. Infecção por vírus
de DNA I. T. II. Villa, Luisa Lina, orientador.

574.88

CDD

Ricardo Cesar Cintra

Integração do DNA de Papilomavírus Humanos no genoma de células derivadas de esfregaços genitais, anais e orais masculinos.

Dissertação apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do Título
de Mestre em Ciências (Bioquímica)

Aprovado em: _____

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

AGRADECIMENTOS

Minha sincera gratidão à minha orientadora, Prof^a. Dr^a. Luisa Lina Villa, que é uma fonte inesgotável de entusiasmo e otimismo. Sua sabedoria me motiva a pensar continuamente, questionar e explorar.

À minha família pelo estímulo a gostar de ciência desde muito novo e por proporcionar um ambiente regado de questionamentos acerca do universo, da natureza e do nosso papel no mundo.

Ao Romano, meu companheiro que acompanhou e me apoiou em todas as etapas desse processo.

Aos meus amigos de trabalho, Maria Luiza Baggio, Lenice Galan e Roberto Carvalho pela paciência com minhas ansiedades e compreensão diante das minhas ausências.

À todos os colegas do ICESP-CTO, João Sobrinho, Neide, Jimena, Emma, Emily, Laura, e Miriam com os quais eu tive a colaboração em diversas etapas do meu trabalho.

À Paola e ao Lucas pela paciência e colaboração compartilhando conhecimento e colaborando com técnicas complementares.

À Antonieta, que me iniciou nas práticas de biologia molecular com muita paciência.

Ao meu “*brother*” querido, Carlos DeOcesano-Pereira que me inspira e sempre está ao meu lado na ciência, na filosofia e na arte.

À Raquel e Stella, pela eficiência burocrática que por diversas vezes tornou a minha vida mais fácil e por serem pessoas extremamente humanas e atenciosas.

À todos os funcionários do Instituto de Química e do Departamento de Bioquímica pelos diversos serviços prestados.

Às minhas queridas Rafa, Gabi e Giana que me confortaram em todos os momentos de insegurança dentro e fora do laboratório.

Ao meu “*partner*” Lanre pela constante troca de conhecimento ao longo de todos esses anos.

À Jéssica e ao Lucas pelo suporte de bioinformática e acolhimento em Ribeirão Preto.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), CNPq e Capes pelo suporte financeiro.

"There's Real Poetry in the Real World, Science is the Poetry of Reality."

Richard Dawkins

RESUMO

Cintra, R.C. **Integração do DNA de Papilomavírus Humanos no genoma de células derivadas de esfregaços genitais, anais e orais masculinos.** 2016. 119 p. Dissertação de Mestrado - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

Infecções pelos Papilomavirus Humanos (HPVs) estão relacionadas com uma série de malignidades nas regiões genitais, anais e recentemente a uma parcela dos tumores na região da cabeça e pescoço. Enquanto muito já fora explorado a respeito dessas infecções em mulheres, os fatores de risco virais e epidemiológicos para persistência da infecção e desenvolvimento de lesão em homens ainda estão sendo estabelecidos. Dessa forma, em meados de 2005, começou o Estudo HIM (*HPV in men*), um estudo multicêntrico envolvendo 4200 homens que tem como principal objetivo um melhor entendimento da história natural das infecções por esses vírus nessa população. Sabe-se que um importante evento no processo neoplásico em lesões do colo do útero é a integração do DNA viral no genoma da célula hospedeira. Embora o mecanismo de integração ainda não esteja claro, parece haver uma relação da progressão da doença com o aumento da proporção de lesões com genomas na forma integrada. Diante do exposto, o objetivo desse trabalho foi explorar as metodologias baseadas em PCR (convencional e em tempo real) e de sequenciamento de última geração (*NGS*) para verificar integração do DNA viral nas células de esfregaços genitais, do canal anal e cavidade oral de homens assintomáticos participantes do Estudo *HIM*. Além disso, teve como objetivo secundário verificar a carga viral nos diferentes sítios anatômicos. Foram inicialmente incluídas 125 amostras das quais 56 puderam ser incluídas nas análises de integração pela metodologia baseada em PCR e 90 amostras foram selecionadas para os ensaios de carga viral. Dentre as amostras testadas pela metodologia de PCR para integração, 11 foram desafiadas pela metodologia baseada em sequenciamento. Entre as 11 amostras sequenciadas, encontramos concordância entre os resultados em apenas 3 e isso pode ser atribuído à sensibilidade diferencial das metodologias diante de baixas cargas virais. Entre as amostras testadas pela metodologia baseada em PCR pudemos verificar que 30,4% das amostras

apresentavam-se na forma puramente epissomal, 10,7% na forma puramente integrada e 58,9% das amostras apresentam as duas formas virais (classificadas como mistas). A amostra que se confirmou integrada pela metodologia de sequenciamento apresentou integração em 4 pontos do DNA humano com a ruptura no gene viral *E2*, conforme amplamente relatado na literatura. Grande variação na carga viral foi encontrada entre as amostras (0,00008 a 4200 cópias/célula), e isso se mostrou diretamente relacionado com as limitações das metodologias para verificar integração. Embora outras relações não puderam ser verificadas em nosso estudo pelo limitado número de amostras analisadas até o momento, estes dados ampliados poderão responder se a carga viral e integração seriam um fator de risco para a persistência da infecção ou o desenvolvimento de lesão em homens.

Palavras-chaves: HPV, estado físico, integração do DNA viral, HPV em homens, carga viral.

ABSTRACT

Cintra, R.C. **Human Papillomavirus DNA integration into cells from genital, anal and oral smears from men.** 2016. 83 p. Master's Dissertation – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Human Papillomavirus (HPV) infections are associated with a range of malignancies in the genital and anal regions and recently, with an appreciable number of head and neck tumors. While much has been explored about these infections in women, viral and epidemiological risk factors for persistent infection and lesions development in men are still being established. Thus, in mid-2005, a multicenter study entitled *HIM study* (HPV in men) - involving 4200 men - which aims at a better understanding of the natural history of infection by these viruses in this population, was started. It is known that a key event in the transformation process in cervical lesions is the integration of viral DNA into the host cell genome. Although the integration mechanism is not yet clear, there seems to be a relationship of the disease progression with an increase of the proportion of lesions with integrated viral genomes. In view of the above, the aim of this study was to explore the methodologies based on PCR (conventional and real-time) and next-generation sequencing (NGS) to verify integration of viral DNA into cells from genital swabs, anal canal and oral cavity of asymptomatic men participating in the HIM study. Furthermore, the secondary objective was to verify the viral load in cells from different anatomical sites. Initially, 125 samples were selected of which 56 could be included in integration tests with the methodology based on PCR and 90 samples were selected for viral load determination. Of the samples tested by PCR method for integration, 11 were subjected DNA sequencing. We found correlation between the results in only 3 of the 11 samples and this can be attributed to the methodologies' differential sensitivity to low viral loads. Among the samples tested by PCR-based methodology, we observed that 30.4% of the samples harbored purely episomal viral DNA, 10.7% were integrated and 58.9% of the samples showed the two viral forms (classified as mixed). The

sample which was confirmed as integrated by sequencing presented integration into 4 human DNA points with the breakpoint within the *E2* viral gene, as widely reported in literature. Large variation in viral load was observed among the samples (0.00008 to 4200 copies / cell), and this was shown to be directly related to the limitations of the methods used to verify integration. Evaluation of a larger number of samples may contribute to define the role of HPV DNA integration as a risk factor for the persistence of HPV infection or development of lesions in men.

Keywords: HPV, physical state, DNA viral integration, HPV in men, viral load.

Lista de Abreviaturas e Siglas

ATCC	<i>American Type Culture Collection</i>
ATPase	Enzima com atividade de hidrólise de ATP
ATP	Trifosfato de Adenosina
BHQ	<i>Black Hole Quencher</i>
CDK	Cinase Dependente de Ciclina, do inglês <i>Cyclin Dependent Kinases</i>
CFS	Sítios Frágeis Comuns, do inglês <i>Common Fragile Sites</i>
Ct	Threshold Cycle
ddPCR	PCR digital em gotas, do inglês <i>Droplet Digital PCR</i>
DEPC	Dietilpirocarbonato
DIN	<i>DNA Integrity number</i>
DMSO	Dimetilsulfóxido
DNA	Ácido Desoxirribonucléico
dNTP	Desoxirribonucleotídeos Trifosfatados
E2BS	Sítio de ligação da proteína E2, do inglês <i>E2 Binding Site</i>
EDTA	Ácido Etilenodiamino Tetra-Acético
EGF	Fator de crescimento epidérmico, do inglês <i>Epidermal Growth Factor</i>
EGFR	Receptor do fator de crescimento epidérmico, do inglês <i>Epidermal Growth Factor Receptor</i>
FAM	Fluoróforo utilizado nas sondas de hidólise, abreviação de Amidita de Fluoresceína
HDAC	Histona Deacetilase
HIM	<i>HPV in men</i>
HIV	Vírus da Imunodeficiência Humana
Hi-Q	Conjunto de reagentes utilizados no sequenciador <i>PGM Ion Torrent</i>
HSPG	Proteoglicano Heparan-Sulfato
hTERT	Transcriptase Reversa Telomerase humana, do inglês <i>human Telomerase Reverse Transcriptase</i>
IARC	Agência Internacional de Pesquisa em Câncer
ICTV	Comitê Internacional de Taxonomia em Vírus
kb	Quilobases
kDa	Quilodáltons (unidade de peso molecular)
LCR	Região Controladora Longa, do inglês <i>Long Control Region</i>
MFI	Intensidade Média de Fluorescência do inglês <i>Median Fluorescence Intensity</i>
MHC	Complexo de Histocompatibilidade Principal, do inglês <i>Major Histocompatibility Complex</i>
mRNA	RNA mensageiro
NGS	Sequenciamento de Nova Geração, do inglês <i>Next Generation Sequencing</i>
NIH	<i>National Institute of Health</i>
pAE	Sítio de poliadenilação precoce
pAL	Sítio de poliadenilação tardio
pb	Pares de bases
PBS	Tampão Fosfato Salino, do inglês <i>Phosphate-Buffered Saline</i>
PC	Pró-proteína Convertase
PCR	Reação em Cadeia da Polimerase
PDZ	Classe de Proteínas com domínio estrutural característico (domínios PDZ)
PeIN	Neoplasia Intraepitelial Peniana do inglês <i>Penile Intraepithelial Neoplasia</i>
PGM	<i>Personal Genome Machine</i>
pRB	Proteína do Retinoblastoma
qPCR	PCR quantitativa
RPA	Proteína de Replicação A, do inglês <i>Replication Protein A</i>
rpm	Rotações por minuto
RNA	Ácido Ribonucléico
STM	Meio de transporte de Espécime, do inglês <i>Specimen Transport Medium</i>
VLP	Partículas Semelhantes a Vírus, do inglês <i>Virus-like Particles</i>
Tris-HCl	Tampão que utiliza Tris (hidroximetil-aminometano) como base e ácido clorídrico.
TE	Tampão que utiliza Tris (hidroximetil-aminometano) e EDTA

URR	Região regulatória a montante, do inglês <i>Upstream Regulatory Region</i>
UV	Ultravioleta
UTR	Região não traduzida, do inglês <i>Untranslated Region</i>

ÍNDICE

SUMÁRIO

ÍNDICE.....	11
1. INTRODUÇÃO	13
1.1. A descoberta do HPV como agente causador de câncer	13
1.2. HPV	14
1.3. Biologia do HPV	18
1.4. Ciclo biológico e o papel das proteínas virais	21
1.5. Potencial Oncogênico e Integração do genoma viral	25
1.6. A infecção por HPV em homens e o estudo <i>HIM</i>	32
2. OBJETIVOS	35
2.1. Objetivo geral.....	35
2.2. Objetivo Específico.....	35
2.3. Objetivo Secundário.....	35
3. MATERIAIS E MÉTODOS	36
3.1. Amostras clínicas	36
3.2. Extração de DNA	38
3.2.1. Digestão com Proteinase K e Purificação pelo método Orgânico de Fenol Clorofórmio seguido de precipitação em Etanol.....	38
3.3. Reação em Cadeia de Polimerase (<i>Polymerase Chain Reaction</i> - PCR)	39
3.3.1. PCR convencional de β -actina - avaliação da integridade do DNA.....	39
3.3.2. PCR convencional para análise do gene <i>E2</i> completo de HPV 16.....	40
3.4. PCR quantitativa em tempo real – determinação da integração do DNA do HPV e carga viral .	43
3.4.1. PCR quantitativa do gene <i>E6</i> – HPV 16 – integração do HPV e carga viral	45
3.4.2. PCR quantitativa do gene <i>E2</i> – HPV 16 – integração do HPV	46
3.4.3. PCR quantitativa do gene da <i>β-globina humana</i> – carga viral.....	47
3.5. <i>Ion Torrent NGS – Next Generation Sequencing</i>	48
3.5.1. Análise Bioinformática	51
4. RESULTADOS	53
4.1. Seleção das amostras	53
4.2. Aplicação da metodologia previamente aplicada no laboratório (desenvolvida por Trevisan et al.)....	54
4.3. Aprimoramento da metodologia adaptada para esta dissertação (desenvolvida por Cintra et al.).	55
4.3.1. Verificando a integração do genoma viral	56
4.3.2. Verificando a carga viral do HPV	60
4.3.3. <i>TargetSeq</i> para o Sequenciamento de Nova Geração (<i>NGS</i>)	63
5. DISCUSSÃO.....	67
6. CONCLUSÕES.....	74

7.	REFERÊNCIAS.....	75
8.	LISTA DE ANEXOS	85

1. INTRODUÇÃO

1.1. A descoberta do HPV como agente causador de câncer

Muitas doenças causadas por vírus conhecidas atualmente têm sido descritas desde a antiguidade. Embora médicos gregos e romanos já discutissem sobre verrugas genitais desde 500 d.C., a atenção era centrada mais no tratamento do que no agente causador. Dessa forma, apenas nos últimos dois séculos, a ciência e a tecnologia desenvolveram instrumentos e métodos que permitiram identificar e caracterizar esses patógenos como agentes capazes de infectar, reproduzir-se e transmitir-se de uma célula hospedeira a outra (Garfield, 1988).

Com o advento da microscopia eletrônica, Maurice Strauss e colaboradores relataram pela primeira vez, em 1949, a observação de estruturas semelhantes a vírus retiradas de papilomas da pele, os quais um ano mais tarde foram identificados pelo mesmo grupo de pesquisadores como o papilomavírus humano – HPV – o agente etiológico das verrugas (Strauss et al., 1949; Garfield, 1988).

A associação entre o HPV e o câncer do colo do útero foi primeiramente sugerido por zur Hausen em 1974 (zur Hausen et al., 1974) e a comprovação da hipótese publicada pela Agência Internacional para Pesquisa em Câncer da Organização Mundial da Saúde em 1995 (IARC Working Group 1995). Ainda na década de 70, foram identificados diversos tipos de HPV em lesões de pele e de mucosas como verrugas genitais, carcinoma de colo do útero e displasias epiteliais, reforçando a importância médica de se estudar o HPV (Doorbar & Sterling, 2001). Dessa forma, o desenvolvimento de técnicas moleculares cada vez mais elaboradas ao longo dos últimos 40 anos tem levado a um rápido aumento na compreensão da biologia do HPV. Muito deste esforço tem incidido sobre a propensão de que certos tipos de HPV podem contribuir para a transformação maligna.

Hoje é largamente aceito que o HPV é o agente causador de pelo menos 99,7% dos cânceres de colo de útero e que infecções persistentes por tipos de alto risco oncogênico são

um fator de risco para o desenvolvimento de lesões precursoras (Walboomers et al., 1999; Schlecht et al., 2001). Os tipos HPV 16 e 18 estão presentes em aproximadamente 70% de todos os casos de câncer de colo uterino (Muñoz et al., 2004). Além disso, o tipo HPV 16 sozinho está relacionado com até 90% de todas as malignidades associadas ao HPV em outros sítios anatômicos e é o tipo mais estudado e melhor caracterizado dos pontos de vista molecular e epidemiológico (Doorbar, 2006; Allen et al., 2010; Schwartz, 2013).

Por constituírem um grupo heterogêneo de vírus, além do câncer do colo uterino, câncer de ânus, câncer de vagina, vulva e pênis, os diferentes tipos de HPV estão relacionados também com lesões benignas de pele, verrugas genitais, papilomatose laríngea e, mais recentemente, ligado a uma parcela dos cânceres de cabeça e pescoço (zur Hausen, 2009; Forman et al., 2012).

1.2. HPV

Atualmente o HPV está classificado na família *Papillomaviridae* (ICTV, 2013). Estão divididos fundamentalmente em 5 gêneros baseado em diferenças na sequência de DNA da região mais conservada do genoma viral (porção do gene *L1*) e considera-se um novo tipo de HPV quando essa sequência nucleotídica diferir em mais de 10% dos tipos já descritos (revisado por Doorbar et al., 2015). Dessa forma, atualmente, foram descritos mais de 170 tipos de HPV dos quais aproximadamente 40 deles infectam o trato anogenital, estes últimos, com tropismo por epitélio mucoso ou cutâneo, são do gênero *Alphapapillomavirus* (de Villiers et al., 2013) (FIGURA 1).

HPV 31, HPV 33, HPV 35, HPV 39, HPV 45, HPV 51, HPV 52, HPV 56, HPV 58, HPV 59) (Muñoz et al., 2003; revisado por Doorbar et al., 2012). Embora a infecção pelo HPV, mesmo pelos tipos de alto risco oncogênico sejam muito comuns, mais de 90% das novas infecções regredem entre 6 – 18 meses (Franco et al., 1999; Castle et al., 2009).

Os HPVs de alto risco oncogênico têm como característica comum uma maior probabilidade de persistir e de integrar o seu material genético no genoma da célula hospedeira durante seu ciclo biológico. Consequência disso é um potencial aumentado de alterações genéticas permanentes que podem conduzir a uma situação de instabilidade genômica e ao próprio processo de transformação maligna (revisado por Chen, 2014). Dessa forma, a Agência Internacional para pesquisa em Câncer, o IARC, classificou esses agentes infecciosos como carcinógenos humanos de tipo I (IARC, 2005) (TABELA 1).

Tabela 1- Porcentagem dos cânceres de colo de útero que são atribuídos à infecção pelos tipos do gênero *Alphapapillomavirus*:

Gênero + Espécie	Tipo	Cancer Cervical Invasivo	Categoria IARC	Carcinoma de Celulas Escamosas	Adenocarcinoma	Tropismo
Alpha 1	HPV 32		3			mucoso
	HPV 42					
Alpha 2	HPV 3					cutâneo
	HPV 10					
	HPV 28		3			
	HPV 29		3			
	HPV 77		3			
	HPV 94					
	HPV 117					
	HPV 125					mucoso
Alpha 3	HPV 61	0,01	3			
	HPV 62		3			
	HPV 72		3			
	HPV 81		3	0,4		
	HPV 83		3	0,4		
	HPV 84		3			
	HPV 86		3			
	HPV 87		3			
	HPV 89		3			
	HPV 102					
	HPV 114					cutâneo
Alpha 4	HPV 2		3			
	HPV 27		3			
	HPV 57		3			mucoso
Alpha 5	HPV 26	0,37	2B	0,22		
	HPV 51	1,25	1	0,75	0,54	
	HPV 69	0,08	2B			
	HPV 82	0,07	2B	0,26		
Alpha 6	HPV 30	0,37	2B			
	HPV 53	0,26	2B	0,04		
	HPV 56	0,84	1	1,09		
	HPV 66	0,08	2B	0,19		
Alpha 7	HPV 18	10,28	1	11,27	37,3	
	HPV 39	1,67	1	0,82	0,54	
	HPV 45	5,68	1	5,21	5,95	
	HPV 59	1,08	1	1,05	2,16	
	HPV 68	1,04	2A	0,37		
	HPV 70	0,11	2B			
	HPV 85		2B			
	HPV 97					
Alpha 8	HPV 7		3		41,62	cutâneo (mucoso)
	HPV 40		3		1,08	
	HPV 43				0,54	
	HPV 91	0,01	3		1,08	mucoso
Alpha 9	HPV 16	61,35	1	54,38		
	HPV 31	3,65	1	3,82	0,54	
	HPV 33	3,83	1	2,06		
	HPV 35	1,94	1	1,27		
	HPV 52	2,71	1	2,25		
	HPV 58	2,22	1	1,72		
	HPV 67	0,31	2B			mucoso
Alpha 10	HPV 6	0,11	3	0,07		
	HPV 11	0,02	3	0,07		
	HPV 13		3			
	HPV 44	0,01	3			
	HPV 74	0,01	3			mucoso
Alpha 11	HPV 34	0,07				
	HPV 73	0,52	2B	0,49		mucoso
Alpha 12						
Alpha 13	HPV 54					
Alpha 14	HPV 71					
	HPV 90		3			
	HPV 106		3			

Os tipos das espécies Alpha-7 e Alpha-9 identificados em vermelho foram classificados como carcinógenos humanos com base em dados epidemiológicos e têm sido os mais estudados. Adaptado de Doorbar, 2015.

1.3. Biologia do HPV

Os HPVs são vírus não envelopados, de simetria icosaédrica com aproximadamente 50 nm de diâmetro e 72 capsômeros que encerram um DNA de fita dupla circular de aproximadamente 8000 pares de bases. O DNA viral encontra-se associado à proteínas semelhantes à histonas e seu genoma é dividido em uma região precoce (*early* – E), uma região tardia (*late* – L) e uma região reguladora (LCR) onde se encontra o início da replicação (de Villiers et al., 2004; Doorbar et al., 2015). Essas três regiões em todos os papilomavírus são separadas por dois sítios de poliadenilação: sítio de poliadenilação precoce (pAE) e sítio de poliadenilação tardio (pAL) além de dois promotores de transcrição: um promotor precoce (p97) e um promotor tardio (p670) (Doorbar, 2006; Kajitani et al., 2012). Os dois sítios de poliadenilação são precedidos por uma UTR (região 5' não traduzida) que quando transcritos, contêm elementos que determinam a eficiência da poliadenilação, eficiência na tradução e meia-vida do mRNA (FIGURA 2) (revisado por Johansson & Schwartz 2012).

A região precoce ocupa aproximadamente 50% do genoma e codifica para 6 genes não estruturais (*E1*, *E2*, *E4*, *E5*, *E6* e *E7*) enquanto a região tardia codifica para dois genes estruturais (*L1* e *L2*) que são as proteínas principais e secundárias do capsídeo. Entre os genes *L1* e *E6* encontra-se a região não codificante denominada LCR (*Long Control Region*) ou URR (*Upstream Regulatory Region*) que compreende cerca de 10% do genoma viral onde existem regiões estimuladoras e repressoras da transcrição dos genes precoces (FIGURA 2) (Doorbar, 2006; Johansson & Schwartz 2012).

Brevemente, o gene *E1* é altamente conservado e codifica para uma DNA helicase viral dependente de ATP necessária para a replicação e amplificação do genoma do vírus (Tsakogiannis et al., 2014). O produto de expressão do gene *E2* tem múltiplas funções atuando como fator de transcrição, cooperando com *E1* no reconhecimento e ligação à origem de replicação e também no processo de partição dos genomas virais durante o processo de

divisão celular (McBride, 2008; Kurg et al., 2010). A proteína E2 ainda é capaz de reconhecer alguns sítios de ligação na LCR do genoma viral promovendo dessa forma a regulação da expressão dos genes *E6* e *E7* (Doorbar, 2006; Kajitani et. al 2012). Já o papel da proteína E4 parece estar relacionado com a instabilidade de citoqueratinas e maturação da partícula viral, promovendo uma melhor movimentação dessas moléculas dentro das células em processo mais avançado de diferenciação, como sugerem alguns estudos (Doorbar et. al, 1991; Roberts et. al, 1997; McIntosh et.al, 2010).

Figura 2 - Organização genômica do HPV-16: O genoma do HPV 16 é constituído por um DNA dupla-fita circular. Os genes são transcritos em uma única direção. Há genes que codificam para proteínas não estruturais (*E1*, *E2*, *E4*, *E5*, *E6* e *E7*), genes que codificam para proteínas estruturais (*L1* e *L2*) e uma região reguladora URR (também chamada de LCR). Na URR encontra a origem de replicação do DNA viral. Os mais importantes promotores e sítios de poliadenilação são indicados por p97, p670, pAE e pAL (Harari et al., 2014).

O produto de expressão do gene *E5* está relacionado com evasão do sistema imune por interferir na alcalinização dos endossomos e moléculas apresentadoras de antígenos, além disso, tem a função de amplificar o estímulo de proliferação pela sua ligação com o receptor do fator de crescimento epidérmico – EGF- (Hwang et al., 1995; Zhang et al., 2003; Disbrow et al., 2005) . Os genes *E6* e *E7* codificam produtos capazes de interagir com diversas proteínas celulares incluindo os mecanismos de proliferação celular, atraso no processo de diferenciação, tolerância aumentada à apoptose e imortalização, favorecendo à infecção produtiva (Fu et al., 2010; Doorbar et al., 2012; Mighty and Laimins, 2014) . O papel de cada uma dessas proteínas no processo oncogênico do HPV será descrito com maiores níveis de detalhamento posteriormente.

Os genes da região tardia tem função estrutural e compõem as proteínas principais e secundárias do capsídeo, respectivamente L1 e L2. A proteína L1 é a proteína mais abundante do capsídeo viral constituindo cerca de 80% do total das proteínas virais. Esta, junto com a proteína L2 coopera no processo de internalização do DNA viral dentro do vírion (revisado por Zhou and Frazer, 1996). Pela proteína L1 ser altamente imunogênica e ser capaz de formar partículas virais vazias (*VLPs –virus-like particles* ou partículas semelhantes a vírus) tem sido utilizada no desenvolvimento de vacinas profiláticas (Villa et al., 2005; revisado por Galloway, 2003).

A regulação da transcrição dos genes virais é complexa e controlada por fatores de transcrição celulares e virais. Todas as espécies de RNA virais são transcritos de forma policistrônica de uma única cadeia do DNA do vírus e estão sujeitas a extensivo e complexo mecanismo de *splicing* alternativo de acordo com o processo de diferenciação da célula infectada (McBride, 2008; Johansson and Schwartz, 2013).

1.4. Ciclo biológico e o papel das proteínas virais

O tecido alvo para a infecção pelo HPV é o epitélio estratificado e seu ciclo biológico está intimamente ligado ao programa de diferenciação celular deste epitélio. No epitélio estratificado normal, as células da membrana basal são as únicas células com potencial de proliferação. Estas células se dividem e as células filhas seguem no processo de diferenciação. Durante esse processo, a célula filha sai do ciclo celular e muda sucessivamente seu padrão de expressão gênica até a diferenciação terminal, quando é descamada do epitélio (Mole et al., 2006; Kajitani et al., 2012).

A infecção inicial pelo HPV começa na camada basal do epitélio, onde o vírus ganha acesso através de possíveis lesões que exponham essa superfície. Na membrana basal, a ligação das proteínas do capsídeo com cadeias do proteoglicano heparan-sulfato (HSPG) levam a uma mudança conformacional que permite a clivagem da porção N-terminal da proteína L2 (proteína secundária do capsídeo). Essa clivagem é realizada pelas pró-proteínas convertases furina e ou PC5/6 (Furina-PC5/6), expondo um epítipo capaz de interagir com a superfície celular e ganhar a entrada na célula (Kines et al., 2009). A precisa natureza de um receptor celular para a internalização do vírus continua controversa embora os receptores de fatores de crescimento e a α -6-integrina estejam envolvidos nesse processo (Doorbar et al., 2006; Surviladze et al., 2012). Uma vez internalizados, os vírions são submetidos ao transporte endossomal onde são desencapsulados e liberados dentro da célula. O complexo proteína L2- DNA garante sua internalização no núcleo celular enquanto a proteína L1 é mantida no endossomo e sujeita à degradação lisossomal (Doorbar et al., 2012).

Dentro do núcleo das células basais (células indiferenciadas com grande atividade mitótica), o ciclo entra na fase de manutenção do genoma viral, onde o vírus na forma epissomal é mantido em um baixo e estável número de cópias (50-100 cópias por célula) e sua amplificação está ligada exclusivamente à progressão do ciclo celular (Moody and Laimins,

2010). Nessa fase, a transcrição dos genes *E1* e *E2* são os requisitos básicos para a utilização da maquinaria de replicação celular. Nas células basais, a proteína E2 desempenha diversos papéis importantes tanto no processo de replicação quanto no processo de segregação genômica viral. E2 tem um domínio de ligação ao DNA que reconhece 4 sequências palindrômicas contidas na LCR viral, sendo uma delas adjacente à origem de replicação. A ligação de E2 a este sítio é necessária para o recrutamento da helicase viral E1 para a origem de replicação, que por sua vez liga-se a outras proteínas necessárias para a replicação do DNA incluindo RPA (*replication protein A*) e DNA polimerase celular (Stenlund, 2003; Doorbar et al., 2006). Nessa camada de células, o genoma viral é então mantido como elementos replicativos extracromossomais sincronizados com o ciclo celular servindo como reservatório de células infectadas (McBride, 2008).

No processo de divisão celular das células basais, a proteína E2 atua na partição dos genomas virais ligando os epissomos à cromatina e garantindo tanto a segregação igualitária nas células filhas quanto a sua localização nuclear. Isso é feito através da interação de E2 com outras proteínas celulares incluindo Brd4 (Whang et al., 2013; Doorbar et al., 2015). Embora as proteínas virais E1 e E2 sejam obrigatórias nessa etapa de manutenção e amplificação do epissomo viral, as atividades regulatórias de E6 e E7 atuam como um fator potencializador do processo de replicação viral por promoverem a proliferação e divisão celular da camada basal (Doorbar et al., 2012) (FIGURA 3).

Conforme as células infectadas se dividem e migram para as camadas superiores do epitélio, seguindo pelas rotas de diferenciação, as células diminuem a expressão de fatores celulares que são requeridos para a replicação do genoma viral e para que o vírus mantenha seu processo replicativo, outras proteínas virais têm suas atividades mais acentuadas. Nesse processo, as proteínas E5, E6 e E7 cooperam no sentido de manter a atividade replicativa viral através de inúmeras interações com proteínas da célula hospedeira (Johansson & Schwartz,

2013). Em suma, a proteína E7 tem a propriedade de se ligar às formas hipofosforiladas das proteínas da família do retinoblastoma (pRb) – p105, p107 e p130 - degradando-as. As duas primeiras (p105 e p107) controlam a entrada no ciclo celular na camada basal enquanto que a última (p130) está envolvida na reentrada no ciclo celular nas camadas superiores e mais diferenciadas do epitélio. Como consequência, há a liberação do fator de transcrição E2F que dirige a expressão de genes da fase S como as enzimas de replicação do DNA. Além disso, E7 ainda mantém a atividade dos complexos ciclina E- Cdk2 e ciclina A-Cdk2 em células já diferenciadas por interagir com as proteínas inibidoras p27 e p21, importantes reguladoras da parada do ciclo celular no processo de diferenciação (Munger et al., 2004; Moody & Laimins, 2010).

Por outro lado, altos níveis de E7 leva a uma indução da proteína supressora p53, o que resultaria na parada no ciclo celular e apoptose. Para contornar esse desfecho, a ação da proteína E6 interfere nesse processo, levando p53 à degradação proteassômica (Scheffner et al., 1990; Doorbar et al., 2015). Durante essa etapa, a proteína E5 coopera nesse processo, ligando-se ao receptor do fator de crescimento epidérmico (EGFR) causando sua dimerização que resulta em um sinal mitogênico contínuo (Bouvard et al., 1994; Venuti et al., 2011). Esses fatores associados mantêm um ambiente favorável para uma eficiente replicação do genoma viral mantendo as células de estratos superiores do epitélio em um estado como a fase de síntese (S) do ciclo celular. É importante lembrar que as propriedades-chaves das proteínas E6 e E7 não são as de promover a proliferação das células da camada basal, mas ao invés disso, manter o estímulo de reentrada no ciclo celular nas camadas intermediárias e superiores do epitélio para promover o aumento do número de cópias do genoma viral (Doorbar et al., 2012).

Como parte do ciclo viral, altos níveis da proteína E2 promovidos pelo número aumentado de cópias do genoma viral decorrentes da fase replicativa, favorece sua

dimerização e ligação ao E2BS que sobrepõem o promotor precoce p97, regulando negativamente a expressão das oncoproteínas E6 e E7. Esse evento permite à célula infectada sua diferenciação em resposta a sinais celulares (McBride, 2008; Thierry, 2009).

A mudança de promotor dependente de diferenciação, ou seja, a ativação do promotor tardio p670, promove uma expressão ainda maior das proteínas E1 e E2 o que contribui ainda mais com a amplificação do genoma viral. Embora o promotor tardio tenha sido ativado pelo processo de diferenciação, o sítio de poliadenilação nessa fase continua sendo o sítio de poliadenilação precoce, sem a expressão das proteínas tardias L1 e L2 (Doorbar, 2006).

A mudança de promotor promove ainda a diferenciação celular até a diferenciação terminal e o acúmulo da proteína E4 promoverá um colapso na rede de filamentos intermediários celulares favorecendo a liberação dos vírions. A diferenciação terminal regula negativamente a atividade do sítio de poliadenilação precoce, resultando na expressão das proteínas tardias L1 e L2 através do sítio de poliadenilação tardio (Doorbar, 1991; Johansson & Schwartz, 2013). Como L1 e L2 são altamente imunogênicas, o controle de expressão dessas proteínas as mantém confinadas apenas nos estratos mais diferenciados onde os vírus são liberados com as células em descamação num ciclo não lítico, contribuindo ainda mais para a evasão imune e persistência viral (Stanley, 2008; Kajitani et al., 2012) (FIGURA 3).

O ciclo biológico dos Papilomavírus leva de 2 a 3 semanas, o tempo necessário para uma célula migrar da camada basal para as camadas mais superficiais do epitélio (Stanley, 2010).

Figura 3 - Eventos-chave que ocorrem durante a diferenciação do epitélio: Expressão das proteínas de acordo com a diferenciação do epitélio infectado com as consequentes estratégias virais explicitadas ao lado direito da figura. Adaptado de (Doorbar, 2006).

Como os HPVs, incluindo o HPV 16, são bem adaptados ao seu hospedeiro e normalmente causam infecções assintomáticas, na maioria dos indivíduos infectados, a persistência definida epidemiologicamente costuma ser menor de que um ano, o que é longa o suficiente para transmissão para outros hospedeiros, mas não tão longa para causar doenças como uma neoplasia. Porém, a persistência a mais longo prazo é a principal determinante de risco para neoplasia intraepitelial. O desenvolvimento de neoplasia geralmente surge em indivíduos que são incapazes de eliminar ou manter o vírus em estado de latência, mantendo a expressão dos oncogenes por anos ou décadas (Vinokurova, 2008; Chow *et al.*, 2009).

1.5. Potencial Oncogênico e Integração do genoma viral

A carcinogênese induzida por HPV envolve tanto mudanças genéticas quanto epigenéticas, culminando na ativação de proto-oncogenes e a inativação de genes supressores de tumor (Doorbar, 2006; Bodily & Laimins, 2011). Porém, um fator importante para o processo oncogênico é o escape do sistema imune e consequente persistência da infecção que é alcançado com a expressão extremamente baixa dos genes de expressão precoce (*E1*, *E2*,

E5, E6, E7) nas células indiferenciadas da camada basal e uma indução da síntese de níveis mais altos dessas oncoproteínas assim que as células se dividem e avançam no programa de diferenciação (Ylitalo, 2000; Kajitani et al., 2012). Embora o mecanismo de evasão seja complexo e envolva modulação da sinalização e reconhecimento de citocinas, a oncoproteína *E5* dos HPV de alto risco tem um papel fundamental no processo de reconhecimento pelo sistema imune já que, ao se ligar à bomba de prótons (Bomba H^+ -ATPase) dos endossomos, diminui sua habilidade em acidificar esse compartimento. Como resultado, interfere negativamente no carregamento e exposição dos antígenos pelos complexos de histocompatibilidade (MHC I e II) na superfície celular (Zheng et. al, 2003; Drisbow et al, 2005; Ashrafi, 2006; Crosbie et al., 2013).

Os diferentes tipos de HPV compartilham uma organização genômica comum e as disparidades entre os tipos de alto risco e baixo risco oncogênicos se baseiam em diferenças entre aminoácidos presentes em suas oncoproteínas e suas habilidades em interagir e levar proteínas do sistema de controle do ciclo celular às vias de degradação como consequência das estratégias de manutenção e replicação do genoma viral (Munger et al., 2004).

No processo de carcinogênese viral mediada pelo HPV, as oncoproteínas *E6* e *E7* desempenham papel central por interagirem diretamente com importantes produtos de genes supressores de tumor. Uma característica digna de nota é a multifuncionalidade dessas oncoproteínas resultante da bem adaptada relação com o hospedeiro e, embora as ações de *E6* e *E7* sejam bem caracterizadas nos processos de degradação de *p53* e proteínas do retinoblastoma como citado anteriormente, estudos recentes têm demonstrado que alterações de vias adicionais modificadas por elas são igualmente importantes para o processo tumorigênico. Esses fatores adicionais são reguladores cruciais da progressão do ciclo celular, manutenção da telomerase, apoptose e estabilidade genômica (Moody & Laimins, 2010; Mesri, 2014).

A proteína E7 dos HPV de alto risco é uma proteína pequena, de aproximadamente 13 kDa, que não possui nenhuma atividade enzimática intrínseca ou de ligação ao DNA mas é capaz de se ligar a diversos fatores celulares além das proteínas do retinoblastoma-pRB (McLaughlin-Drubin & Münger, 2009; Moody & Laimins, 2010). Um importante papel da atividade de E7 é sua capacidade de interagir com histonas deacetilases (HDACs) – enzimas chaves na regulação epigenética - em domínios distintos daqueles que o faz com pRB e através disso evitar a formação da heterocromatina em promotores específicos. Isso tem sido demonstrado importante na manutenção do genoma viral e na contribuição para a manutenção da habilidade proliferativa e longevidade celular (Longworth & Laimins, 2004). Além disso, a proteína E7 tem mostrado aumentar os níveis da fosfatase CDC25A, através da ativação de seu promotor. Essa fosfatase específica induz a desfosforilação do resíduo de tirosina da Cdk2 e conseqüentemente promove sua ativação que é requerida na transição da fase G1/S (Nguyen et al., 2001). Dessa forma a atenuação do processo de senescência celular induzida por oncogene também é mediada por E7 e é fortemente sinalizada pelo acúmulo da proteína p16^{INK4A}, que em situação fisiológica bloqueia a fosforilação de várias cinases dependentes de ciclinas e contrapõem a fosforilação das proteínas do retinoblastoma (White et al., 2012). Assim sendo, o acúmulo de p16^{INK4A} é largamente utilizado como biomarcador para a identificação de displasias em biopsias de epitélio de colo uterino como sinalizador da expressão de E7 (Tornosello et al., 2013; Bergeron et al., 2015).

A proteína E6 por sua vez é uma pequena proteína de aproximadamente 18 kDa que além da promoção da degradação de p53 diretamente conforme explicitado anteriormente, interfere em sua atividade supressora de tumor por se ligar e desativar enzimas que ativam e estabilizam p53 através de acetilação como é o caso de p300 e ADA3 (Moody & Laimins, 2010; Wallace & Galloway, 2015). Outra importante ação de E6 é degradação de um conjunto de proteínas coletivamente chamadas de “*PDZ partners*”. Essas proteínas se ligam ao domínio

carboxi-terminal conservado de 4 aminoácidos da proteína E6 através de seus motivos PDZ e são degradadas. Por outro lado, essa interação aumenta a estabilidade da oncoproteína viral, mantendo um alto nível dessa proteína na célula (Thomas et al., 2008, Nicolaides et al., 2011). As proteínas com domínio PDZ estão relacionadas com a progressão do ciclo celular, transdução de sinal além da regulação na polaridade da célula e sua interação com a proteína E6 pode criar um ambiente favorável para a manutenção do genoma viral (Nicolaides et al., 2011).

Uma ação imprescindível nos processos de imortalização e transformação celular também atribuída a E6 é a sua interação com MYC na ativação da transcrição do gene hTERT que codifica a subunidade catalítica da telomerase permitindo a ativação da enzima (Wallace & Galloway, 2015).

Embora a ação dessas oncoproteínas tenha o potencial de facilitar o processo tumorigênico, parece pouco provável que isso seja o objetivo central dessas interações já que células tumorais são não permissivas para a replicação do HPV. Portanto o que favorece a conversão maligna parece ser a desregulação da expressão dessas oncoproteínas que ocorre após a ruptura do genoma viral no processo de integração (Arias-Pulido et al., 2006).

Diversos trabalhos relacionam a integração do genoma viral com o desenvolvimento de lesões precursoras de câncer. Em lesões de baixo grau, o DNA viral existe preferencialmente na forma epissomal (plasmídeos extracromossômicos) enquanto que, nas lesões malignas como neoplasias intraepiteliais cervicais de grau III (que incluem carcinoma *in situ*), o DNA do HPV está predominantemente integrado nos cromossomos do hospedeiro (Peitsaro et al., 2002; Yu et al., 2005; IARC, 2007; Groves & Coleman, 2015). Durante o acompanhamento de diversas mulheres com lesão de colo de útero observou-se que ambas as formas do genoma viral (epissomal e integrada), na maioria das vezes coexistiam, havendo uma relação da progressão da doença e o aumento da proporção de lesões com genomas na

forma integrada (Park, 1997; Vinokurova et al., 2008).

Um fator importante relacionado com uma maior frequência da integração do genoma viral é a persistência da infecção associada a certos tipos de HPV. A duração mediana de infecções por HPV de alto risco oncogênico é de aproximadamente 7-9 meses, sendo mais duradouras que às infecções por tipos não oncogênicos (Franco et al., 1999; Giuliano et al., 2002). Além disso, infecções persistentes por HPV de alto risco oncogênico, em especial o HPV 16, apresentam maior carga viral, outro fator que também parece favorecer a integração do DNA do vírus no genoma do hospedeiro (Li et al., 2008; Trevisan, 2013).

No processo de integração, ocorre uma quebra do DNA viral para sua integração no genoma hospedeiro mudando então seu estado físico da forma episomal (circular) para a forma integrada (linear). Estudos que mapearam os genomas completos do HPV em amostras de câncer cervical observaram que a quebra do genoma do vírus geralmente ocorre na proximidade do gene *E2* e isso afeta tanto a expressão do gene viral quanto a do genoma celular, conforme descrito abaixo (Kalantari et al., 1998; Pett et al., 2004; Arias-Pulido et al., 2006; Moody & Laimins, 2010). Com a diminuição da expressão de *E2* ocorre um aumento na atividade imortalizante do HPV, pois níveis mais altos das proteínas *E6* e *E7* são produzidos (Pett et al., 2004; Doorbar, 2006). Além disso, características conservadas desse processo são descritas na forma de uma cópia integrada (mais comum) ou múltiplas repetições em tandem com até várias centenas de cópias do DNA viral e pelo menos um segmento do genoma integrado contendo a região regulatória (LCR) e parte da região precoce, incluindo os genes *E6* e *E7* que codificam os dois principais oncogenes (FIGURA 4). Normalmente, transcritos quiméricos de DNA genômico da célula hospedeira com sequências integradas do DNA viral parecem ter maior estabilidade além de serem positivamente controlados por fatores de transcrição celulares ao invés dos virais (Rusan et al., 2015; Groves, 2015). Embora o processo de integração pareça favorecer o processo tumorigênico, ainda é

controverso em qual momento da infecção ele ocorre. Alguns trabalhos defendem que ele é um processo tardio já que a maior parte das lesões pré-malignas não apresentam essa alteração, enquanto outros demonstram ser um evento precoce e que a progressão tumoral é dependente de outros fatores subjacentes ao processo de integração (Peitsaro et al., 2002; Li et al., 2008; Collins et al., 2009; Raybould, 2011).

Figura 4 - Estado físico do genoma viral: O lado esquerdo da figura mostra a organização genômica do HPV 16 em sua forma epissomal. À direita temos as formas mais comuns de integração encontradas na neoplasia cervical mostrando no padrão superior a deleção de parte do gene *E2* ou no segundo padrão, a integração concatamizada (menos comum), a qual a atividade transcritora parece ser restrita apenas a sequência mais a jusante do sítio de integração mantendo todo o restante do genoma integrado reprimido por mecanismos epigenéticos. Adaptado de Groves and Coleman, 2015.

A proteína codificada pelo gene *E2* regula a transcrição dos oncogenes *E6* e *E7* através da sua dimerização e interação com o DNA viral especificamente em duas das sequências palindrômicas 5'-AACCG-N₄-CGGTT-3' encontradas na LCR que, por sua vez, sobrepõe o promotor precoce p97. Essa sobreposição evita a ligação dos fatores de transcrição por impedimento estérico neste promotor e consequentemente suprime sua atividade (Pett et al., 2004; Dellarole, 2010). (FIGURA 5).

Figura 5 - LCR Viral e os sítios de ligação da protein E2 (E2BS): (A) Visão esquemática da LCR de um genoma prototípico de um *Alphapapillomavirus*. Mostrando as regiões flangeadoras das ORFs L1 e E6, os sítios de ligação de E2 (BS1, BS2, BS3 e BS4), o sítio de ligação de E1, o sítio de ligação de Sp1 e o TATA box. (B) Sequência consenso para a ligação de Sp1, TATA box característica e sequências palindrômicas consenso para a ligação da proteína E2. (C) Modelo do complexo de dois homodímeros de E2 em dois sítios de ligação separados por apenas um par de bases. (D) Sequências nucleotídicas palindrômicas do sítio de ligação duplo que sobrepõem o promotor precoce. (Dellarole et al., 2010).

Embora esse evento pareça ser aleatório, mais da metade da ocorrência da integração dos tipos de alto risco oncogênico (HPV 16 e HPV 18) ocorrem em sítios frágeis comuns do genoma celular - “*common fragile sites*” (CFS) - (Yu et al., 2005; IARC, 2007; Thomas et al., 2012). Os CFSs são locais de maior instabilidade genômica, geralmente associados à deleções e intervalos que ocorrem quando as células são expostas à replicação sob estresse e a maioria dessas regiões contém um ou mais genes supressores de tumor (Yu et al., 2005; Drusco et al., 2011). Dessa forma o processo de integração pode ter dois papéis na transformação celular, seja pela expressão constitutiva das oncoproteínas virais ou pela mutagênese insercional que leva a uma desregulação da expressão gênica da célula. No processo de mutagenese insercional, genes supressores de tumor podem ser regulados negativamente e oncogenes

celulares podem ser superexpressos (Xu et al, 2013; Parfenov et al., 2014; Rusan et al., 2015). Além disso, diversos estudos sugerem uma extensiva alteração do padrão epigenético atrelado ao processo de integração que pode contribuir nas mudanças da expressão gênica celular e estar relacionado com a severidade da doença (Snellenberg et al., 2012; Chaiwongkot et al., 2013).

1.6. A infecção por HPV em homens e o estudo *HIM*

Embora a causa mais comum de mortalidade associada ao HPV seja o câncer de colo do útero e a maior parte da atenção dos pesquisadores esteja voltada em direção do entendimento da doença e seus precursores, a infecção por HPV em homens é de grande importância dado que o contato sexual é o primeiro modo de transmissão para a mulher e um entendimento dessas infecções em homens é, portanto, crítico para a redução do seu risco de transmissão. Além disso, entender a dinâmica que rege o ciclo biológico viral na região anogenital e cavidade oral no homem é essencial posto que, similar à mulher, os homens sofrem as consequências desta infecção na forma de condilomas, neoplasias e câncer.

A importância do “fator masculino” na etiologia do câncer de colo uterino em mulheres fora sugerido antes da identificação de um vírus sexualmente transmissível, o HPV, como causa desses tumores. Embora as pesquisas em mulheres tenham avançado muito nos últimos anos, o progresso no entendimento da prevalência e história natural da infecção genital por HPV em homens tem sido limitado.

Publicações sobre infecção por HPV em homens descrevem principalmente sua relação com o risco de desenvolvimento de câncer de ânus em homens homossexuais e/ou coinfectados com HIV (revisado por Palefsky, 1998; Critchlow *et al.*, 1998); outros observam um risco reduzido de aquisição e transmissão de HPV em homens circuncidados (Castellsague *et al.*, 2002) ou ainda estudos realizados com maridos de mulheres com câncer de colo de útero (Van Doornum *et al.*, 1994; Wikstrom *et al.*, 2000).

Embora esses estudos tenham contribuído de alguma forma para o entendimento dessas infecções em homens, estudos maiores são necessários para o entendimento da história natural dessas infecções assim como os fatores de risco para o desenvolvimento de lesões nas regiões anogenitais e cavidade oral. Dessa forma, para ampliar os conhecimentos acerca do HPV em homens, está sendo conduzido desde 2005 um estudo prospectivo de grande porte, multicêntrico (Brasil, México e Estados Unidos), o *Estudo da Infecção por HPV em Homens* (HIM – *HPV in men*), para acompanhamento de aproximadamente 4.000 homens de 18-44 anos a cada seis meses durante 4 anos. Também estão em seguimento 500 homens com idade entre 45 e 70 anos. O estudo é financiado pelo NIH-USA (Instituto Nacional de Saúde dos Estados Unidos) sendo os pesquisadores principais nos Estados Unidos a Dra. Anna Giuliano do *H. Lee Moffit Cancer Center* em Tampa, Florida; no Brasil a Dra. Luisa Lina Villa, na ocasião no Instituto Ludwig de Pesquisa sobre o Câncer em São Paulo; e no México o Dr. Eduardo Lazcano-Ponce do Instituto Nacional de Saúde Pública do México em Cuernavaca. Os homens recrutados nos três países passaram por critérios de elegibilidade que excluía indivíduos que já tivessem tido verrugas genitais, câncer nos sítios anatômicos analisados e infecção pelo vírus da imunodeficiência humana (HIV) (Giuliano et al., 2008). Em cada visita, além das amostras coletadas, são preenchidos questionários que versam sobre o uso de tabaco e álcool, hábitos alimentares, comportamento sexual, histórico familiar de câncer e sobre atividade física.

Os objetivos específicos do Estudo *HIM* são: (1) determinar a incidência e persistência das infecções do pênis por 37 tipos de HPV que comumente afetam o trato anogenital; (2) estudar a resposta imune humoral à infecção por HPV e; (3) identificar fatores de risco para aquisição, persistência e remissão de infecções tipo-específicas em homens.

Até o momento, mais de 55 trabalhos já foram publicados, sendo que as principais conclusões deste estudo de coorte até o momento são que a prevalência da infecção por HPV

foi de 65,2%, sendo 12% por tipos de alto risco oncogênico e 17,8% por infecção múltipla (baixo e alto risco oncogênico) e que essa taxa foi maior entre os participantes recrutados no Brasil (72,3%) em comparação aos Estados Unidos (61,3%) e ao México (61,9%) (Giuliano et al., 2008). Com relação ao desenvolvimento de lesão na região genital, este estudo observou que 5% dos homens infectados pelo HPV desenvolveram algum tipo de lesão (seja condiloma ou neoplasia) sendo que 2% dos homens infectados pelo tipo HPV 16 desenvolveram lesões neoplásicas genitais (PeIN). Nenhum outro tipo de HPV de alto risco oncogênico evoluiu para neoplasia no período de tempo avaliado de aproximadamente 24 meses (Sudenga et al., 2016). Um manuscrito detalhando a prevalência e incidência de verrugas genitais em homens brasileiros está em submissão (Silva et al., 2016; manuscrito anexo a esta dissertação).

Quanto aos fatores de risco para persistência da infecção na região anal, foi verificado que homens que relatam sexo com homens infectados por HPV de alto risco oncogênico apresentaram uma maior tendência a infecções mais persistentes (Nyitray et al., 2016). Já as infecções orais foram raras dentro desse estudo e a maioria parece ter sido eliminada em um período menor que 12 meses (Kreimer et al., 2013).

Diante do exposto, nos perguntamos se integração do genoma viral em amostras de homens saudáveis em sua grande maioria, seria um fator de risco para a persistência e desenvolvimento de lesão assim como observado na história natural dessa infecção na cérvix uterina. Para tal, testamos metodologias que pudessem definir com precisão a integração do DNA de HPV no genoma celular.

2. OBJETIVOS

2.1. Objetivo geral

Caracterizar a integração do DNA de HPV no genoma de células de diferentes sítios anatômicos (esfregaços genitais, esfregaços de canal anal e enxague oral) de homens sem qualquer patologia evidente.

2.2. Objetivo Específico

Explorar as metodologias de PCR (convencional e em tempo real) e *Next Generation Sequencing* para verificar a integração do HPV. Caracterizar a integração do DNA do HPV em espécimes de diferentes sítios anatômicos (esfregaços genitais, esfregaços anais e enxague oral) de homens assintomáticos.

2.3. Objetivo Secundário

Determinar a carga viral nos diferentes sítios anatômicos.

3. MATERIAIS E MÉTODOS

3.1. Amostras clínicas

Todos os voluntários passaram por um teste de elegibilidade e assinaram um termo de consentimento livre e esclarecido aprovado pelas devidas comissões de ética em pesquisa das instituições envolvidas para o uso das amostras (ANEXO A).

Antes de cada visita os voluntários são instruídos a não terem práticas sexuais por 24 horas, não lavar as regiões genitais por 12 horas e não se alimentar por 30 minutos que antecedem a coleta de espécimes. São realizados esfregaços com *swabs* umedecidos com soro fisiológico do corpo do pênis, saco escrotal, glândula peniana e borda anal. Além disso, um enxague oral também é realizado para coleta de células orais. Os *swabs* com espécimes genitais e anais são acondicionados em uma solução tamponante - STM (*Specimen Transport Medium*) Qiagen®. Os enxagues orais são realizados com um produto comumente utilizado como enxaguatório oral sem álcool - “Plax”®.

Os esfregaços do corpo do pênis, glândula e saco escrotal em solução de STM são centrifugados a 4°C e reunidos em um único tubo enquanto o esfregaço de borda anal passa pelo mesmo procedimento e é acondicionado em tubo separado. Os enxagues orais são lavados duas vezes com PBS e depois de descartada a solução sobrenadante, as células são ressuspensas em 1mL de PBS para serem acondicionadas em um único tubo. Essas amostras são armazenadas a -80°C e transportadas ao *H. Lee Moffit Cancer Center* em Tampa, Florida, até que a detecção e genotipagem de HPV sejam realizadas.

Os espécimes genitais e orais coletados nos diferentes centros de recrutamento são aliquotados e uma parcela do material tem seu DNA genômico extraído e é testado pela metodologia de *Linear Array*® (*Roche Molecular System Inc.*, Indianapolis, Indiana, US) capaz de detectar a presença de uma ampla gama de DNA viral de HPV genitais. Para tal empregam-se os iniciadores consenso PGMY09/11 capazes de amplificar um fragmento de

450 pb do gene *L1* (Gravitt et al., 2000) seguido de hibridização em tiras com sondas específicas para 37 tipos de HPV de alto e baixo risco comumente encontrados em amostras genitais femininas. Das amostras testadas, aquelas positivas para HPV 16 são selecionadas para serem testadas quanto à integração do HPV e sua carga viral.

Adicionalmente, foram incluídas amostras positivas para HPV 16 pela metodologia de Luminex já que outro ramo do mesmo estudo estava utilizando essa metodologia para tipagem de HPV e resolvemos ver como essas amostras se comportavam para os testes de integração (Nunes et al., 2016). Além disso, todas as amostras passariam pelo ensaio de qPCR em tempo real para ensaios de carga viral e dessa forma, pudemos acrescentar esse ensaio como uma metodologia adicional de tipagem de HPV (nesse caso HPV-16, já que é um ensaio específico, segundo Gravitt et al., 2003).

Um total de 125 amostras foi inicialmente selecionado para este estudo, sendo 69 amostras de esfregaços genitais, 44 amostras de canal anal e 12 amostras de cavidade oral conforme fluxograma abaixo.

Figura 6 - Fluxograma das amostras inicialmente selecionadas para os ensaios de integração e carga viral

3.2. Extração de DNA

3.2.1. Digestão com Proteinase K e Purificação pelo método Orgânico de Fenol Clorofórmio seguido de precipitação em Etanol

As amostras de DNA foram purificadas através da metodologia de Fenol-Clorofórmio para remover proteína e contaminantes residuais. Uma alíquota de 200µL de uma suspensão de células de esfregaços genitais/anais em STM ou enxague oral em PBS foram centrifugadas a 12.000 rpm por 15 minutos. Após terem o sobrenadante descartado foram acrescentadas de 200µL de tampão de digestão (50mM de Tris-HCl pH 7,5; 1mM de EDTA; 0,5% de Triton X-10) e 200µL de Tampão de Proteinase K (200µg/mL de Proteinase K liofilizada solubilizada em TE – 10mM de Tris-HCl; 1mM de EDTA). Foram incubadas por 48 horas a 55°C em banho seco com agitação. **Após 24 horas foram acrescentadas de 50µL de Tampão de Proteinase K no caso de células remanescentes em suspensão.

Ao término da digestão foi adicionado igual volume de Fenol: Clorofórmio: Álcool Isoamílico na proporção 25:24:1 em cada tubo e agitado vigorosamente (Vortex) por 1 minuto. Após terem sido centrifugadas por 15 minutos a 12.000 rpm o sobrenadante de cada amostra foi transferido para um tubo limpo.

Cada tubo contendo o sobrenadante da etapa anterior foi adicionado de 10% do volume de uma solução de acetato de sódio 3M e 2½ do volume de Etanol 100% gelado(<0°C). As amostras ficaram precipitando por um período mínimo de 12 horas a -20°C. Após esse período as amostras foram centrifugadas a 12.000 rpm a 4°C por 15 minutos para favorecer a formação do *pellet*. O etanol foi gentilmente removido. O *pellet* resultante foi lavado com etanol 70% a 4°C através de uma centrifugação a 12.000 rpm por 15 minutos por 2 vezes. Após a segunda lavagem, o etanol foi cuidadosamente removido e os tubos foram invertidos por 15 minutos para a retirada do etanol remanescente. Na última etapa do procedimento, o DNA foi ressuspenso em 50 µL de água MiliQ autoclavada.

As amostras de DNA foram quantificadas por sua densidade óptica a 260nm, e sua pureza atestada pela razão 260/280nm e 260/230nm. As medidas foram realizadas no espectrofotômetro NanoDrop2000® (*Thermo Scientific*, Wilmington, DE).

3.3. Reação em Cadeia de Polimerase (*Polymerase Chain Reaction* - PCR)

3.3.1. PCR convencional de β -actina - avaliação da integridade do DNA

O DNA foi analisado quanto à sua integridade realizando-se uma PCR convencional utilizando como alvo o gene da de β -actina humana. Esse parâmetro se baseia no fato de que o embora o DNA extraído possa estar em estados diferentes de integridade, existam fragmentos íntegros na faixa de 850 pares de bases, tamanho similar aos fragmentos que verificarão o estado físico do genoma viral. Apenas amostras que amplificaram esse fragmento passaram para a etapa subsequente de ensaios.

Inicialmente, 50 ng do DNA extraído das amostras biológicas fora utilizado na reação de PCR empregando-se os iniciadores para o gene da β -actina humana, que gera um fragmento de 850 pb. Resumidamente, 35 ciclos de amplificação foram efetuados em um volume de reação de 25 μ L contendo uma concentração final de 1x Taq Gold PCR Buffer® (*Applied Biosystems*, Foster City, California, USA), 200 μ M de cada dNTP, 2mM de MgCl₂, 1U de AmpliTaq Gold® DNA polimerase (*Applied Biosystems*, Foster City, California, USA), 4% de DMSO (*Sigma-Aldrich*, Inc, St. Louis. Missouri, USA), 250nM de cada iniciador (*primer*) (sequência dos iniciadores no final da seção). Cada ciclo consta de 1 minuto de desnaturação a 95°C, 1 minuto de anelamento a 58°C e 1 minuto de alongamento da cadeia a 72°C. O primeiro ciclo foi precedido por 5 minutos de desnaturação a 95°C e o último ciclo foi estendido por 7 minutos de alongamento a 72°C no termociclador GeneAmp® PCR System 9700 (*Applied Biosystems*, Foster City, California, USA) . Foram incluídos controles positivos e negativos na reação. O controle positivo utilizado foi DNA da linhagem celular CaSki ATCC® CRL- 1550™ (Yee et al., 1985) que contém o gene da β -

actina humana; como controle negativo foi incluído na reação de PCR um tubo com a mistura de reação, porém sem DNA. Os “*amplicons*” (produtos de amplificação) foram analisados por eletroforese em géis de agarose 0,8% corados com GelRed™ (*Biotium Inc*, Califórnia, EUA) em transiluminação UV (*BD Bioscience*, Oxford, UK).

Além desta análise da β -actina por PCR, foi sugerido um ensaio utilizando a 4200 TapeStation® (*Agilent Technologies Inc.*, Santa Clara, CA, USA) que também pode ser utilizada como controle de qualidade para diferentes finalidades. O equipamento é um analisador de fragmentos que avalia a integridade do DNA através de um eletroferograma exibindo para cada amostra qual é a média do tamanho dos fragmentos contidos e o grau de degradação da mesma.

<i>Primer Forward – β-actina humana</i>	5'-GTCCTTTGGGCGCTAACTG-3'
<i>Primer Reverse – β-actina humana</i>	5'-GATGGGGTACTTCAGGGTGA-3'

3.3.2. PCR convencional para análise do gene *E2* completo de HPV 16

Para verificar a integração do genoma viral analisa-se o gene *E2*, frequentemente ausente ou parcialmente ausente no processo de integração conforme citado na introdução. Avalia-se também a região imediatamente a montante o gene *E2*, reportada em alguns estudos também estar ausente ou parcialmente ausente (Cricca et al., 2009; Rusan et al., 2015). Dessa forma, foram realizadas duas reações para analisar 25% do gene *E1* e o gene *E2* completo do HPV 16. O primeiro gera um produto de 877 pb para analisar a porção 5' de *E2*, e o outro um produto de 766 pb para analisar a porção 3' de *E2*. A ausência de amplificação de ambos os produtos reflete a perda completa do gene *E2*, a qual foi classificada como Integração tipo I. A perda da porção 5' e da porção 3' foram classificadas como integração do tipo II e do tipo III respectivamente. Amplificações em ambas as reações indicam amostras

contendo a forma epissomal (Trevisan et al., 2016; MANUSCRITO SUBMETIDO) (FIGURA 7).

Figura 7 - Fluxograma para verificar o Estado Físico do Genoma Viral.

Inicialmente, 2µL do DNA extraído das amostras biológicas, independente da sua concentração foram utilizadas na reação de PCR empregando-se os iniciadores (sequência dos iniciadores no final da seção) para os diferentes fragmentos dos genes acima citados. Resumidamente, as duas reações tem condições parecidas modificando-se apenas os iniciadores, e constou de 35 ciclos de amplificação que foram realizados em um volume de reação de 25µL contendo uma concentração final de 1x Taq Gold PCR *Buffer*® (*Applied Biosystems*, Austin, Texas, USA), 200µM de cada dNTP, 2mM de MgCl₂, 1U de AmpliTaq Gold® DNA polimerase (*Applied Biosystems*, Austin, Texas, USA), 4% de DMSO (*Sigma-*

Aldrich, Inc, St. Louis. Missouri, USA) e 300nM de cada iniciador (*primer*) no termociclador GeneAmp® PCR System 9700 (*Applied Biosystems*, Foster City, California, USA). Cada ciclo incluiu 1 minuto de desnaturação a 95°C, 1 minuto de anelamento a 60°C e 1 minuto de alongamento da cadeia a 72°C. O primeiro ciclo foi precedido por 5 minutos de desnaturação a 95°C e o último ciclo foi estendido por 7 minutos de alongamento a 72°C. Foram incluídos controles positivos e negativos na reação. O controle positivo utilizado foi DNA da linhagem celular CaSki ATCC® CRL- 1550™ que contem HPV 16 com os genes *E2* e *E6* presentes; como controle negativo foi utilizado DNA da linhagem celular SiHa ATCC® HTB-35™ que contem HPV 16 integrado com ausência do gene *E2*. Os “*amplicons*” foram analisados por eletroforese em géis de agarose 0,8% corados com GelRed™ (*Biotium Inc*, Califórnia, EUA) em transiluminação UV (BD Bioscience, Oxford, UK).

<i>Primer Forward E2 5'</i>	5'-TGAAATTTCTGCAAGGGTCTG-3'
<i>Primer Reverse E2 5'</i>	5'-TCAACTTGACCCTCTACCACAG-3'

<i>Primer Forward E2 3'</i>	5'-TCAGTAACTGTGGTAGAGGGTCAA-3'
<i>Primer Reverse E2 3'</i>	5'-ACAAAAGCAGCGGACGTATT-3'

Para amostras que geraram os dois fragmentos alvo, discriminamos as que contêm a forma puramente epissomal do genoma daquelas que contem a forma mista, isto é, parte do DNA viral integrado e parte do genoma como plasmídeos extracromossômicos, sua forma epissomal. Isso é alcançado através da análise da carga de *E2* e *E6* por PCR quantitativa em tempo real conforme parte do protocolo. Uma carga de *E2* igual a *E6* significa a presença exclusiva da forma epissomal enquanto uma carga menor de *E2* em relação a *E6* indica a coexistência de ambos estados, epissomal e integrado.

3.4. PCR quantitativa em tempo real – determinação da integração do DNA do HPV e carga viral

Para os ensaios quantitativos, foi utilizada a tecnologia de PCR em tempo real que se utiliza da atividade 5'-exonucleásica da Taq Polimerase, detecção da fluorescência em tempo real com reagentes TaqMan® (*Applied Biosystems*, Foster City, California, USA) e uma plataforma que contém um termociclador com sistema ótico para excitação e captação da fluorescência - 7500 *Real-Time PCR System* (*Applied Biosystems*, Foster City, California, USA). As quantificações absolutas foram alcançadas comparando o *Ct* (*Cycle Threshold*) das amostras clínicas com o *Ct* de amostras com um número conhecido de número de cópias dos genes de interesse em uma curva padrão. Padrões de DNA foram preparados com quantidades pré-selecionadas de um plasmídeo referência com inserto do genoma completo do HPV16 (gentilmente cedido pelo Dr. E.M. de Villiers, *The German Cancer Research*, Heidelberg, DE) (o mais concentrado com 10^6 cópias/ μL) em um *background* constante de DNA de *Escherichia coli* Strain B (*Sigma-Aldrich*, Inc, St. Louis. Missouri, USA) ($2,5\mu\text{g/mL}$) para concentração final de $2,5 \times 10^5$ cópias de HPV 16/ μL . Foi preparado então, uma diluição seriada de 1:10 de modo a se obter uma curva padrão com quantidades finais de $2,5 \times 10^5$, $2,5 \times 10^4$, $2,5 \times 10^3$, $2,5 \times 10^2$, $2,5 \times 10^1$ e 2,5 cópias de HPV 16/ μL . Para a quantificação absoluta de β -globina, foi utilizada a mesma estratégia acima com quantidades conhecidas de um plasmídeo referência, fornecido gentilmente por F. Taddeo (*Merck & Co*, West Point, PA, USA), contendo um inserto de 254 pares de bases do gene da β -globina humana (10^6 cópias/ μL) para gerar um padrão contendo $2,5 \times 10^5$, $2,5 \times 10^4$, $2,5 \times 10^3$, $2,5 \times 10^2$, $2,5 \times 10^1$ e 2,5 cópias do fragmento do gene de β -globina/ μL . A quantidade de padrão utilizada em cada ponto da curva foi de 4 μL e todos os 6 pontos da curva foram realizados em triplicata para todas as reações de quantificação. Todas as reações quantitativas de PCR em tempo real foram realizadas em placas ópticas de 96 poços MicroAmp® (*Applied Biosystems*, Foster City, California, USA).

Com relação aos cálculos utilizados para a elaboração da curva padrão, deve-se levar em consideração o número de Avogrado ($6,023 \times 10^{23}$ moléculas – de pb-) e o peso molecular médio de um par de nucleotídeos que é de 660 Daltons. Dessa forma, conclui-se que um par de bases contem aproximadamente $1,096 \times 10^{-21}$ g. Utiliza-se o tamanho do genoma completo de HPV 16 inserido no plasmídeo e realiza-se o cálculo de sua massa (inserto+plasmídeo). Em seguida, calcula-se a massa de plasmídeos que contém o número de cópias desejado e divide-se a massa necessária pelo volume a ser pipetado por reação. Por último prepara-se uma diluição seriada do DNA plasmidial. No caso desse trabalho utilizamos o plasmídeo pBR322 que contem 4361pb + o genoma completo do HPV 16 que contem aproximadamente 7906 pb totalizando 12267pb. Aplicando-se a fórmula da massa $m = (12267)(1,096 \times 10^{-21} \text{ g})$ temos que um genoma completo (HPV 16 + plasmídeo) pesa $1,34 \times 10^{-17}$ g. Para o ponto mais concentrado da minha curva padrão, desejo 10^6 cópias de HPV 16, e dessa forma, multiplicando a massa de um plasmídeo por 10^6 obtem-se que a massa total necessaria é de $1,34 \times 10^{-11}$ g. Como o volume utilizado na reação é de 4 uL, conclui-se que a solução de plasmídeos deve conter uma concentração final de $3,36 \times 10^{-3}$ ng/uL. A partir dessa solução é realizada uma diluição seriada para 6 pontos de curva padrão (o mesmo é realizado para a curva de globina).

Para a validação da quantificação absoluta das curvas padrão foi utilizado o ensaio de ddPCR (*droplet digital PCR*) que é capaz de quantificar precisamente a amplificação da sequencia alvo sem a necessidade de curvas padrão e utiliza os mesmos pares de iniciadores e sondas utilizados para as reações descritas nesta dissertação. De forma sucinta, esta metodologia utiliza os reagentes ddPCR SuperMix® (*Bio-Rad Laboratories Inc*, Califórnia, USA) que adicionados dos iniciadores, sondas e amostra resulta no total de 20uL. Procede-se com a fragmentação dessa amostra em aproximadamente 20000 gotas submersas em óleo pela QX200 Droplet Generator® (*Bio-Rad Laboratories Inc*, Califórnia, USA) e a captação de

fluorescência total após 45 ciclos de amplificação de PCR que é analisada pelo *software Quanta-Soft®* (Bio-Rad Laboratories Inc, Califórnia, USA) que corrige esses números de acordo com a distribuição de Poisson. Para que a acurácia desse ensaio seja garantida, são consideradas apenas as gotas que apresentam volume e integridade compatíveis com o esperado e ser exigido um mínimo de 10.000 gotas para os cálculos (Pinheiro et al., 2011).

3.4.1. PCR quantitativa do gene E6 – HPV 16 – integração do HPV e carga viral

O ensaio de PCR quantitativa para o gene *E6* do HPV16 serviu para determinar a integração do HPV, se puramente episomal ou misto, e ainda para a quantificação absoluta da carga viral. As condições de ciclagem e sequência dos iniciadores/sonda foram realizadas com base em um protocolo publicado por Gravitt et al., 2003 e constou de 55 ciclos de amplificação que foram realizados em um volume de reação de 25µl contendo uma concentração final de 1x Taq Gold PCR Buffer® (Applied Biosystems, Austin, Texas, USA), 200µM de cada dNTP, 2mM de MgCl₂, 5U de AmpliTaq Gold® DNA polimerase (Applied Biosystems, Austin, Texas, USA), 300nM de cada iniciador (*primer*) (sequência dos iniciadores e sonda no final da seção), 200nM da sonda de hidrólise marcada com 5'-FAM e 3'-BHQ e 1x CXR® Reference Dye (Promega, Madison, WI, USA) na plataforma 7500 Real-Time PCR System (Applied Biosystems, Foster City, California, USA). Cada ciclo incluiu 15 segundos de desnaturação a 94°C, e 30 segundos de anelamento e de alongamento da cadeia a 55°C. O primeiro ciclo foi precedido por 10 minutos de desnaturação a 95°C. Foram utilizados 2µL da amostra e dos respectivos controles e adicionados mais 2µL de água MilliQ (correspondendo ao volume utilizado nas curvas padrão) todos previamente homogeneizados. O controle positivo utilizado foi DNA da linhagem celular CaSki ATCC® CRL-1550™ que contem HPV 16 (400-600 cópias/célula) e DNA da linhagem celular SiHa

ATCC® HTB-35™ que contem HPV 16 integrado (1-2 cópias por célula) e ausência do gene *E2*. Como controle negativo foi usado o *background* constante de DNA de *Escherichia coli* Strain B (*Sigma-Aldrich*, Inc, St. Louis. Missouri, USA) (2,5µg/mL).

<i>Primer Forward E6</i>	5'-ATGACTTTGCTTTTCGGGAT-3'
<i>Primer Reverse E6</i>	5'-CTTTGCTTTTCTTCAGGACA -3'
<i>Probe E6</i>	5'-FAM-ACGGTTTGTGTATTGCTGTTCTAA-BHQ1-3'

3.4.2. PCR quantitativa do gene E2 – HPV 16 – integração do HPV

O ensaio de PCR quantitativa para o gene *E2* do HPV16 serviu para determinar o estado físico do HPV, se puramente episomal ou misto (FIGURA 6). As condições de ciclagem e sequência dos iniciadores/sonda foram realizadas segundo Peitsaro, et al., 2002 e constou de 40 ciclos de amplificação que são realizados em um volume de reação de 25µl contendo uma concentração final de 1x PCR Master Mix TaqMan® (*Applied Biosystems*, Foster City, California, USA), 300nM de cada iniciador (*primer*) (sequência dos iniciadores e sonda no final da seção), 50nM da sonda de hidrólise marcada com 5'-FAM e 3'-BHQ na plataforma 7500 *Real-Time PCR System* (*Applied Biosystems*, Foster City, California, USA). Cada ciclo incluiu 15 segundos de desnaturação a 95°C, e 1 minuto de anelamento e de alongamento da cadeia a 60°C. O primeiro ciclo foi precedido por 2 minutos para a ativação da Uracil-DNA-glicosilase a 50°C e 10 minutos de desnaturação a 95°C. Foram utilizados 2µL da amostra e dos respectivos controles e adicionados mais 2µL de água MilliQ (correspondendo ao volume utilizado nas curvas padrão) todos previamente homogeneizados. O controle positivo usado foi DNA da linhagem celular CaSki ATCC® CRL-1550™ que contem HPV 16 integrado contendo o gene *E2* intacto repetidas vezes e como controle negativo foi usado DNA da linhagem celular SiHa ATCC® HTB-35™ que contem HPV 16

integrado e ausência do gene *E2*. O *background* constante de DNA de *Escherichia coli* Strain B (*Sigma-Aldrich*, Inc, St. Louis. Missouri, USA) (2,5µg/mL) também foi utilizado como controle negativo extra.

<i>Primer Forward E2</i>	5'-AACGAAGTATCCTCTCCTGAAATTATTAGGC -3'
<i>Primer Reverse E2</i>	5'-CCAAGGCGACGGCTTTG-3'
<i>Probe E2</i>	5'-FAM-CACCCCGCCGCGACCCATA-BHQ1-3'

3.4.3. PCR quantitativa do gene da β -globina humana – carga viral

O ensaio de PCR quantitativa para o gene β -globina humana tem como objetivo verificar a carga viral do HPV 16 em cópias por célula. Cada célula somática humana, portanto diplóide, contém 2 cópias do gene da β -globina e para se obter o número de cópias por célula utilizou-se o valor da quantificação absoluta de *E6* e dividiu-se pela metade da quantidade absoluta obtida deste ensaio de β -globina (obtido através da curva padrão com plasmídeos contendo como inserto parte desse gene, conforme descrito em 3.4.). As condições para a reação e sequência dos iniciadores/sonda foram gentilmente fornecidas por F. Taddeo (*Merck & Co. Inc.*, West Point, PA, USA) e constaram de 40 ciclos de amplificação em um volume de reação de 25µl contendo uma concentração final de 1x PCR Master Mix TaqMan® (*Applied Biosystems*, Foster City, California, USA), 300nM de cada iniciador (*primer*) (sequência dos iniciadores e sonda no final da seção), 100nM da sonda de hidrólise marcada com 5'-FAM e 3'-BHQ na plataforma *7500 Real-Time PCR System* (*Applied Biosystems*, Foster City, California, USA). Cada ciclo incluiu 15 segundos de desnaturação a 94°C, e 1 minuto de anelamento e de alongamento da cadeia a 60°C. O

primeiro ciclo foi precedido por 2 minutos de ativação da Uracil-DNA-glicosilase a 50°C e 10 minutos de desnaturação a 95°C. Foram utilizados 2µL da amostra e dos respectivos controles e adicionados mais 2µL de água MilliQ (correspondendo ao volume utilizado nas curvas padrão) todos previamente homogeneizados. Os controles positivos usados são DNA da linhagem celular CaSki ATCC® CRL-1550™ que contem HPV 16 (400-600 cópias/célula) e DNA da linhagem celular SiHa ATCC® HTB-35™ que contem HPV 16 integrado (1-2 cópias por célula) e ausência do gene *E2*. Como controle negativo foi usado o *background* constante de DNA de *Escherichia coli* Strain B (*Sigma-Aldrich*, Inc, St. Louis. Missouri, USA) (2,5µg/mL).

<i>Primer Forward β-globin humana</i>	5'-GCTTCTGACACAACCTGTGTTTCAC -3'
<i>Primer Reverse β-globin humana</i>	5'-GGCCTCACCACCAACTTCAT-3'
<i>Probe β-globin humana</i>	5'-FAM-CACGTTACCTTGCCCCACAGG-BHQ -3'

3.5. *Ion Torrent NGS – Next Generation Sequencing*

Com o intuito de explorar uma abordagem mais moderna e que tem se mostrado bastante acurada para os ensaios de integração em diversas amostras de tumores HPV positivos, foi customizado um *TargetSeq* para sequenciamento *Next Generation Sequencing* (Xu et al., 2013; Rusan et al., 2015; Hu et al., 2015). Essa abordagem se baseia no enriquecimento das amostras através da captura de fragmentos contendo sequências de HPV 16 a partir do DNA genômico humano. Para isso, a etapa inicial foi a construção de uma biblioteca de DNA genômico utilizando o Kit de preparo de biblioteca *Ion Xpress™* (*Life Technologies*, Carlsbad, CA, USA) e também o Kit de adaptadores Ion com *barcode* 1-16, que permitiu a análise em paralelo de até 16 amostras em uma mesma corrida.

Após o preparo das bibliotecas, seguiu-se à captura da porção de DNA que continha sequências de qualquer parte do genoma do HPV16 através de um kit de captura customizado *TargetSeq*TM (Life Technologies, Carlsbad, CA, USA), que contém sondas oligonucleotídicas biotiniladas complementares ao DNA viral. Após incubação, o DNA hibridizado foi capturado com esferas magnéticas cobertas com estreptavidina. Caso a amostra contivesse o genoma viral integrado, as sequências do DNA viral conteriam em suas bordas regiões de inserção do genoma humano. Todos os procedimentos foram realizados de acordo com as instruções do fabricante. (FIGURA 8).

O material enriquecido foi sequenciado no Chip 316, que contém de 300Mb a 1Gb, com a química Hi-Q no sequenciador *Ion PGM*TM (Life Technologies), um sequenciador de nova geração de alta capacidade de geração de dados que se baseia na liberação de íons H⁺ durante a ação da polimerase, os quais ativam uma camada sensora nos chips semicondutores, produzindo um sinal elétrico. Essa química de sequenciamento permite leituras de 200 pb, que devem ser longas o suficiente para que algumas das leituras abranjam regiões tanto do genoma viral quanto do genoma humano, o que permitiu definir se há pontos de inserção.

Figura 8 - Estratégia de sequenciamento baseado na captura prévia de fragmentos do genoma viral: (A) DNA genômico total é fragmentado, ligado a adaptadores e amplificado para a construção de uma biblioteca de DNA. (B) Após a separação de fragmentos do tamanho desejado, o DNA é incubado com sondas biotiniladas que se hibridizam com sequências do HPV 16. (C) Após dois ciclos de hibridização, o material total é incubado com esferas magnéticas cobertas de estreptavidina. (D) Através de uma rack magnética o DNA capturado nas esferas é separado do material não ligado. (E) Apenas o DNA genômico que contém sequências de HPV-16 na forma de quimeras é purificado. (F) O material passa por um ciclo de amplificação através de PCR. (G) O material amplificado segue para o sequenciamento na plataforma ION TORRENT. Adaptado de <http://www.genomics.agilent.com/article.jsp?pageId=3083>.

3.5.1. Análise Bioinformática

A análise bioinformática contou com a colaboração do grupo do Prof^o. Wilson Araújo Junior do Laboratório de Genética Molecular e Bioinformática da Faculdade de Medicina de Ribeirão Preto/USP e o desenvolvimento do *pipeline* estabelecido inclusive servirá para as análises de futuras corridas.

Os dados brutos obtidos do sequenciador *PGM Ion Torrent* em formato fastq foram submetidos à avaliação de qualidade através do programa FastQC . O FastQC é uma ferramenta de controle de qualidade para dados de sequenciamento em alta escala mostrando regiões com potenciais problemas que afetam a acurácia da sequência (Andrews, 2010). De todos os parâmetros observados, o mais importante e utilizado como critério para a seleção dos fragmentos foi o parâmetro de qualidade da sequência por base (*Per base Sequence Quality*) que tem o *Phred Score* como medida de qualidade da identificação das bases nucleotídicas geradas por qualquer forma de sequenciamento automatizado de DNA.

Baseado nos resultados, foram estabelecidos os critérios iniciais para a inclusão das *reads* através do PRINSEQ (Schmieder and Edwards, 2011). Dessa forma, os fragmentos selecionados foram novamente avaliados pelo FastQC e submetidos à verificação de integração através do programa Virus-Clip (Ho et al., 2015).

O Virus-Clip é uma ferramenta de detecção de integração baseada em sequências parcialmente mapeadas no genoma viral e que representariam a fusão quimérica entre o genoma humano e o genoma viral. Esse programa foi escolhido por aceitar sequenciamentos do tipo *single-end* e ter se mostrado mais vantajoso quando comparado a outros programas com o mesmo propósito. Sua execução é realizada em várias etapas resumidamente listadas e esquematizadas a seguir (FIGURA 9). Primeiro ele executa o mapeamento das *reads* ao genoma referência viral através do *Burrows-Wheeler Aligner* (BWA-MEM) que gera os alinhamentos parciais (Li and Durbin, 2010). Então com o uso do SAMTools são analisados

os fragmentos parcialmente mapeados no formato *Sequencing Alignment/Map* (SAM) através da *CIGAR flag* (Li et al., 2009). As *reads* que mapearam parcialmente são então submetidas ao mapeamento contra o genoma humano de referência (nesse caso hg19) através da versão autônoma do *BLASTN* (disponível em <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>). Por fim é utilizado o *ANNOVAR* que faz a anotação da região genômica humana afetada (Wang and Hakonarson, 2010). A anotação da integração do genoma viral é realizada manualmente no *RefSeq* a partir da posição nucleotídica fornecida pelo BWA (disponível em <http://www.ncbi.nlm.nih.gov/refseq/>). A partir desses dados, foi construído o *Circos Plot*, uma maneira elegante e intuitiva de mostrar relações entre diferentes genomas (Krzywinski et al., 2009).

Figura 9 - Estratégia bioinformática

4. RESULTADOS

4.1. Seleção das amostras

Partimos do critério que para se avaliar integração as amostras deveriam ser positivas para os três métodos de detecção e ser capaz de amplificar o fragmento de 850 pb do gene da β -actina humana utilizado como controle de qualidade. Dessa forma, foi incluído nos ensaios um total de 56 amostras, sendo 39 amostras de esfregaços genitais, 13 amostras de canal anal e 04 amostras de cavidade oral, conforme fluxograma abaixo. De todas as 125 amostras positivas para Luminex inicialmente incluídas nesse trabalho, 100 foram concordantes com os resultados de qPCR e 61 com o Linear Array.

4.2. Aplicação da metodologia previamente aplicada no laboratório (desenvolvida por Trevisan et al.)

No nosso laboratório foi desenvolvida uma metodologia para testes de integração e carga viral baseada em amostras de cérvix uterina ou biopsias de regiões anatômicas diversas. Estas amostras, embora heterogêneas quanto à qualidade e quantidade de DNA, superam em muito a qualidade encontrada em esfregaços de pele da região genital masculina e esfregaços de canal anal. Isso se refletiu logo nos primeiros ensaios de *β-actina* para verificar a qualidade do DNA. De todas as amostras testadas mais da metade resultaram em bandas fracas por eletroforese em gel de agarose ou ausência de amplificação mesmo com maiores quantidades de DNA. As amostras de enxague oral foram as únicas a terem 100% de amplificação.

Como o número de amostras e a quantidade de DNA extraída eram limitados e este processo composto por múltiplas etapas, houve a necessidade de se fazer um aprimoramento das reações para que fossem utilizadas as menores quantidades de amostra para o máximo de resultados.

Dessa forma uma mesma metodologia “otimizada” poderia ser utilizada para os três sítios anatômicos sempre se baseando no aprimoramento do material mais escasso e/ou de menor qualidade, nesse caso, nas amostras de DNA do canal anal conforme observado no ensaio da *β-actina* (850pb) e confirmadas pelo *BioAgilent* (FIGURA 10).

Figura 10 Resultado do BioAnalyzer: As amostras C1, D1 e E1 são provenientes de esfregaços genitais; F1, G1, H1 e A2 são de cavidade oral e as amostras B2, C2, D2 e E2 são oriundas de canal anal; as amostras F2 e H2 são amostras controle consideradas com grau razoável de integridade enquanto que a amostra controle G2 apresenta um certo grau de degradação. DIN (*DNA Integrity Number* é uma escala de 1 a 10 que qualifica o DNA de altamente degradado a altamente intacto). No caso das amostras obtidas para este estudo, apenas as amostras de cavidade oral foram avaliadas com algum conceito de DIN enquanto que as amostras de canal anal e esfregaços genitais tinham qualidade ainda inferior.

4.3. Aprimoramento da metodologia adaptada para esta dissertação (desenvolvida por Cintra et al.)

De acordo com os critérios adotados para a inclusão das amostras para os ensaios de integração, seguimos para o processo de padronização. Durante esse processo e com o objetivo de “otimizar” cada uma das etapas (por conta da baixa qualidade e/ou quantidade de DNA) submeti amostras de DNA de linhagens celulares contendo genomas completos de HPV 16 em diluições seriadas a todas as reações, modificando cada variável por vez. Assim foi possível determinar quais eram as melhores condições para cada reação e indiretamente medir a sensibilidade de cada uma delas.

4.3.1. Verificando a integração do genoma viral

Para a reação de *β -actina humana*, que amplifica um fragmento de 850 pb, comecei com o estudo das temperaturas de anelamento dos iniciadores no intuito de melhorar a sensibilidade sem prejudicar a especificidade. Além disso, diferentes condições de ciclagem foram testadas. Diferentes concentrações de $MgCl_2$, iniciadores, dNTP e DMSO foram testadas e os dois fatores que mais geraram impacto foram a redução de temperatura de anelamento em 2°C e a mudança da enzima para a AmpliTaq Gold® DNA polimerase (*Applied Biosystems*, Austin, Texas, USA).

Os resultados deste ensaio estão descritos na tabela abaixo e refletem as amostras incluídas nas análises de integração conforme fluxograma na sessão 4.1.

Tabela 2 – Resultados do ensaio da β -actina - controle de qualidade - 850pb

SITIO ANATÔMICO	TOTAL DE AMOSTRAS	POSITIVAS
Esfregaços Genitais	42	39 (92,8%)
Canal Anal	15	13 (86,6%)
Cavidade Oral	4	4 (100%)

Na etapa subsequente, as condições também foram “otimizadas” e novamente a troca da enzima foi o diferencial no aumento da sensibilidade, além disso, a presença do DMSO em mais de 4% inibe a reação, que foi o contrário do observado na reação da *β -actina humana*, que tem sua sensibilidade reduzida na ausência do DMSO. Os alvos nessa reação são 25% do gene *E1* e 100% do gene *E2* do HPV 16 em dois fragmentos para ser verificada se o genoma viral está integrado ou episomal. Conforme citado anteriormente, a ausência de amplificação de um ou dos dois fragmentos indicaria a perda dessa parte do genoma como ocorre no processo de integração.

Embora apenas as amostras que tiveram o fragmento de *β -actina humana* amplificados fossem consideradas para análise, todas as amostras negativas foram testadas nesse ensaio também (já que estamos em processo de padronização, é interessante saber como o resultado anterior se confirma nesse teste) e curiosamente 4 amostras negativas na etapa anterior geraram fragmentos nesta etapa. Duas delas geraram os dois fragmentos alvo e as outras duas, geraram apenas um dos fragmentos. De acordo com essa metodologia, poderíamos inferir que as amostras negativas para o gene da *β -actina humana*, positivas para os dois fragmentos de *E2* conteriam a forma epissomal, porém as amostras que resultaram em apenas um fragmento positivo não podem ser afirmadas integradas já que o ensaio anterior não foi capaz de confirmar a integridade de DNA para a faixa de 850 pb. Os resultados dos padrões de integração e amplificação conforme descritos no fluxograma da figura 6 seguem abaixo:

Tabela 3 - Ensaio para verificar Integração (25% do gene *E1* e 100% do gene *E2*)

SITIO ANATÔMICO	Integrado Tipo I	Integrado Tipo II	Integrado Tipo III
Esfregaços Genitais (39)	0	0	2
Amplificaram ambos = 37			*Amplificou apenas 5'
Canal Anal (13)	1	0	1
Amplificaram ambos = 11	*Ausência de ambas		*Amplificou apenas 5'
Cavidade Oral (4)	2	0	0
Amplificaram ambos = 2	*Ausência de ambas		

Os resultados dessa primeira etapa de amplificações, por PCR convencional, indicariam que teríamos 15,4% amostras anais integradas, 5,1% das amostras genitais integradas e 50% das amostras orais integradas (ilustrados também na figura 11).

Na etapa seguinte o objetivo principal é verificar se as amostras que amplificaram os dois fragmentos contem a forma puramente epissomal ou se contem a coexistência de formas

epissomais e integradas. Isso é alcançado realizando a quantificação absoluta por PCR quantitativa em tempo real baseado em curvas padrão conforme mencionado em materiais e métodos. Inúmeras variáveis determinarão se as curvas padrão terão a acurácia e reprodutibilidade a que elas se propõem. Dessa forma, foi possível confirmar com o ensaio de ddPCR que o número de cópias estimado pelas curvas que amplificam os genes *E6*, *E2* e de β -globina estavam de acordo com os cálculos realizados para a elaboração das respectivas curvas.

São realizadas então, duas reações para cada amostra, uma reação para quantificação do gene *E6* (frequentemente presente em amostras integradas ou epissomais) e uma reação para a quantificação do gene *E2* (frequentemente ausente no processo de integração) sendo que a razão E2/E6 é utilizada para diferenciar entre essas duas categorias.

Para avaliar o poder de discriminação entre amostras que continham o estado puramente epissomal daquelas que continha o estado misto foi feito um ensaio com diferentes concentrações de plasmídeos contendo o inserto completo de HPV 16 (representando a forma epissomal) e linhagem celular SiHa contendo apenas o gene *E6* (representando a forma integrada) simulando uma proporção de integração de 0, 25, 50, 75 e 100% para cada ponto em triplicata. Esse ensaio foi repetido três vezes e na tabela 4 encontram-se os resultados com as médias e seus respectivos desvios.

Tabela 4 - Categorização da relação E2/E6

% Epissomal/Integrado	Categoria	Valores médios E2/E6 + Desvio
100% Epissomal	Epissomal	1,02±0,06
75% Epissomal/25% Integrado	Misto	0,64±0,05
50% Epissomal/50% Integrado	Misto	0,36±0,04
25% Epissomal/75% Integrado	Misto	0,18±0,03
100% Integrado	Integrado	ZERO

Assim, para a determinação do estado puramente epissomal, e para o intervalo de confiança de 95%, foi considerado o intervalo de 0,917 até 1,13 para a relação de E2/E6. Amostras cujas relações E2/E6 foram menores que 0,917 foram classificadas como mistas.

Além disso, um ensaio multiplex seria o ideal já que mostraria para uma mesma amostra os valores para as duas variáveis em questão no mesmo experimento. Porém, depois de diversas tentativas, concluímos que as reações têm condições muito diferentes, impossibilitando essa padronização. A reação que avalia o gene *E6* tem condições muito peculiares (descrita em materiais e métodos) para abranger todas as variantes de HPV 16. Seu “*amplicon*” consta de 233 pb e a sonda de hidrólise é distante das duas extremidades dos iniciadores, o que não é recomendado para uma PCR quantitativa em tempo real, por isso não é possível utilizar os reagentes TaqMan® (*Applied Biosystems*, Foster City, California, USA) como na reação para o gene *E2*. Modificações nas quantidades de dNTP, iniciadores, sondas e ciclagem foram verificados sem sucesso. O mesmo foi realizado para os ensaios de carga viral, também sem sucesso, com o gene da *β -globina humana* amplificando concomitantemente com o gene *E6* já que além de ter o resultado da carga viral em uma mesma reação teríamos uma referencia ativa conforme recomenda Bustin et al., 2009.

Os padrões das amostras classificadas como mistas ou puramente epissomais estão descritos na tabela abaixo e ilustradas na figura 11.

Tabela 5 - Amostras Mistas – Ensaio de qPCR e razão E2/E6 de HPV 16

Amostras com 2 fragmentos (+)	Misto	Epissomal
Esfregaços Genitais (37)	24	13
Canal Anal (11)	8	3
Cavidade Oral (2)	1	1

Dessa forma verificou-se que aproximadamente 65% das amostras de esfregaço genital, 72% das amostras de canal e 50% das amostras de cavidade oral continham uma relação E2/E6 menor do que 0,917, o que as classifica como amostras mistas, isto é, apresenta integração do genoma viral e genomas na forma episomal na mesma amostra.

A linearidade das curvas de calibração (curvas padrão) de todas as reações de PCR quantitativas foi muito boa, com eficiência superior a 95% e um coeficiente de correlação médio (r^2) de 0,998.

Figura 11 - Resultados do perfil de integração do HPV 16 nos diferentes sítios anatômicos baseado nas metodologias de PCR convencional e em tempo real: Nesse histograma temos 5,1% das amostras de esfregaço genital integradas, 15,4% das amostras de esfregaço anal integradas e 50% das amostras de enxague oral com DNA integrado. As barras vermelhas indicam amostras que se apresentam com estado misto da infecção.

4.3.2. Verificando a carga viral do HPV

Para a verificação da carga viral, foi utilizado o resultado da quantificação do gene *E6* de HPV 16 da etapa anterior e uma reação de PCR quantitativo em tempo real para o gene da *β-globina humana*.

Para a padronização desse ensaio foi utilizado DNA de esfregaços genitais HPV negativos em uma concentração constante de 10ng/μL e diferentes quantidades de plasmídeos contendo o genoma completo de HPV 16 simulando a carga viral com 6 pontos (½, 1, 5, 10, 50 e 100 cópias/célula). Como controle de precisão e reprodutibilidade em estimar a carga viral foi utilizado DNA extraído das linhagens celulares de carcinoma cervical CaSki ATCC® CRL-1550™, que contem HPV 16 (400-600 cópias/célula), e DNA da linhagem celular SiHa ATCC® HTB-35™, que contem HPV 16 integrado (1-2 cópias por célula) (FIGURAS 12 e 13). O resultado de cópias por célula é alcançado dividindo-se o valor da quantificação do gene *E6* de HPV 16 pela metade do valor da quantificação da *β-globina humana*.

Figura 12 - Padronização do ensaio de carga viral: Padrões submetidos ao ensaio de qPCR realizados com diluições seriadas de plasmídeos contendo um inserto completo do HPV 16 nas cargas virais de ½, 1, 5, 10, 50 e 100 cópias por célula representadas da direita para a esquerda no “Amplification plot”.

Figura 13 - Curvas coincidentes entre os padrões de 1 cópia por célula e amplificação de DNA da linhagem celular SiHa: A utilização dessa linhagem celular foi utilizada como controle de precisão e reprodutibilidade dos ensaios de carga viral.

Para os ensaios de carga viral foram consideradas todas as amostras positivas pela metodologia de qPCR, totalizando 61 amostras de esfregaços genitais, 30 amostras de canal anal e 9 amostras de cavidade oral. Dessa forma pode se verificar que a carga viral variou enormemente tanto entre os diferentes sítios anatômicos quanto no mesmo sítio anatômico, conforme tabela 6.

Tabela 6 - Carga viral em cópias/célula dos diferentes sítios anatômicos

Sítio Anatômico	Carga viral Mínima	Carga viral Máxima
Esfregaços Genitais (61)	0,030	1770
Canal Anal (30)	0,001	14500
Cavidade Oral (9)	8×10^{-5}	0,3

A média da carga viral dos controles baseada no gene *E6* foi de 1,5 cópias por célula (95% IC: 1,13-1,87) na linhagem celular SiHa (esperado entre 1-2 cópias) e 497,7 cópias por células (95% IC: 407,9-587,4) em CaSki (esperado 400-600 cópias).

4.3.3. TargetSeq para o Sequenciamento de Nova Geração (NGS)

As amostras sequenciadas foram previamente selecionadas contemplando todos os resultados obtidos pela metodologia de PCR/qPCR, sendo que 6 amostras apresentavam-se como integradas, duas epissomais, duas mistas e uma indeterminada (amostra 1). De qualquer maneira esta última amostra não tinha o fragmento de β -actina amplificado como controle de qualidade e só foi incluída por apresentar resultados inconclusivos nas amplificações do gene *E2* (TABELA 7).

Amostras		Actina	E2 - 5'	E2 - 3'	L. Array	qPCR	Luminex	β	Carga Viral	E2/E6
01	10389gen	neg	pos	neg	pos	18	132	1,5	12,00	xxx
02	22877gen	pos	pos	neg	pos	2,7	32	8	0,34	I
03	10861gen	pos	pos	neg	pos	320	65	7	45,71	I
04	10415gen	pos	pos	pos	pos	12	21	10	1,20	M
05	22516gen	pos	pos	pos	pos	28	100	12	2,33	E
06	11112gen	pos	pos	pos	pos	791	180	29	27,28	E
07	31579ac	pos	pos	pos	pos	134	154	9	14,89	M
08	22299ac	pos	pos	neg	pos	10	81	37	0,27	I
09	22172ac	pos	neg	neg	pos	4,3	23	15	0,29	I
10	30577oc	pos	neg	neg	pos	4,6	59	7260	0,0006	I
11	22175oc	pos	neg	neg	pos	7,4	34	4500	0,0016	I

Tabela 7 - Lista das amostras selecionadas para sequenciamento. Actina = ensaio de β -actina (850 pb); E2 - 5' e E2 - 3' = análise de integração (pos=positivo/ neg=negativo). Metodologias de detecção de HPV com seus respectivos valores de detecção no caso do Linear Array pos=positivo; qPCR número de cópias de *E6* de HPV 16 por reação e no caso do Luminex o valor dado em MFI (Intensidade média de fluorescência). β =quantificação absoluta da β -globina (dividida por 2) por reação; Carga Viral em cópias por célula refletindo o valor obtido em qPCR dividido pelo valor da β -globina e integração viral (xxx=indeterminado; I= integrado; M= misto; E= epissomal).

De todas as amostras sequenciadas, apenas quatro delas apresentaram fragmentos contendo sequências de HPV (amostras 03, 05, 06 e 07) sendo que destas, 1 delas apresentou-se integrada (amostra 03) e está representada no *circos plot* com as linhas roxas (Figura 15). A amostra que se apresentou integrada confirmou o resultado da metodologia baseada em PCR, com perda da porção 3' do fragmento de *E2*. O exato ponto de deleção foi dentro do gene *E2*, conforme esperado, na posição 3391 do genoma do HPV. A integração foi detectada em 4

pontos do genoma humano sendo duas delas em regiões intrônicas dos genes *PAX5* e *GPR158* e duas em regiões intergênicas. Além das representações das integrações mostradas no *circos plot* (FIGURA 15), a representação de um dos fragmentos quiméricos encontrados que apresenta a inserção no gene *PAX5* está mostrado na figura 14.

AMOSTRA 10861 GEN

Figura 14 - Representação do fragmento quimérico da fusão entre o gene *E2* de HPV 16 e o cromossomo 9 humano, dentro do gene *PAX5*.

Não foram detectados fragmentos com sequências de HPV nas outras amostras que apresentaram o padrão integrado pela metodologia de PCR. Das outras amostras contendo fragmentos com sequências de HPV detectados pelo sequenciamento, duas amostras apresentaram o padrão epissomal pela metodologia de PCR (amostras 05 e 06) e uma das que apresentavam o estado misto (amostra 07). Em uma das amostras que fora classificada como mista pela metodologia de PCR não foi possível verificar fragmentos contendo sequências de HPV (amostra 04). A concordância entre as duas metodologias está representada na tabela 7.

Como controle positivo da detecção de integração foi utilizado DNA da linhagem SiHa cujos pontos de deleção e inserção estão bem estabelecidos e estão representados pela linha azul clara no *circos plot* (Figura 15). O padrão de integração da linhagem SiHa ocorre com ruptura no gene *E2* do HPV 16 e inserção no cromossomo 13 humano (Diao et al., 2015).

Tabela 8 – Concordância entre as metodologias baseadas em PCR e NGS

Amostras Seleccionadas	PCR+NGS	Sem Concordância
Integradas (6)	1	5
Epissomais (2)	2	0
Mistas (2)	0	2
Indeterminada (1)	**	**

Figura 15 - Circus Plot de Integração: Estão representados todos os cromossomos do genoma humano e o genoma completo do HPV 16 em forma de círculo e as linhas diametrais representam os pontos de inserção entre os genes *E2* e as respectivas inserções no genoma humano (linhas roxas representando a amostra integrada; linha azul representando o controle -SiHa).

O gene *GPR1585* é estimulado por andrógenos e codifica um receptor que ativa a proliferação celular em câncer de próstata e está relacionado também com a diferenciação

neuroendócrina (Patel et al., 2015). O gene *PAX 5* por sua vez é um gene que codifica para família de fatores de transcrição que são importantes reguladores do desenvolvimento e alterações em sua expressão estão relacionadas à transformações neoplásicas. Está localizado no cromossomo 9 na região 9p13 e a desregulação de sua expressão está relacionada com certos tipos de linfoma (Torlakovic et al., 2002).

Embora essa abordagem seja mais moderna e robusta, algumas limitações e dificuldades dessa metodologia são dignas de nota. Os reagentes têm um custo bastante elevado e os prazos para a customização também são longos. Quando todo o material foi recebido, encontramos dificuldade com a disponibilidade do equipamento e de técnicos com conhecimento específico para a metodologia de *TargetSeq*. Outra limitação do método é o tempo de preparo das amostras. Caso o trabalho seja realizado em tempo integral, o tempo médio para o preparo de um grupo de 16 amostras para uma corrida (incluindo preparo de bibliotecas, incubação com sondas e PCR de emulsão) é em média 5 dias e não existem etapas intervenientes que possibilitem avaliar o rendimento individual de cada etapa. Apesar dessas limitações, espera-se que com futuras corridas e com suporte do serviço de bioinformática da Faculdade de Medicina de Ribeirão Preto seja possível aprimorar essas estratégias ou mesmo estabelecer novas abordagens de NGS para as questões de integração.

5. DISCUSSÃO

Embora diversos estudos tenham demonstrado que as oncoproteínas virais E6 e E7 podem interagir com inúmeras proteínas celulares e dessa forma contribuir para a instabilidade genômica, apenas essas interações não são suficientes para a transformação maligna e por isso, alterações genéticas e mecanismos adicionais são requeridos para esse processo (Saha et al., 2010; Crosbie et al., 2013). A integração viral entra nesse contexto de forma dupla, tanto com a perda do produto de expressão do gene *E2* pela ruptura do genoma viral que culmina na expressão constitutiva dos oncogenes virais *E6* e *E7* quanto pela mutagenese insercional que o processo provoca (Yu et al., 2005; Schmitz et al., 2012; Akagi et al., 2014).

O entendimento do delicado balanço entre os fatores do hospedeiro e do vírus podem fornecer informações fundamentais para a identificação precoce dessas infecções e até mesmo dirigir o tratamento e prognósticos de suas formas mais avançadas. Mesmo que o encontro de um biomarcador ideal ainda esteja longe de ser alcançado, pistas que subjazem o processo de integração são de suma importância, portanto além do processo de integração, a precisão do local de inserção no genoma humano tem sido alvo de estudo por inúmeros trabalhos com o uso de sequenciamento de nova geração (Yu et al., 2005; Xu et al., 2013; Rusan et al., 2015; Bodelon et al., 2016).

Embora as lesões neoplásicas em homens sejam mais raras mesmo em países onde o risco é mais alto, como no caso do Brasil, o câncer de pênis está associado a uma alta morbidade e mortalidade e, portanto as diretrizes para estratégias de prevenção e rastreio devem ser cuidadosamente elaboradas (Forman et al., 2012; Ingles et al., 2014). Além disso, infecções na região anal mesmo entre homens heterossexuais não são incomuns, necessitando de mais dados sobre os fatores de risco virais para este sítio anatômico (Nyitray et al., 2015).

Adicionalmente, a crescente frequência de tumores na região de cabeça e pescoço atribuídas pelo HPV é três vezes mais alta em homens do que em mulheres (Marur et al., 2010).

Considerando que a prevalência da infecção pelo HPV na região genital masculina é alta, são poucos os trabalhos que avaliam a integração e carga viral, especialmente em homens assintomáticos. Dessa forma, um estudo com um grande número de homens acompanhados por um longo período de tempo tem um grande potencial de fornecer informações de diferentes aspectos, visando uma melhor compreensão do papel dos homens na transmissão e manutenção dessas infecções, além de um melhor entendimento nos mecanismos envolvidos no desenvolvimento de lesões neoplásicas nesses indivíduos.

Há muito têm sido discutidas as diferentes metodologias de detecção de HPV e seus significados clínicos. A sensibilidade desses testes tanto para a detecção do maior número possível de tipos virais quanto a menor quantidade deles contidas na amostra, tem sido muito importante para estudos epidemiológicos em grande escala (Gheit et al., 2007). Nesse contexto está a metodologia baseada em Luminex, que tem sido utilizado com sucesso para a tipagem de HPV e se mostrou o mais sensível dos três testes realizados, corroborando com outros trabalhos que o compara com as metodologias baseadas em hibridação reversa e PCR (Schmitt et al., 2005; Clifford et al., 2016). Nesse trabalho, porém, nos deparamos com uma dificuldade técnica e não foi possível a análise de integração e carga viral a partir de amostras positivas exclusivamente por Luminex. Embora alguns trabalhos tenham confirmado a sensibilidade analítica do Luminex, esse teste não foi clinicamente avaliado, não foi desenvolvido para rastreio e pode ser erroneamente interpretado dependendo do estudo em questão (Clifford et al., 2016). Por outro lado, tentar verificar como diferentes metodologias de detecção de HPV se comportam em ensaios de integração foi uma proposta pioneira e os dados de concordância e a discussão de múltiplas metodologias é um manuscrito em preparação pelo nosso grupo de pesquisa.

Enquanto muitos estudos com mulheres já observaram uma forte associação entre a carga viral, persistência e gravidade da lesão, estudos relacionando a carga viral em homem envolve um número menor de indivíduos (grande parte soropositivos para HIV) e existe muita variação tanto do sítio anatômico analisado quanto aos métodos utilizados para estimá-la. Alguns estudos tem observado uma menor carga viral nas infecções por HPV em homens com relação às mulheres enquanto que ainda não foi estabelecida uma associação entre alta carga viral e persistência e progressão de lesões na população masculina. (Francheschi et al., 2002; Bleeker et al., 2006; Pierangeli et al., 2008; Alvarez et al, 2010).

Para a determinação da carga viral do presente estudo, escolhemos a mesma metodologia utilizada pelo nosso laboratório para amostras de cérvix uterina e os resultados são reportados em cópias por células. A metodologia para coleta das amostras da região genital poderia em parte explicar a alta variação da carga viral já que a extensão da área coletada é bastante ampla e abrange corpo do pênis, glândula e saco escrotal, porém, as variações da carga viral do canal anal e cavidade oral foram muito amplas. De qualquer maneira isso não foi um fator limitante para a determinação da carga viral que pode ser verificada em 80% das amostras, entretanto talvez tenha sido limitante para os testes de integração. Embora isso tenha sido por diversas vezes evidenciado neste trabalho, a alta variação da carga viral também é relatada por outros autores (Pierangeli et al., 2008; Flores et al., 2009; Alvarez et al, 2010).

É importante ressaltar que um fator importante das medidas de carga viral para este trabalho foi a determinação da sensibilidade dos testes utilizados tanto na detecção do HPV, conforme mencionado anteriormente, quanto na sensibilidade em determinar integração. Além disso, devemos ainda lembrar que existe uma heterogeneidade muito grande entre as amostras por se tratar de três diferentes sítios anatômicos. Dessa forma, a “otimização” das reações de PCR convencional e em tempo real foi importante para que se utilizasse o mínimo

de material para o máximo de resultados já que este trabalho emprega duas metodologias com múltiplas etapas que necessita de uma grande quantidade de DNA e com os resultados dependentes de sua qualidade. Além disso, espécimes biológicos devem ser utilizados com cautela para que possam ser explorados com novas tecnologias, permitindo comparações com metodologias previamente utilizadas e eventualmente validadas.

Com relação aos ensaios de integração, este estudo utilizou primeiramente uma metodologia baseada em PCR e posteriormente esses dados puderam ser comparados com a metodologia de sequenciamento de ultima geração. Uma vantagem da nossa metodologia baseada em PCR é que analisamos o gene *E2* inteiro mais uma porção do gene *E1* e não porções do gene *E2* como muitos trabalhos têm relatado (Peitsaro et al., 2002; Cañadas et al., 2010; Ramanakumar et al., 2010). Por outro lado, pela escassez de material e heterogeneidade das amostras quanto à integridade, poderíamos obter resultados de falso casos de integração devido à dificuldade em amplificar fragmentos de aproximadamente 800 pb.

Diversos trabalhos tem demonstrado que a frequência de integração é maior em lesões mais graves e que uma vez integrado, a replicação viral fica limitada à replicação celular, estabelecendo um ciclo produtivo incompleto em que partículas virais não chegam a ser formadas (Berg and Stenlund, 1997; Das et al., 2010). Como consequência, ocorre uma diminuição da transmissão devido à baixa carga viral. Essa mesma diminuição pode ser um fator limitante para as análises de integração e por outro lado, avaliar amostras com baixa carga viral é de grande importância já que podem conter a forma integrada.

Neste trabalho, se considerarmos as amostras mistas como amostras integradas, então teremos uma alta prevalência desse evento já que aproximadamente 70% das amostras mostraram algum grau de integração. Essa observação pode ser explicada com base em trabalhos que exploram a ideia de que a integração é um evento precoce no processo infeccioso de determinados tipos de HPV e que pode ocorrer em vários sítios cromossômicos

(Peitsaro et al., 2002; Yu et al., 2005; Collins et al., 2009). Entretanto, em células persistentemente infectadas podem ocorrer integrações em sítios específicos que contribuiriam para o processo tumorigênico, estes sítios por outro lado seriam enriquecidos pelo processo de seleção clonal e dessa forma vem a idéia de *hotspots* de integração descrito em diversos trabalhos (Akagi et al., 2014; Hu et al., 2015; Bodelon et al., 2016). *Hotspots* são geralmente determinados em estudos que utiliza NGS e deve ser um sítio onde eventos são recorrentes entre diferentes amostras e de preferência, observado por diferentes grupos de estudos. Por outro lado, amostras que contenham apenas eventos de integração isolados numa única célula sem a subsequente seleção clonal não seriam prontamente identificadas.

Analisando as amostras selecionadas para o sequenciamento, observamos que em amostras que continham menor carga viral, não foi possível capturar fragmentos contendo sequências virais durante a hibridização. Isso se deve ao fato das múltiplas etapas do preparo das bibliotecas e se mostrou como uma limitação da metodologia. Durante o preparo das bibliotecas, o DNA é fragmentado, e apenas fragmentos de aproximadamente 200 pares de bases são capturados. Se nessa etapa, tivermos poucos fragmentos de HPV e estes não estiverem com o tamanho adequado, serão eliminados antes da incubação com as sondas.

Isso pode se refletir também na análise de amostras que são classificadas como integradas ou mistas pela metodologia de qPCR já que com a diminuição do número de cópias virais, a sensibilidade e acurácia da metodologia diminui, podendo levar a falsos resultados de integração.

Embora este trabalho não pretenda explorar os mecanismos de integração, pela observação obtida da única amostra que se mostrou integrada pelas duas metodologias exploradas, deve se mencionar uma recente elucidação desse mecanismo baseada em regiões de microhomologias entre o genoma viral e celular que predis põem à ruptura na região do gene *E2* (Akagi et al., 2014; Hu et al., 2015). Adicionalmente, foi observado o envolvimento

da maquinaria de reparo celular na amplificação do genoma viral, corroborando com o fato de que os processos de recombinação homóloga podem cooperar ou ter algum papel no processo de integração do genoma viral (Kadaja et al., 2009). Outra observação importante foi que em regiões onde ocorrem as inserções são observadas translocações, rearranjos e amplificações tanto de sequências do genoma da célula hospedeira quanto do genoma viral, o que justificaria em parte a inserção da mesma região gênica do genoma do HPV em diferentes cromossomos da célula hospedeira (Xu et al., 2013; Parfenov et al., 2014; Rusan et al., 2015). Além disso, uma recente publicação que analisa dados de sequenciamento de larga escala de diversos estudos envolvendo integração de HPV observou uma tendência de integração no cromossomo 9 em indivíduos assintomáticos (Bodelon et al., 2016).

Dentre as várias formas de avaliar qualidade do sequenciamento, a mais comum e importante para este estudo foi obtida através dos gráficos de Qualidade da Sequência por base (ANEXO A). Esta variável tem como parâmetro o *Phred Score* ou *Q Score* que é a representação da fidelidade do equipamento em discriminar corretamente a base nucleotídica chamada. É definido como uma propriedade logaritmicamente relacionada à probabilidade de erro de reconhecimento da base e calculado a partir da expressão $Q = -10 \log_{10} P$. Assim uma sequência correta é o primeiro passo para o mapeamento acurado evitando conclusões imprecisas (Andrews, 2010).

A qualidade do sequenciamento depende de vários fatores que podem ser controlados, como por exemplo, a escolha dos reagentes, preparo das bibliotecas e correta utilização dos protocolos, e fatores não controlados como a fonte das amostras. Amostras de DNA de diferentes sítios anatômicos já foram repetidamente mostradas com alta heterogeneidade de integridade ao longo desse trabalho, e esta variável poderia ser um fator limitante para o processo de sequenciamento já que a primeira etapa de preparo das bibliotecas é a fragmentação física do DNA. Dessa forma, a padronização dessa metodologia deve ser focada

no quanto cada tipo de amostra deve ser exposta à fragmentação física de acordo com o sítio anatômico. Isso pode inclusive ser estimado com resultados obtidos pelo Bioanalyzer.

Outra variável que se mostrou importante para o sequenciamento foi a carga de HPV na amostra. Embora células cujo DNA viral encontra-se integrado tenham baixa ou nenhuma replicação do genoma viral (Woodman et al., 2007), a sensibilidade das duas metodologias parecem convergir nesse ponto já que em 5 amostras classificadas como integradas pela metodologia de PCR apresentavam baixas cargas virais e não foram identificados fragmentos com sequências de HPV. Dessa forma a ampliação desse teste com um número maior de amostras pode inclusive estimar com maior segurança um cutt-off para a realização desse tipo de ensaio.

Embora este trabalho não tem como objetivo ser um estudo epidemiológico e sim o desenvolvimento de uma ferramenta para futuras análises epidemiológicas, estes dados, testados em um número maior de amostras, servirão de base para a discussão do risco da presença de integração ou de genomas livres no risco de desenvolvimento de lesão em homens.

6. CONCLUSÕES

- A integração só pode ser verificada em amostras biológicas positivas para HPV pela metodologia de Linear Array (amplificação de um fragmento de 450 pb do gene *L1* de diversos tipos de HPV);
- A integridade do DNA é um fator importante para se determinar a integração viral por estes métodos;
- Dos três sítios anatômicos analisados o DNA de amostras genitais está mais bem preservado;
- Embora amostras de cavidade oral tenha um DNA mais preservado e em maior quantidade, a carga viral é extremamente baixa e limitante para análise de integração;
- Houve concordância entre as metodologias para verificar integração apenas em amostras com cargas virais acima de 2,33 cópias por célula;
- Além de determinar a integração o ensaio baseado em sequenciamento (NGS) é capaz de informar com precisão o local de inserção do genoma viral, podendo inclusive identificar formas menos convencionais de integração que mantém intacta a região do gene *E2*;
- Houve uma alta frequência de formas virais classificadas como mistas (contendo genomas livres e integrados);
- A carga viral foi extremamente variável entre os três sítios anatômicos variando entre 8×10^{-5} até 14500 cópias por célula;
- As amostras positivas para qPCR e Luminex mas negativas para Linear Array tinham carga viral $< 0,28$ cópias/célula para canal anal, $< 0,3$ cópias/célula para cavidade oral e $< 0,15$ cópias/célula para a região genital – refletindo a sensibilidade dos diferentes métodos;
- Carga viral extremamente baixa pode ser um fator limitante para verificar integração.

7. REFERÊNCIAS

AKAGI, Keiko et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. **Genome research**, v. 24, n. 2, p. 185-199, 2014.

ALLEN, Clint T. et al. Human papillomavirus and oropharynx cancer: biology, detection and clinical implications. **The Laryngoscope**, v. 120, n. 9, p. 1756-1772, 2010.

ALVAREZ, Jennifer et al. Episomal and integrated human papillomavirus type 16 loads and anal intraepithelial neoplasia in HIV-seropositive men. **Aids**, v. 24, n. 15, p. 2355-2363, 2010.

ARIAS-PULIDO, Hugo et al. Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. **Journal of clinical microbiology**, v. 44, n. 5, p. 1755-1762, 2006.

ASHRAFI, G. Hossein et al. E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain. **International journal of cancer**, v. 119, n. 9, p. 2105-2112, 2006.

BERG, Mikael; STENLUND, Arne. Functional interactions between papillomavirus E1 and E2 proteins. **Journal of virology**, v. 71, n. 5, p. 3853-3863, 1997.

BERGERON, Christine et al. The clinical impact of using p16INK4a immunochemistry in cervical histopathology and cytology: An update of recent developments. **International journal of cancer**, v. 136, n. 12, p. 2741-2751, 2015.

BLEEKER, Maaïke CG et al. Flat penile lesions: the infectious “invisible” link in the transmission of human papillomavirus. **International journal of cancer**, v. 119, n. 11, p. 2505-2512, 2006.

BODILY, Jason; LAIMINS, Laimonis A. Persistence of human papillomavirus infection: keys to malignant progression. **Trends in microbiology**, v. 19, n. 1, p. 33-39, 2011.

BOUVARD, Veronique et al. The human papillomavirus type 16 E5 gene cooperates with the E7 gene to stimulate proliferation of primary cells and increases viral gene expression. **Virology**, v. 203, n. 1, p. 73-80, 1994.

BUSTIN, Stephen A. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. **Clinical chemistry**, v. 55, n. 4, p. 611-622, 2009.

CANADAS, María-Paz et al. Human papillomavirus 16 integration and risk factors associated in anal samples of HIV-1 infected men. **Sexually transmitted diseases**, v. 37, n. 5, p. 311-315, 2010.

CASTLE, Philip E. et al. Short term persistence of human papillomavirus and risk of cervical precancer and cancer: population based cohort study. **Bmj**, v. 339, p. b2569, 2009.

CHAIWONGKOT, Arkom et al. Differential methylation of E2 binding sites in episomal and integrated HPV 16 genomes in preinvasive and invasive cervical lesions. **International Journal of Cancer**, v. 132, n. 9, p. 2087-2094, 2013.

CHEN, Yan et al. Viral carcinogenesis: factors inducing DNA damage and virus integration. **Cancers**, v. 6, n. 4, p. 2155-2186, 2014.

CLIFFORD, Gary M. et al. Comparison of two widely-used HPV detection and genotyping methods: GP5+/6+ PCR followed by reverse line blot hybridization and multiplex type-specific E7 PCR. **Journal of clinical microbiology**, p. JCM. 00618-16, 2016.

COLLINS, Stuart I. et al. Disruption of the E2 gene is a common and early event in the natural history of cervical human papillomavirus infection: a longitudinal cohort study. **Cancer research**, v. 69, n. 9, p. 3828-3832, 2009.

CROSBIE, Emma J. et al. Human papillomavirus and cervical cancer. **The Lancet**, v. 382, n. 9895, p. 889-899, 2013.

CRICCA, Monica et al. Disruption of HPV 16 E1 and E2 genes in precancerous cervical lesions. **Journal of virological methods**, v. 158, n. 1, p. 180-183, 2009.

DAS, Damayanti et al. Association of viral load with HPV16 positive cervical cancer pathogenesis: causal relevance in isolates harboring intact viral E2 gene. **Virology**, v. 402, n. 1, p. 197-202, 2010.

DE VILLIERS, Ethel-Michele et al. Classification of papillomaviruses. **Virology**, v. 324, n. 1, p. 17-27, 2004.

DELLAROLE, Mariano; SÁNCHEZ, Ignacio E.; DE PRAT GAY, Gonzalo. Thermodynamics of cooperative DNA recognition at a replication origin and transcription regulatory site. **Biochemistry**, v. 49, n. 48, p. 10277-10286, 2010.

DIAO, Ming-Kun et al. Integrated HPV genomes tend to integrate in gene desert areas in the CaSki, HeLa, and SiHa cervical cancer cell lines. **Life sciences**, v. 127, p. 46-52, 2015.

DOORBAR, John et al. Human papillomavirus molecular biology and disease association. **Reviews in medical virology**, v. 25, n. S1, p. 2-23, 2015.

DOORBAR, John et al. The biology and life-cycle of human papillomaviruses. **Vaccine**, v. 30, p. F55-F70, 2012.

DOORBAR, John. Molecular biology of human papillomavirus infection and cervical cancer. **Clinical science**, v. 110, n. 5, p. 525-541, 2006.

DOORBAR, J. et al. Specific interaction between HPV-16 E1–E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. 1991.

DRUSCO, Alessandra et al. Common fragile site tumor suppressor genes and corresponding mouse models of cancer. **BioMed Research International**, v. 2011, 2010.

DISBROW, Gary L.; HANOVER, John A.; SCHLEGEL, Richard. Endoplasmic reticulum-localized human papillomavirus type 16 E5 protein alters endosomal pH but not trans-Golgi pH. **Journal of virology**, v. 79, n. 9, p. 5839-5846, 2005.

FLORES, Roberto et al. Correlates of human papillomavirus viral load with infection site in asymptomatic men. **Cancer Epidemiology Biomarkers & Prevention**, v. 17, n. 12, p. 3573-3576, 2008.

FORMAN, David et al. Global burden of human papillomavirus and related diseases. **Vaccine**, v. 30, p. F12-F23, 2012.

FRANCESCHI, S. et al. Prevalence and determinants of human papillomavirus genital infection in men. **British journal of cancer**, v. 86, n. 5, p. 705-711, 2002.

FRANCO, Eduardo L. et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. **Journal of Infectious Diseases**, v. 180, n. 5, p. 1415-1423, 1999.

FU, Leiping et al. Degradation of p53 by human Alphapapillomavirus E6 proteins shows a stronger correlation with phylogeny than oncogenicity. **PloS one**, v. 5, n. 9, p. e12816, 2010.

GALE, Nina; ZIDAR, Nina. Benign and potentially malignant lesions of the squamous epithelium and squamous cell carcinoma. In: **Pathology of the Head and Neck**. Springer Berlin Heidelberg, 2006. p. 1-38.

GARFIELD, E. ALL SORTS OF WARTS-SEPARATING FACT FROM FICTION. 1. ETIOLOGY, BIOLOGY, AND RESEARCH MILESTONES. **Current contents**, n. 9, p. 3-11, 1988.

GHEIT, Tarik et al. Development of a sensitive and specific multiplex PCR method combined with DNA microarray primer extension to detect Betapapillomavirus types. **Journal of clinical microbiology**, v. 45, n. 8, p. 2537-2544, 2007.

GIULIANO, Anna R. et al. The human papillomavirus infection in men study: human papillomavirus prevalence and type distribution among men residing in Brazil, Mexico, and the United States. **Cancer Epidemiology Biomarkers & Prevention**, v. 17, n. 8, p. 2036-2043, 2008.

GRAVITT, P. E. et al. Improved amplification of genital human papillomaviruses. **Journal of clinical microbiology**, v. 38, n. 1, p. 357-361, 2000.

GRAVITT, Patti E. et al. Reproducibility of HPV 16 and HPV 18 viral load quantitation using TaqMan real-time PCR assays. **Journal of virological methods**, v. 112, n. 1, p. 23-33, 2003.

GROVES, Ian J.; COLEMAN, Nicholas. Pathogenesis of human papillomavirus-associated mucosal disease. **The Journal of pathology**, v. 235, n. 4, p. 527-538, 2015.

HARARI, Ariana; CHEN, Zigui; BURK, Robert D. Human papillomavirus genomics: past, present and future. In: **Human Papillomavirus**. Karger Publishers, 2014. p. 1-18.

HO, Daniel WH; SZE, Karen MF; NG, Irene OL. Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. **Oncotarget**, v. 6, n. 25, p. 20959, 2015.

HU, Zheng et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. **Nature genetics**, v. 47, n. 2, p. 158-163, 2015.

HWANG, Eun-Seong; NOTTOLI, Timothy; DIMAIO, Daniel. The HPV16 E5 protein: expression, detection, and stable complex formation with transmembrane proteins in COS cells. **Virology**, v. 211, n. 1, p. 227-233, 1995.

IARC - INTERNATIONAL AGENCY FOR RESEARCH ON CANCER, WORLD HEALTH ORGANIZATION. IARC monographs on the evaluation of carcinogenic risks to humans. Vol. 90. Human papillomaviruses. Lyons, France: International Agency for Research on Cancer, 2007.

IFTNER, Thomas; VILLA, Luisa Lina. Human papillomavirus technologies. **Journal of the National Cancer Institute. Monographs**, n. 31, p. 80-88, 2002.

INGLES, Donna J. et al. Human papillomavirus virus (HPV) genotype-and age-specific analyses of external genital lesions among men in the HPV Infection in Men (HIM) Study. **Journal of Infectious Diseases**, p. jiu587, 2014.

JOHANSSON, Cecilia; SCHWARTZ, Stefan. Regulation of human papillomavirus gene expression by splicing and polyadenylation. **Nature reviews Microbiology**, v. 11, n. 4, p. 239-251, 2013.

KADAJA, Meelis et al. Mechanism of genomic instability in cells infected with the high-risk human papillomaviruses. **PLoS Pathog**, v. 5, n. 4, p. e1000397, 2009.

KAJITANI, Naoko et al. Productive lifecycle of human papillomaviruses that depends upon squamous epithelial differentiation. **Receptor-independent/-associated viral tropism**, v. 3, n. 152, p. 5, 2012.

KALANTARI, Mina et al. Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. **International journal of gynecological pathology**, v. 17, n. 2, p. 146-153, 1998.

KREIMER, Aimee R. et al. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. **Cancer Epidemiology Biomarkers & Prevention**, v. 14, n. 2, p. 467-475, 2005.

KREIMER, Aimée R. et al. Incidence and clearance of oral human papillomavirus infection in men: the HIM cohort study. **The Lancet**, v. 382, n. 9895, p. 877-887, 2013.

KRZYWINSKI, Martin et al. Circos: an information aesthetic for comparative genomics. **Genome research**, v. 19, n. 9, p. 1639-1645, 2009.

KURG, Reet et al. Human papillomavirus E2 protein with single activation domain initiates HPV18 genome replication, but is not sufficient for long-term maintenance of virus genome. **Virology**, v. 408, n. 2, p. 159-166, 2010.

LI, Heng; DURBIN, Richard. Fast and accurate long-read alignment with Burrows–Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589-595, 2010.

LI, Heng et al. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009.

LI, Wei et al. The physical state of HPV16 infection and its clinical significance in cancer precursor lesion and cervical carcinoma. **Journal of cancer research and clinical oncology**, v. 134, n. 12, p. 1355-1361, 2008.

LONGWORTH, Michelle S.; LAIMINS, Laimonis A. The binding of histone deacetylases and the integrity of zinc finger-like motifs of the E7 protein are essential for the life cycle of human papillomavirus type 31. **Journal of virology**, v. 78, n. 7, p. 3533-3541, 2004.

MARUR, Shanthi et al. HPV-associated head and neck cancer: a virus-related cancer epidemic. **The lancet oncology**, v. 11, n. 8, p. 781-789, 2010.

MCINTOSH, Pauline B. et al. E1⁺E4-mediated keratin phosphorylation and ubiquitylation: A mechanism for keratin depletion in HPV16-infected epithelium. **J Cell Sci**, v. 123, n. 16, p. 2810-2822, 2010.

MCLAUGHLIN-DRUBIN, Margaret E.; MÜNGER, Karl. The human papillomavirus E7 oncoprotein. **Virology**, v. 384, n. 2, p. 335-344, 2009.

MESRI, Enrique A.; FEITELSON, Mark A.; MUNGER, Karl. Human viral oncogenesis: a cancer hallmarks analysis. **Cell host & microbe**, v. 15, n. 3, p. 266-282, 2014.

MIGHTY, Kristen K.; LAIMINS, Laimonis A. The role of human papillomaviruses in oncogenesis. In: **Viruses and Human Cancer**. Springer Berlin Heidelberg, 2014. p. 135-148.

MOLE, S. et al. Regulation of splicing-associated SR proteins by HPV-16. **Biochemical Society Transactions**, v. 34, n. 6, p. 1145-1147, 2006.

MOODY, Cary A.; LAIMINS, Laimonis A. Human papillomavirus oncoproteins: pathways to transformation. **Nature Reviews Cancer**, v. 10, n. 8, p. 550-560, 2010.

MÜNGER, Karl et al. Mechanisms of human papillomavirus-induced oncogenesis. **Journal of virology**, v. 78, n. 21, p. 11451-11460, 2004.

MUÑOZ, Nubia et al. Against which human papillomavirus types shall we vaccinate and screen? The international perspective. **International Journal of Cancer**, v. 111, n. 2, p. 278-285, 2004.

MUÑOZ, Nubia et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. **New England Journal of Medicine**, v. 348, n. 6, p. 518-527, 2003.

NGUYEN, Don X.; WESTBROOK, Thomas F.; MCCANCE, Dennis J. Human papillomavirus type 16 E7 maintains elevated levels of the cdc25A tyrosine phosphatase during deregulation of cell cycle arrest. **Journal of virology**, v. 76, n. 2, p. 619-632, 2002.

NICOLAIDES, Lietta et al. Stabilization of HPV16 E6 protein by PDZ proteins, and potential implications for genome maintenance. **Virology**, v. 414, n. 2, p. 137-145, 2011.

NUNES, Emily Montosa et al. Diversity of beta-papillomavirus at anogenital and oral anatomic sites of men: The HIM Study. **Virology**, v. 495, p. 33-41, 2016.

NYITRAY, Alan G. et al. The natural history of genital human papillomavirus among HIV-negative men having sex with men and men having sex with women. **Journal of Infectious Diseases**, p. jiv061, 2015.

NYITRAY, Alan G. et al. Incidence, duration, persistence, and factors associated with high-risk anal HPV persistence among HIV-negative men having sex with men: A multi-national study. **Clinical Infectious Diseases**, p. ciw140, 2016.

PARFENOV, Michael et al. Characterization of HPV and host genome interactions in primary head and neck cancers. **Proceedings of the National Academy of Sciences**, v. 111, n. 43, p. 15544-15549, 2014.

PARK, Jong Sup et al. Physical status and expression of HPV genes in cervical cancers. **Gynecologic oncology**, v. 65, n. 1, p. 121-129, 1997.

PATEL, Nitin et al. Expression and functional role of orphan receptor GPR158 in prostate cancer growth and progression. **PloS one**, v. 10, n. 2, p. e0117758, 2015.

PATEL, Ravi K.; JAIN, Mukesh. NGS QC Toolkit: A Platform for Quality Control of Next-Generation Sequencing Data. In: **Encyclopedia of Metagenomics**. Springer US, 2015. p. 544-548.

PEITSARO, Panu; JOHANSSON, Bo; SYRJÄNEN, Stina. Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique. **Journal of clinical microbiology**, v. 40, n. 3, p. 886-891, 2002.

PETT, Mark R. et al. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. **Cancer research**, v. 64, n. 4, p. 1359-1368, 2004.

PINHEIRO, Leonardo B. et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. **Analytical chemistry**, v. 84, n. 2, p. 1003-1011, 2011.

RAMANAKUMAR, Agnihotram V. et al. Human papillomavirus (HPV) types 16, 18, 31, 45 DNA loads and HPV-16 integration in persistent and transient infections in young women. **BMC infectious diseases**, v. 10, n. 1, p. 1, 2010.

ROBERTS, Sally et al. Mutational analysis of the human papillomavirus type 16 E1--E4 protein shows that the C terminus is dispensable for keratin cytoskeleton association but is

involved in inducing disruption of the keratin filaments. **Journal of virology**, v. 71, n. 5, p. 3554-3562, 1997.

ROUTH, Andrew; JOHNSON, John E. Discovery of functional genomic motifs in viruses with ViReMa—a Virus Recombination Mapper—for analysis of next-generation sequencing data. **Nucleic acids research**, v. 42, n. 2, p. e11-e11, 2014.

RUSAN, Maria; LI, Yvonne Y.; HAMMERMAN, Peter S. Genomic landscape of human papillomavirus-associated cancers. **Clinical cancer research**, v. 21, n. 9, p. 2009-2019, 2015.

SAHA, Abhik et al. Tumor viruses and cancer biology: Modulating signaling pathways for therapeutic intervention. **Cancer biology & therapy**, v. 10, n. 10, p. 961-978, 2010.

SCHEFFNER, Martin et al. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. **Cell**, v. 63, n. 6, p. 1129-1136, 1990.

SCHLECHT, Nicolas F. et al. Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia. **Jama**, v. 286, n. 24, p. 3106-3114, 2001.

SCHMIEDER, Robert; EDWARDS, Robert. Quality control and preprocessing of metagenomic datasets. **Bioinformatics**, v. 27, n. 6, p. 863-864, 2011.

SCHMITT, Markus et al. Bead-based multiplex genotyping of human papillomaviruses. **Journal of clinical microbiology**, v. 44, n. 2, p. 504-512, 2006.

SCHMITZ, Martina et al. Loss of gene function as a consequence of human papillomavirus DNA integration. **International journal of cancer**, v. 131, n. 5, p. E593-E602, 2012.

SNELLENBERG, Suzanne et al. Methylation status of the E2 binding sites of HPV16 in cervical lesions determined with the Luminex® xMAP™ system. **Virology**, v. 422, n. 2, p. 357-365, 2012.

STANLEY, Margaret. Pathology and epidemiology of HPV infection in females. **Gynecologic oncology**, v. 117, n. 2, p. S5-S10, 2010.

STANLEY, Margaret. Immunobiology of HPV and HPV vaccines. **Gynecologic oncology**, v. 109, n. 2, p. S15-S21, 2008.

STRAUSS, Maurice J. et al. “Crystalline” Virus-Like Particles from Skin Papillomas Characterized by Intranuclear Inclusion Bodies. **Experimental Biology and Medicine**, v. 72, n. 1, p. 46-50, 1949.

SUDENGA, Staci L. et al. Genital human papillomavirus infection progression to external genital lesions: the HIM study. **European urology**, v. 69, n. 1, p. 166-173, 2016.

SURVILADZE, Zurab; DZIDUSZKO, Agnieszka; OZBUN, Michelle A. Essential roles for soluble virion-associated heparan sulfonated proteoglycans and growth factors in human papillomavirus infections. **PLoS Pathog**, v. 8, n. 2, p. e1002519, 2012.

THIERRY, Françoise. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. **Virology**, v. 384, n. 2, p. 375-379, 2009.

THOMAS, M. et al. Human papillomaviruses, cervical cancer and cell polarity. **Oncogene**, v. 27, n. 55, p. 7018-7030, 2008.

TORLAKOVIC, Emina et al. The value of anti-pax-5 immunostaining in routinely fixed and paraffin-embedded sections: a novel pan pre-B and B-cell marker. **The American journal of surgical pathology**, v. 26, n. 10, p. 1343-1350, 2002.

TORNESELLO, Maria Lina et al. Viral and cellular biomarkers in the diagnosis of cervical intraepithelial neoplasia and cancer. **BioMed research international**, v. 2013, 2013.

TREVISAN, Andrea et al. Human papillomavirus type 16 viral load measurement as a predictor of infection clearance. **Journal of General Virology**, v. 94, n. 8, p. 1850-1857, 2013.

TREVISAN Andrea et al. Coexistence of episomal and integrated HPV16 genomes in cervical cells. for the **Ludwig-McGill Study Group. 2016** (em submissão).

TSAKOGIANNIS, D. et al. Nucleotide polymorphisms of the human papillomavirus 16 E1 gene. **Archives of virology**, v. 159, n. 1, p. 51-63, 2014.

VENUTI, Aldo et al. Papillomavirus E5: the smallest oncoprotein with many functions. **Molecular cancer**, v. 10, n. 1, p. 1, 2011.

VINOKUROVA, Svetlana et al. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. **Cancer research**, v. 68, n. 1, p. 307-313, 2008.

VILLA, Luisa L. et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. **The lancet oncology**, v. 6, n. 5, p. 271-278, 2005.

WALBOOMER, J. A. N. et al. HUMAN PAPILLOMAVIRUS IS A NECESSARY CAUSE OF INVASIVE CERVICAL CANCER. **VVORLDWIDE.J. pathol**, v. 189, p. 12-19, 1999.

WALLACE, Nicholas A.; GALLOWAY, Denise A. Novel Functions of the Human Papillomavirus E6 Oncoproteins. **Annual review of virology**, v. 2, p. 403-423, 2015.

WANG, Kai; LI, Mingyao; HAKONARSON, Hakon. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. **Nucleic acids research**, v. 38, n. 16, p. e164-e164, 2010.

WANG, Xin et al. Recruitment of Brd4 to the human papillomavirus type 16 DNA replication complex is essential for replication of viral DNA. **Journal of virology**, v. 87, n. 7, p. 3871-3884, 2013.

WHITE, Elizabeth A. et al. Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses. **Proceedings of the National Academy of Sciences**, v. 109, n. 5, p. E260-E267, 2012.

WOODMAN, Ciaran BJ; COLLINS, Stuart I.; YOUNG, Lawrence S. The natural history of cervical HPV infection: unresolved issues. **Nature Reviews Cancer**, v. 7, n. 1, p. 11-22, 2007.

XU, Bo et al. Multiplex identification of human papillomavirus 16 DNA integration sites in cervical carcinomas. **PloS one**, v. 8, n. 6, p. e66693, 2013.

YEE, Carole et al. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. **The American journal of pathology**, v. 119, n. 3, p. 361, 1985.

YLITALO, Nathalie et al. A prospective study showing long-term infection with human papillomavirus 16 before the development of cervical carcinoma in situ. **Cancer research**, v. 60, n. 21, p. 6027-6032, 2000.

YU, Tingxi et al. The role of viral integration in the development of cervical cancer. **Cancer genetics and cytogenetics**, v. 158, n. 1, p. 27-34, 2005.

ZUR HAUSEN, Harald et al. Attempts to detect virus-specific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus. **International Journal of Cancer**, v. 13, n. 5, p. 650-656, 1974.

ZHANG, Benyue et al. The E5 protein of human papillomavirus type 16 perturbs MHC class II antigen maturation in human foreskin keratinocytes treated with interferon- γ . **Virology**, v. 310, n. 1, p. 100-108, 2003.

8. LISTA DE ANEXOS

ANEXO A – Carta de Aprovação da Comissão de Ética em Pesquisa

ANEXO B - Gráfico da qualidade de Sequência por Base da amostra que foi encontrada
integração

ANEXO C – Súmula Curricular

ANEXO D – Manuscrito Submetido para o *Brazilian Journal of Infectious Diseases (BJID)*
intitulado *HPV-related external genital lesions among men residing in Brazil*.

ANEXO A – Carta de Aprovação do Comitê de Ética em Pesquisa

CIÊNCIA

O Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo, em 08 de Julho de 2016, **TOMOU CIÊNCIA** do Projeto de Pesquisa intitulado: **"História Natural da Infecção por HPV em Homens – Estudo HIM"** apresentado pelo **Departamento de Centro de Investigação Translacional em Oncologia**, e por não haver óbice ético a proposta foi **Aprovada**. O projeto possui o registro de aprovação **número 028/04**, pelo CEP do Centro de Referência e Treinamento DST/AIDS – **CRT-DST/AIDS**, e também pela **CONEP** com número de **Registro 11.741**, parecer de Aprovação número 916/2005.

Pesquisador (a) Responsável: **Profa. Dra. Luisa Lina Villa**
Pesquisador (a) Executante: **Dr. Ricardo Cesar Cintra**

CEP-FMUSP, 08 de Julho de 2016

Profa. Dra. Maria Aparecida de Azevedo Koike Folgueira
Coordenador
Comitê de Ética em Pesquisa

**ANEXO B - Gráfico da qualidade de Sequência por Base da amostra que
foi encontrada integração**

ANEXO B – Representação gráfica da qualidade média por base – Verde – Boa qualidade; Laranja – Razoável qualidade; Vermelho – Qualidade ruim.

Para cada posição de base um *Box plot* é desenhado contendo uma reta vermelha com o valor da mediana e uma linha azul representando a qualidade média. O eixo y mostra as variáveis de qualidade (*quality scores*) enquanto o eixo x mostra a posição na leitura (*position in read*). Como essa representação é baseada no *Phred Score*, quanto mais alta a posição do Box, melhor a qualidade. A maioria das plataformas tende a reduzir de qualidade ao prosseguir do sequenciamento e as *reads* tendem à cor alaranjada. As principais razões para esta queda de qualidade são a degradação dos reagentes ao decorrer da corrida ou pela má qualidade da biblioteca.

Este gráfico é a representação gráfica da qualidade de sequenciamento da amostra que foi confirmada integração pelas duas metodologias.

ANEXO C – Súmula Curricular

Dados Pessoais

Nome: Ricardo Cesar Cintra

Nascimento: 05/05/1980; Naturalidade: São Paulo; Nacionalidade: Brasileiro.

Endereço: Laboratório de Inovação em Câncer – Centro Translacional em Oncologia - Centro de Medicina Nuclear, Universidade de São Paulo, Rua Doutor Ovídio Pires de Campos, 872 3º andar – Cerqueira Cesar – Cep 05403-911 São Paulo, SP. Fone: (011) 2661-7554.

E-mail: ricardocintra@usp.br; ricardotiamat@gmail.com

Formação:

- Bacharel em Ciências Biológicas pela Universidade São Judas Tadeu, São Paulo, Brasil.
- Estudante de Mestrado em Bioquímica, Instituto de Química, Departamento de Bioquímica, Universidade de São Paulo. Orientação Prof^a. Dr^a Luisa Lina Villa

Histórico profissional:

2009 - atual **Centro de Referência e Treinamento em DST/AIDS**
Unidade de Pesquisa
“Estudo HIM – História Natural da Infecção por HPV em Homens”

Desde 2009 participo ativamente no recrutamento e manutenção da coorte do Estudo *HIM*. As atividades no centro de referência envolvem aplicação de TCLE (Termo de Consentimento Livre e Esclarecido), questionários de frequência alimentar, orientação sobre Doenças Sexualmente Transmissíveis, alimentação do banco de dados *Moffitt Cancer Center*.

2013 - atual **INCT – HPV**
 Professora Dr^a Luisa Lina Villa
 Projeto: “Estado físico do Papilomavírus Humano em amostras anogenitais e orais masculinas”

Em meados de 2013, iniciei as atividades Instituto Nacional de Ciência e Tecnologia das doenças do Papilomavírus Humano – INCT – HPV- no aprimoramento e desenvolvimento de tecnologias baseadas em biologia molecular para a prevenção, detecção e tratamento das doenças causadas pelo HPV. Essa interação conduziu esta dissertação de mestrado e outros projetos em andamento.

Participação em Congresso:

X Congresso da Sociedade Brasileira De DST e VI Congresso Brasileiro de Aids – DST realizado nos dias 17, 18, 19 e 20 de maio de 2015 em São Paulo.

ANEXO D – Manuscrito Submetido para o *Brazilian Journal of Infectious Diseases (BJID)* intitulado ***HPV-related external genital lesions among men residing in Brazil.***

August 15, 2016

Submit to: Brazilian Journal of Infectious Diseases (BJID) (JCR 1.3)

HPV-related external genital lesions among men residing in Brazil

Roberto J. C. Silva¹, Staci L. Sudenga², Laura Sichero³, Maria Luiza Baggio³, Lenice Galan⁴, Ricardo Cintra⁵, B. Nelson Torres², Mark Stoler⁶, Anna R. Giuliano² and Luisa L. Villa^{3,7}

Affiliations:

¹ Centro de Referência e Treinamento DST/Aids, São Paulo, Brazil.

² Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.

³ Molecular Biology Laboratory, Center for Translational Investigation in Oncology, Instituto do Câncer do Estado de São Paulo, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

⁴ Ludwig Institute for Cancer Research, São Paulo branch, São Paulo, Brazil.

⁵ Department of Biochemistry, Chemistry Institute, Universidade de São Paulo, São Paulo, Brazil.

⁶ University of Virginia Health System, Charlottesville, VA, USA.

⁷ Department of Radiology and Oncology, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil.

Correspondence to:

Luisa Lina Villa, PhD

Laboratory of Innovation in Cancer

Center for Translational Investigation in Oncology

Instituto do Câncer do Estado de São Paulo, Hospital das Clínicas

Faculdade de Medicina, Universidade de São Paulo, Brazil.

Rua Dr. Ovídio Pires de Campos, 872 - Cerqueira César

São Paulo (SP) - CEP: 05403-911 - Brazil

Phone: +5511-26617554

E-mail: l.villa@hc.fm.usp.br

Word count: Abstract: 251; Text:

Running Title: External genital lesions in Brazilian men.

Footnote Page

Funding Statement: This work was supported by the Ludwig Institute for Cancer Research; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [Grant number 08/57889-1 to LLV]; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [Grant number 573799/2008-3 to LLV]. The infrastructure of the HIM Study cohort was supported through a grant from the National Cancer Institute (NCI), National Institutes of Health (NIH) (R01 CA098803 to ARG). Merck & Co. provided financial support for the data analysis to Fundação Faculdade de Medicina da Universidade de São Paulo on behalf of LLV.

Conflict of Interest

ARG (IISP39582) and SLS (IISP53280) are current recipient of grant funding from Merck, and ARG and LLV are members of the Merck Advisory Board for HPV prophylactic vaccines. RJCS received travel grants from Merck, Sharp & Dohme. No conflicts of interest are declared for any of the remaining authors.

Ethical approval

Ethical approval was given by the University of South Florida (IRB# 102660), Centro de Referência e Treinamento em DST/AIDS, São Paulo, Brazil, and Instituto Nacional de Salud Pública de México.

Abstract

Our aim was to determine the incidence of external genital lesions (EGLs), specifically histologically confirmed condyloma (genital warts) and Penile Intraepithelial Neoplasia (PeIN) and genital HPV infection to EGLs among healthy men aged 18-73 residing in Brazil. Subjects included 1,118 men enrolled in the HIM study between July 2005 and June 2009. At each visit, EGLs were biopsied and subjected to pathological evaluation. HPV status in genital swabs and biopsies was determined by Linear Array and INNO-LiPA, respectively. Age-specific EGLs incidence and the proportion and median time to EGL development were estimated. Kaplan-Meier cumulative incidence rates at 6, 12 and 24 months were determined. During follow-up, 73 men developed an incident EGL. Men could develop multiple EGLs and there were 36 men with condyloma, 27 men with suggestive of condyloma, 6 men with PeIN, and 20 men with non-HPV lesions. HPV-positive men who developed EGLs were younger ($p=0.002$) than men that did not develop lesions. Among the 815 men with an HPV infection, 4% progressed to an EGL with the same HPV detected in the lesion. During follow up, 15.7% genital HPV-6 and HPV-11 infections progressed to condyloma (9 months median time progression for HPV-6 versus 6.8 months for HPV-11). Approximately 1% of HPV-16 infections progressed to PeIN with a median time to progression of 25 months. HPV types covered by the 4-valent HPV vaccine were detected in 82.3% and 83.3% of condyloma and PeIN, respectively. The high burden of HPV and high frequency of progression to disease underscores the need to offer HPV prophylactic vaccination to men to reduce the overall burden of infection and diseases caused by HPV.

Keywords: Men, genital lesions, PeIN, Condyloma, HPV.

Introduction

Human papillomavirus (HPV) is the most common sexually transmitted infection worldwide with sexual activity constituting the major risk factor for infection [de Martel et al., 2012]. Most infections are asymptomatic and clear spontaneously within two years. However, persistence of high-risk HPV is a strong predictor of cervical and other HPV-related cancers development. To date, more than 200 HPV genotypes have been identified, 40 of which are routinely detected in the anogenital specimens of both females and males [Bouvard et al., 2009]. Genital HPV infection is very common in men [Giuliano et al., 2008] and infections may progress to external genital lesions (EGLs), mostly condylomata acuminata (genital warts or condyloma) and penile intraepithelial neoplasia (PeIN) that are thought to precede penile cancer (Cubilla et al., 2004).

Condyloma is a common clinical outcome of HPV infection in males, mainly among men aged 25-29 years [Anic et al., 2012]. Nononcogenic HPV-6 and HPV-11 are the etiologic agents of over 90% of condylomas [IARC, 2012]. Although about one third of these warts spontaneously regress, recurrence is common, therefore, treatment results in high medical costs [Steben and Garland, 2014]. The incidence of condyloma among Brazilian men is poorly known. While there are some reports on HPV infection at the male genitals [Nicolau et al., 2005; Rombaldi et al., 2006; Silva et al., 2011], the incidence of HPV infections and condyloma in Brazilian men are limited to few publications [Moreira et al., 2014; Giuliano et al., 2011a; Sudenga et al., 2016a].

Penile cancer occurs more frequently in men aged 50-70 years, with a global estimate of 22,000 cases per year [de Martel et al., 2012]. Though this cancer is rare, incidence rates are higher in less developed countries representing up to 10% of male tumors in some parts of Africa, Asia and South America [Parkin et al., 2008]. In Brazil, this neoplasia accounts for about 2% of all male cancers, where both the North and

Northeast regions are mostly affected [Favorito et al., 2008]. Both benign and malignant HPV-related tumors in men are potentially preventable with the 4-valent HPV vaccine [Giuliano et al., 2011b; Villa et al., 2005].

The purpose of this study was to estimate EGL incidence and assess the rate of progression from HPV infection to histopathologically confirmed EGLs in otherwise healthy men from Brazil participating in an international HPV natural history study, the *HPV Infection in Men (HIM) Study*.

Materials and Methods

Study design

The HIM Study includes over 4,000 men aged 18-70 years enrolled between 2005 and 2009 in São Paulo (Brazil), Cuernavaca (Mexico), and Tampa, Florida (USA). Study methods and design have been described in detail [Giuliano et al., 2008]. In brief, every six months participants underwent interview, a physical exam, and laboratory analysis of anogenital specimens for HPV detection. In 2009 a biopsy and pathology protocol was implemented allowing for pathological confirmation and HPV genotyping of lesions. Men who had two or more study visits after the implementation of this protocol were included in the present study.

In São Paulo, the largest city of Brazil, a total of 1,443 men were recruited at technical and higher education institutions, in communities, companies and military corporations, through leaflets posters, letters and e-mails addressed to participants of other studies. Men were also invited through some radio and internet ads and during counseling sessions on a public STD (sexually transmitted disease) clinic. . Men in treatment for STDs or condyloma, and HIV-positive individuals were excluded from participating in the study. Retention rates in Brazil were the highest among all three sites participating at the HIM study: 80% of all enrolled men completed 10 visits of

follow-up in approximately 5 years, according to the study protocol (6 months interval visits). The study was approved by the ethical review boards of the Institutional Review Boards at the University of South Florida (Tampa, FL, USA), the Ludwig Institute for Cancer Research (São Paulo Branch, Brazil), Centro de Referência e Treinamento em DST/Aids (São Paulo, Brazil), and the Instituto Nacional de Salud Publica (Cuernavaca, Mexico) and informed consent was obtained from all participants.

Clinical specimens, HPV detection and genotyping

At each visit, three different pre-wetted Dacron swabs were used to sample the external genitalia (coronal sulcus, glans penis, penile shaft) and scrotum of participants, and were later combined to form a single sample [Giuliano et al., 2008]. Specimens underwent DNA extraction using the Qiagen Media kit (Qiagen, Venlo, The Netherlands), and HPV detection and typing was conducted using the Linear Array® HPV Genotyping Test (Roche Molecular Diagnosis, Pleasanton, USA), capable of identifying 37 HPV types, classified as high-risk (HR-HPV: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68) or low-risk (LR-HPV: 6, 11, 26, 40, 42, 53, 54, 55, 61, 62, 64, 66, 67, 69, 70, 71, 72, 73, 81, 82, IS39, 83, 84, 89).

Furthermore, at each clinic visit participants underwent clinical examination under 3x light magnification by a trained clinician in order to detect EGLs. Whenever a lesion was observed, a tissue sample was obtained by shave excision, fixed and sent for histopathological evaluation. All EGLs that appeared to be HPV-related or had an unknown etiology based on visual inspection were sent for HPV testing. EGLs were categorized as condyloma, suggestive of condyloma ([very early condylomas that have some but have not yet developed all of the pathological features of condyloma](#)), PeIN, or not HPV-related (such as skin tags, seborrheic keratosis, chronic balanitis, Molluscum contagiosum), according to criteria previously published [Sudenga et al.,

2016a]. PeIN lesions were further categorized in PeIN I (low-grade squamous intraepithelial lesion [SIL]), PeIN II (high-grade SIL), PeIN II/III (high-grade SIL), and PeIN III (high-grade SIL). DNAs from formalin-fixed, paraffin-embedded (FFPE) tissues were extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen, Venlo, The Netherlands), and HPV detection and genotyping was conducted using the INNO-LiPA HPV Genotyping Extra assay (Fujirebio, Göteborg, Sweden) which detects 28 viral types (HR-HPV: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68; or LR-HPV: 6, 11, 26, 40, 43, 44, 53, 54, 66, 69, 70, 71, 73, 74, 82). DNAs from swabs and tissues were considered adequate and included in this study if tested positive for β -globin or any HPV genotype.

Statistical analysis

EGL incidence

All men with prevalent lesions were excluded from the biopsy cohort. Age-specific analyses were conducted among men who developed an incident EGL, and age groups were stratified as 18-30, 31-44, and 45-73 years. For EGL incidence analyses, only the first detected EGL was considered. Time to incident EGL was calculated from biopsy cohort baseline date to the date of first EGL detection. Person-time incidence and 95% confidence intervals (CIs) were calculated based on the number of events modeled as a Poisson variable for the total number of person-months. Kaplan-Meier curves for EGL incidence were generated, and the incidence of EGL over time across the 3 age groups was compared using the log-rank test.

To evaluate HPV distribution in EGLs, all prevalent and incident lesions were included. In addition to specific HPV types, infections positive for ≥ 1 type were included in the any HPV group, those positive for ≥ 1 high-risk HPV type were included in the high-risk HPV group, and those positive for ≥ 1 low-risk types were included in

the low-risk HPV group. Independent analyses were conducted for high-risk and low-risk infections. Additionally, EGLs that were positive for ≥ 1 high-risk type and ≥ 1 low-risk type were included in the both HR/LR HPV group.

HPV progression to EGL

Sociodemographic and sexual behavior characteristics were described for HPV positive men who did or did not develop an EGL during follow-up, and comparisons were performed using the Monte Carlo estimation of exact Pearson chi-square tests.

Men with an incident or prevalent genital HPV infection and without a prevalent condyloma or PeIN lesion at the biopsy protocol baseline visit were included in the analyses designed to assess the rate of progression and the proportion of infections progressing to disease. HPV infection was reported by genotype or grouped (any, HR-HPV, LR-HPV, and vaccine (HPVs 6/11/16/18)). The classification of any HPV type was defined as a positive test result for at least one of 25 (HPV types 43/44/74 are not detected by the Linear Array assay) HPV genotypes detected by INNO-LiPA. HPV infections with single or multiple HR-HPV types were classified as HR and those with at least one LR-HPV type were classified as low risk.

Time-to-event approach was applied to assess the time from type-specific genital HPV positivity to EGL incidence harboring the same HPV type within the lesion. The analytical unit for this study is infection. HPV genital infections that did not progress to EGL were censored at the last visit. The 6-, 12-, and 24-month cumulative incidence of EGLs and median time to EGL development for individual genital HPV types was estimated using the Kaplan-Meier method. For grouped genital HPVs, we adjusted for within-subject correlation using the clustered Kaplan-Meier method as men could have been infected with multiple HPV types within a defined group. The overall EGL

incidence rate during the study period was also calculated. Multiple HPV types could be detected in a single EGL, and a man could develop multiple EGLs.

Results

EGL incidence

Overall 1,134 Brazil HIM Study participants aged 18-73 years were included in the biopsy cohort. Sixteen men with a prevalent EGL were excluded from the EGL incidence analysis (Figure 1). Among the remaining 1,118 men, 73 developed at least one incident EGL during follow up. Men primarily developed condyloma (n=36), suggestive of condyloma (n=27), PeIN (n=6), or non-HPV related EGLn=20).

Table 1 presents EGL incidence stratified by age. There was a decline in the 12-month cumulative incidence for any EGL by age (3.4%, 1.9%, and 0.9% for ages 18-30, 31-44, and 45-73 years, respectively; $p=0.02$; Figure 2). When stratified by lesion type, the 12-month cumulative condyloma incidence was higher among men <30 years of age (1.8% among men aged 18-30, $p=0.06$) compared to older men. In contrast, among men who developed PeIN, the 12-month cumulative incidence remained constant across categories of age (0.6%, 0.0%, and 0.0% for ages 18–30, 31–44, and 45–73 years, respectively; $p=0.38$ across the entire follow-up period)

Any HPV DNA was detected in 87.3% of condyloma, and 83.5% were positive for low-risk viral types HPV-6 and HPV-11 were detected in 48.1% and 36.7% of condylomas, respectively (Figure 3). Less frequently detected low-risk types include HPV 26, 40, 44, 54, 66, 74. Among 48 EGLs categorized as suggestive of condyloma, 72.9% were positive for any HPV, and the most common type detected was HPV-6 (58.3%). Among the six prevalent and incident PeIN lesions, 100% were HPV positive and 66.7% were positive for ≥ 1 HR-HPV genotype. HPV-16 and HPV-18 were detected in 2 (33.3%) and 1 (16.7%) of the PeIN lesions, respectively. Additionally two other

HR-HPVs were also detected in PeIN lesions: HPV-39 was found in 2 samples whereas HPV-68 was detected in one sample (data not shown). Low-risk HPV types were mostly found in PeIN lesions as co-infections with a high-risk HPV, however HPV-6 was detected as a single infection in one PeIN I and one PeIN III lesions. PeIN lesions occurred at the coronal sulcus (two cases), the glans penis (two cases), the meatus and the shaft left dorsal (1 case each) (table 1). At least one of the HPV types in the 4-valent HPV vaccine was detected in 82.3%, 70.8%, and 83.3% of condyloma, suggestive of condyloma, and PeIN lesions, respectively.

HPV progression to EGL

Among HPV-positive men, EGL incidence was associated only with younger age ($p=0.002$) (Table 2). No other sociodemographic or sexual behavior variable was associated with EGL incidence.

The HPV-6 positive condyloma incidence rate among men with a previously detected HPV-6 infection was 5.9 per 1,000 persons-month (Table 3). Among HPV-6 infections, 7.34% progressed to HPV-6-positive condyloma during the first 6 months of follow-up. For HPV-11, the incidence rate for HPV-11-positive condyloma was higher than HPV-6 at 6.7 per 1,000 persons-month and the 6-month cumulative incidence was also higher at 10.07% (Figure 4). The HPV-16 positive PeIN incidence rate was 0.4 per 1,000 pm among men with a prior HPV-16 infection. By 24 months of follow-up 0.79% of men with a genital HPV-16 infection progressed to an HPV-16-positive PeIN.

Discussion

The *HIM study* is the first international comparison of the distribution of 37 HPV genotypes in the genitals of men enrolled in three different countries. Among these individuals, HPV prevalence at the genitals was higher in Brazil (72.3%), than in the

USA (61.3%) and Mexico (61.9%) [Giuliano et al., 2008]. Although HPV infection is a common finding in men, when considering all HIM participants, only 5% of these incident viral infections progressed to an EGL during follow-up [Sudenga et al., 2016a]. We herein present data comparing EGL incidence and progression of HPV infection to EGL among men enrolled in Brazil. During follow-up 73 men from Brazil developed an incident EGL including 36 condyloma, 6 PeIN, and 20 non-HPV related EGLs (e.g. skin tags, benign keratosis). HPV positive men who developed EGLs were younger ($p=0.002$) than men that did not develop lesions. HPV infection was detected in 815 men; 4% of infections progressed to an EGL. During the follow up, 15.7% genital HPV-6 and HPV-11 infections progressed to condyloma (9 months median time progression for HPV-6 versus 6.8 months for HPV-11). Approximately 1% of HPV-16 infections progressed to PeIN with a median time to progression of 25 months. HPV types covered by the 4-valent HPV vaccine were detected in 82.3%, 70.8% and 83.3% of condyloma, suggestive of condyloma and PeIN, respectively.

We previously observed that for all men enrolled in the HIM Study, factors strongly associated with condyloma development included age, high lifetime number of female or male sexual partners [Anic et al., 2012]. In the present analysis restricted to Brazilian men, we observed that condyloma incidence significantly differed only by age ($p=0.02$): younger men (<30) developed more genital warts either when considering the whole population [Sudenga et al., 2016b] or solely HPV positive individuals. Nevertheless, it should not be ignored that individuals aged 45-70 years continued to develop condyloma during follow-up which highlights that men remain susceptible to developing HPV-related EGLs throughout their lives. In the current study HPV DNA was detected in 87.3% of condyloma biopsies, in line with previous results of 91% HPV prevalence in condyloma samples of women enrolled in the placebo arm of a vaccine trial [Garland et al., 2009]. HPV-6 and HPV-11 were the most prevalent viral types

detected in condylomas in agreement with previous reports [Brown et al., 1999; Chan et al., 2009; Forman et al., 2012; Ingles et al., 2015], although HPV-6 detection surpassed that of HPV-11 (48.1% versus 36.7%) in these samples.

About 16% of HPV-6 and HPV-11 infections progressed to a condyloma with the same viral type during follow-up; however median time to progression was higher among HPV-6 positive men (9 mo for HPV-6 versus 6.8 mo for HPV-11). Our results are in line with previous estimates of 2 weeks to 8 months of the time required from initial HPV-6 or HPV-11 infections to progression to condyloma [Steben and Garland, 2014].

Although PeIN was a rare event in this population, we were able to detect and genotype 6 prevalent and incident PeIN lesions. Histopathologically, three cases were classified as PeIN III, one case as PeIN II and two cases as PeIN I. All PeIN lesions were HPV-positive; HPV-16 was detected in two cases, in agreement with previous studies reporting 88%-100% of samples positive for HPV DNA, and about 40% of PeIN positive for HPV-16 [Alemany et al., 2016]. It is interesting to notice that genital swabs collected prior to HPV-16 positive PeIN harbored multiple HPV infections. Among Brazilian men enrolled in the HIM study, 1.1% of HPV-16 infections progressed to HPV-16-positive PeIN and median time to progression was 25.6 months. Among low-risk HPV genital infections, one HPV-6 and one HPV-73 infections progressed to PeIN. However, whereas the HPV-6 infection progressed to a PeIN within 6.7 months, the HPV-73 infection progression occurred after 30.5 months. Similar results are observed when pooling data across the three participating countries of the HIM study, namely Brazil, Mexico and USA: 1.6% of HPV-16 infections progressed to PeIN in a median time of 19 months [Sudenga et al., 2016a]. In addition, in one of the PeIN cases, we detected exclusively HPV-6 and time from infection to progression was 6.7 months. The presence of low-risk HPVs in high-grade precancerous anogenital lesions has been also

observed by others [Aynaud et al., 1994; Krustup et al., 2009; Rubin et al., 2001]. Nevertheless, the limited number of PeIN detected limits further considerations.

Considering the HIM Study cohort as a whole, during all 6 years of follow-up, 10 incident PeIN lesions were detected, 2 from the USA and 4 from each Mexico and Brazil [Sudenga et al., 2016a]. Regional differences concerning the rates of HPV-related tumors in these countries are reported. In Brazil the incidence rate of penile cancer is 0.7-2.3/100,000 inhabitants, one of world's highest incidence rates for this neoplasia [Bruni et al., 2015]. The city of Recife in Pernambuco, Northeast region of the country has an incidence rate of penile cancer of 4.6/100,000 inhabitants. Unfortunately, there is still very incomplete data available in regards to HPV prevalence in penile cancer and its precursor lesions in Brazil. In one study, high-risk HPV was detected in 2 cases of condyloma and 2 cases of PeIN (1 PeIN I and 1 PeIN III), whereas just one case of PeIN I was HPV negative by hybrid-capture [Giraldo et al., 2008].

As with any epidemiological study, there are limitations and strengths to the present study. Although most previous reports of HPV incidence in EGLs were based on detection from the surface of the lesion, this analysis took advantage of HPV detection within biopsy tissues, which is more likely to represent the viral type present in the lesion [Pierce-Campbell et al., 2013]. Further strengths of this study include the clinical data from a large sample size of men followed for genital disease in Brazil with very high retention rates (>80% at 5y of follow-up). Moreover, men were followed-up every 6-months to allow for early detection of lesions, which were available for histopathological evaluation and HPV genotyping. Nevertheless, our study may be underestimating HPV infection progression to condyloma, which could have occurred during the 6 months between study visits.

Taken together, approximately 16% of genital HPV-6 and HPV-11 infections progressed to condylomas with the same HPV type while less than 1% of genital HPV-

16 infections progressed to an HPV-16-positive PeIN among Brazilian men. In addition, men continue to be at risk of EGLs across all age groups. This study emphasizes the need for vaccine-related immunity in males and for that immunity to be long lasting across the lifespan to ensure a reduction in the overall burden of infection and diseases caused by HPV.

The 4-valent vaccine offers protection against HPVs 6, 11, 16 and 18 has been demonstrated to be efficacious in the reduction of HPV-related diseases in men [Giuliano et al., 2011b]. In our series, HPV vaccine types were detected in 82.3% and 83.3% in condyloma and PeIN cases, respectively. To date, few countries globally have national vaccine programs that include male vaccination, specifically the USA, Austria, Israel, Switzerland and Australia. As male HPV infection significantly contributes to infection and subsequent cervical disease in women and disease in men, our data support the implementation in Brazil of gender-neutral national immunization policies, crucial to reduce the overall HPV infection burden and ultimately, disease in both women and men.

Acknowledgements

This research was supported in part by research funding from Merck Sharp & Dohme Corp. The opinions expressed in this paper are those of the authors and do not necessarily represent those of Merck Sharp & Dohme Corp.

The authors would like to thank the HIM Study Teams in the USA (Moffitt Cancer Center, Tampa, FL), Brazil (Centro de Referência e Treinamento em DST/AIDS, Fundação Faculdade de Medicina, Instituto do Câncer do Estado de São Paulo, Ludwig Institute for Cancer Research, São Paulo) and Mexico (Instituto Mexicano del Seguro Social, Instituto Nacional de Salud Pública, Cuernavaca).

References

(To be formatted according to Journal's instructions, after we receive all comments from authors)

- Aleman L, Cubilla A, Halec G, et al. Penile Cancer: Role of Human Papillomavirus in Penile Carcinomas Worldwide. *Eur Urol*. 2016;pii: S0302-2838(15)01215-4.
- Anic GM, Lee JH, Villa LL, et al. Risk factors for incident condyloma in a multinational cohort of men: the HIM study. *J Infect Dis*. 2012;205:789-93.
- Aynaud O, Ionesco M, Barrasso R. Penile intraepithelial neoplasia. Specific clinical features correlate with histologic and virologic findings. *Cancer*. 1994;74(6):1762-7.
- Bouvard V, Baan R, Straif K, et al. A review of human carcinogens-Part B: biological agents. *Lancet Oncol*. 2009;10(4):321-2.
- Brown DR, Schroeder JM, Bryan JT, Stoler MH, Fife KH. Detection of multiple human papillomavirus types in Condylomata acuminata lesions from otherwise healthy and immunosuppressed patients. *J Clin Microbiol*. 1999;37(10):3316-22.
- Bruni L, Barrionuevo-Rosas L, Albero G, et al. ICO Information Centre on HPV and Cancer (HPV Information Centre). Human Papillomavirus and Related Diseases in Brazil. Summary Report 2015;12-23.
- Chan PK, Luk AC, Luk TN, et al. Distribution of human papillomavirus types in anogenital warts of men. *J Clin Virol*. 2009;44:111-4.
- Cubilla AL¹, Velazquez EF, Young RH. Epithelial lesions associated with invasive penile squamous cell carcinoma: a pathologic study of 288 cases. *Int J Surg Pathol*. 2004;12(4):351-64.
- de Martel C, Ferlay J, Franceschi S, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol*. 2012;13:607-15.
- Favorito LA, Nardi AC, Ronalsa M, Zequi SC, Sampaio FJ, Glina S. Epidemiologic study on penile cancer in Brazil. *Int Braz J Urol*. 2008;34: 587-91.
- Forman D, de Martel C, Lacey CJ, et al. Global burden of human papillomavirus and related diseases. *Vaccine*. 2012;30 Suppl 5:F12-23.
- Garland SM, Steben M, Sings HL, et al. Natural history of genital warts: analysis of the placebo arm of 2 randomized phase III trials of a quadrivalent human papillomavirus (types 6, 11, 16, and 18) vaccine. *J Infect Dis*. 2009;199(6):805-14.
- Giraldo PC, Eleutério J Jr, Cavalcante DI, Gonçalves AK, Romão JA, Eleutério RM. The role of high-risk HPV-DNA testing in the male sexual partners of women with HPV-induced lesions. *Eur J Obstet Gynecol Reprod Biol*. 2008;137(1):88-91.
- Giuliano AR, Lazcano-Ponce E, Villa LL, et al. The human papillomavirus infection in men study: human papillomavirus prevalence and type distribution among men residing

in Brazil, Mexico, and the United States. *Cancer Epidemiol Biomarkers Prev.* 2008;17:2036-43.

Giuliano AR, Lee JH, Fulp W, et al. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. *Lancet.* 2011a;377(9769):932-40.

Giuliano AR, Palefsky JM, Goldstone S, et al. Efficacy of quadrivalent HPV vaccine against HPV Infection and disease in males. *N Engl J Med.* 2011b;364:401-11.

IARC Monographs. A review of human carcinogens. Part B: Biological agents / IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, (2012: Lyon, France).

Ingles DJ, Pierce Campbell CM, Messina JA, et al. HPV Genotype- and Age-specific Analyses of External Genital Lesions Among Men in the HPV Infection in Men (HIM) Study. *J Infect Dis.* 2015;211(7):1060-7.

Krustrup D, Jensen HL, van den Brule AJ, et al. Histological characteristics of human papilloma-virus-positive and -negative invasive and in situ squamous cell tumours of the penis. *Int J Exp Pathol.* 2009;90(2):182-9.

Moreira ED Jr, Giuliano AR, Palefsky J, et al. Incidence, clearance, and disease progression of genital human papillomavirus infection in heterosexual men. *J Infect Dis.* 2014;210(2):192-9.

Nicolau SM, Camargo CG, Stávale JN, et al. Human papillomavirus DNA detection in male sexual partners of women with genital human papillomavirus infection. *Urology.* 2005;65(2):251-5.

Parkin DM, Almonte M, Bruni L, Clifford G, Curado MP, Pineros M. Burden and trends of type-specific human papillomavirus infections and related diseases in the Latin America and Caribbean region. *Vaccine.* 2008;26 Suppl 11:L1-15.

Pierce Campbell CM, Messina JL, Stoler MH, et al. Cutaneous human papillomavirus types detected on the surface of male external genital lesions: a case series within the HPV Infection in Men Study. *J Clin Virol.* 2013;58(4):652-9.

Rombaldi RL, Serafini EP, Villa LL, et al. Infection with human papillomaviruses of sexual partners of women having cervical intraepithelial neoplasia. *Braz J Med Biol Res.* 2006;39(2):177-87.

Rubin MA, Kleter B, Zhou M, et al. Detection and typing of human papillomavirus DNA in penile carcinoma: evidence for multiple independent pathways of penile carcinogenesis. *Am J Pathol.* 2001;159:1211-8.

- Silva RJ, Casseb J, Andreoli MA, Villa LL. Persistence and clearance of HPV from the penis of men infected and non-infected with HIV. *J Med Virol*. 2011;83(1):127-31.
- Steben M, Garland SM. Genital warts. *Best Pract Res Clin Obstet Gynaecol*. 2014;28(7):1063-73.
- Sudenga SL, Ingles DJ, Pierce Campbell CM, et al. Genital Human Papillomavirus Infection Progression to External Genital Lesions: The HIM Study. *Eur Urol*. 2016a;69(1):166-73.
- Sudenga S, N Torres B, Fulp W, Silva R, Villa L, Lazcano Ponce E, Ingles D, Messina J, Abrahamsen M, Baggio ML, Salmeron J, Quiterio M, Giuliano AR. Country specific HPV-related genital lesions, among men residing in Brazil, Mexico and the United States: Him study. (abstract SS 12-8). In: Program and abstracts: Eurogin 2016 (Salzburg), 2016b.
- Villa LL, Costa RL, Petta CA, et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol*. 2005;6(5):271-8.

Figure 1. Men included in the HIM Study Brazil Biopsy Cohort

Figure 2. Kaplan–Meier curve showing differences in cumulative incidence of external genital lesions (EGLs) over time by age group. P values were determined using the log-rank test and denote differences across the entire follow-up period. Values <0.05 are considered statistically significant.

Figure 3. Prevalence of mucosal Human Papillomavirus (HPV) DNA within the tissue of prevalent and incident external genital lesions (EGLs) among 89 men from Brazil

Figure 4. Kaplan–Meier curves showing progression from HPV to condyloma. P values were determined using the log-rank test and denote differences across the entire follow-up period. Values <0.05 are considered statistically significant.

Table 1. Age-specific incidence of pathologically confirmed condyloma among men from Brazil

	Any ^a	Condyloma	Pathological Diagnosis Suggestive ^b	PeIN ^c	Other ^d
All ages (n=1118) ^e					
Men with incident EGL, no.	58	36	27	6	20
Person-months	41109	41760	41743	42407	42079
Incidence rate ^f (95% CI)	1.69 (1.31-2.19)	1.03 (0.75-1.43)	0.78 (0.53-1.13)	0.17 (0.08-0.38)	0.57 (0.37-0.88)
12-month Incidence	2.2 (1.4-3.3)	1.1 (0.6-2)	1 (0.6-1.9)	0.2 (0-0.7)	0.7 (0.3-1.4)
18-30 y (n=345)					
Men with incident EGL, no.	25	17	11	3	12
Person-months	12504	12794	12822	13092	12895
Incidence rate ^f (95% CI)	2.4 (1.62-3.55)	1.59 (0.99-2.56)	1.03 (0.57-1.86)	0.28 (0.09-0.85)	1.12 (0.63-1.97)
12-month Incidence	3.4 (1.9-6.1)	1.8 (0.8-4.1)	1.2 (0.5-3.2)	0.6 (0.2-2.4)	1.2 (0.5-3.2)
31-44 y (n=550)					
Men with incident EGL, no.	29	16	15	3	6
Person-months	20322	20646	20611	20968	20851
Incidence rate (95% CI)	1.71 (1.19-2.46)	0.93 (0.57-1.52)	0.87 (0.53-1.45)	0.17 (0.06-0.53)	0.35 (0.16-0.77)
12-month Incidence	1.9 (1-3.5)	0.9 (0.4-2.3)	1.1 (0.5-2.5)	0.0 (0.0-0.0)	0.6 (0.2-1.8)
45-73 y (n=223)					
Men with incident EGL, no.	4	3	1	0	2
Person-months	8283	8320	8310	8347	8333
Incidence rate (95% CI)	0.58 (0.22-1.54)	0.43 (0.14-1.34)	0.14 (0.02-1.03)	0.0 (0.0-0.0)	0.29(0.07-1.15)
12-month Incidence	0.9 (0.2-3.8)	0.5 (0.1-3.3)	0.5 (0.1-3.3)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
P-value^e	0.02	0.06	0.10	0.38	0.02

^aMen with ≥ 1 incident, pathologically confirmed HPV-related EGL throughout the study period. For men with >1 EGL, incidence rates for the Any EGL category are determined; for the first detected lesion; thus, men may contribute fewer person-months in this category than for specific pathological diagnoses; ^bIncludes lesions suggestive but not diagnostic of HPV infection or condyloma; ^cData concerning PeIN specimens were previously described (Sudenga et al., 2016b). ^dIncludes various HPV-unrelated benign skin conditions, such as seborrheic keratosis and skin tags; ^eAlthough the initial cohort included 1134 men, 16 men with prevalent EGLs were excluded in this analysis.; ^fSpecified as the number of cases per 100 person-years; ^fDetermined using the log-rank test and correspond to overall differences in EGL incidence across the entire follow-up period, by age group. Values < 0.05 are considered statistically significant. CI, confidence interval; PeIN, penile intraepithelial neoplasia (I–III).

Table 2. Comparison of sociodemographic and sexual behavior characteristics between human papillomavirus-positive men who did and did not develop an external genital lesion (EGL) during follow-up in the HIM study in Brazil

		No EGL Incidence	Any EGL Incidence	P*
		N (%)	N (%)	
Age				0.002
	18-30	240 (30.8%)	19 (54.3%)	
	31-44	392 (50.3%)	15 (42.9%)	
	45-74	148 (19%)	1 (2.9%)	
Race				0.90
	White	469 (60.1%)	23 (65.7%)	
	Black	230 (29.5%)	10 (28.6%)	
	Asian/PI	14 (1.8%)	0 (0%)	
	Other	58 (7.4%)	2 (5.7%)	
	Refused	9 (1.2%)	0 (0%)	
Ethnicity				0.68
	Hispanic	193 (24.7%)	8 (22.9%)	
	Non-Hispanic	570 (73.1%)	27 (77.1%)	
	Refused	17 (2.2%)	0 (0%)	
Years of Education				0.82
	Completed 12 Years or Less	379 (48.6%)	15 (42.9%)	
	13-15 Years	150 (19.2%)	8 (22.9%)	
	Completed at Least 16 Years	247 (31.7%)	12 (34.3%)	
	Refused	4 (0.5%)	0 (0%)	
	Missing	0 (0%)	0 (0%)	
Marital Status				0.23
	Single	302 (38.7%)	19 (54.3%)	
	Married/Cohabiting	381 (48.8%)	12 (34.3%)	
	Divorced/Separated/Widowed	95 (12.2%)	4 (11.4%)	
	Refused	2 (0.3%)	0 (0%)	
	Missing			
Circumcised				0.82
	Not Circumcised	645 (82.7%)	28 (80%)	
	Circumcised	135 (17.3%)	7 (20%)	
Vaginal Condom Use				0.21
	No Sex	105 (13.5%)	1 (2.9%)	
	Always	140 (17.9%)	6 (17.1%)	
	Sometimes	297 (38.1%)	18 (51.4%)	
	Never	207 (26.5%)	9 (25.7%)	
	Missing	31 (4%)	1 (2.9%)	
Anal Condom Use				0.12
	No Sex	397 (50.9%)	13 (37.1%)	
	Always	130 (16.7%)	10 (28.6%)	
	Sometimes	121 (15.5%)	8 (22.9%)	
	Never	126 (16.2%)	4 (11.4%)	
	Missing	6 (0.8%)	0 (0%)	
Smoking Status				0.48
	Current	161 (20.6%)	10 (28.6%)	
	Former	239 (30.6%)	9 (25.7%)	

	Never	380 (48.7%)	16 (45.7%)	
	Missing	0 (0%)	0 (0%)	
Alcohol per Month				1.00
	0	155 (19.9%)	7 (20%)	
	1-30	314 (40.3%)	14 (40%)	
	>30	285 (36.5%)	13 (37.1%)	
	Missing	26 (3.3%)	1 (2.9%)	
Sexual Orientation ^a				0.61
	MSW	41 (5.3%)	1 (2.9%)	
	MSM	210 (26.9%)	10 (28.6%)	
	MSMW	501 (64.2%)	24 (68.6%)	
	Missing	28 (3.6%)	0 (0%)	
Total Number of Female Partners				0.38
	0-1	74 (9.5%)	4 (11.4%)	
	2-9	167 (21.4%)	4 (11.4%)	
	10-49	380 (48.7%)	21 (60%)	
	50+	131 (16.8%)	6 (17.1%)	
	Refused	28 (3.6%)	0 (0%)	
Total Number of Male Partners				1.00
	0	514 (65.9%)	24 (68.6%)	
	1-9	164 (21%)	7 (20%)	
	10+	87 (11.2%)	4 (11.4%)	
	Missing	15 (1.9%)	0 (0%)	

^aMSW: Men who only have sex with women; MSM: Men who ever have sex with men; MSMW: Men who have sex both with men and women

Table 3. Incidence of condyloma^a and penile intraepithelial neoplasia (PeIN)^a by HPV type detected in the lesion^b among men with the same HPV type detected at the genitals^c among men from Brazil

HPV Type ^d	Incidence Rate ^g (95% CI)	Cumulative Incidence (%)		
		6m (95% CI)	12m (95% CI)	24m (95% CI)
Condyloma ^{a, e}				
Any	0.5 (0.4-0.7)	0.70 (0.42-1.15)	1.05 (0.69-1.59)	1.74 (1.23-2.44)
High Risk	0.1 (0.1-0.3)	0.08 (0.01-0.57)	0.17 (0.04-0.67)	0.39 (0.14-1.05)
16	0.4 (0.1-1.5)	0.0 (0.0-0.0)	0.64 (0.09-4.43)	1.36 (0.34-5.35)
18	0.4 (0.1-2.7)	1.14 (0.16-7.79)	1.14 (0.16-7.79)	1.14 (0.16-7.79)
45	0.4 (0.1-3.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	2.13 (0.30-14.16)
52	0.2 (0.0-1.3)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Low Risk	1.0 (0.7-1.5)	1.48(0.88-2.49)	2.17 (1.41-3.35)	3.44 (2.39-4.93)
6	5.9 (3.7-9.4)	7.34 (3.73-14.15)	10.27 (5.81-17.79)	17.83 (11.35-27.39)
11	6.7 (3.3-13.3)	10.07 (4.31-22.53)	12.31 (5.72-25.43)	17.59 (9.12-32.4)
40	0.6 (0.1-4.0)	1.69 (0.24-11.43)	1.69 (0.24-11.43)	1.69 (0.24-11.43)
53	0.2 (0-1.3.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.77 (0.11-5.33)
66	0.4 (0.1-1.5)	0.0 (0.0-0.0)	1.3 (0.33-5.09)	1.3 (0.33-5.09)
Vaccine ^h	2.4 (1.6-3.4)	3.37 (2.01-5.63)	4.67 (3-7.24)	7.61 (5.29-10.90)
Penile Intraepithelial Neoplasia ^{a, f}				
Any	0.1 (0.0-0.2)	0.05 (0.01-0.34)	0.05 (0.01-0.34)	0.11(0.03-0.44)
High Risk	0.1 (0.0-0.2)	0 (0-0)	0 (0-0)	0.11 (0.02-0.78)
16	0.4 (0.1-1.5)	0 (0-0)	0 (0-0)	0.79 (0.11-5.5)
Low Risk	0.1 (0.0-0.3)	0.11 (0.02-0.77)	0.11 (0.02-0.77)	0.11 (0.02-0.77)
6	0.3 (0.0-2.0)	0.92 (0.13-6.33)	0.92 (0.13-6.33)	0.92 (0.13-6.33)
73	0.4 (0.1-3.0)	0 (0-0)	0 (0-0)	0 (0-0)
Vaccine ^h	0.2 (0.1-0.7)	0.25 (0.03-1.74)	0.25 (0.03-1.74)	0.57 (0.14-2.27)

CI = confidence interval;

^aNewly acquired, pathologically confirmed condyloma and PeIN. ^bDNA detected using INNO LiPA. ^cDNA detected using Linear Array. ^dPrevalent and incident genital HPV infections. ^eHPV types 33/35/39/51/56/58/59/68/26/54/69/70/71/73/82 did not progress to a condyloma lesion; therefore, incidence rates and cumulative incidence could not be calculated. ^fHPV types 18/31/33/35/39/45/51/52/56/58/59/68/11/26/40/53/54/66/69/70/71/82 did not progress to a PeIN; therefore, incidence rates and cumulative incidence could not be calculated. ^gIncidence rate is cases per 1000 person-months. ^hVaccine HPV types 6/11/16/18.