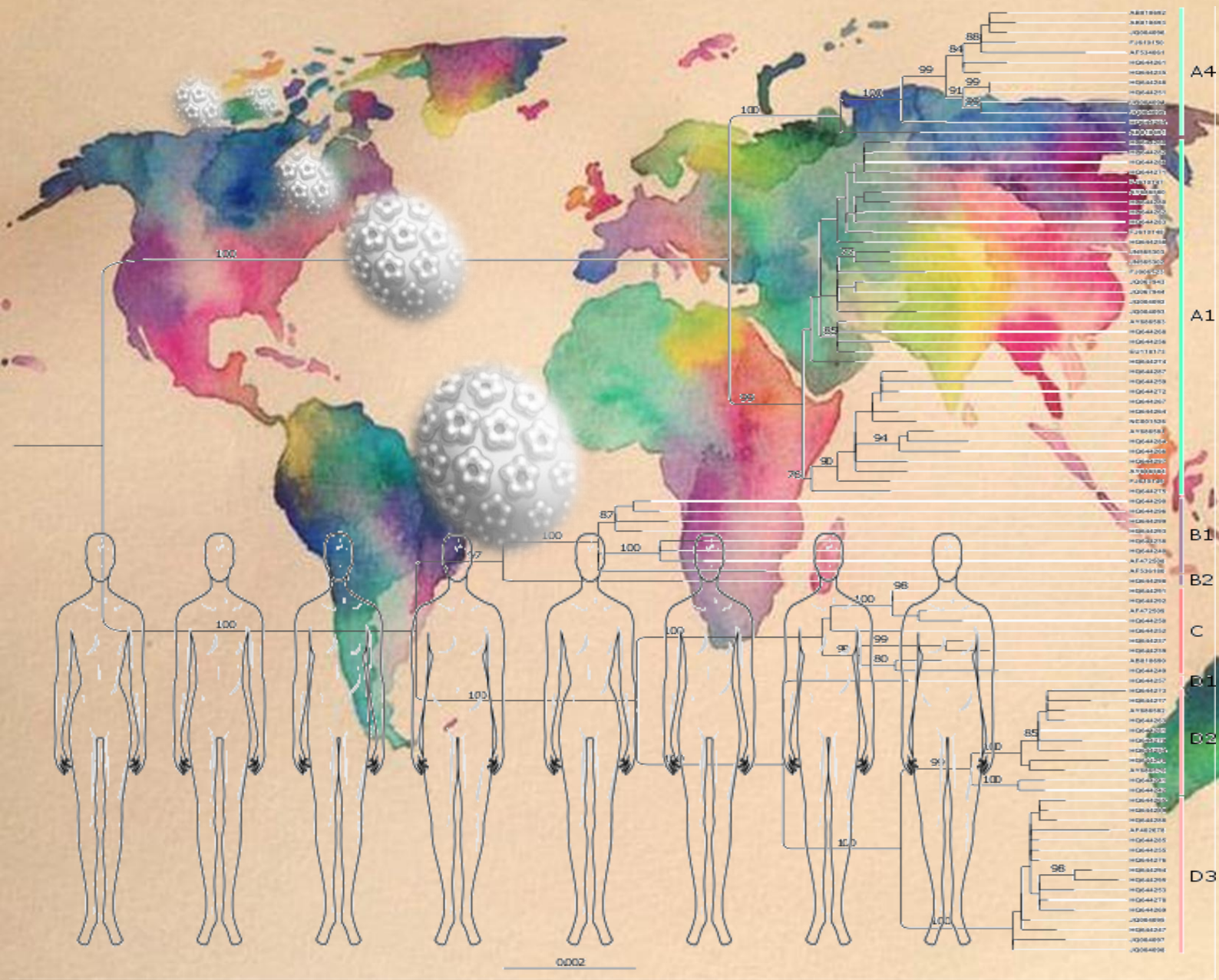
Genetic diversity of human papillomavirus infections in anogenital warts and cancers

Sara Nicolàs Pàrraga

Genetic diversity of human papillomavirus infections in anogenital warts and cancers



Sara Nicolàs Pàrraga

Universitat de Barcelona

Facultat de Medicina

Programa de doctorat en Biomedicina

Genetic diversity of human papillomavirus infections in anogenital warts and cancers

Memòria presentada per Sara Nicolàs Pàrraga per optar al grau de doctora per la Universitat de Barcelona.

Vist i plau del director	Vist i plau del co-director	Vist i plau del tutor	La interessada,
Ignacio González Bravo	Francesc Xavier Bosch i José	Víctor Moreno Aguado	Sara Nicolàs Pàrraga

Barcelona 2018

The work described in this thesis was done at Unitat d'Infecció i Càncer (UNIC), Institut d'investigació biomèdica de Bellvitge (IDIBELL)-Institut Català d'Oncologia (ICO), Barcelona, Catalonia (head dr. Francesc Xavier Bosch)

Source of funding: This work was financially supported by the *Fundación Dexeus* for Women's Health grant, the European Society of Clinical Microbiology and Infectious Diseases (ESCMID), the Agència de Gestió d'Ajuts Universitaris i d' Investigació, AGAUR, Generalitat de Catalunya (2014SGR1077), and the Fondo de Investigaciones Sanitarias (FI12/00142 and FIS Grant PI11/02096), grants from the Spanish Ministry for Science and Innovation (MICINN Grant CGL2010-16713); Red Temática de Investigación Cooperativa en Cancer (RTICC Grant RD06/0020/0095); the Lilly Foundation (Premio de Investigación Biomédica Preclínica 2012 to FXB); and the Institut d'Investigacio Biomédica de Bellvitge – IDIBELL. None of the aforementioned agencies had any role in the interpretation of the results or in the preparation of this manuscript.

Acknowledgments

First I want to thank my thesis directors, Dr. I. G. Bravo and Dr. F. Xavier Bosch. It has been a pleasure to do my PhD in your group. I appreciate all the help, advices and ideas you have given to me and the time spent following the work during my research. I am grateful to Dra. Silvia de Sanjosé and Dra. Laia Alemany for encouraging me at personal and professional scale at Institut Català d'Oncologia (ICO). I thank the members of Program on Research and Epidemiology of Cancer (PREC) group for the multifariously professional contributions. I would also like to extend my thanks to Dra. Montse Torres for your, enthusiasm and good mood and mainly for the laughs inside and outside science. I really appreciate it. To Dra. Marta Félez Sanchez-Ocaña, my PhD mate, for the good times spent. I am grateful also to the master students, postdocs, technicians and administrative stuff from Unitat d'Infeccions and cancer laboratory (UNIC) and from PREC that have worked with me and have given me always a hand.

I want to thank Dra. Paz Cañadas group, an excellent example of a successful team. Thank you for introducing me to the HPV field, for giving me the possibility to continue researching while learning more everyday of molecular biology and its clinical interpretation.

I want to thank other PhD students for the support given to me, especially during the tough times in the PhD pursuit.

I would like to thank my friends; thank you girls! Thank you for your advices, for your positive energy, and for the laughs all together.

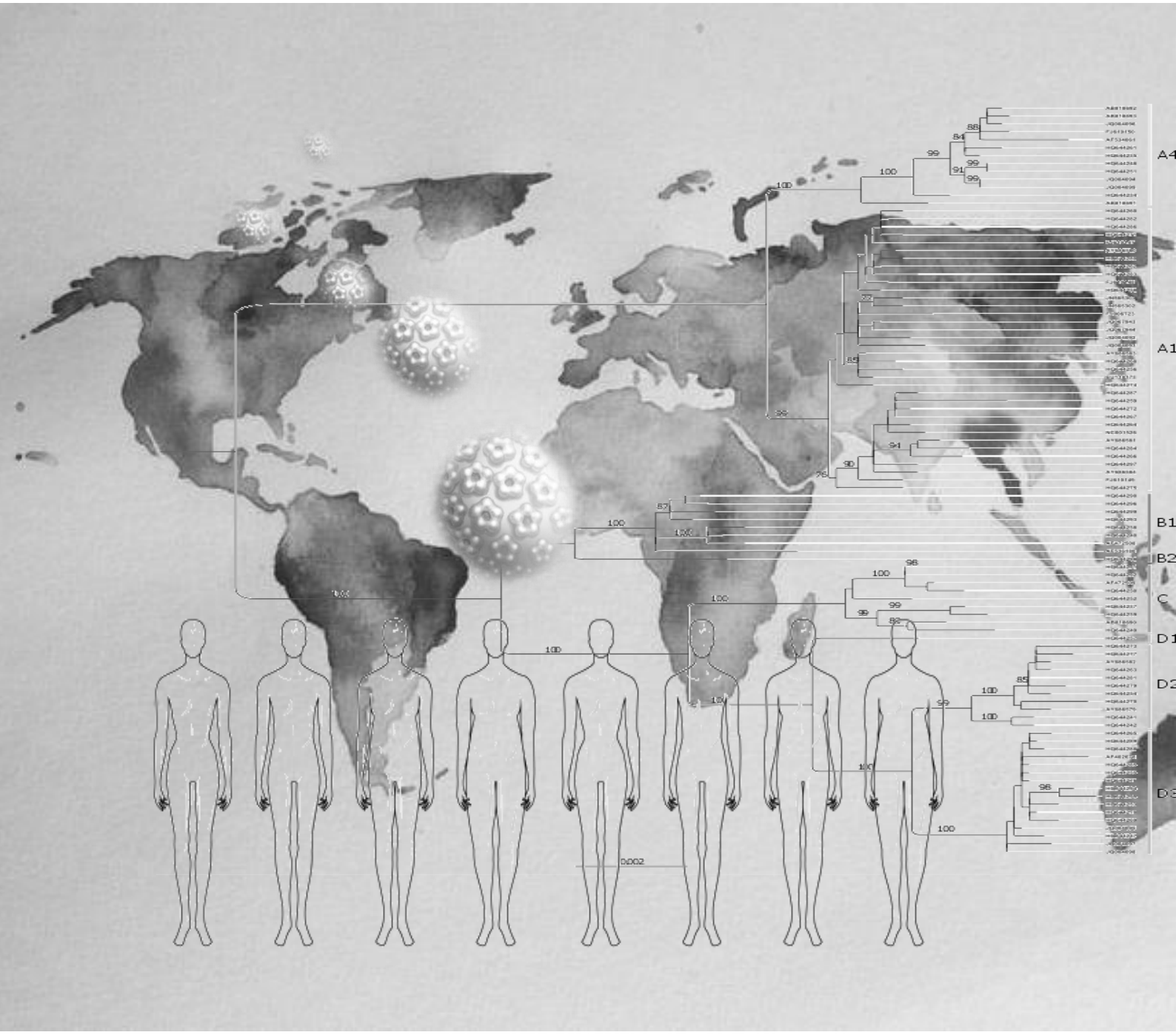
I would like to thank my flat mates; Thank you for those evenings and endless talks sat in the sofa.

I am especially grateful to German; for your encouragement, help, sensitivity and understanding. Thank you for making me realize it was worth to keep on striving until the end. Thank you for being there!

Finally, I would like to thank my family; my mum, my dad and my brother. Thank you for trusting me, and for your unconditional support and for always making me smile.

Sara

EXECUTIVE SUMMARY



Low risk Human Papillomaviruses (LR-HPVs) 6 and 11 are the main causative agents of benign proliferative lesions such genital warts (GWs) and recurrent respiratory papillomatosis (RRP) (Aubin et al. 2008; Ball et al. 2011; Garland et al. 2009; Prétet et al. 2008). High Risk HPV (HR-HPV) 16 is the most oncogenic type and is responsible for invasive cancers (IC) of cervix (ICC), vulva (IVuC), vagina (IVaC), penis (IPeC) and anus (IAnC) (Alemany et al. 2014, 2015, 2016; Larsson et al. 2013; de Sanjose et al. 2010). The well-established connection between HPV16 infection and IC is observed for the most prevalent ICC histological presentations namely squamous cell carcinomas (SCC), adenocarcinomas (ADC) and adenosquamous cell carcinomas (ADSC) (de Sanjose et al. 2010). All these three HPVs (HPV6, HPV11 and HPV16) are the most prevalent in their associated pathological outcomes (Alemany et al. 2014, 2015, 2016; Aubin et al. 2008; Ball et al. 2011; Garland et al. 2009; Larsson et al. 2013; Prétet et al. 2008; de Sanjose et al. 2010). At level of HPV variants, previous studies have addressed differential HPV6 and HPV11 lineage distributions in GWs and RRP, but mainly at a national or regional level (Kocjan *et al.*, 2009). Although without large sample size, some studies describe differential prevalence of HPV variants among distinct pathologies (Danielewski et al. 2013; Jelen et al. 2014). For HPV6, it has been observed a higher presence of HPV6_B1 variants in GWs compared to RRPs (Flores-Díaz et al. 2017b). HPV16 variant distribution has been mainly focused on the uterus cervix and less on other anatomical sites (Cornet et al. 2012; Yamada et al. 1997), what emphasizes the still wanting research of HPV16 viral lineages in other anogenital cancer location (non-cervical cancers). Some studies show that HPV16 variants in anogenital cancers are largely the same regardless of cancer anatomical locations (ICC, IVuC, IVaC, IAnC and IPeC) , showing increased prevalence of HPV16 A1-3 for all IC of squamous nature (de Koning, Quint, and Pirog 2008; Larsson et al. 2013; Ouhoummane et al. 2013; Tornesello et al. 2008; Zuna et al. 2011). HPV16 variants have been widely studied in SCC (Cornet et al. 2012; Zuna et al. 2011), as this histological type remains the most prevalent ICC (Vinh-Hung et al. 2007; Vizcaino et al. 2000). Nonetheless, data available is poorer for other cervical cancer histological presentations such ADC and ADSC. Although few studies had addressed other glandular histologies, they had been restricted to local geographic origin or small populations (Burk et al. 2003; Lizano 2006; Qmichou et al. 2013; Tornesello et al. 2011). Previously described background, shows an increased prevalence of HPV16_A1-3 variants in SCCs (Zuna et al. 2011) while an enhanced presence of HPV16_D in ADCs (Burk et al. 2003; Mirabello et al. 2016; Quint et al. 2010). Regarding HPV16 intratype variability, it has grown the interest of researchers on the T350G *E6* polymorphism as data shows an increased oncogenic potential of those variants containing the 350G allele (Grodzki et al. 2006; Jackson et al.

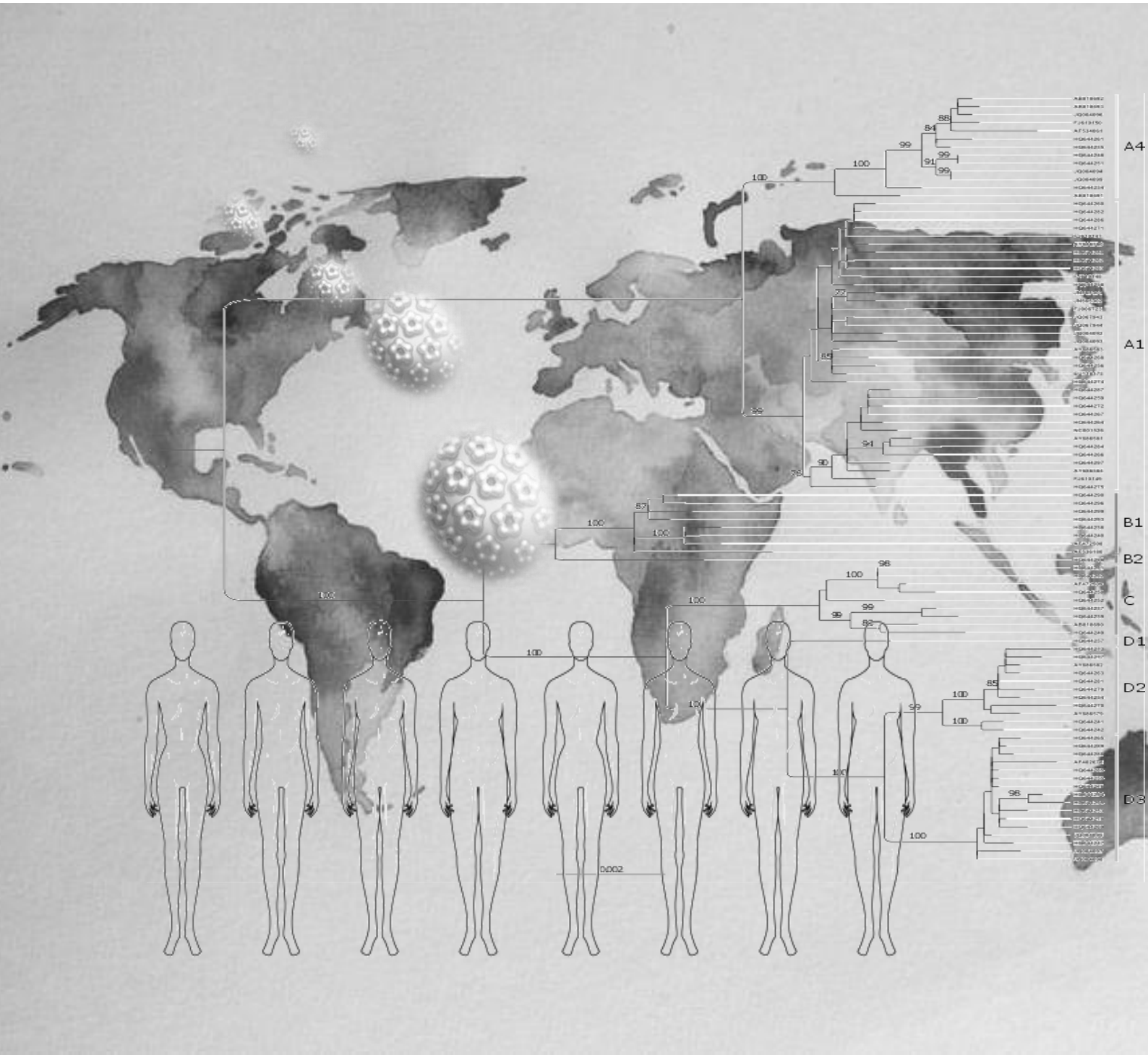
2016; Zehbe et al. 1998, 2001). Finally, most of the data published do not show results regarding age at tumour diagnosis or use different age definitions such age at enrollment (Mirabello et al. 2016).

According to the HPV variant context, the works presented in this thesis try to provide all together information about LR-HPV6 and 11 and HR-HPV16 variant diversity among different lesions, cancers and among different geographical regions with a worldwide spectrum, analyzing large number of HPV monoinfected samples, strategy that prevents possible added bias introduced by co-infections or multi-infections. Through these manuscripts we present for the first time the relative contributions of variant differential abundance, geographical origin, cancer anatomical locations and IC histological presentations to the observation of differential prevalence distribution of HPV16 variants. Through our cases data, we show concordant results with previous published works, observing an increased prevalence of HPV6 B1 variants in GWs, of HPV16 A1-3 variants in anogenital cancers of squamous nature and of HPV16 D variants in ADC. We further show determinate geographical structure of HPV16 variants largely based on the dominance of HPV16 A1-3 variants in Europe, the virtually exclusive presence of HPV16 B and C variants in Africa, the increased prevalence of HPV16 A4 variants in Asia and the enrichment of HPV16 D variants in the Americas. For the most oncogenic-related polymorphism, the E6-T350G, we further show different allelic frequencies according to geographical location independently of the anogenital cancer analyzed, revealing an enhanced 350G allele frequency in isolates from Central-South America compared with Europe. Additionally, we confirme the worldwide trend of cervical cancers to be diagnosed significantly earlier than other anogenital cancers and at histological level, we further present that ADC are diagnosed earlier (mid-forties) than SCC (mid-fifties).

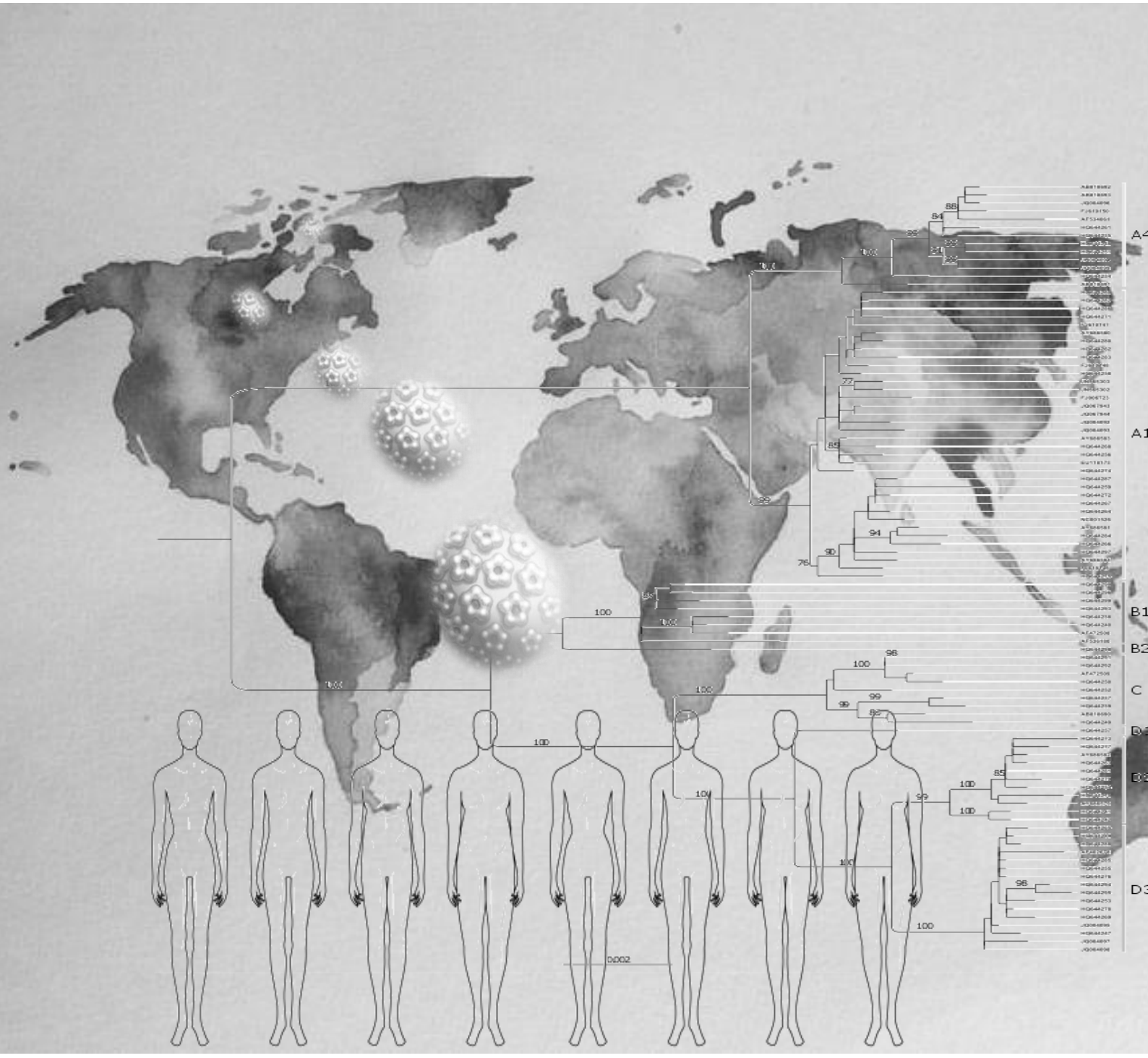
According to the previous context, our results complement and may expand those communicated. The current data suggests that the outcome of the virus-host interaction depends on the combination of phylogeny (i.e. the individual genetic background of both virus and patient) and ontogeny (i.e. the differential susceptibility of different tissues) and for HPV16, provide integrating knowledge of variant-specific differential risk that may impact the future screening algorithms, helping to ensure proper early detection of, for example, elusive ADCs. Furthermore, our data emphasizes the necessity of developing deep analyses in HPV variant field. Additionally, the monitoring of initial steps in viral colonization of anogenital mucosas and the follow-up of differential viral persistences according to patient genetic background, may be of remarkable importance. All these research might ultimately provide answers about the extent of the differential fitness of

HPV viral lineages and will help to understand the virus-host interplay for the most oncogenic HPVs. Additionally, in vaccinated women follow-up, tracing HPV variants may need to keep more attention as, although not described any evidence, they may be indeed, a possible causative agent of not expected pathological outsiders.

TABLE OF CONTENT



1. INTRODUCTION



1.1 Background

Papillomaviruses (PVs) infect the epithelia of vertebrates causing different types of benign lesions, neoplasias or asymptomatic infections (Doorbar et al. 2012, 2015; zur Hausen 2009). The *Papillomaviridae* family contains more than 300 viruses (https://pave.niaid.nih.gov/) from which more than 200 infect humans (Human Papillomaviruses, HPVs). HPVs have been classified by the International Agency for Research in Cancer (IARC) in distinct risk groups (1, 2A, 2B and 3) depending on the association between infection and the development of cancer, specially invasive cervical cancer (ICC). HPVs classified in group 1 are carcinogenic for humans and are classically named "high-risk HPVs" (HR-HPVs). Within this group, HPV16 is the main representative as it is the etiological factor of over 50% of ICCs worldwide (de Sanjose et al. 2010). HPVs classified in groups 2A and 2B are respectively classified as probably and possibly carcinogenic for humans. HPVs classified in group 3 are not classifiable as to their carcinogenicity to humans, and some of them cause benign hyperproliferative lesions such genital warts (GWs) or Recurrent Respiratory Papillomatosis (RRP) (Donne et al. 2010). These later viruses are historically named "low risk HPVs" (LR-HPVs). Among others, the main representatives are HPV6 and HPV11 (Wiatrak et al. 2004) as they are the most prevalent in the mentioned pathologies (Garland et al. 2009; Gissmann et al. 1983) with a prevalence around 1% of the sexually active population (Brentjens et al. 2002).

Regarding taxonomy, based on *L1* gene nucleotide identity, PVs are classified in different phylogenetic levels covering from "*Genera*", with 60% of nucleotide similarity, to "*Variant*"and "Subvariant" sharing between 90 to 98% and 98.5 to 99% of *L1* nucleotide identity, respectively (De Villiers et al. 2004). Variants of either LR-HPVs and HR-HPVs have been well-characterized (Burk et al. 2011; Burk, Harari, and Chen 2013) and several studies have focused on HPV6, HPV11 and HPV16 variants (Flores-Díaz et al. 2017a, 2017b; Larsson et al. 2012, 2013; Tornesello et al. 2008, 2011). At within-variant level, polymorphic sites have grown on interest (Heinzel et al. 1995; Jelen et al. 2014; Khouadri et al. 2006; Swan et al. 2006) and specially certain Single Nucleotide Polymorphisms (SNPs) located in HR-HPVs oncogenes, some of them described to be associated with an increased variant oncogenic potential (Cornet, Gheit, and Iannacone 2013; Grodzki et al. 2006) One of the best described SNPs has been the T350G (L83V) polymorphism located in the HPV16-*E6* oncogene (Cornet et al. 2013; Grodzki et al. 2006).

Current "hot-topics" in HPV variant literature address differential variant prevalence between geographical regions and their association with the anatomical location of the lesion, the nature of the infected tissue or the pathological outcome (i.e. GWs, RRPs or

anogenital cancers) (Cornet et al. 2013; Heinzel et al. 1995; Kocjan et al. 2011; Kocjan, Seme, and Poljak 2008; Larsson et al. 2012, 2013; Tornesello et al. 2008; Xi et al. 1998). Research on HPV16 variants and their differential association with distinct anogenital cancers and their histologies, is a subject of clinical and possibly of public health importance because of the standardisation and growing coverage of screening procedures and vaccination.

The present thesis and the manuscripts included therein pretend to characterise the HPV6, HPV11 and HPV16 variant diversity and analyse their prevalence in a comprehensive set of HPV-related pathologies: GWs, RRP and invasive cancers of the cervix (ICC), vulva (IVuC), vagina (IVaC), penis (IPeC) and anus (IAN), encompassing different countries. Furthermore, at level of cervical invasive cancer histology, this work aims at describing the HPV16 genomic diversity in squamous cell carcinoma (SCC), adenosquamous carcinoma (ADSC) and adenocarcinoma (ADC). For HPV16, the work additionally evaluates the allele frequencies for the long-studied *E6*-T350G polymorphism across anogenital cancers and geographical origin. Finally, this research explores variation in age at tumour diagnosis depending on anogenital cancer locations and on ICC histological presentation. This last assessment evaluates whether cervical cancers, and specially the glandular histological presentation (*i.e.* ADC and ADSC) may be diagnosed earlier than non-cervical cancers and squamous cell carcinoma respectively.

With this commitment, we resorted to the repository of anatomopathology specimens from different invasive cancers and from different benign lesions, already studied at the Institut Català d'Oncologia (ICO) to assess the contribution of HPVs infection to disease burden (Alemany et al. 2014, 2015, 2016; de Sanjosé et al. 2013). We identified and used specimens containing genetic material on a single HPV, *i.e.* monoinfections. We identified the most suitable genomic regions for assessing the intratype genetic diversity and explored the viral diversity using a maximum-likelihood based phylogenetic algorithm (Berger and Stamatakis 2011; Stamatakis 2006). We finally assessed the contribution of different demographic and clinical variables to the differential prevalence of HPV6, HPV11 and HPV16 variants through generalised linear models.

The results from HPV6 and 11 demonstrated distinct variant distribution for HPV6 in GWs and RRPs (Manuscript 1). For HPV16 we showed differential variant prevalence as a function of the geographical origin and of anatomical location for a number of anogenital cancers. We additionally explored E6-T350G polymorphic site, potentially related with increased carcinogenic potential, and showed distinct allelic frequencies for the T350G allele according to geographical region. We showed further the impact of anogenital cancer location and ICC histological presentation on the age at cancer diagnosis

(Manuscript 2). Finally, for HPV16 we showed differential variant prevalence as a function of ICC histologic presentations and of geographic origin (Manuscript 3).
Overall our results highlight the different biological interplay between viruses and hosts, and exemplify the fundamental and applied importance and utility of the *genotype*genotype* interaction studies.

1.2 A Brief history of HPV research: origins of LR and HR-HPVs

PVs history dates back to mid 19th century, when Domenico Rigoni-Stern observed in 1842 that cervical cancer was relatively common in women living in the city but was rare in catholic nuns who lived in the countryside convents (DiMaio 2015; Scotto and Bailar 1969). This pattern was similar to the one identified in sexually transmitted infections. The search for the underlying infecting agents lasted decades and had many false culprits (*e.g*. *Human herpes virus*, *Treponema pallidum* or *Chlamydia trachomatis*) (DiMaio 2015). Around 1933, microscopic examination of horn-like skin protuberances in rabbits from the plains of the Midwestern United States revealed highly keratinized warts or papillomas, *i.e.* benign tumors of epithelial cells that contained large numbers of virus particles (Shope and Hurst 1933). This was officially the first description of PVs and of the associated diseases. Other researchers observed that the warts produced by the virus on laboratory rabbits occasionally progressed to squamous cell carcinoma, presenting carcinogenic potential (Rous and Beard 1935). Years after, in 1965, the first reports describing the double-stranded circular DNA of HPV appeared (Crawford 1965; KLUG and FINCH 1965). In 1967, Rowson and Mahy characterized the various forms of warts and papillomas produced by the "human wart virus" (Rowson and Mahy 1967). In 1972 in Poland, Stefania Jablonska proposed the association between HPVs and skin cancer in Epidermoplasia Verruciformis (EV) patients, thus correlating for the first time skin cancer with viral infection in humans (Jablonska, Dabrowski, and Jakubowicz 1972). During 1975 and 1976, Harald zur Hausen proposed the hypothesis that HPVs played a main role in cervical cancer and suggested that cervical cancer might arise from infections with the virus previously found in condylomata acuminata (zur Hausen 1977; zur Hausen et al. 1975). A couple of years after, in 1978, Stefania Joblonska and Gerard Orth, at the Pasteur Institute in Paris, discovered HPV5 in skin cancer (Jablonska et al. 1978; Orth et al. 1978) and from 1980 to 1983, HPV6 and HPV11 DNA was first isolated from genital warts (GWs) and laryngeal papilomatosis (LPs) (Gissmann and zur Hausen 1980; Gissmann et al. 1982, 1983; de Villiers, Gissmann, and zur Hausen 1981). Zur Hausen and colleagues went on studying HPVs and in 1983-1984 they isolated HPV16 DNA from cancer biopsies (Dürst et

al. 1983). Subsequently, the same group isolated HPV18 DNA from cancer biopsies as well from several cervical cancer derived cell lines (among them HeLa cells) (Boshart et al. 1984). Within the same period the presence of HPV16 DNA in precursor lesions of anogenital cancers, Bowenoid papulosis (Ikenberg et al. 1983) and in cervical intraepithelial neoplasias (Crum et al. 1984) was demonstrated. Global epidemiological studies characterized the persistent infection by HPV16, 18 and a few other HPVs as major risk factor for cervical cancer (Bosch et al. 1992; Muñoz et al. 1992). From then on, a vast amount of research regarding HPVs and anogenital cancers including vulva, vagina, penis and anus were developed (Bezerra et al. 2001; Madsen et al. 2008; Rubin et al. 2001; Stanley et al. 2012). In 2008, the Nobel Prize for Medicine was awarded to Dr. Zur Hausen for his discovery of HPVs causing cervical cancer.

1.3 PVs Taxonomy and Epidemiology

1.3.1 PVs classification: from *Genera* to *Variants*

PVs infect the skin and mucosa of mammals, but they have also been found in birds (Osterhaus, Ellens, and Horzinek 1977; Prosperi et al. 2016; Terai, DeSalle, and Burk 2002) , turtles (Herbst et al. 2009) snakes (Lange et al. 2011) and fish (Peters and Watermann 1979), probably infecting all amniotes (Bravo, de Sanjosé, and Gottschling 2010). Although most PVs cause asymptomatic infections, some PVs can provoke benign lesions, and malignant transformations in a small number of cases. Benign neoplastic lesions associated to PVs have been identified in a number of animal hosts: bats (RaPV1), cats (FcaPV2 and FcaPV3)(García-Pérez et al. 2014; Lange et al. 2009; Munday et al. 2013) , dogs (CPV1, CPV3, and CPV7) (Delius et al. 1994; Lange et al. 2009; Tobler et al. 2006), horses (BPV1, BPV2 and EcPV2) (Bogaert et al. 2012; Nasir and Campo 2008; Scase et al. 2010), rodents (McPV2) (Nafz et al. 2008), rabbits (SfPV1) (Giri, Danos, and Yaniv 1985), or sheep (OvPV3) (Alberti et al. 2010). To date, around 350 PVs integrate the *Papillomaviridae* family (Figure 1) (https://pave.niaid.nih.gov/).

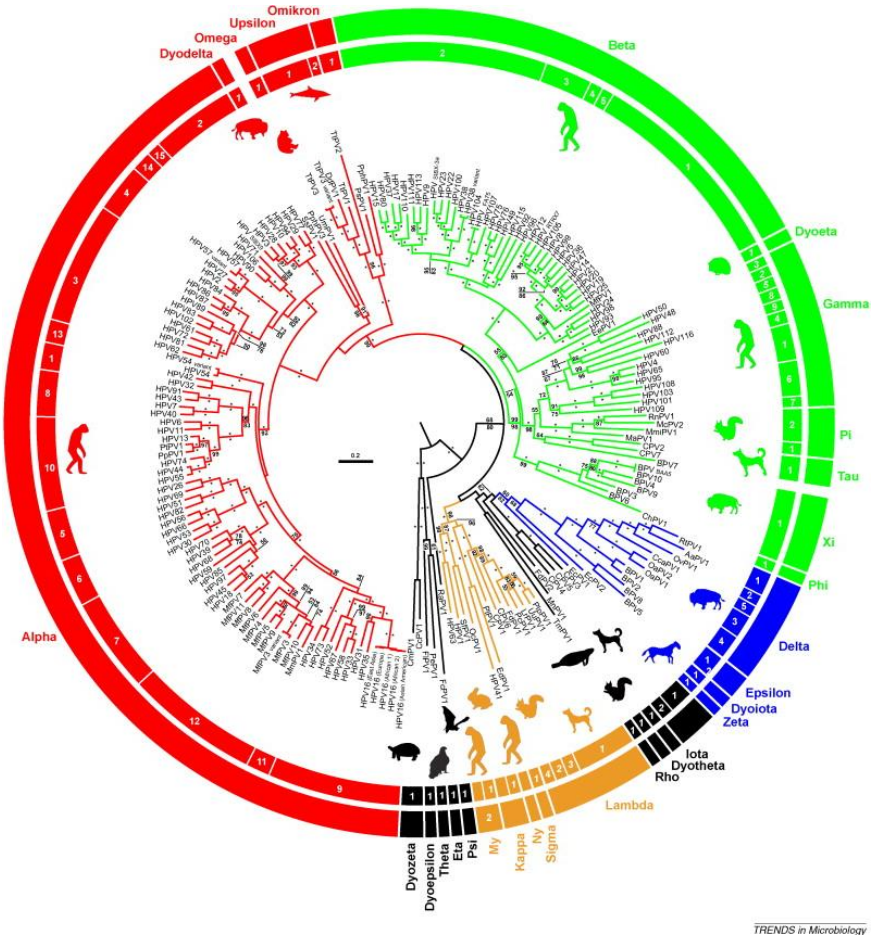


Figure 1: Phylogenetic tree containing the sequences of 186 PVs: Best-known maximum likelihood phylogenetic tree for PVs constructed using the concatenated *E1–E2–L1* genes calculated with RAxML (http://www.kramer.in.tum.de/exelixis/software.html). Color codes highlight the four PV supertaxa. PVs that cannot be assigned yet with confidence to a supertaxon are labeled in black. Silhouettes represent the hosts infected by the corresponding viruses (Bravo et al. 2010)

The *L1* ORF is the most conserved gene within the genome at the nucleotide level (Mengual-Chuliá et al. 2016) and has been used for the classification of new PV sequences (De Villiers et al. 2004). The PVs section of the International Committee on Taxonomy of Viruses (ICTV) has chosen nucleotide identity in the *L1* gene as yardstick for PV classification (Bernard et al. 2010; De Villiers et al. 2004) (Table 1).

PVs taxonomy	
PVs Phylogenetic levels	***L1* gene nucleotide identity (%)**
Genera	≤ 60%
Specie	60-70%
Type	70-90%
Variant	90-98%
Subvariant	98.5-99%

Table 1: PVs taxonomic classification based on the *L1* gene sequence nucleotide identity threshold (Burk et al. 2011, 2013)

According to these guidelines, more than 200 HPVs have been completely sequenced and classified into five genera: *Alphapapillomaviruses* (AlphaPVs), *Betapapillomaviruses* (BetaPVs), *Gammapapillomaviruses* (GammaPVs), *Mupapillomaviruses* (MuPVs) and *Nupapillomaviruses* (NuPVs)(https://ki.se/en/labmed/international-hpv-reference-center) (Figure 2). This classification shows certain correlation between life-cycle characteristics and association with disease (Doorbar et al. 2012). HPVs within AlphaPVs display essentially mucosal tropism, albeit certain members of the genus infect skin and cause common warts, *i.e.* HPV2 and 57. (Figure 2) (Doorbar et al. 2012).

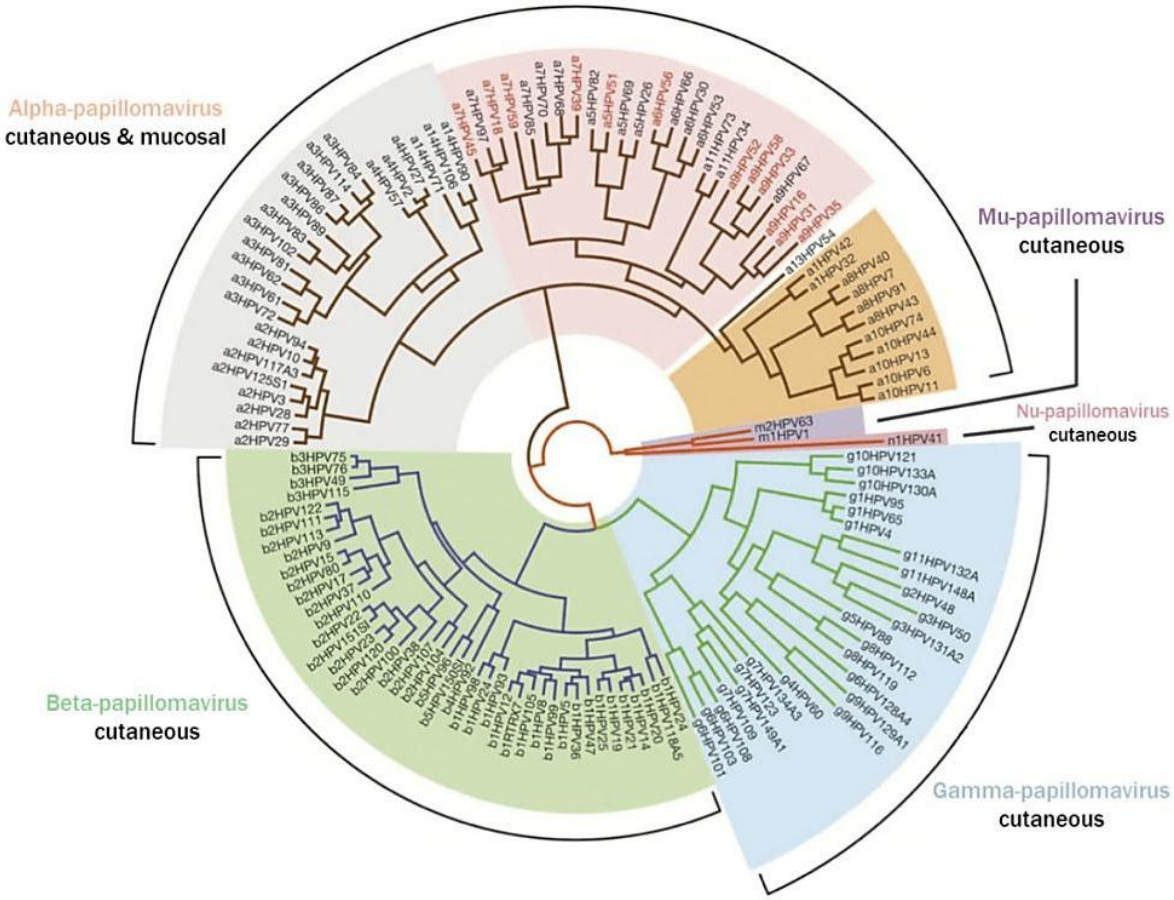


Figure 2: **Phylogenetic tree for HPVs based in *E1E2L2L1* DNA sequences.** HPVs comprise five evolutionary groups with different epithelial tropisms and disease associations. AlphaPVs include the low-risk mucosal types that cause genital warts (orange branch) and the high-risk mucosal types (pink branch) that can cause cervical neoplasias and cancer. Cutaneous HPVs comprise some AlphaPVs (Grey branch) and the other four groups, BetaPVs (Green branch), GammaPVs (Blue branch) MuPVs (Purple branch) and NuPVs (Orange branch)(Doorbar et al. 2012).

AlphaPVs is the most studied genus, as it includes mucosal HR-HPVs (*e.g* HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59) that are associated with anogenital and oropharyngeal carcinomas, as well as mucosal LR-HPVs (*e.g*. HPV6 and 11) associated with non-malignant lesions.

1.3.2 HPV6, HPV11 and HPV16 variants classification

HPV6 variants reveal two distinct lineages, named A and B. Whereas linage A is less diverse, encompassing two clades that differ around 0.2% at the nucleotide level, linage B is more diverse and encompasses three variant lineages namely B1, B2 and B3, with inter-lineage differences ranging 0.4-0.7%. These three lineages are equally distant from the A linage, with a difference of approximately 1.5% of the *L1* nucleotide sequence (Burk et al. 2011) (Figure 3)

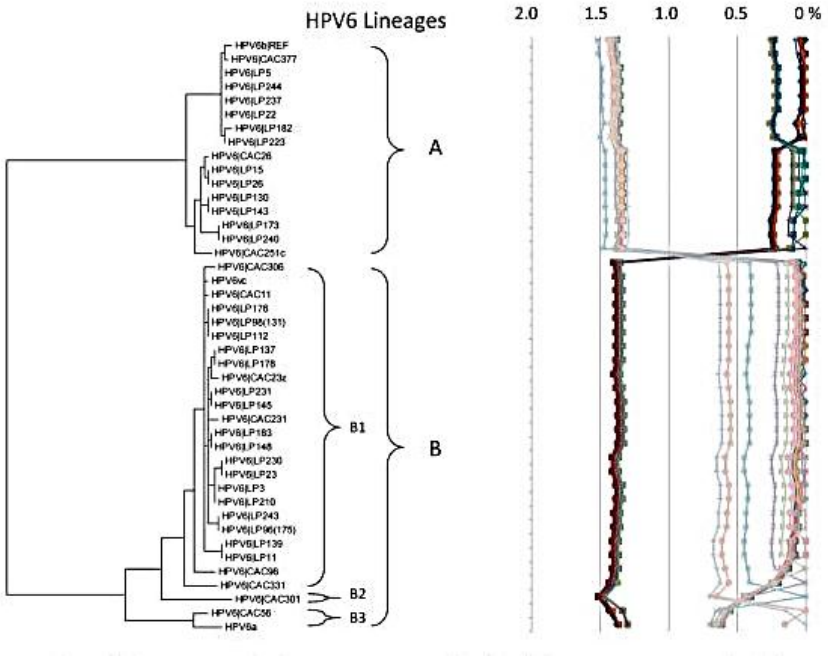


Figure 3: HPV6 variant tree topology and pair-wise comparisons of individual complete genomes: A maximum likelihood tree from global alignment of 43 complete genome nucleotide sequences of HPV6. Distinct variant lineages (termed A and B) and sublineages (termed B1, B2 and B3) are classified according to the topology and nucleotide sequence differences from > 1% to < 10%, and > 0.5% to < 1% ranges respectively. The percent nucleotide sequence differences is calculated for each isolate compared to all other isolates of the same type based on the complete genome nucleotide sequences. Values for each comparison of a given isolate are connected by lines and the comparison to self is indicated by the 0% difference point (Burk et al. 2011).

HPV11 variants are less diverse than their HPV6 counterparts. They reveal two distinct clades that differ at a nucleotide level approximately 0.4%. Based on these small nucleotide sequence differences, the two clades cannot be classified as different variants linages and have been defined instead as variants sublineages A1 and A2 (Burk et al. 2011) (Figure 4).

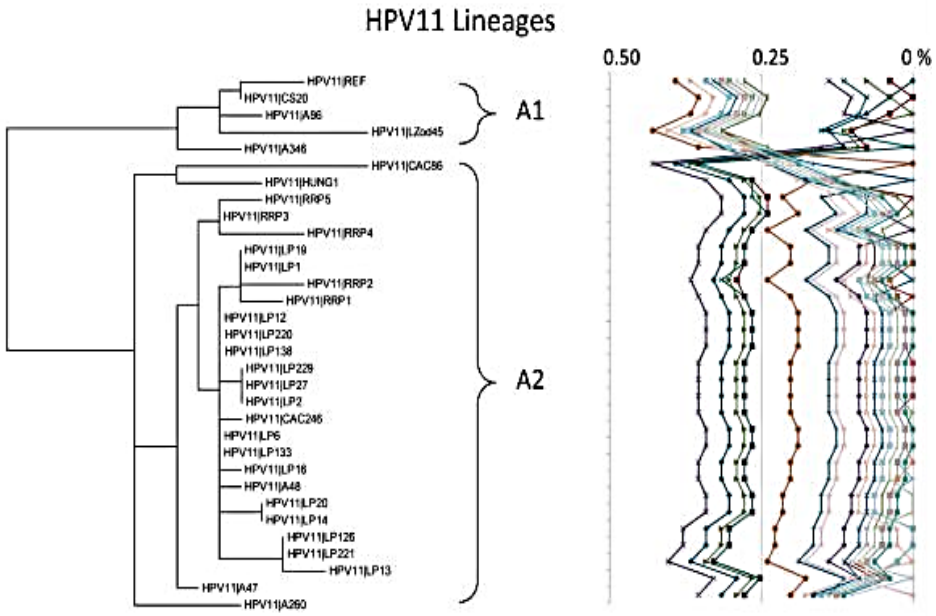


Figure 4: HPV11 variant tree topologies and pair-wise comparisons of individual complete genome. A maximum likelihood phylogenetic tree from global alignment of 32 complete genome nucleotide sequences of HPV11. Distinct sublineages (termed A1 and A2) are inferred from the tree topology and nucleotide sequence differences in the range of ~ 0.5%. The percent nucleotide sequence differences are calculated for each HPV11 isolate compared to all other HPV11 isolates based on the complete genome nucleotide sequences. Values for each comparison of a given isolate are connected by lines and the comparison to self is indicated by the 0% difference point (Burk et al. 2011).

HPV16 variants encompass four different linages, named A, B, C and D (Figure 5). Whereas A clade is the most heterogeneous, including subvariants A1, A2, A3 that differ 0.2-0.4% at the nucleotide level and A4 that differ 0.5% from A1-3 counterparts, the other clades are less diverse: B variants are subclassified in B1 and B2 presenting intra-linage variation of 0.2-0.3% and D variants are subclassified in D1, D2 and D3 subvariants showing intra-linage difference of 0.2 to 0.4% (Burk et al. 2013) At inter-linage level the differences are higher. HPV16_A clade shows an average difference of 1.5% of the nucleotide sequence with B/C/D linages, whereas B/C/D linages are more closer phylogenetically, showing around 1% of differences (Chen, de Freitas, and Burk 2015; Harari, Chen, and Burk 2014) (Figure 5) (Harari et al. 2014).

The classification proposed by Burk and coworkers categorizes variants and subvariants with a consecutive alpha-numeric codes (a letter for variant and a digit for subvariant) as they increase in divergence with the "prototype" (Burk et al. 2011; Harari et al. 2014) (Figure 5). Previous HPV16 variant nomenclatures included potentially misleading geographical references (e.*g.* "European") or ill-defined arbitrary classifications (*e.g.* "prototype" or "non-prototype"). The use of a geography-based nomenclature conveys a message of a close match between differential HPV16 variants prevalence and geography, which is not justified by the best available data (Cornet et al. 2012).

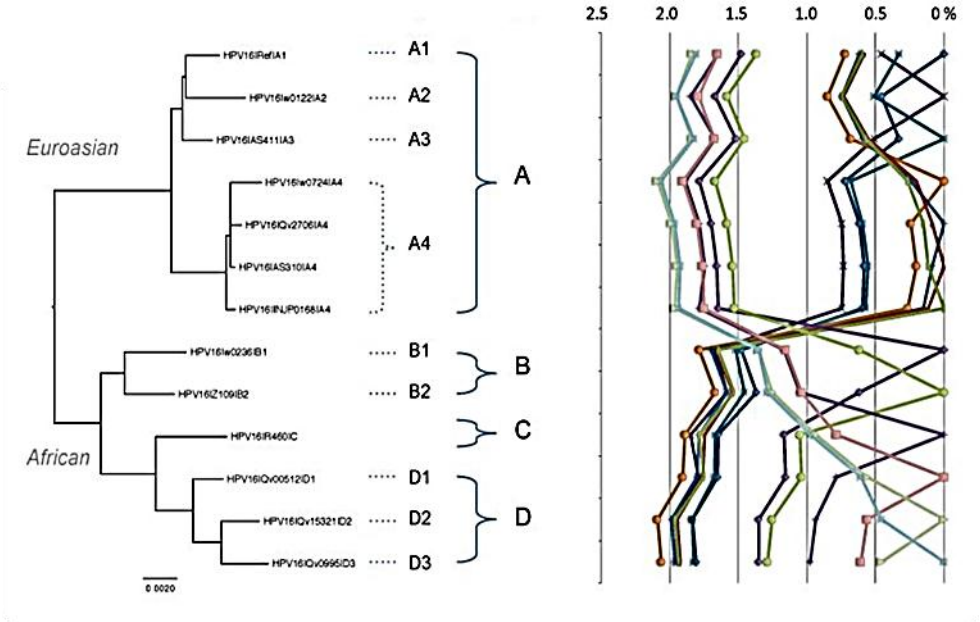
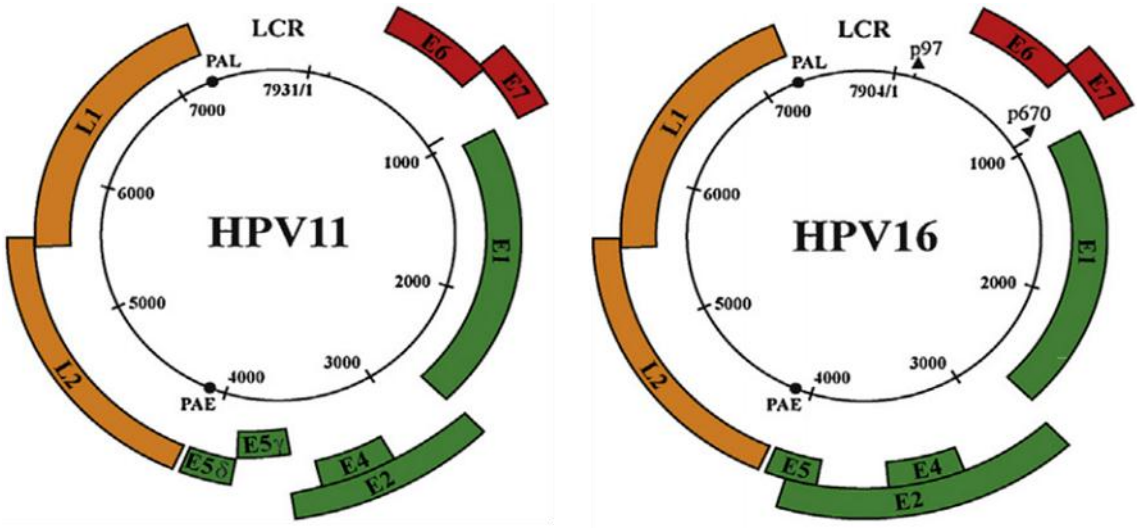


Figure 5: HPV16 variant tree topology and pairwise comparisons of individual complete genomes: A maximum likelihood (ML) tree showing HPV16 variant lineages (termed A/B/C/D) and sublineages (termed A1/A2/A3/A4) classified according to the topology and nucleotide sequence differences from > 1% to < 10%, and > 0.5% to < 1%, respectively . The percent nucleotide sequence differences based on the nucleotide sequences (complete genome) are shown in the panel to the right of the phylogeny. Values for comparison from an isolate are connected by lines and the comparison to self is indicated by the 0% difference point. Symbols and colored lines are used to distinguish each isolate. The scale bar at the bottom of the tree represents nucleotide change of 0.002 per site. Note the clustering in the right hand graph that depicts nucleotide sequence differences. (Harari et al. 2014).

1.4 HPV structure, genome organization and protein functions

PVs are small, non-enveloped double-stranded DNA viruses with circular genome of around 8 kbp, encapsulated inside an icosahedral proteic capsid of 55 nm in diameter (Doorbar et al. 2012; Nebesio, Mirowski, and Chuang 2001). The viral genome includes one coding and one non-coding regions. The coding region contains early genes (E) *E1*, *E2*, *E4*, *E5*, *E6* and *E7* and Late genes (L) *L1* and *L2* (Figure 6). The non-coding region contains regulatory elements located in the Long Control Region (LCR) also named upstream regulatory region (URR) (Figure 6) (Doorbar 2013; Egawa and Doorbar 2017). PVs encode a group of core proteins that were present early on during evolution, and that are conserved in sequence and in function between PVs: E1, E2, L2 and L1 (Doorbar et al. 2012) . The E4 protein may also be a core protein that has evolved to meet epithelial specialization (Doorbar 2017). The accessory proteins have evolved in different PV lineages during adaptation to different epithelial niches (Bravo et al. 2010). The sequence and function of these genes are divergent between types (Table 5). In general, these accessory proteins (E5, E6 and E7) are involved in modifying the cellular environment to facilitate virus life cycle completion, contributing to virulence and pathogenicity (Figure 6, Table 5). Differently to Alpha and BetaPVs, protein functions of Gamma, Mu and NuPVs have not been investigated to the same depth (Doorbar 2017).

(a)



(b)

HPV16 genes	Functions
E1	ATP-Dependent DNA helicase
E2	Viral DNA replication and transcription
E4	Virus release and/or transmission
E5*	Growth factor signaling pathways and immune avoidance
E6*	Block apoptosis and cell proliferation
E7*	Cell cycle arrest
L1	Major constituent of viral capsid
L2	Minor constituent of viral capsid

Figure 6: (a) HPV16 and HPV11 genome organization and their protein functions: Rectangles representing the circular HPV genome represent the positions of the ORFs. The early ORFs (*E1*, *E2*, *E4* and *E5* (in green) and *E6* and *E7* (in red)) are expressed from the different promoters at different stages during epithelial cell differentiation. Early (E) genes encode proteins necessary for viral replication and cell transformation produced during early infection The late ORFs (*L1* and *L2* (in orange)) encode structural proteins produced late in infection. The LCR contains the replication origin as well as post-transcriptional control sequences that contribute to viral gene expression. Scale bar is in kilo-base pairs. (b) HPV16 gene functions. * Different functions of Early proteins between LR and HR proteins (Doorbar 2013; Egawa and Doorbar 2017).

Early genes encode for proteins involved in the regulation of viral DNA replication and transcription (Tyring 2000) (Figure 6). The *E1* gene encodes for a protein with helicase activity preparing the viral genome for replication, while the E2 protein serves as a master transcriptional regulator for viral promoters located primarily in the Long Control Region (Figure 6) (Doorbar 2005). Expression of *E4* increases dramatically during the late phase of infection and the protein encoded is thought to facilitate virion release into the environment by disrupting intermediate filaments of the keratinocyte cytoskeleton

(Figure 6) (Doorbar et al. 1986). E5 is a small hydrophobic protein that destabilizes the function of many membrane proteins in the infected cell. In LR-HPVs such HPV6 and HPV11, the intergenic region Int*E2L2* encodes for E5γ and E5δ (Figure 6). In oncogenic HPVs E5 activates the signal cascade initiated by epidermal growth factor upon ligand binding (Doorbar 2005). E6 and E7 proteins have been defined as oncoproteins because of their molecular activities. In HR-HPVs E6 and E7 proteins induce cellular immortalization by interfering with the proteins involved in cell cycle regulation (Yim and Park 2005): E7 targets pRb and E6 targets p53, leading in both cases to protein degradation (Longworth and Laimins 2004). p53 depletion prevents apoptosis whereas pRb depletion keeps off cell cycle detention. However, in LR-HPVs, E6 does not degrade p53 nor PDZ proteins, and E7 only destabilizes p130 which does not drive to immortalization nor transformation events. Table 5 represents E6, E7 and E5 different functions between LR HPVs and HR HPVs. Late proteins L1 and L2 are respectively the major and minor constituents of the viral capsid (Doorbar 2005). When overexpressed in various eukaryotic cells, L1 monomers self-assemble to form virus-like particles (VLPs)(Hernandez et al. 2011). The non-coding region is formed by LCR. LCR encloses *cis*-regulatory elements important for viral replication and gene control expression (Bernard 2002; Chow, Broker, and Steinberg 2010; Doorbar 2005) (Figure 6).

	High-Risk Alpha	Low-Risk Alpha
E6	Encodes E6* products	Does not encodes E6* products
	Binds E6AP	
	Degradation of p53 and PDZ-proteins	No degradation of p53 and PDZ proteins
	Inhibition of p53 transactivation and acetylation	
	Inhibits Notch pathway via p53	Not known
	Inhibition of interferon response	Weak inhibition of interferon response
	Degradation of BAK	
	Activation of telomerase	No activation of telomerase
E7	Destabilizes pRB (p105), p107 and p130	Destabilizes p130
	Induction of cell cycle entry and DNA synthesis, role in genome amplification	
	Supression of STAT-1 function	No supression of STAT-1 function
	Immortalization and transformation	No immortalization and transformation
E5	Stimulation of EGFR signaling pathways	Not known
	Downregulation of MHC	

Table 5: Differential functions for E5, E6 and E7 proteins between LR and HR-HPVs: E6, E7 and E5 accessory proteins have evolved in each HPV type during adaptation to different epithelial niches. The sequence and function of these genes are divergent between LR and HR AlphaPV types (Doorbar 2017). *1 (White et al., 2014), *2 (Cordano et al., 2008)

1.4.1 LR and HR- HPVs present different viral cycle strategies

HPVs infect cells in the basal layer of the stratified epithelia (Doorbar 2005) , presumably gaining access to the target cells through microlesions (Doorbar 2005). The viral life cycle is synchronized with the keratinocyte differentiation program, proceeding through different successive phases such as genome amplification, viral assembly and viral release (Doorbar 2005). These stages are accompanied by a concomitant shift of the expression pattern, shifting from the expression of the early genes to the expression of the late genes (Mcmurray et al. 2001) (Figure 7).

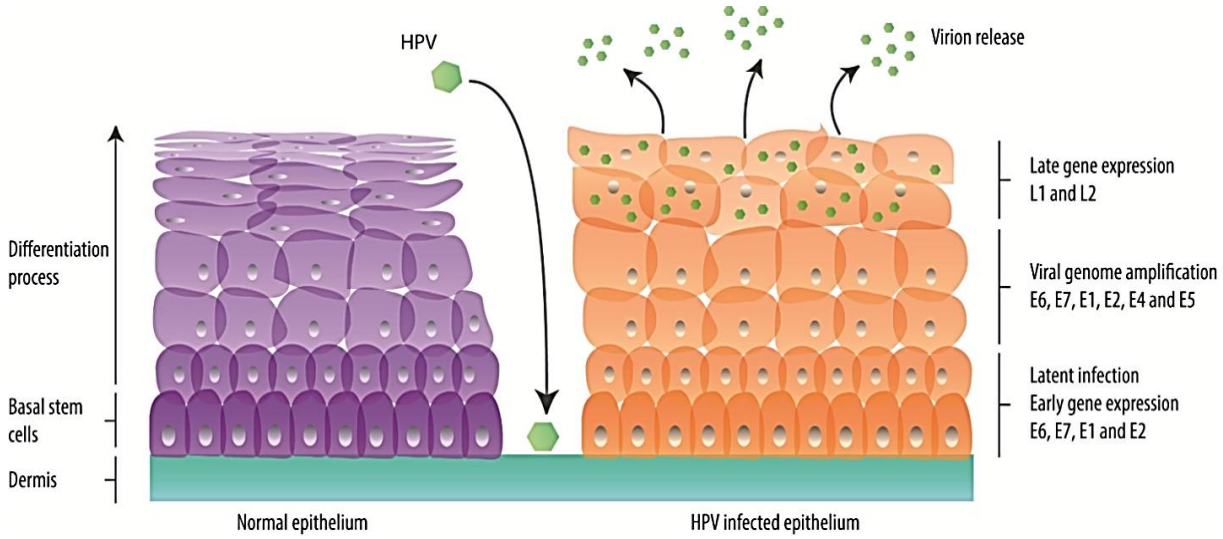


Figure 7: HPV life cycle. HPVs establish latent infection in the basal cells of the differentiating epithelium when it becomes exposed through microwounds. Upon infection (right), the viral genome is established as low-copy episomal circular nuclear plasmids. As epithelial cells goes through differentiation, HPV-positive cells induces the productive phase of the viral life cycle. The expression of *E6* and *E7* reinforced by *E1* and *E2* deregulates cell cycle control, pushing cells into S phase, allowing viral genome amplification in cells that normally would have exit the cycle. During the last phase of the infection, L1 and L2 proteins self-assemble into virions that encapsidate the viral genomes, and are eventually shed from the uppermost layers (green hexagons) (Moody and Laimins 2010).

LR and HR HPVs present different viral cycle strategies (Doorbar 2016) and different DNA damage response (Santegoets et al. 2012). Once infected the cells from the basal layers, LR-HPVs persist for long term as stem-like cells (Doorbar 2016). LR-*E6* and *E7* restore a replication-competent environment in the infected post-mitotic basal cells, but do not increase their proliferation mediating a slow cell division in order to settle into the epithelium surrounding the basal layer (Figure 8). This process is thought to be dependent on the ability of *E7* to bind the Rb family member p130, In the basal cells, *E1* and *E2* levels rise as a result of late promoter activation allowing the viral and cellular gene products to work together to amplify the nuclear HPV episomes (Bodily et al., 2013; Flores and Lambert, 1997; Ozbun and Meyers, 1997; Ozbun and Meyers, 1998; Ruesch et al., 1998; Spink and Laimins, 2005).

In contrast, HR-HPVs drive proliferation of the infected cells through *E6* and E7 oncoproteins expression increasing rapidly the proportion of proliferating cells in the low layers of the epithelium. Simultaneously to the cell cycle re-entry, DNA integration events occur often at specific sites in the *E1-E2* region disrupting *E2* sequence and allowing the deregulation of transcription activities, resulting in overexpression of viral oncoproteins E6 and E7. (Figure 8). This overexpression leads *E6* and *E7* to bind p53 and pRb tumor suppressors inducing an uncontrolled cell cycle (Figure 8). Thus, *E6* and *E7*, which are expressed in the early stages of the infection process are largely responsible for the changes related to the process of malignancy. Finally, virion assembly requires the timed expression of *L2* and *L1* post-genome amplification, and the accumulation of the viral *E4* protein involved in virus release and survival (Doorbar, 2013).

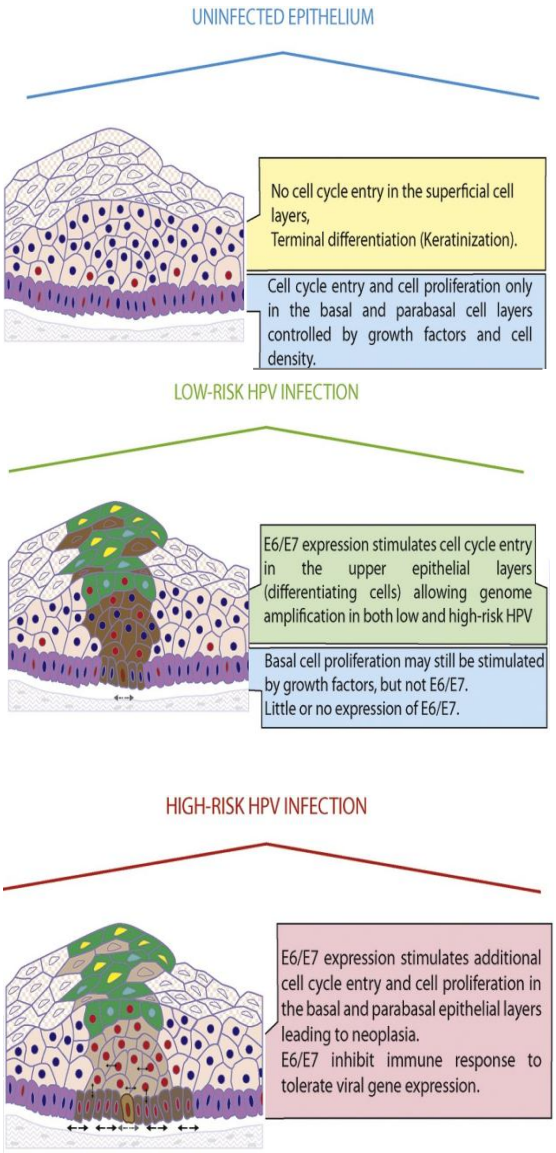


Figure 8: LR and HR-HPV *E6* and *E7* gene expression: (a) In uninfected epithelium, cell cycle entry (red nuclei) and cell division in the basal/parabasal cell layers is controlled by cell density and growth factors. In the suprabasal layer, cells exit the cell cycle and start to terminally differente (keratinization). (b) In lesions caused by LR-HPVs, it is thought that basal cell proliferation is largely regulated by the presence of growth factors. The primary role of the HPV E6/E7 proteins in these lesions is to drive cell cycle entry above the basal layer in order to facilitate HPV genome amplification (red nuclei in mid epithelial layers). This is thought to be dependent on the ability of *E7* to bind the Rb family member p130. Little or no *E6/E7* expression is thought to occur in the basal layer, and the precise role of these proteins in basal cells is not known. E6/E7 may limit keratinocyte differentiation in the basal layer, or increase the population of infected cells to drive "Papillomatosis" and to retain the reservoir of infection. (c) In high-risk Alpha HPV infections, expression of the high-risk E6/E7 proteins in the basal layer leads to cell proliferation and evasion from host immune surveillance. In these cell populations, malignant transformation may be developed. (Doorbar 2013; Egawa and Doorbar 2017)

The propensity of LR-HPVs to cause neoplasia and cancers appear to be very low and is often considered to be negligible in the general population (Brentjens et al. 2002). Rarely HPV11 has been associated to certain cancers (Venkatesan, Pine, and Underbrink 2012). A number of cases have nevertheless been described in individuals persistently infected by non oncogenic HPVs and who cannot properly control their active infections, as in patients suffering from RRP (Venkatesan et al. 2012). Nonetheless, the concept of persistent

deregulated viral gene expression as a general cause of HPV-induced cancers extends beyond just the extensively studied HR-AlphaPVs (Doorbar et al. 2012) . In either case, cancer development takes time and is associated with the host inability to keep a check on viral gene expression (Figure 9).

1.4.2 Disease evolution following HPV infection

1.4.2.1 Immune response, chronic infection, latency and clearance

Once HPV infection has occurred, the process of lesion formation begins (Schiffman et al. 2016) (Figure 9). In an immune competent host, both LR-and HR-HPVs can persist for months or years, causing chronic productive lesions that shed virus from their surface layers over a prolonged period (Figure 9). To achieve this, HPVs have developed a number of key adaptations, which allow them to persist in infected epithelial cells, even in the face of an active adaptive immune system (Christensen 2016) . HPVs could avoid immune detection by limiting viral gene expression in the epithelial basal and parabasal layers to very low levels (Christensen 2016). Indeed, small number of viral proteins are required for basal cell genome maintenance, a situation that restricts the presentation of viral antigens on MHC class I and the stimulation of adaptive cell-mediated immunity (Doorbar et al. 2015). In fact, for many LR-HPVs, viral gene expression in the epithelial basal cell is extremely difficult to detect. The elevated viral gene expression that is essential for viral genome amplification and virion synthesis is typically delayed until the infected cell reaches the mid- or upper epithelial layers where T-cells and dendritic cells are less abundant and where immune surveillance is thus less efficient (Kanodia, Fahey, and Kast 2007). In addition to down-regulate canonical MHC class I levels (Seliger, Ritz, and Ferrone 2006), HPVs may also delay the adaptive immune response by inhibiting retention of Langerhans cells (Cromme, Meijer, et al. 1993; Cromme, Snijders, et al. 1993). Since HPVs do not cause a lytic infection and they shed virus particles only from the epithelial surface, the opportunity for Langerhans cells to sample and appropriately present viral antigens becomes restricted (Guess and McCance 2005; Matthews et al. 2003). Even so, lesion regression, when it eventually occurs, appears to depend on the cross-priming of epithelial-specific dendritic cell with viral antigens and the subsequent activation of a T cell response in the draining lymph node (Figure 9) (Hibma 2012). It is generally believed that these HPV-specific defenses, by stalling the adaptive immune response, act to prolong the duration of infection.

Most HR-HPV infections eventually clear as a result of a host cell-mediated immune response in less than two years, with effective immune recognition leading to T-cell

homing and T-cell infiltration at the site of infection (Doorbar et al. 2015) (Figure 12). HPVs-specific helper CD4 T cells that can recognize epitopes on the E2 and E6 proteins have been reported to be important in the clearance of low-grade HPV-induced disease (Dillon et al. 2007; Welters et al. 2003), with a CD4 response to *E7* being more important in the control of high-grade neoplasia (Peng et al. 2007). Clearly, the precise nature of the host response to infection depends on which viral antigens are seen by the immune system, as well as the time lapse between infection and eventual immune detection. The concept of genome maintenance in the absence of significant viral gene expression underlies the concept of viral latency and the immunological mechanisms that allow chronic inapparent active HPV infection and the long-term shedding of virus particles from apparently normal epithelial tissue (Maglennon and Doorbar 2012) (Figure 9). Indeed, a comparable chronic inapparent infection can also operate after the immune regression of HPV productive lesions. Clearly, there is a balance between the strength of the host immune response and the virus ability to stimulate host immunity at low level over a prolonged period of time while still producing sufficient viruses into the environment. Any reduction in the level of immune-surveillance, such as that which may occur during aging or following treatment with immune suppressive drugs, can allow more extensive viral gene expression and even the appearance/reappearance of papillomas or neoplasia (Figure 9)

1.4.2.2 Disease persistence: progression from neoplasia to cancer

The cell type specificity of viral and host gene functions contributes to increased efficiency of the immune system in some tissues in comparison to others (immunoprivilege) (Virgin, Wherry, and Ahmed 2009). The lack of tissue immunoprivilege is an important contributor to maintenance of chronic viral infection, allowing the HPV persistence in epithelial cells (Virgin et al. 2009) The inefficiency of the immune system to properly detect infection is generally characterised by only low numbers of circulating antigen-specific T cells and an abundance of CD25-positive T-regulatory cells, which produce an intraepithelial cytokine environment that restricts T cell trafficking, inhibiting the control of the infection (Hibma 2012). In this immune-tolerant environment, disease persistence is facilitated, and deregulated viral gene expression can go unchecked (Doorbar et al. 2015). For the oncogenic HPVs, the loss of control of key viral genes, particularly the *E6* and *E7* genes that regulate cell cycle entry, cell proliferation, and differentiation, leads to a dramatic change in lesion phenotype (Middleton et al. 2003)(Figure 9). During the ordered productive life cycle, these genes are carefully regulated and act to subtly control the basal cells commitment to differentiation and the suprabasal cells ability to enter the

cell cycle for genome amplification (Doorbar 2005). As their expression increases in the basal layer, normal cellular controls are progressively compromised, leading to neoplasia, with the ability of the virus to complete the life cycle and to produce infectious virions at the epithelial surface being progressively lost (Figure 9). The situation is exacerbated because the viral proteins that drive neoplastic progression, at least amongst oncogenic, also contribute to immune evasion (Hibma 2012). Persistent deregulated gene expression, as occurs in CIN3 and following viral genome integration, can lead to the accumulation of secondary genetic changes in the infected host cell and development of cancer (Figure 9) (Doorbar et al. 2012).

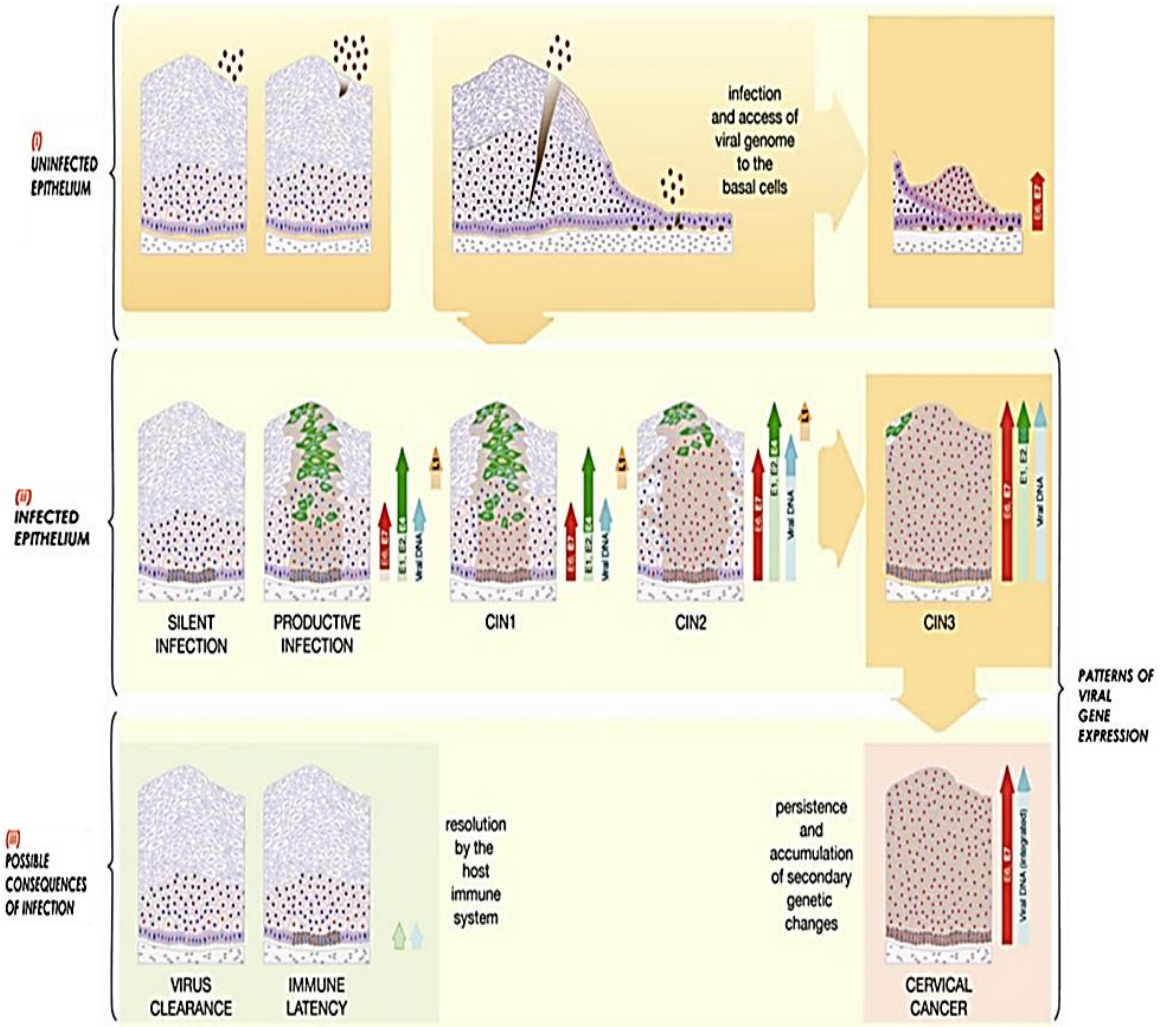


Figure 9: HR-HPV Infection and possible consequences. (i) Infection requires the entry of HPV virions into the mitotically active epithelial cells of the basal layer, which in stratified epithelium is thought to require a microwound. In the columnar cell layers, infection is thought to be facilitated by the proximity of the target cell to the epithelial surface, which may allow the virus to access a cell type that is unable to support the full productive life cycle (right) (ii) Following infection (shown in (i)), expression from the viral genome can sometimes be suppressed (*e.g.* by genome methylation), leading to a 'silent' infection in which the viral genomes are retained in the basal layer without apparent disease. Infection may alternatively lead to an ordered pattern of viral gene expression leading to virus synthesis and release from the upper epithelial layers (productive infection or CIN1), or to deregulated viral gene expression and high-grade neoplasia (CIN2/CIN3). Persistent high-grade disease is associated with cancer progression. Cells in cycle are indicated by the presence of red nuclei. Cells expressing *E4* are shown in green, while those expressing *L1* are shown in yellow. The brown shading on the diagrammatic representations of the epithelium identify all the cells (differentiated and un-differentiated) that contain viral genomes. (iii) In most cases, HPV infections are resolved as a result of a cell-mediated immune response (left). This may lead to viral clearance or to viral latency and the persistence of viral episomes in the epithelial basal layer without life-cycle completion. Persistent deregulated gene expression, as occurs in CIN3 and following viral genome integration, can lead to the accumulation of secondary genetic changes in the infected host cell and development of cancer (right). This is facilitated by over-expression of the high-risk E6 and E7 proteins. Cells in cycle are shown by red nuclei. Red arrow indicates *E6* and *E7* expression, green arrow indicates *E1*, *E2* and *E4* expression, blue arrow indicates viral DNA and yellow arrow indicates *L1* expression (Doorbar et al. 2012)

1.5 Pathologies associated with LR and HR-HPVs

Lesions caused by LR-HPVs (*i.e* benign, hyperproliferative lesions including nongenital and anogenital skin warts, oral and laryngeal papillomas, and anogenital mucosal condylomata) are self-limiting and are generally cleared by the host immune system. This is also the case for the HR-HPVs that produce only asymptomatic infection in most individuals (Bzhalava et al. 2013; Egawa and Doorbar 2017; Garland et al. 2009; Mounts and Kashima 1984). However, among susceptible populations, infections by LR-HPVs can be refractory to treatment, and show problematic pathologies (*e.g.* RRP). This situation can be sometimes associated with the development of cancers (Lübbe et al. 1996). In order to decrease mainly GWs incidence, the quadrivalent HPV vaccine included both HPV6 and 11 along with HPV16 and 18 (Braaten and Laufer 2008) . Currently, after more than ten years of high coverage vaccination, a robust reduction in GWs burden is observed among young women in certain geographical regions(Baandrup et al. 2013; Harrison et al. 2014; Read et al. 2011).

HR-HPVs have been the subject of most research and are associated *in vivo* with asymptomatic infections that are generally acquired in young adults following the onset of sexual activity (Forman et al. 2012; de Martel et al. 2012). Incidence of infection peaks before the age of 25 and then declines as these asymptomatic infections are cleared by the host immune system (Forman et al. 2012). Thus, most HR-HPVs infections persist for months or perhaps years, but are generally resolved without causing serious disease (Moscicki et al. 2012). Persistent or long-term infection, rather than transient infection by a subset of HR-HPVs, is generally considered to lead to malignant anogenital tumors, including cancers of the anus, penis, vulva, vagina, and cervix (Alemany et al. 2014, 2015, 2016; Larsson et al. 2013; de Sanjose et al. 2010) Further, a proportion of oropharyngeal cancers is also attributable to chronic HR-HPVs infections (Castellsagué et al. 2016).

1.5.1 LR-HPV pathologies: genital warts (GWs)

Genital and anal warts (GWs) are the most easily recognizable sign of genital infection by HPVs (Ball et al. 2011). GWs are flat keratotic warts or benign tumors caused mainly by HPV6 and HPV11 (Handsfield 1997). Although a wide variety of HPVs can cause GWs, HPV6 and HPV11 account for about 85% of all cases (Aubin et al. 2008; Garland et al. 2009; Prétet et al. 2008)(Figure 10)

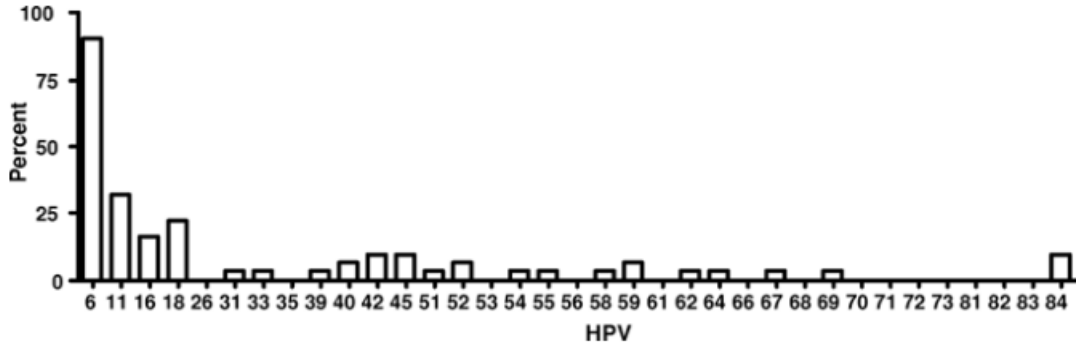


Figure 10: Distribution of HPVs in wart tissue as determined by Linear Array (Ball et al. 2011)

These two HPVs are the most common non-oncogenic HPVs found in the female genital tract in general population (Bruni et al. 2010). However, co-infections by oncogenic and non-oncogenic HPVs are detected in large fraction of these lesions (45%)(Sturegård et al. 2013), which could partially explain the increased risk of Cervical Intraepithelial Neoplasia (CIN) and ICC in women with GWs (Blomberg et al. 2012).The incubation period for GWs is short in the majority of lesions, being of 3-5 months after infection among women (Garland et al. 2009; Winer et al. 2005) and in one year among men (Anic and Giuliano 2011). The risk of suffering GWs is closely associated to sexual behavior, being the number of sexual partners the main risk factor (Munk et al. 1997). The highest incidence rate for GWs is at 20-24 years for women, which correlates well with the peak of HPVs infection in the female genital tract (Bruni et al. 2010)and at 20-29 years for men (Kliewer et al. 2009). Most people acquiring GWs-associated HPVs clear the infection rapidly without ever developing warts or any other symptoms. However, people may transmit the viruses to others even if they do not display overt symptoms of infection(Braaten and Laufer 2008).

1.5.2 LR-HPV pathologies: Recurrent Respiratory Papillomatosis

Recurrent Respiratory Papillomatosis (RRP) is a neoplastic disease of the upper respiratory tract, usually the larynx, which affects mainly pediatric but also adult populations (Larson and Derkay 2010; Syrjänen 2010). RRP is characterized by solitary or multiple benign hyperproliferative papillary tumors that appear in the respiratory tract (Venkatesan et al. 2012). HPV6 and HPV11 are the causative agents of most RRP cases(Omland et al. 2014). Albeit with a very low prevalence, HPV16 has also been detected in these lesions (Mounts and Kashima 1984) (Table 6).

	HPV6	HPV11	HPV11+6[Δ]	LR+HR HPV[§]	HPV negative	Total
Juvenile-onset RRP	25 (51.0)	14 (28.6)	7 (14.3)	2 (4.1)	1 (2.0)	49 (100.0)
Adult-onset RRP	109 (63.3)	26 (15.1)	8 (4.7)	17 (9.9)	12 (6.9)	172 (100.0)

Table 6: Distribution of HPV genotype profile in juvenile and adult onset RRP, n (%). The difference in HPV profile between juveniles and adults suffering RRP. [Δ]Infection with both HPV 6 and 11. [§]HPV 6 or 11 (LR-HPV) in co-infection with one or two high-risk HPVs (HR-HPV). High-risk HPVs (HR-HPV) comprise HPV 33, 45, 18, 16, 31 or 35. (Omland et al. 2014)

Despite the benign nature of RRP, lesions tend to grow and extend throughout the entire respiratory tract, eventually causing severe airway obstruction (Buchinsky et al. 2008). As these warts can frequently recur, repetitive surgery is required in some cases to remove them in order to avoid interfere with breathing. In extremely rare cases the lesions can progress to cancer, and malignant transformation has been described in approximately 5% of the RRP cases (Hobbs and Birchall 2004). For this reasons, RRP can result in considerable mortality rates and morbidity that includes dysphonia, dyspnea and in some cases, complete obstruction of the airways (Wiatrak et al. 2004).

1.5.3 HR-HPV pathologies: Invasive anogenital carcinomas

HR-HPVs cause cancers at various epithelial sites, including the cervix, vulva, vagina, penis and anus (Alemany et al. 2014, 2015, 2016; Larsson et al. 2013; de Sanjose et al. 2010). HPV16, the most oncogenic HPV type, is the most frequently detected in anogenital ICs worldwide (Bruni et al. 2010; Bzhalava et al. 2013), and is also responsible for more than 50% of all ICC all over the world and for even higher fractions of other HPV associated anogenital carcinomas such as vulvar, vaginal, penile and anal IC (Alemany et al. 2014, 2015, 2016; de Sanjosé et al. 2013) (http://www.hpvcentre.net). For the cervical transformation zone (TZ), the particular vulnerability for the development of HR HPV-associated neoplasia is thought to result from the presence of a unique type of epithelial cell, known as the cervical reserve cell or cuboidal cell (Herfs et al. 2012), that normally gives rise to either the columnar epithelium of the endocervix and cervical glands but, under some circumstances, can give rise to the stratified cells of the cervical transformation zone (Reich and Regauer 2015)(Martens et al. 2004) (Figure 11).

1.5.3.1: Invasive cervical carcinoma

ICC is the second most common cancer affecting females worldwide (Maxwell Parkin et al. 2001; Waggoner 2003) being responsible for approximately 266,000 deaths per year (http://globocan.iarc.fr/Default.aspx). Around 88% of the global burden occurs in developing countries (Ferlay et al. 2013): 53.000 in Africa, almost 32.000 in Central-South America and Caribbean and *ca.* 160.000 in Asia (Ferlay et al. 2013). The causative role of chronic HPV infection in the pathogenesis of virtually all cervical cancers is well established, being HPV16 the most prevalent type in ICC and representing the aetiological cause of more than a half of cases worldwide (61%) followed by HPV18 (10%) and 45 (6%) (de Sanjose et al. 2010).

HPV16 induced ICC and its precursor lesions are normally located in the squamous columnar junction (SCJ) near the TZ (Figure 11) (Crum 2000; Herfs et al. 2012; Yang et al. 2015). The epithelial changes that allow reserve cells to develop into a stratified TZ typically occur at puberty. Since estrogen levels increase during early adolescence, the cervical orifice becomes dilated, exposing the endocervical columnar epithelium to the ectocervix. This area of columnar cells of the ectocervix forms a cervical eroded area; this process is termed "cervical ectropion" (or cervical erosion). This eroded region is then, exposed to the acid environment of the vagina and through a process of squamous metaplasia, it is transformed into stratified squamous epithelium (Herfs et al. 2012) Such metaplastic changes can, however, occur throughout a woman's life whenever the conversion of columnar endocervical epithelial cells to a multi-layered epithelium is required (Schiffman et al. 2016). This region of replacement called the TZ, leads to the proximal migration of the SCJ towards ectocervix, region that defines the boundary between the squamous-lined ectocervix and the columnar lined endocervix and varies its location depending on hormone production, essentially synthesised by the endocervix (Bosch et al. 2002; Crum et al. 1984; Ferenczy and Franco 2002; Herfs et al. 2012)(Herfs et al. 2012). Indeed, a second group of vulnerable cuboidal cells have also been identified more precisely at the SCJ (Yang et al. 2015). (Figure 11). Thus, TZ is maintained by specialized cell called reserve cell and probably by a group of cuboidal cells located at the SCJ (Herfs et al. 2012). These cells seem to respond differently to signals from adjacent epithelial cells and from the dermis, when compared to the more conventional epithelial stem cells that colonize the stratified layers of the ectocervix (Yang et al. 2015) (Figure 11). According to these models, viral gene expression is deregulated at these specific sites following infection (Schiffman et al. 2016).

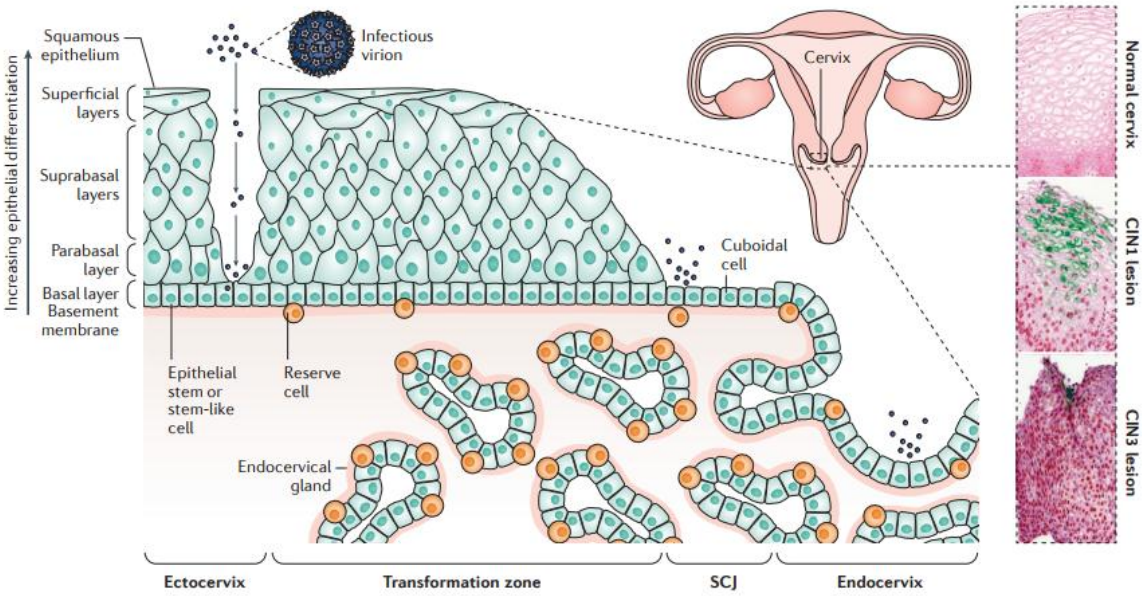


Figure 11: HPV infection and the transformation zone. Most cancers at the cervix arise at the transformation zone and adjacent endocervix, a region that initially comprises columnar epithelium and at puberty, undergoes metaplasia to form a fully differentiated squamous epithelial layer. The stratified layers of the ectocervix are thought to be maintained by 'conventional' epithelial stem cells that are located in the basal layer. The stratified layers of the transformation zone, and the single layer of columnar cells that line the endocervix, are thought to be maintained by the cervical reserve cells. Although reserve cells are typically abundant at sites of metaplasia, there is ruther the involvement of a second type of stem-like cell with cuboidal appearance that are located more precisely at the squamocolumnar junction (SCJ). Current thinking suggests that productive high-risk HPV infection is favoured at the ectocervix and that lesion formation begins from infection of an epithelial stem cell (reserve or cuboidal stem cell) at the transformation zone (TZ) or endocervix. The immunohistochemistry images on the right show a normal cervix, a low-grade clinical lesion pathologically labelled as cervical intraepithelial neoplasia grade 1 (CIN1) and a high-grade CIN3 lesion stained to detect the HPV *E4* protein (green) and the cell cycle marker minichromosome maintenance protein complex (red) (Schiffman et al. 2016)

The strong linkage between chronic infection by HR-HPVs and the development of ICC is observed for different histological presentations. The different histological subtypes of the disease have been classified by the World Health Organization (WHO) (https://www.iarc.fr/en/publications/pdfs-online/pat-gen/bb4/bb4-chap5.pdf).

1.5.3.1.1 Squamous cell carcinoma

Squamous cell carcinoma (SCC) is an epithelial invasive cancer that affects the squamous cells: flat, skin like cells that cover the ectocervix (Schiffman et al. 2016) . SCC arises at the SCJ between the squamous epithelium of the ectocervix and the columnar epithelium of the cervix and approximately 80-85% of the ICC are SCC (Kosary 1994).Nowadays, thanks to the current screening procedures and the increased detection of premalignant lesions, SCC has decreased slightly whereas other histological types of cancer are increasing its prevalence (Vinh-Hung et al. 2007).

1.5.3.1.2 Adenocarcinoma (ADC)

Cervical adenocarcinoma (ADC) originates from glandular precursor lesions of the endocervical mucosa (Schiffman et al. 2016). Among ICC, this glandular histological type accounts approximately for the 10-15% of the whole invasive cancers (Kosary 1994). Over the past 40 years, relative and absolute incidence of adenocarcinoma has increased, especially among younger women aged between their twenties and their forties (Baek et al. 2014; Castellsagué et al. 2006; Smith et al. 2000; Vinh-Hung et al. 2007) . Some studies have reported that ADC has worse prognosis than SCC with nearly 10-20% differences in 5 year overall survival rates (Davy et al. 2003).

HPV16 is the most prevalent type in ADC, being the aetiological cause of 50% of all of them closely followed by HPV18 (32%) and 45 (12%) (de Sanjose et al. 2010).

1.5.3.1.3 Adenosquamous cell carcinoma (ADSC)

Adenosquamous cell carcinoma (ADSC) is a mixed histological type (Yan et al. 2008). It contains malignant glandular and squamous components consisting of intermingled ADC and SCC (Yan et al. 2008) (Burk et al. 2003). ADSC occurs in 2-3% of patients with cervical cancer and its incidence is increasing along with that of ADC (Baek et al. 2014; Vinh-Hung et al. 2007). Similar to ADC, patients with ADSC have been suggested to present poor prognoses than those with SCC after radical hiterectomy (Lai et al. 1999). There is some controversy with the ADC and ADSC diagnosis. Initially it was considered as a subtype of ADC, however, several authors have observed that cervical ADSC is a clinic-pathological factor which influences the prognosis (Yan et al. 2008). Thus, in recent years it has been classified as a different type of ICC (https://www.iarc.fr/en/publications/pdfs-online/pat-gen/bb4/bb4-chap5.pdf).

HPV16 is the most prevalent type in ADSC, being the aetiological cause of 39% cases worldwide followed by HPV18 (32%) and 45 (12%) (de Sanjose et al. 2010)

1.5.3.2 Invasive vulvar carcinomas (IVuC)

Vulvar tumors are relativity rare, representing only the 4% of all anogenital cancers diagnosed each year (de Sanjosé, Bruni, and Alemany 2014). Vulvar cancer is diagnosed using two histological categories with different risk factor: basaloid, warty and verrucous cancers on the one side, and keratinizing squamous cell-cancer on the other side (Kosary 1994) (Ansink 1996). The keratinizing type appears in elder women and they are rarely associated with HPVs DNA (de Sanjosé et al. 2013) (IARC 2012). HPV16 is the most prevalent type being the aetiological cause of 72.5% of all vulvar carcinomas worldwide, followed by HPV33 (6.5%) and 18 (4.6%) (de Sanjosé et al. 2013).

1.5.3.3 Invasive vaginal carcinomas (IVaC)

Vaginal carcinomas are also uncommon, representing only 2% of all gynecological tumors (Kosary 1994). Women with vaginal cancer present higher risk of developing other anogenital cancers, particularly cervical cancer (Melkert et al. 1992; de Sanjosé et al. 2014). At worldwide level, HPVs DNA is found in 70% of vaginal invasive carcinomas. HPV16 is the most prevalent type detected in 57% of HPV positive carcinomas, followed by HPV18, HPV31 and HPV33 with the same proportion (5.0%) (Alemany et al. 2014)

1.5.3.4 Invasive anal carcinomas (IAnC)

Anal cancers are rare malignancies arising in the anal canal, and the developmental stages are similar to those of cervical tumors (Ouhoummane et al. 2013). Similarly to cervix, anus also present a squamous columnar junction (SCJ) that joins with the rectal mucosa at the anal transformation zone (Yang et al. 2015). Although, both cervical and anal SCJ present stem cell like similar traits (Yang et al. 2015), the anal and cervical TZ differ in their microanatomy (Yang et al. 2015). Indeed, cervical SCJ cells are monolayered, present a direct contact with the basal membrane and show an immune-phenotype distinct tumour (Yang et al. 2015). IAnC increasing trend in incidence has been reported in the last decades in both men and women (Giuliano et al. 2008). Most (≥80%) precancerous lesions have also been linked to HPV infection and HPV16 is the most prevalent type being the aetiological cause of 75.8% of all anal carcinomas worldwide, followed by HPV18 (3.4%) and 33 (2.3%)(Alemany et al. 2015)

1.5.3.5 Invasive penile carcinomas (IPeC)

Penile cancer is responsible for less than 0.5% of cancers in men (Alemany et al. 2016; de Sanjosé et al. 2014). Cervical and penile cancer present similar geographical incidence and the concordance of these two malignancies in married couples suggests common etiology (Chan et al. 2012). HPV DNA is detectable in 33% of all penile cancers being HPV16 the most prevalent, representing 62.9% of all penile carcinomas worldwide followed by HPV6 (3.6%) and 35 (2.7%) (Alemany et al. 2016)

1.6 HPV genetic variability and its association with anatomical location and pathological outcome of the infection

HPVs present intra-type polymorphisms heterogeneously distributed along the DNA sequence (Chen et al. 2009, 2011). As HPV evolutionary rate is slow, estimated to be 10^{-8} base substitutions per site per year (Chen et al. 2009), other factors such co-divergence of HPV with separate but closely related ancestral Hominin populations with

subsequent host-switch events (Pimenoff et al. 2016) have been suggested to be fundamental factors for nucleotide shifts among HPV genomes and thus, for HPV variants rise (i.e HPV16 variants)(Burk et al. 2013).

The extent of HPV genome variability, the interaction among positive and negative selection pressures, the genetic drift and the biological outcome of such genetic variation are essential to understand the evolution of infection, disease, malignization and carcinogenesis in order to improve, for example, preventive strategies such as triage or prophylactic vaccination. Nonetheless, HPVs genotypic determinants and its association with anatomical location or lesion, are currently, largely unclear (Bernard, Calleja-Macias, and Dunn 2006). Historically, the interest for HPV variants has been twofold: on one hand, the knowledge on the mutation-selection balance for the evolution of HPVs, still scarce, and the knowledge on the forces that have differentially driven the expansion and success of certain HPV lineages in terms of different variants (Gottschling et al. 2011); on the other hand, variants within the same type may show different biological features, which may appear for instance as differences in tropism (Danielewski et al. 2013).

1.6.1 HPV6 and HPV11 variants and association with pathological outcome

To date, a limited number of studies have investigated the differential association of HPVs variants in pathologies such GWs and RRP (Chansaenroj et al. 2012; Combrinck et al. 2012; Flores-Díaz et al. 2017a, 2017b; Gabbott et al. 1997; Gáll et al. 2011; Heinzel et al. 1995; Jelen et al. 2014; Kocjan et al. 2008, 2009, 2011; Maver et al. 2011). Only few studies explore the association of HPV6 and/or HPV11 linages throughout these pathologies. (Combrinck et al. 2012; Flores-Díaz et al. 2017a, 2017b; Jelen et al. 2014; Mounts and Kashima 1984; Mounts, Shah, and Kashima 1982). Differential virulence of lineages within HPV6 has been suggested as a reason for variations in the effectiveness of RRP therapies (Combrinck et al. 2012; Mounts and Kashima 1984; Mounts et al. 1982). A similar interpretation has been proposed for the enhanced prevalence of HPV6_B1 variants in anogenital lesions (Danielewski et al. 2013)(Komloš et al. 2013), and a recent and unique case-control study has described, for the first time, that infections with HPV6_B1 are associated with an increased risk of GW development (Flores-Díaz et al. 2017b). In anogenital epithelial samples, HPV11_A2 are the most common isolates (Danielewski et al. 2013). A case-control study described that these variants are the most frequently detected in GWs and its preceding genital swab, but no particular association between HPV11 variants and an increased risk for GW development was detected (Flores-Díaz et al. 2017a).

The overall findings lead us to question why HPV6 and HPV11 variants seem to have differential distribution according to the infection outcome and whether may either show *a priori* different potential to induce lesions in different epithelia.

1.6.2 HPV16 variants and association with anogenital cancer

A vast amount of research studies analyzing differential pathogenesis for distinct HPV16 variants have been developed, mainly in cervical lesions (Schiffman et al. 2010; Zuna et al. 2011). Published data describe that HPV16 non-A1 variants, present an enhanced risk for persistence and progression to cervical cancer (Burk et al. 2003; Schiffman et al. 2010; Zuna et al. 2009). Berumen and colleagues, show an increased oncogenic potencial of HPV16_D variants compared to HPV16_A1 for cervical cancer (Berumen et al. 2001). However, there is not an strict consensus and other authors point towards the opposite, reporting a less aggressive behavior of HPV16 non-A1 variants and an increased risk of death in women with HPV16_A1 positive cervical cancer (Zuna et al. 2011). In vulvar squamous cell carcinoma, it has been reported the predominance of HPV16_A1 although in small geographically localized cohorts (Larsson et al. 2012). Sarah E. Tan and colleagues describe the total predominance of HPV16_A1-3 in IVuC in remote Indigenous communities in Arnhem Land (Tan et al. 2013). Koning and colleagues (2008) described similar results in small sample set of North American population (de Koning et al. 2008). In vaginal tumors, Larsson and colleagues reported a similar prevalence of the HPV16_A (Larsson et al. 2013). In contrast to other anatomical sites, in penile cancers, Tornesello and coworkers found an increased frequency of HPV16 non_A1, especially HPV16_D variants, in small Italian population (Tornesello et al. 2008) but oppositely, Lopez-Romero and coworkers found an important contribution of HPV16_A1 in a Central-South American cohort (85%). (López-Romero et al. 2013). In anal carcinoma, Ouhoummane and colleagues describe a high prevalence of A1-3 (Called European)(Ouhoummane et al. 2013).

Scarce literature has focused on HPV16 variant distribution according to the cancer histology Certain authors propose that HPV16 genomic variation might partially shape the pathogenesis of cervical cancer histological outcome (Mirabello et al. 2016). However, there are many clues on the mechanistic factors. Some published studies have evidenced an association between HPV16_D with adenocarcinoma (Berumen et al. 2001; Burk et al. 2003; Quint et al. 2010; Rabelo-Santos et al. 2006). Nonetheless, this studies are based on small sample size and present some result discrepancies(De Boer et al. 2005) (Lizano 2006). A recent work published by Mirabello and coworkers with huge numbers, present a

significant association between cervical carcinomas of glandular histological nature (ADC) and HPV16_D2 and D3 and A4, providing therefore, certain evidence of an existing linkage among HPV16 variants and histological presentation(Mirabello et al. 2016). In contrast to Mirabello and collegues, Chopjitt and colleagues had previously reported an association between HPV16_A4 subvariant with an increased risk of squamous histology in precancerous lesions and in cancer outcome (Chopjitt et al. 2009).

1.7 HPV6, HPV11 and HPV16 genetic variability and its geographical association

Differential prevalence of HPV variants has been suggested to correlate with geographic origin and ethnicity (Bernard et al. 2006). Immigrant populations, depending on their respective ethnic origins, have been described to contain particular mixtures of variants (Bernard et al. 2006). For example, the colonization of the Americas by Europeans, Africans and other human populations is reflected in the composition of their HPV16 variants (Ho et al. 1993). Indeed, at shallower evolutionary timescale, it has been proposed that HPV16 has co-diverged with modern human populations (Bernard 1994) and some authors claim for a partial HPV16 phylogeographical match (Bernard et al. 2006; Cornet et al. 2013; Jelen et al. 2014; Yamada et al. 1997). According to this scenario, nomenclatures based on geographical origin have been proposed (Swan et al. 2006). However, this geographical match is not shared for other HPVs: for example, no signature for virus-host co-evolution based on phylogeography has been detected for HPV6 nor HPV11, two of the most frequently associated HPVs with benign lesions in humans (Flores-Díaz et al. 2017b; Heinzel et al. 1995; Jelen et al. 2014; Kocjan et al. 2011, 2008; Matos et al. 2013). Additionally local studies have reported that the total repertory of HPV6 variants can be found in samples coming from a single country (Kocjan et al., 2009). Contrary, for HPV16, the partial phylogeographical match is largely based on the virtual absence of HPV16_B and C out of Sub-Saharan Africa, on the enrichment of HPV16_D in the Americas and on HPV16_A4 in Asia(Cornet et al. 2012; Pimenoff et al. 2016; Tornesello et al. 2004). (Figure 12) Overall, the existing literature suggests that more complex scenarios are required to understand the origin and spread of the PVs at different evolutionary scales (Gottschling et al. 2011).

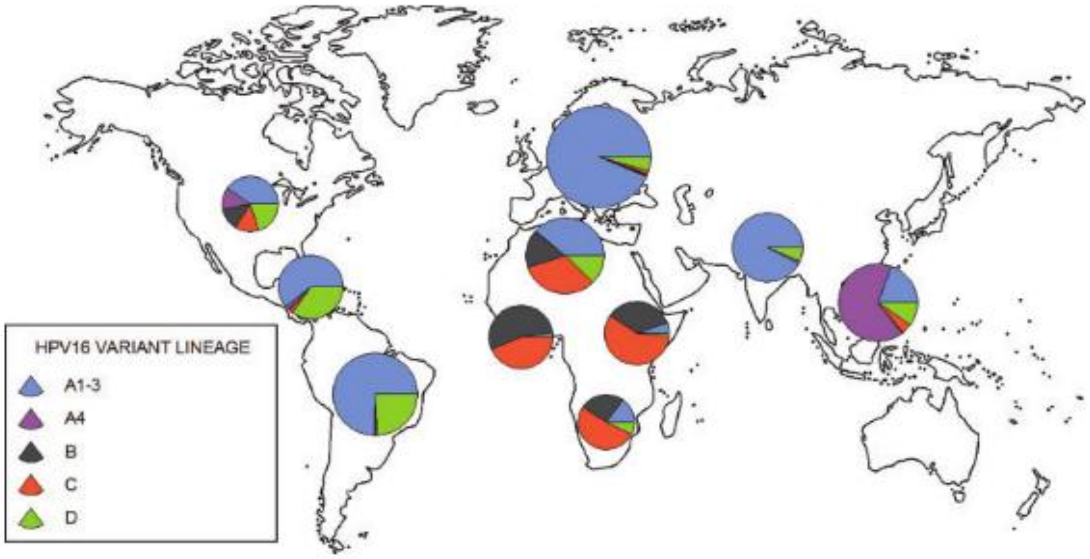


Figure 12: Phylogeographic distribution of the 1,680 HPV16 sequences encompassing the LCR, *E6* and *L2* genome loci. Each sequence was assigned to a specific HPV16 variant lineage (see the color coding for A1–3, A4, B, C, and D variants). The size of the pie charts is proportional to the number of sequences from the corresponding geographic region (Pimenoff et al. 2016)

1.8 T350G polymorphism in HPV16

A large body of experimental research on the differential biological activities of HPV16 variants has focused on the *E6* gene, especially on the T350G polymorphism, corresponding to the L83V amino acid substitution within the *E6* oncoprotein (Togtema et al. 2015; Zehbe et al. 2009). Initially, the literature described the T350 allele as the "prototype", found in the "European" HPV16 variant, and the G350 allele as the "non-prototype", found in "non-European" HPV16 variants (Zehbe et al. 1998). However, posterior data found both alleles in different HPV16 variant lineages, proposing T350G polymorphism not to be an HPV16 variant-specific marker (Cornet et al. 2012). Thus, the still common practice of analysing individual *E6* polymorphisms instead of more informative haplotypes, and to automatically assign them to a particular variant, may also be misleading. The importance of studying the T350G *E6* polymorphism stems from the fact that it has been associated with an increased infection persistence, (Togtema et al. 2015) increased risk of progression and enhanced oncogenic potential in cervical cancer (Grodzki et al. 2006). Melisa Togtema and coworkers, observed that the 350T is more prevalent in low-grade lesions than 350G, which is more present in high-grade lesions, and confers a 2-fold higher risk for both viral persistence and progression to high grade lesions (Grodzki et al. 2006). Case-control studies analysing the polymorphism distribution at geographical level reported the T350G polymorphism to be associated with

an increased risk of developing cervical cancer in Central South America but not in Europe or in Asia (Cornet et al. 2013).

Published data showed that T350G HPV16_A1-3 demonstrate a better ability to enhance human telomerase reverse transcriptase (*hTERT*) immortalized cells and better capacity to maintain it at higher levels than other variants such as HPV16_D2 or D3 (Togtema et al. 2015). In the literature, multiple mechanisms have been implicated in the ability of *E6* to increase immortalized cells (McMurray and McCance 2004). For example, it has been suggested that *E6* co-localizes with *c-Myc* at *Ebox* elements of the *hTERT* promoter (Veldman et al. 2003). In addition, it has also been shown that the *E6*/*E6AP* complex targets repressors of the *hTERT* promoter, for ubiquitin-mediated proteasomal degradation (Gewin et al. 2004). Therefore, it is possible that *E6* variants may differentially exploit these mechanisms to modulate immortalized cells expression (McMurray and McCance 2004). Despite the large body of experimental research based on the differential biological activities of HPV16 variants and focused especially on the *E6* T350G SNP, the necessity of investigating more deeply this particular polymorphism is warranted.

1.9 HPV16-associated invasive cancers and age at tumour diagnosis

Cancer registry data show that cervical cancers are diagnosed earlier than other anogenital cancers associated with HPVs (https://www.cdc.gov/cancer/hpv/statistics/age.htm) (Alemany et al. 2015, 2016; L. Alemany et al. 2014; de Sanjosé et al. 2013) (Figure 13). Furthermore, at global level, cancers associated to highly oncogenic HPVs such HPV16, HPV18 or HPV45 are diagnosed earlier than cancers produced by other HPVs (de Sanjose et al. 2010).

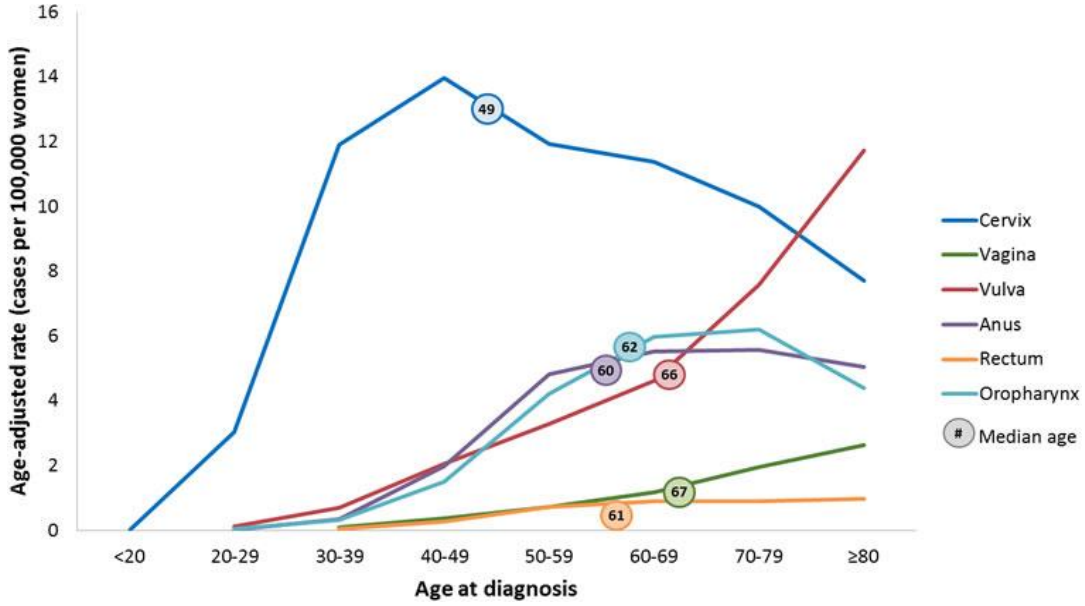


Figure 13: Rates by age group for HPV-associated cancers in the United States during 2008–2012. The rates shown are the number of women in each age group diagnosed with HPV-associated cancer for every 100,000 women.The chart also shows that the median age at diagnosis (the age at which half were older and half were younger), is 49 years for HPV-associated cervical cancer, 67 for HPV-associated vaginal cancer, 66 for HPV-associated vulvar cancer, 60 among women for HPV-associated anal cancer, 61 among women for HPV-associated rectal cancer, and 62 among women for HPV-associated oropharyngeal cancers. (https://www.cdc.gov/cancer/hpv/statistics/age.htm).

However, uneven HPV contribution in different anogenital cancers could be a confounding factor when describing age at tumour diagnosis patterns only for HPV16 type(Alemany et al. 2015, 2016; de Sanjose et al. 2010). For example, HPV16 contribution is higher in non-cervical cancers than in cervical ones, except in vagina (Table 7)

Invasive anogenital cancer	HPV+ cases (%)	HPV16 + cases (%)
Cervix (ICC)	80-90[1]	60-65[1]
Vulva (IVuC)	30-40[2]	70-75[2]
Vagina (IVaC)	70-80[3]	55-60[3]
Penis (IPeC)	30-40[4]	60-65[4]
Anus (IAnC)	85-90[5]	75-80[5]

Table 7: HPV positive cases and HPV16 positive cases for invasive anogenital cancers of cervix, vulva, vagina, penis and anus. [1](de Sanjose et al. 2010); [2](de Sanjosé et al. 2013);[3](Alemany et al. 2014, 2015); [4](Alemany et al. 2016);[5](Alemany et al. 2015)

A few studies tackle the question of whether different HPV16 variants are associated with early or delayed presentation of ICC (Alfaro et al. 2016; Berumen et al. 2001). Ana Alfaro and colleagues show that the median age of patients positive for HPV16_D2 is around the early forties, whereas patients positive for HPV16_A1/2 or D3 are around their early fifties and middle fifties, respectively (Alfaro et al. 2016) (Figure 14).

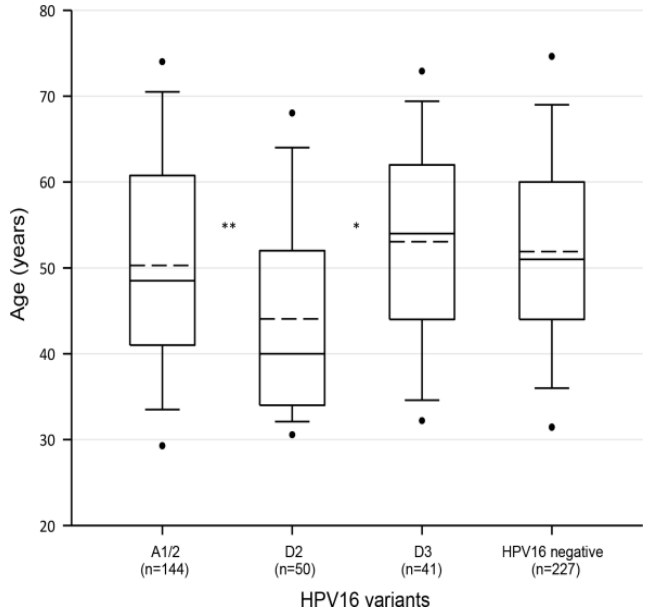
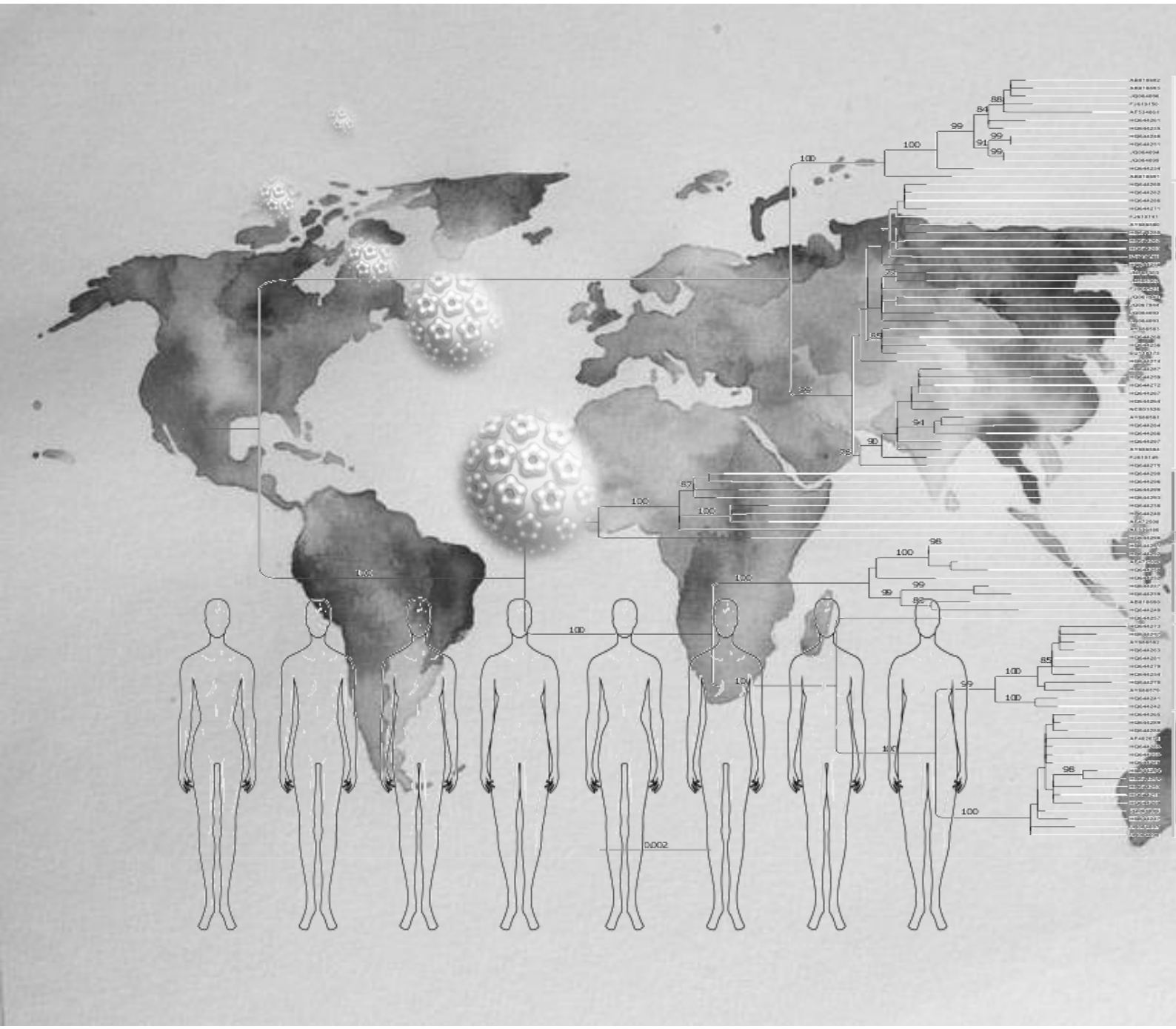


Figure 14: Age distribution of CC patients classified by HPV16 variants. Box plots show the age distribution of patients classified by HPV16 variant. The upper and lower boundaries of the boxes represent the 75th and 25th percentiles, respectively. The black and dotted lines within the boxes represent the median and mean values, respectively, and the whiskers represent the minimum and maximum values that lie within 1.5× the interquartile range from the end of the box. Values outside this range are represented by black circles. The statistical significance for the differences in the median age between the D2 group and the other groups is determined by the Mann-Whitney U Test. The box labeled as HPV16 negative includes samples positive for HPVs other than HPV16 and HPV-negative samples (Alfaro et al. 2016).

Furthermore, within cervical cancer and the different histological presentations, the literature suggests that cervical ADCs are diagnosed in younger women than cervical SCCs (Pérez, Cid, Iñarrea, Pato, Lamas, Couso, Gil, Alvarez, et al. 2014; dos Reis et al. 2007; Vinh-Hung et al. 2007). However, other large studies do not find differences in age between glandular and squamous ICCs (Mirabello et al. 2016), In this latter case, however, the analyses were performed using "age at enrollment" instead of "age at tumor diagnosis" as focal variable, which may lead to changes in the patterns, given the long time period between viral infection and cancer development (Mirabello et al. 2016). At variant level, there are also some trends in age at cancer diagnosis according to cervical histological presentations. Alfaro and colleagues show that patients with ADC associated to HPV16_D2

present slightly younger ages at dignosis (≤ 49 years) than those with SCC (≥ 50 years) (Alfaro et al. 2016). Additional research on age at cancer diagnosis and the association with other factors such tissue-specific characteristics (Crum 2000) genetics or patient lifestyle factors (Alfaro et al. 2016) are warranted to understand better age at cancer diagnosis patterns.

2. HYPOTHESES AND OBJECTIVES



2.1 Hypotheses

2.1.1 Manuscript 1

- HPV6 variants may be differentially associated with GWs or RRP lesions.
- HPV11 variants may be differentially associated with GWs or RRP lesions.

2.1.2 Manuscript 2

- HPV16 variants may be differently associated with invasive tumors of cervix, vulva, vagina, penis and anus
- HPV16 variants may display different geographical distribution in anogenital cancers
- HPV16 variants may present specific association with T350G polymorphism, presumably related with an increased oncogenic potential.
- HPV16 ICC might be diagnosed earlier (younger patients) than other anogenital invasive tumors (i.e. vulva, vagina, penis and anus).

2.1.3 Manuscript 3

- HPV16 variants may be differently associated with cancers with distinct histological presentation of invasive cervical cancers: squamous cell carcinoma (SCC), adenosquamous cell carcinoma (ADSC) and adenocarcinoma (ADC)
- HPV16 variants may display distinct geographical distribution according to the isolate origin
- HPV16 glandular invasive cervical cancers (ADSC and ADC) may be diagnosed earlier (younger patients) than invasive cervical cancers with squamous nature (SCC)

2.2 Objectives

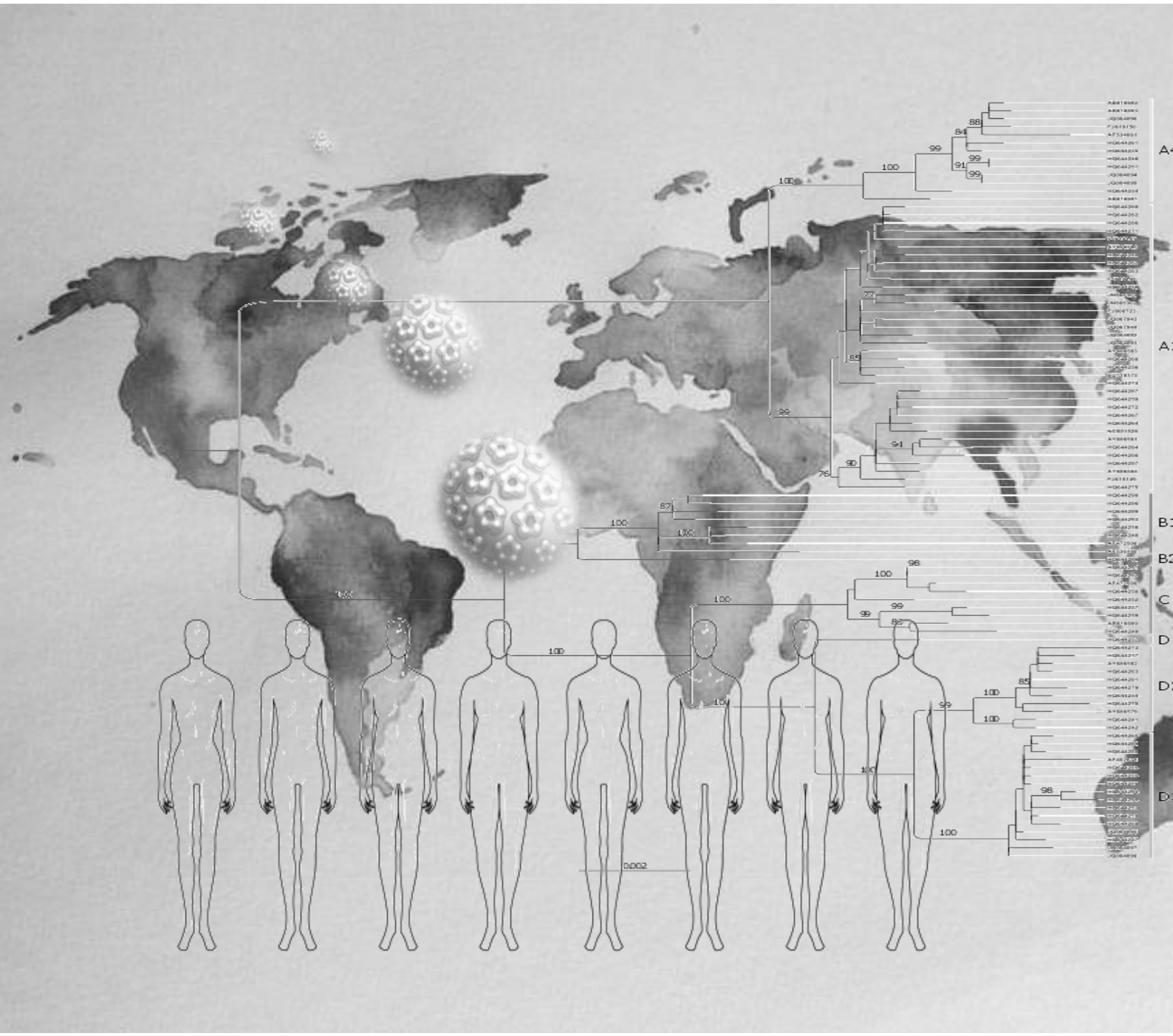
2.2.1 At an operative level in manuscript 1, manuscript 2 and manuscript 3

- To identify the most informative regions at nucleotide level for HPV6 , HPV11 and HPV16 genomes and to design PCR primers to perform viral variant identification based on Sanger sequencing.

2.2.2 At research level in manuscript 1, manuscript 2 and manuscript 3

- To describe the HPV6 and HPV11 lineages present in a large worldwide GWs and RRP sample set and to analyse the differential prevalence of the viral lineages identified.
- To describe HPV16 lineages present in a large worldwide sample set in cervical, vulvar, vaginal, penile and anal invasive cancers and to analyse the differential prevalence of the viral lineages identified according to anatomical location, geographical origin and histological presentation of the cancer.
- To describe patterns on age at tumour diagnosis depending on the anogenital cancer type and histological presentation
- To describe the allelic frequencies of the T350G polymorphism according to anatomical location and geographical origin of the cancer

3. MANUSCRIPTS



Phylogenetically related, clinically different: human papillomaviruses 6 and 11 variants distribution in genital warts and in laryngeal papillomatosis

J. M. Godınez 1,2,†, S. Nicolas-Parraga 1,2,†, V. N. Pimenoff 1,2, B. Mengual-Chulia 1,2, N. Muñoz 3, F. X.Bosch 1,2, G. I.Sanchez 4, J. McCloskey 5 and I. G. Bravo 1,2

1) Infections and Cancer Laboratory, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), 2) Bellvitge Institute of Biomedical Research
(IDIBELL), Barcelona, Spain, 3) Cancer Institute of Colombia, Bogota, Colombia, 4) Infection and Cancer Group, School of Medicine, University of Antioquia,
Medellin, Colombia and 5) Sexual Health Clinic, Royal Perth Hospital, School of Laboratory Medicine, University of WA, Perth, Australia

Abstract

Genital warts (GWs) and laryngeal papillomatosis (LP) are two usually benign pathologies related to infection with human papillomaviruses (HPVs), mainly HPV6 and HPV11. The aim of this work was to describe the genetic diversity of HPV6 and HPV11 isolates found in GWs and LPs, and to analyse the differential involvement of viral variants in either lesion. A total of 231 samples diagnosed as GWs (n = 198) or LP (n = 33) and caused by HPV6 or HPV11 monoinfections were analysed. The phylogenetic relationships of the retrieved viral sequences were explored. We have identified the long control region and the intergenic E2–L2 region as the two most variable regions in both HPV6 and HPV11 genomes. We have generated new HPV6 (n = 166) or HPV11 (n = 65) partial sequences from GWs and LPs lesions spanning both regions and studied them in the context of all available sequences of both types (final n = 412). Our results show a significant ($p < 0.01$) differential presence of HPV6 variants among both pathologies, with HPV6 B variants being preferentially found in GW versus LP samples. No differential involvement of HPV11 variants was observed. Our findings suggest that different HPV6 variants may either show differential tropism or have different potential to induce lesions in different epithelia.

Keywords: Genital warts, human papillomaviruses, laryngeal papillomatosis, phylogeny, recurrent respiratory papillomatosis, Tissue tropism, variants
Original Submission: 25 June 2013; **Revised Submission:** 2 September 2013; **Accepted:** 30 September 2013
Editor: G. Antonelli
Article published online: 9 December 2013
Clin Microbiol Infect 2014; **20**: O406–O413
doi: 10.1111/1469-0691.12420

Corresponding author: J. M. Godınez, Infections and Cancer Laboratory, Cancer Epidemiology Research Programme, Institut Catala d' Oncologia, L'Hospitalet de Llobregat, ICO-IDIBELL, Gran Via 199-203, 08907 Barcelona, Spain
E-mail: jmgodinez@iconcologia.net
† Both authors contributed equally to this work.

Introduction

Papillomaviruses are small, non-enveloped viruses with a circular double-stranded DNA genome of around 8000 bp [1]. More than 250 complete papillomavirus genomes have been described, infecting human and non-human hosts (http://pave.niaid.nih.gov/#home). Human papillomaviruses (HPVs) are the causative agents of cancer of the cervix, and are also involved in cancers of the penis, anus, vagina, vulva, and head and neck, as well as in other benign, wart-like lesions [2]. Based on this association to cervical cancer, HPVs have been epidemiologically stratified into three risk groups: carcinogenic, probably and possibly carcinogenic, and not carcinogenic to humans [3]. Alphapapillomaviruses HPV6 and HPV11 belong to the non-carcinogenic group, being the most common non-oncogenic HPVs found in cervical specimens in the general population [4].

HPV6 and HPV11 are the causative agents in some conspicuous lesions, namely anogenital warts (GWs) and

laryngeal papillomatosis (LP). GWs are benign tumours of the epithelium caused by papillomavirus infection, mainly with HPV6 and HPV11 (85% of the cases) [5]. Co-infections by oncogenic and non-oncogenic types are commonly detected in a high proportion of anogenital warts (45%), which have been proposed as a partial explanation of the increased risk of cervical intraepithelial neoplasia and invasive cervical carcinoma in women with GWs [6]. GWs are closely associated with sexual behaviour, with number of sexual partners being the main risk factor [7]. The highest incidence rate for GWs in women is at 20-24 years, which correlates well with the peak of papillomavirus infection in the female genital tract [4]. In men, the incidence peak occurs at 20-29 years of age [2].

Laryngeal papillomatosis, or recurrent respiratory papillomatosis, is a neoplastic disease of the airways mainly caused by HPV6 and HPV11, although HPV16 has also been identified in a few cases [8]. It represents the most common benign tumour of the larynx in infants and children [9]. Some studies have identified infection with HPV11 as being associated with more aggressive disease and higher recurrence of lesions [8,10], and malignant transformation of lesions has been described in approximately 5% of cases [11]. The clinical complications of this pathology include dysphonia, dyspnoea and, in serious cases, complete obstruction of the airways [12].

Papillomavirus variants are defined as viral sequences sharing >98% identity in the nucleotide sequence in the L1 gene [13]. Based on this criterion, HPV6 and HPV11 variant lineages have been described [14]. Several studies have addressed the genetic diversity of HPV6 and HPV11 [15-17], and some of them have aimed to establish a link between genetic variation and differential outcome of the infection [8,18].

The aim of this study was to analyse first the genetic diversity of HPV6 and HPV11 sequences retrieved from two different but related pathologies, namely GWs and LP. Further, the phylogenetic relationship of all HPV6 and HPV11 sequences and tissue-dependent distribution of the variants were analysed.

Methods

Samples

Samples analysed in this project originate from two different formalin-fixed paraffin-embedded (FFPE) sample repositories. GWs were obtained from the Surgical Genital Wart Biobank established in 1995 at the Sexual Health Clinic at Royal Perth Hospital, Perth, Australia. These samples include FFPE surgery specimens excised from patients who required surgical resection of anal and/or perianal GWs [19]. One hundred and forty-three HPV6 and sixty-four HPV11 single infected samples from the first surgical event of each patient were included.

Laryngeal papillomatosis samples originated from a multicentre study of cases diagnosed between 1985 and 2009, in the cities of Cali and Medellin, Colombia [10]. Forty-one HPV6 and eleven HPV11 single-infected samples, each from a different patient, were included. Detailed information about the samples included is shown in the Supplementary material, Table S1.

Presence of HPV DNA in the samples was assessed by using the SPF$_{10}$-DEiA-LiPA protocol (version 1; Laboratory Biomedical Products, Rijswijk, the Netherlands). The SPF$_{10}$ system targets a 65-base pair region of L1 gene of a broad spectrum of Alphapapillomaviruses. HPV-positive samples were identified and genotyped by amplicon hybridization (DEiA) and reverse hybridization line probe assay, LiPA25. The detected viruses were HPV6, 11, 16, 18, 31, 33, 34, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 58, 59, 56, 66, 68, 70 and 74.

Selection of the most informative genomic regions

Fragmentation of genetic material in the FFPE samples prevented us from obtaining full-genome sequences. To select the most informative regions for the study, the variability of different regions of the viral genomes was assessed.

HPV6 and HPV11 complete genome unique sequences were obtained via GenBank. The different open reading frames (ORFs: E6, E7, E1, E2, L1 and L2), the long control region (LCR), and the intergenic E2-L2 region (IntE2L2) were extracted and aligned. This intergenic region spans the E5a and E5b ORFs of HPV6 and HPV11 [20]. All sequences were aligned at amino acid level (except the non-coding LCR), back-translated and concatenated to obtain full-genome reference alignments. For each of the alignments, phylogenetic relationships were inferred under a maximum likelihood framework using RAxML v7.2.8 (http://www.exelixis-lab.org/) [21], using the GTR+Γ4 model, and the number of required bootstrap cycles was determined with the -autoMRE command [22]. The well-resolved phylogenetic trees obtained were further employed to compute tree-guided, model-based pairwise genetic distances between taxons (f x command in RAxML).

PCR and sequencing

DNA was extracted by incubation of the material with 250 μL of proteinase K buffer (10 mg/mL proteinase K in 50 mM Tris-HCl, pH 8.0) overnight at 56°C. The samples were later incubated at 95°C for 8 min to inactivate proteinase K and were stored at −20°C until use.

Based on the pairwise distance results, the LCR and the IntE2L2 were chosen as amplification targets. Different type-specific PCR systems were designed for the amplification

of the samples. One primer set per region per genome was initially designed. For HPV6 samples not amplifiable by the primer set because of amplicon length, targeted regions were obtained by amplification of overlapping fragments. Primer sequences and amplified regions are shown in the Supplementary material, Table S2.

PCR products were sequenced at the Genoscreen facilities (Lille, France) in both strands.

Phylogenetic analyses

We have applied an Evolutionary Placement Algorithm (EPA) [23] to the inference of the phylogenetic relationships of the short fragments generated, in the context of the whole HPV6 or HPV11 variability. This methodology had been successfully applied for the analyses of short papillomavirus DNA sequences [24]. The reference tree described above, inferred using the genomic information of all full-length HPV6 or HPV11 variants, was used as scaffold.

The final set of reference sequences contained: for HPV6, 38 sequences, 8047 nucleotides and 172 alignment patterns; for HPV11, 26 sequences, 7878 nucleotides and 77 alignment patterns. These sets included sequences obtained from different pathologies (GWs, LP, cervical and lung samples) in different regions (Slovenia, Sweden and Thailand). Detailed information on the sequences is shown in Table S1.

Sequences obtained from our samples, and those partial sequences retrieved from GenBank were included and aligned with the reference sequences. Genome alignments were chopped to the length of the larger partial sequence in the alignments. Final alignment included 253 sequences, 1432 nucleotides and 343 alignments patterns for HPV6; 159 sequences, 1452 nucleotides and 157 alignment patterns for HPV11. The EPA algorithm was performed as implemented in RAxML v7.2.6, using the GTR+Γ4 model.

The results of the variants distribution in different pathologies were compared by means of the Pearson's chi-squared test for count data as implemented in R.

Results

LCR and IntE2L2 are the most informative regions

We studied the distribution of pairwise genetic distances for the different genomic regions among full-length genomes of HPV6 and HPV11. The most variable regions in both HPVs were IntE2L2, E2 and LCR; and E6 for HPV11 only (see Supplementary material, Table S3). The accumulated frequencies of the pairwise distances for each ORF and the values of the 95th centile are depicted in Fig. 1. The study of E6 was discarded because we chose to use the same regions for the study of both HPV6 and HPV11, and the E6 gene was not informative for HPV6 variability [25]. E2 was discarded because we aimed to maximize the number of sequences from other studies for our combined analyses, and the only E2 sequences available in the GenBank were those of the full-length genomes. Hence, the most informative regions, IntE2L2 and LCR, were chosen for further analyses.

HPV6 variants analysis

One hundred and sixty-eight GW and LP samples were successfully amplified and sequenced. All newly generated sequences fitted into the previously described clades [14]

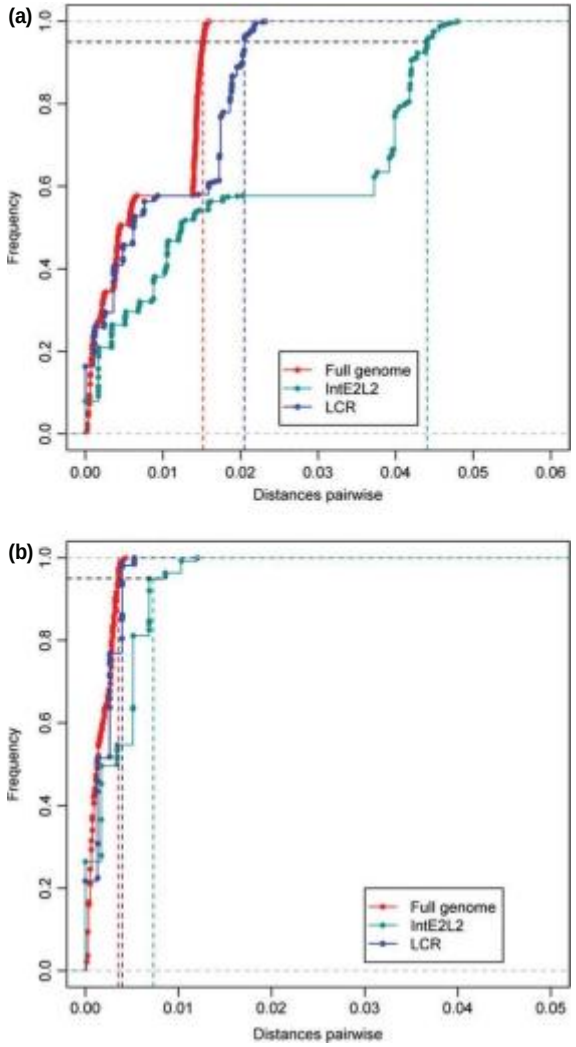


FIG. 1. (a) Human papillomavirus strain 6 (HPV6) pairwise distances calculated for full-genome, LCR and IntE2L2 of reference sequences; (b) HPV11 pairwise distances calculated for full-genome, LCR and IntE2L2 of reference sequences. Note that both plots are represented at the same scale. Dashed lines represent the distances pairwise values for 95th centile.

(Fig. 2). Likelihood weights for the ascription of each individual sequence to each clade/subclade are shown in the Supplementary material, Table S4. Detailed information on the distribution of the sequences in the different clades is presented in Table 1.

Ten partial sequences were excluded for the analyses of the global dataset because no information regarding anatomical site of the lesion was available. Finally, the analysis conducted with 243 sequences showed that 32 out of 85 LP samples belonged to clade A (37.7%), and the remaining sequences belonged to clade B (n = 53, 62.3%) [B1: 30 (35.3%); B2: 6 (7.1%); B3: 17 (20.0%)]. Among all HPV sequences identified in GWs, ten (6.3%) belonged to clade A, and 148 (93.7%) to clade B [B1: 108 (68.4%); B2: 34 (21.5%); B3: 6 (3.8%)] (Table 1). Fig. 2(b) displays the generated tree, including the whole set of sequences.

The differential presence of the different variants in GWs and in LP was further analysed. In both GWs and LP HPV6 B variants were the most common, 93.6% and 62.4%, respectively.

TABLE 1. Distribution of the analysed human papillomavirus 6 (HPV6) sequences into phylogenic clades

	GWs		LP	
Clade	**No. samples**	**%**	**No. samples**	**%**
A	10 (7)	6.33	32 (14)	37.65
B	148 (135)	93.67	53 (10)	62.35
B1	108 (97)	68.35	30 (·)	35.30
B2	34 (33)	21.52	6 (1)	7.05
B3	6 (5)	3.80	17 (9)	20.00
Total	158 (142)	100.00	85 (24)	100.00

The comparison of the distribution of HPV6 variants among both pathologies, genital warts (GWs) and laryngeal papillomatosis (LP), shows a statistically significant difference (chi-squared test p <0.01).
Numbers in brackets correspond to new sequences generated from the samples collections described.

However, a significant difference (chi-squared test; p <0.01) was observed between the distribution of variants between the two types of lesions. While in GWs almost two-thirds of sequences belonged to subclade B1, in LP we found an increased contribution of A (38%) and B3 variants (20%).

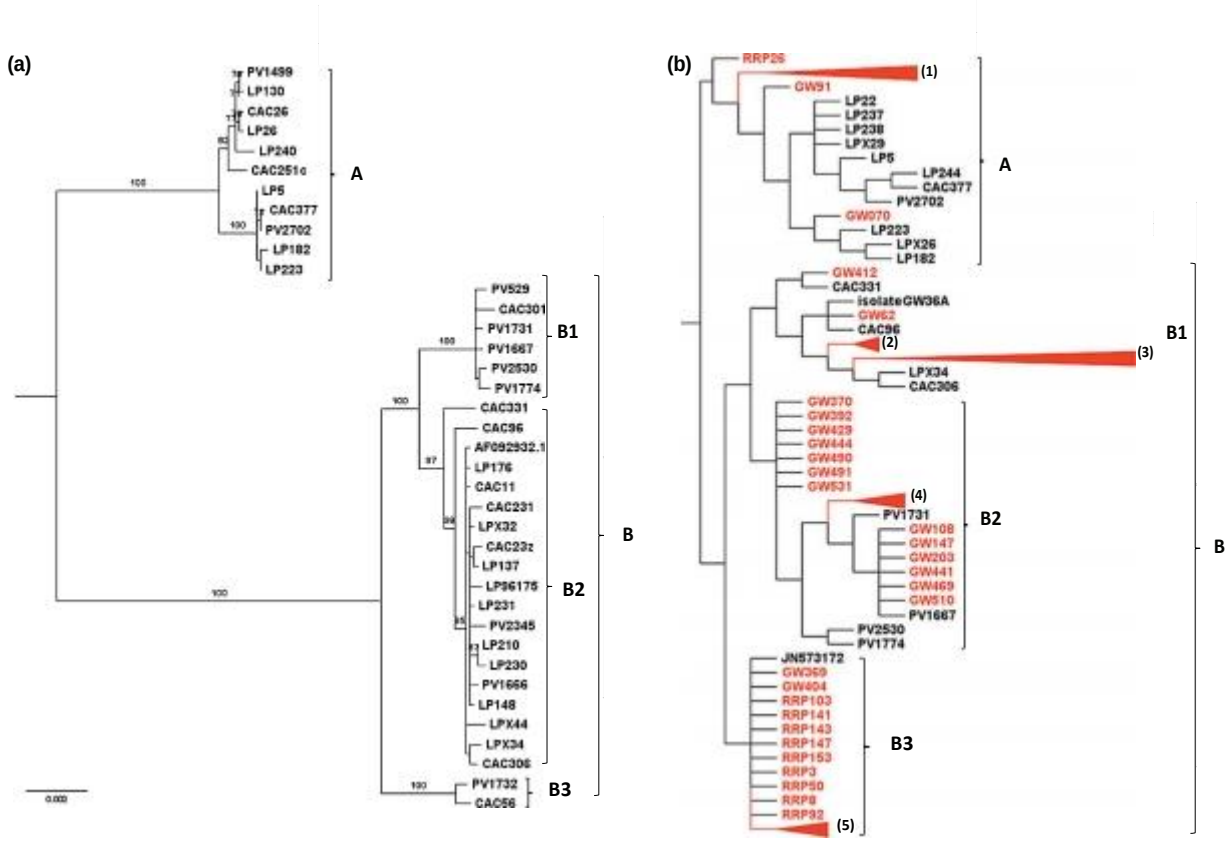


FIG. 2. (a) Midpoint rooted best-known maximum likelihood (ML) phylogenetic tree of human papillomavirus strain 6 (HPV6) isolates using 38 unique full-length genome sequences retrieved from GenBank. HPV6 variants are classified into two clades, A and B. Only bootstrap values over 70 supporting each branch of the generated tree are represented; (b) projection of the LCR/IntE2L2 HPV6 sequences analysed in this study onto the scaffold of the best-known full-length ML tree using the Evolutionary Placement Algorithm approach. Some branches have been collapsed for better presentation, the collapsed branches include: (1) 6 GWs/20 LPs/2 Cervix; (2) 32GWs; (3) 66 GWs/45 LP/2 Cervix; (4) 23 GWs/4 LPs; (5) 4 GWs/12 LP. Sequences newly generated in this study are shown in red. An uncollapsed version of the tree is available from the authors under request.

HPV11 variants analysis

Sixty-five samples were successfully amplified and sequenced. All HPV11 sequences fitted into the previously described clades (Fig. 3). Likelihood weights for the assignment of each individual sequence to each clade/subclade are shown in Table S4.

Detailed information on the distribution of the sequences in the different clades is presented in Table 2.

Eleven partial HPV11 sequences retrieved from GenBank were not included in the final analyses because of unknown origin of the material. Fig. 3(b) shows the generated tree, including the complete set of sequences. In both GWs and LP the vast majority of sequences belonged to the HPV11 A2 clade, with no difference between variant distributions depending on the different types of lesion (chi-squared test p 0.493).

TABLE 2. Distribution of the analysed human papillomavirus 11 (HPV11) sequences into phylogenic clades

	GWs		LP	
Clade	No. samples	%	No. samples	%
A1	6 (2)	5.50	3 (1)	7.5
A2	103 (54)	94.50	37 (8)	93.5
Total	109 (56)	100.00	40 (9)	100.00

The comparison of the distribution of HPV11 variants among both pathologies, genital warts (GWs) and laryngeal papillomatosis (LP), shows no statistically significant difference (chi-squared test p 0.493).
Numbers in brackets correspond to new sequences generated from the sample collections described.

Discussion

Both GWs and LP are benign proliferative lesions caused mainly by HPV6 and HPV11 [2]. Both types of lesions present similar clinical features, such as high recurrence and need of long-term treatments, and have been linked to the development, in a small proportion of patients, of malignant neoplasms [11]. Previous studies have addressed differential HPV6 and HPV11 genotype distributions in GWs and LP, but mainly at a national/regional level [18,26,27]. Here we provide the first study analysing HPV6 and HPV11 variant distribution in two different pathologies, with a large number of samples, GWs (n = 198) and LP (n = 33). The combination with all available

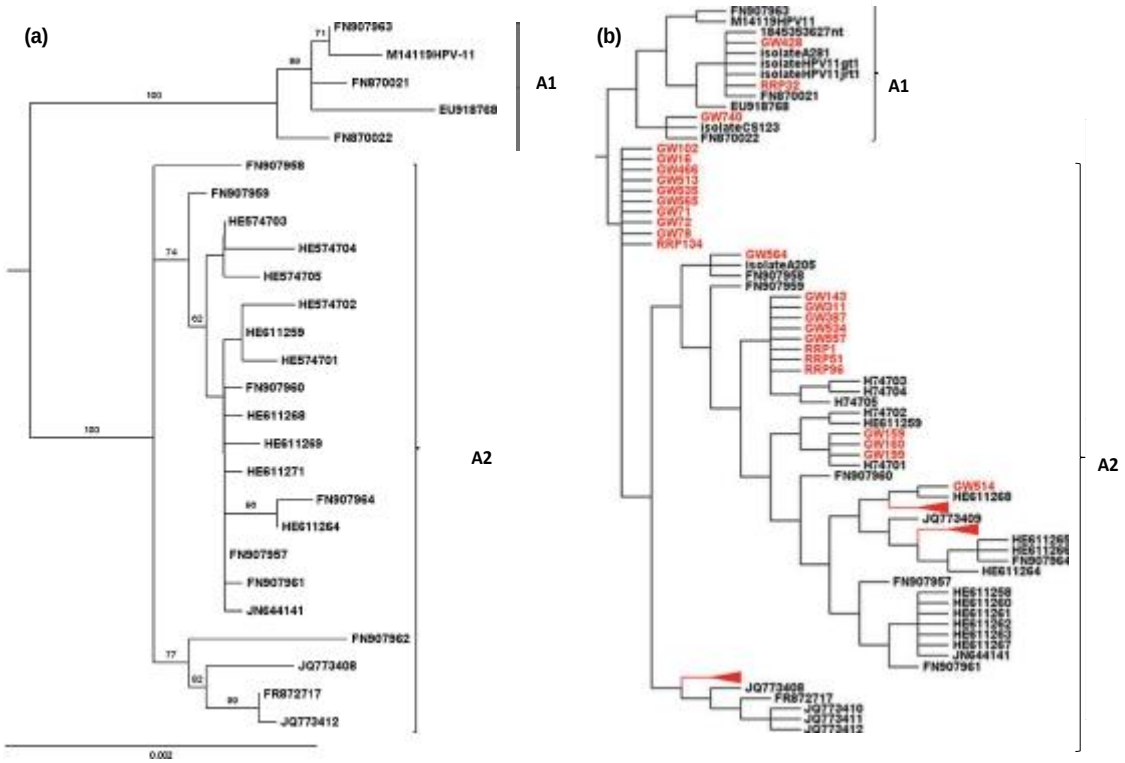


FIG. 3. (a) Midpoint rooted best known maximum likelihood (ML) phylogenetic tree of human papillomavirus strain 11 (HPV11) variants using 26 unique full-length genome sequences retrieved from GenBank. The HPV11 variants are classified into one clade (A) and two sub-clades (A1, A2). Only bootstrap values over 70 supporting each branch of the generated tree are represented; (b) projection of the LCR/IntE2L2 HPV11 sequences analysed in this study onto the scaffold of the best-known full-length ML tree. Some branches have been collapsed for better presentation, the collapsed branches include: (1) 31 GWs/6 LPs; (2) 26 GWs/4 LP/2 Cervix; (3) 18 GWs/1 LPs. Sequences newly generated in this study are shown in red. An uncollapsed version of the tree is available from the authors under request.

sequences in the GenBank generates a final dataset of 253 samples for HPV6, and 159 samples for HVP11 sequences, encompassing ten countries (Australia, China, Colombia, Germany, Hungary, Slovenia, South Africa, Sweden, Thailand, USA) and five continents.

We have identified the most suitable genomic regions for assessing intratype genetic diversity, which for HPV6 and HPV11 are the LCR and the IntE2L2 regions (Table S3). Similar data on the heterogeneous rate of variation throughout the papillomavirus genomes had been previously described. Among coding regions, the E5 genes are the fastest evolving ORFs, and E1 and L1 are the more slowly evolving genes [20,28]. The LCR itself, devoid of the selective pressures for protein encoding, accumulates changes more than twice as fast as the L1 or the E1 genes [28]. Our results confirm therefore that the general trend of variation accumulation is conserved also at shallower levels within Papillomaviridae.

HPV6 and HPV11 are close relatives, and belong together in Alphapapillomaviruses, species 10. Genetic diversity is about four times greater among HPV6 isolates than among HPV11 isolates, as concluded after the analyses of all available full-length genomes for both viruses (Wilcoxon's test, $p < 0.01$) (Fig. 1). These diversity values fit well the described taxonomic definition: for HPV6 variants based on nucleotide similarities in the L1 gene, intravariant differences are around 0.7% and intervariant differences are around 1.5%; for HPV11, no different variants are described and nucleotide differences are below 0.5% [14]. The EPA approach [23] allowed us to assign all partial sequences into the different clades defined using the full-length genome sequences (Fig. 2, Fig. 3). The most important finding of our study is that HPV6 A and B variants are not equally distributed in GWs and LP ($p < 0.01$). Specifically, HPV6 B variant isolates are preferentially found in GWs compared with LP. Furthermore, the contribution of subclades within HPV6 B variants is also different in GWs and in LP ($p < 0.01$). No distribution difference could be observed in our data for HPV11 variants. It could be argued that the observed differences arise from a geographical bias for the origin of the samples analysed. However, previous research did not identify geographical origin as an important component of viral diversity for HPV6 and HPV11. Heinzel and co-workers communicated a global study of these two HPVs [15], including 19 samples containing HPV6 and ten samples containing HPV11. More recently, de Matos et al. presented data on the phylogenetic relations of HPV6 variants using 117 sequences from South America, Europe and South Africa [29], suggesting no evidence of a geographical distribution of HPV variants in these lesions. Further, more local studies have reported that the whole repertoire of HPV6 variants can be found in isolates originating from a single country [25].

It was not the aim of this study to assign a differential risk for certain variants, as this is not a case-control study. The low prevalence of both studied HPVs in a healthy population worldwide (0.5% for HPV6; 0.2% for HPV11) [4] makes it difficult to obtain a sample size with enough statistical power for a meaningful comparative study.

In summary, in this study we demonstrate the differential presence of papillomavirus variants in different pathologies, with variants B1 of HPV6 being more prevalent in GWs than in LP. The current state of knowledge therefore supports our finding of a preferential involvement of HPV6 B1 variants in GWs, which may reflect a biological difference in the interplay between viruses and the different mucosal epithelia. Such differences may arise from either a preferential tropism or a differential viral fitness and potential to induce lesions between anogenital and laryngeal mucosa. Similar scenarios have been proposed for HPV16 variants differentially enriched in vulvar cancer compared with cervical cancer [30]. Further research on the prevalence of the different variants within HPV6 in healthy tissue from both locations will be necessary to solve this question.

Acknowledgements

The authors want to thank Dr Gustavo Cuello, MD, otorhinolaryngologist in Medellın, Colombia, for his help with initial sample collection and data processing. The authors thank Catalina Bonet for her assistance in statistical analysis and Stephanie Bedhomme for critical review of the results and analyses. This work was supported by grants from the Spanish Ministry for Science and Innovation (MICINN Grant CGL2010-16713); Agencia de Gestio d'Ajuts Universitaris i de Recerca (Catalan Government, AGAUR Grant 2009 SGR 1026); Red Tematica de Investigacion Cooperativa en Cancer (RTICC Grant RD06/0020/0095); Spanish Ministry for Economy and Competitivity (PFIS Grant FI12/0142 and FIS Grant PI11/02096); the Lilly Foundation (Premio de Investigacion Biomedica Preclınica 2012 to FXB); and the Institut d'Investigacio Biomedica de Bellvitge - IDIBELL. None of the aforementioned agencies had any role in the interpretation of the results or in the preparation of this manuscript.

Authors' Contribution

JMG conceived and coordinated the study, analysed the data and drafted the manuscript. SN performed experiments and analysed the data. VP performed initial HPV detection. BM analysed the data. GIS and NM obtained and characterized LP

samples. JMc obtained and characterized GW samples. FXB performed epidemiological analysis. IGB designed the analyses and supervised the study. All authors contributed to, read and approved the final manuscript.

Transparency Declaration

NM is member of the HPV Global Board of Merck. FXB is member of the speakers' bureau from GlaxoSmithKline and Sanofi Pasteur MSD.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Descriptive table with the information concerning the analysed sequences with HPV6 and HPV11 infection.

Table S2. Primers for HPV6 and HPV11 amplification and sequencing.

Table S3. 95th centile, 10th centile and Median Absolute Deviation (MAD) values of the pairwise distances for each viral open reading frame.

Table S4. Likelihood weights for assignment of human papillomavirus variants to the different clades.

Data S1. Human papillomavirus (HPV) reference trees generated for the analyses.

Data S2. Sequences not submitted to GenBank, because total length is below 200 bp.

References

1. Doorbar J, Quint W, Banks L et al. The biology and life-cycle of human papillomaviruses. Vaccine 2012; 30 (suppl 5): F55-F70.
2. Forman D, de Martel C, Lacey CJ et al. Global burden of human papillomavirus and related diseases. Vaccine 2012; 30 (suppl 5): F12-F23.
3. Bouvard V, Baan R, Straif K et al. A review of human carcinogens - part B: biological agents. Lancet Oncol 2009; 10: 321-322.
4. Bruni L, Diaz M, Castellsague X, Ferrer E, Bosch FX, de Sanjose S. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. J Infect Dis 2010; 202: 1789-1799.
5. 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: 805-814.
6. Blomberg M, Friis S, Munk C, Bautz A, Kjaer SK. Genital warts and risk of cancer: a Danish study of nearly 50 000 patients with genital warts. J Infect Dis 2012; 205: 1544-1553.
7. Munk C, Svare EI, Poll P, Bock JE, Kjaer SK. History of genital warts in 10,838 women 20 to 29 years of age from the general population. Risk factors and association with Papanicolaou smear history. Sex Transm Dis 1997; 24: 567-572.
8. Mounts P, Kashima H. Association of human papillomavirus subtype and clinical course in respiratory papillomatosis. Laryngoscope 1984; 94: 28-33.
9. Syrjanen S. Current concepts on human papillomavirus infections in children. APMIS 2010; 118: 494-509.
10. Sanchez GI, Jaramillo R, Cuello G et al. Human papillomavirus genotype detection in recurrent respiratory papillomatosis (rrp) in Colombia. Head Neck 2012; 35: 229-234.
11. Hobbs CG, Birchall MA. Human papillomavirus infection in the etiology of laryngeal carcinoma. Curr Opin Otolaryngol Head Neck Surg 2004; 12: 88-92.
12. Wiatrak BJ, Wiatrak DW, Broker TR, Lewis L. Recurrent respiratory papillomatosis: a longitudinal study comparing severity associated with human papilloma viral types 6 and 11 and other risk factors in a large pediatric population. Laryngoscope 2004; 114: 1-23.
13. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology 2004; 324: 17-27.
14. Burk RD, Chen Z, Harari A et al. Classification and nomenclature system for human alphapapillomavirus variants: general features, nucleotide landmarks and assignment of HPV6 and HPV11 isolates to variant lineages. Acta Dermatovenerol Alp Panonica Adriat 2011; 20: 113-123.
15. Heinzel PA, Chan SY, Ho L et al. Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. J Clin Microbiol 1995; 33: 1746-1754.
16. Kocjan BJ, Jelen MM, Maver PJ, Seme K, Poljak M. Pre-vaccination genomic diversity of human papillomavirus genotype 6 (HPV 6): a comparative analysis of 21 full-length genome sequences. Infect Genet Evol 2011; 11: 1805-1810.
17. Kocjan BJ, Seme K, Poljak M. Detection and differentiation of human papillomavirus genotypes HPV-6 and HPV-11 by FRET-based real-time PCR. J Virol Methods 2008; 153: 245-249.
18. Combrinck CE, Seedat RY, Randall C, Roodt Y, Burt FJ. Novel HPV-6 variants of human papillomavirus causing recurrent respiratory papillomatosis in southern Africa. Epidemiol Infect 2012; 140: 1095-1101.
19. McCloskey JC, Metcalf C, French MA, Flexman JP, Burke V, Beilin LJ. The frequency of high-grade intraepithelial neoplasia in anal/perianal warts is higher than previously recognized. Int J STD AIDS 2007; 18: 538-542.
20. Bravo IG, Alonso A. Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. J Virol 2004; 78: 13613-13626.
21. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006; 22: 2688-2690.
22. Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol 2010; 17: 337-354.
23. Berger SA, Stamatakis A. Aligning short reads to reference alignments and trees. Bioinformatics 2011; 27: 2068-2075.
24. Godınez J, Heideman DA, Gheit T et al. Differential presence of papillomavirus variants in cervical cancer: an analysis for HPV33, HPV45 and HPV58. Infect Genet Evol 2013; 13: 96-104.
25. Kocjan BJ, Poljak M, Cimerman M et al. Prevaccination genomic diversity of human papillomavirus genotype 6 (HPV 6). Virology 2009; 391: 274-283.
26. Komlos KF, Kocjan BJ, Kosorok P et al. Tumor-specific and gender-specific pre-vaccination distribution of human papillomavirus types 6 and 11 in anogenital warts and laryngeal papillomas: a study on 574 tissue specimens. J Med Virol 2012; 84: 1233-1241.
27. Donne AJ, Hampson L, Homer JJ, Hampson IN. The role of HPV type in recurrent respiratory papillomatosis. Int J Pediatr Otorhinolaryngol 2010; 74: 7-14.

28. Garcia-Vallve S, Iglesias-Rozas JR, Alonso A, Bravo IG. Different papillomaviruses have different repertoires of transcription factor binding sites: convergence and divergence in the upstream regulatory region. BMC Evol Biol 2006; 6: 20.
29. de Matos RP, Sichero L, Mansur IM et al. Nucleotide and phylogenetic analysis of human papillomavirus types 6 and 11 isolated from recurrent respiratory papillomatosis in Brazil. Infect Genet Evol 2013; 16C: 282-289.
30. Larsson GL, Helenius G, Andersson S, Sorbe B, Karlsson MG. Prognostic impact of human papilloma virus (HPV) genotyping and HPV-16 subtyping in vaginal carcinoma. Gynecol Oncol 2013; 129: 406-411.

SUPPORTING INFORMATION
Table S1. Description of analyzed sequences

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	GW019	Australia	1996	B	B1	LCR	KC820175	GW
6	GW026	Australia	1997	B	B1	LCR	KC820176	GW
6	GW026	Australia	1997	B	B1	IntE2L2	KC908039	GW
6	GW027	Australia	1997	B	B2	LCR	KC820177	GW
6	GW028	Australia	1997	B	B2	LCR	KC820178	GW
6	GW029	Australia	1997	B	B1	LCR	KC820179	GW
6	GW029	Australia	1997	B	B1	IntE2L2	KC908040	GW
6	GW031	Australia	1997	B	B1	LCR	KC820180	GW
6	GW034	Australia	1997	B	B1	LCR	KC820139	GW
6	GW042	Australia	1998	B	B1	LCR	KC820181	GW
6	GW042	Australia	1998	B	B1	IntE2L2	KC908041	GW
6	GW048	Australia	1998	B	B1	LCR	KC820182	GW
6	GW051	Australia	1998	A		LCR	KC820183	GW
6	GW051	Australia	1998	A		IntE2L2	KC908042	GW
6	GW055	Australia	1998	B	B1	LCR	KC820184	GW
6	GW055	Australia	1998	B	B1	IntE2L2	KC908043	GW
6	GW063	Australia	1999	B	B1	LCR	KC820185	GW
6	GW063	Australia	1999	B	B1	IntE2L2	KC908044	GW
6	GW070	Australia	1999	A		LCR	KC820186	GW
6	GW076	Australia	1999	B	B1	LCR	KC820187	GW
6	GW092	Australia	2000	B	B2	LCR	KC820188	GW
6	GW098	Australia	2000	A		LCR	KC820189	GW
6	GW098	Australia	2000	A		IntE2L2	KC908045	GW
6	GW104	Australia	2000	B	B2	LCR	KC820174	GW
6	GW104	Australia	2000	B	B2	IntE2L2	KC908038	GW
6	GW108	Australia	2000	B	B2	LCR	KC820190	GW
6	GW108	Australia	2000	B	B2	IntE2L2	KC908046	GW
6	GW110	Australia	2000	B	B1	LCR	KC820140	GW
6	GW113	Australia	2000	B	B1	LCR	KC820191	GW
6	GW115	Australia	2000	B	B2	LCR	KC820192	GW
6	GW117	Australia	2000	B	B1	LCR	KC820193	GW
6	GW118	Australia	2000	A		LCR	KC820141	GW
6	GW118	Australia	2000	A		IntE2L2	KC908009	GW
6	GW120	Australia	2000	A		IntE2L2	KC908047	GW
6	GW123	Australia	2000	B	B1	IntE2L2	KC908048	GW
6	GW128	Australia	2000	B	B1	LCR	KC820194	GW
6	GW128	Australia	2000	B	B1	IntE2L2	KC908049	GW
6	GW135	Australia	2001	B	B1	LCR	KC820195	GW
6	GW135	Australia	2001	B	B1	IntE2L2	KC908050	GW
6	GW137	Australia	2001	B	B2	LCR	KC820196	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	GW140	Australia	2001	B	B1	LCR	KC820150	GW
6	GW140	Australia	2001	B	B1	IntE2L2	KC908014	GW
6	GW146	Australia	2001	B	B3	LCR	KC820197	GW
6	GW147	Australia	2001	B	B2	LCR	KC820198	GW
6	GW147	Australia	2001	B	B2	IntE2L2	KC908051	GW
6	GW151	Australia	2001	B	B1	LCR	KC820199	GW
6	GW153	Australia	2001	B	B1	LCR	KC820200	GW
6	GW153	Australia	2001	B	B1	IntE2L2	KC908052	GW
6	GW161	Australia	2002	B	B1	LCR	KC820201	GW
6	GW164	Australia	2002	B	B1	LCR	KC820202	GW
6	GW165	Australia	2002	A		LCR	KC820203	GW
6	GW165	Australia	2002	A		IntE2L2	KC908053	GW
6	GW168	Australia	2002	B	B2	IntE2L2	KC908054	GW
6	GW172	Australia	2002	B	B1	LCR	KC820204	GW
6	GW172	Australia	2002	B	B1	IntE2L2	KC908055	GW
6	GW174	Australia	2002	B	B1	LCR	KC820151	GW
6	GW174	Australia	2002	B	B1	IntE2L2	KC908015	GW
6	GW177	Australia	2002	B	B1	LCR	KC820152	GW
6	GW177	Australia	2002	B	B1	IntE2L2	KC908016	GW
6	GW179	Australia	2002	B	B1	LCR	KC820205	GW
6	GW179	Australia	2002	B	B1	IntE2L2	KC908056	GW
6	GW203	Australia	2003	B	B2	LCR	KC820206	GW
6	GW203	Australia	2003	B	B2	IntE2L2	KC908057	GW
6	GW206	Australia	2003	B	B1	LCR	KC820207	GW
6	GW206	Australia	2003	B	B1	IntE2L2	KC908058	GW
6	GW216	Australia	2003	B	B1	LCR	KC820208	GW
6	GW216	Australia	2003	B	B1	IntE2L2	KC908059	GW
6	GW224	Australia	2003	B	B2	LCR	KC820153	GW
6	GW224	Australia	2003	B	B2	IntE2L2	KC908017	GW
6	GW23	Australia	2003	B	B1	LCR	KC820148	GW
6	GW233	Australia	2003	B	B3	LCR	KC820258	GW
6	GW239	Australia	2003	B	B1	LCR	KC820259	GW
6	GW244	Australia	2004	B	B1	LCR	KC820142	GW
6	GW248	Australia	2004	B	B1	IntE2L2	KC908060	GW
6	GW251	Australia	2004	B	B1	LCR	KC820154	GW
6	GW251	Australia	2004	B	B1	IntE2L2	KC908018	GW
6	GW258	Australia	2004	B	B2	LCR	KC820155	GW
6	GW258	Australia	2004	B	B2	IntE2L2	KC908019	GW
6	GW264	Australia	2004	B	B1	LCR	KC820209	GW
6	GW264	Australia	2004	B	B1	IntE2L2	KC908061	GW
6	GW271	Australia	2004	B	B1	LCR	KC820210	GW
6	GW271	Australia	2004	B	B1	IntE2L2	KC908062	GW
6	GW279	Australia	2004	B	B2	LCR	KC820156	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	GW279	Australia	2004	B	B2	IntE2L2	KC908020	GW
6	GW280	Australia	2004	B	B1	LCR	KC820211	GW
6	GW280	Australia	2004	B	B1	IntE2L2	KC908063	GW
6	GW282	Australia	2004	B	B1	LCR	KC820212	GW
6	GW284	Australia	2004	B	B1	LCR	KC820137	GW
6	GW287	Australia	2004	B	B1	LCR	KC820213	GW
6	GW287	Australia	2004	B	B1	IntE2L2	KC908064	GW
6	GW289	Australia	2004	B	B1	LCR	KC820214	GW
6	GW289	Australia	2004	B	B1	IntE2L2	KC908065	GW
6	GW295	Australia	2004	B	B1	LCR	KC820215	GW
6	GW295	Australia	2004	B	B1	IntE2L2	KC908066	GW
6	GW313	Australia	2005	B	B2	LCR	KC820143	GW
6	GW313	Australia	2005	B	B2	IntE2L2	KC908010	GW
6	GW319	Australia	2005	B	B1	LCR	KC820157	GW
6	GW319	Australia	2005	B	B1	IntE2L2	KC908021	GW
6	GW326	Australia	2005	B	B2	LCR	KC820158	GW
6	GW326	Australia	2005	B	B2	IntE2L2	KC908022	GW
6	GW329	Australia	2005	B	B1	LCR	KC820138	GW
6	GW33	Australia	1997	B	B1	IntE2L2	KC908067	GW
6	GW331	Australia	2005	B	B2	LCR	KC820159	GW
6	GW331	Australia	2005	B	B2	IntE2L2	KC908023	GW
6	GW362	Australia	2006	B	B1	LCR	KC820216	GW
6	GW362	Australia	2006	B	B1	IntE2L2	KC908068	GW
6	GW363	Australia	2006	B	B1	LCR	KC820217	GW
6	GW363	Australia	2006	B	B1	IntE2L2	KC908069	GW
6	GW365	Australia	2006	B	B1	LCR	KC820218	GW
6	GW366	Australia	2006	B	B1	LCR	KC820219	GW
6	GW366	Australia	2006	B	B1	IntE2L2	KC908070	GW
6	GW369	Australia	2006	B	B3	LCR	KC820144	GW
6	GW370	Australia	2006	B	B2	LCR	KC820220	GW
6	GW370	Australia	2006	B	B2	IntE2L2	KC908071	GW
6	GW377	Australia	2006	B	B2	LCR	KC820160	GW
6	GW377	Australia	2006	B	B2	IntE2L2	KC908024	GW
6	GW380	Australia	2006	B	B1	LCR	KC820221	GW
6	GW380	Australia	2006	B	B1	IntE2L2	KC908072	GW
6	GW383	Australia	2006	B	B1	LCR	KC820222	GW
6	GW389	Australia	2006	B	B1	LCR	KC820161	GW
6	GW389	Australia	2006	B	B1	IntE2L2	KC908025	GW
6	GW39	Australia	1997	B	B1	LCR	KC820225	GW
6	GW391	Australia	2006	B	B2	LCR	KC820162	GW
6	GW391	Australia	2006	B	B2	IntE2L2	KC908026	GW
6	GW392	Australia	2006	B	B2	LCR	KC820223	GW
6	GW392	Australia	2006	B	B2	IntE2L2	KC908073	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	GW397	Australia	2006	B	B2	LCR	KC820224	GW
6	GW400	Australia	2006	B	B1	LCR	KC820226	GW
6	GW400	Australia	2006	B	B1	IntE2L2	KC908074	GW
6	GW403	Australia	2006	B	B1	LCR	KC820145	GW
6	GW403	Australia	2006	B	B1	IntE2L2	KC908011	GW
6	GW404	Australia	2006	B	B3	LCR	KC820227	GW
6	GW404	Australia	2006	B	B3	IntE2L2	KC908075	GW
6	GW406	Australia	2007	B	B1	IntE2L2	KC908076	GW
6	GW406	Australia	2007	B	B1	LCR	KC820228	GW
6	GW408	Australia	2007	B	B1	LCR	KC820146	GW
6	GW411	Australia	2007	B	B1	LCR	KC820229	GW
6	GW411	Australia	2007	B	B1	IntE2L2	KC908077	GW
6	GW412	Australia	2007	B	B1	LCR	KC820163	GW
6	GW412	Australia	2007	B	B1	IntE2L2	KC908027	GW
6	GW415	Australia	2007	B	B1	LCR	KC820230	GW
6	GW415	Australia	2007	B	B1	IntE2L2	KC908078	GW
6	GW419	Australia	2006	B	B1	LCR	KC820164	GW
6	GW419	Australia	2006	B	B1	IntE2L2	KC908079	GW
6	GW420	Australia	2007	B	B2	LCR	KC820231	GW
6	GW420	Australia	2007	B	B2	IntE2L2	KC908028	GW
6	GW429	Australia	2007	B	B2	LCR	KC820232	GW
6	GW429	Australia	2007	B	B2	IntE2L2	KC908080	GW
6	GW432	Australia	2007	B	B1	LCR	KC820233	GW
6	GW432	Australia	2007	B	B1	IntE2L2	KC908081	GW
6	GW437	Australia	2007	B	B1	LCR	KC820165	GW
6	GW437	Australia	2007	B	B1	IntE2L2	KC908082	GW
6	GW440	Australia	2007	B	B1	LCR	KC820234	GW
6	GW440	Australia	2007	B	B1	IntE2L2	KC908029	GW
6	GW441	Australia	2007	B	B2	LCR	KC820235	GW
6	GW441	Australia	2007	B	B2	IntE2L2	KC908083	GW
6	GW444	Australia	2007	B	B2	IntE2L2	KC908084	GW
6	GW445	Australia	2007	B	B1	IntE2L2	KC908085	GW
6	GW447	Australia	2007	B	B1	LCR	KC820236	GW
6	GW447	Australia	2007	B	B1	IntE2L2	KC908086	GW
6	GW455	Australia	2008	B	B1	LCR	KC820237	GW
6	GW455	Australia	2008	B	B1	LCR	KC908087	GW
6	GW456	Australia	2008	B	B1	LCR	KC820166	GW
6	GW456	Australia	2008	B	B1	IntE2L2	KC908030	GW
6	GW468	Australia	2008	B	B3	LCR	KC820238	GW
6	GW469	Australia	2008	B	B2	LCR	KC820239	GW
6	GW469	Australia	2008	B	B2	IntE2L2	KC908088	GW
6	GW477	Australia	2008	B	B1	LCR	KC820167	GW
6	GW477	Australia	2008	B	B1	IntE2L2	KC908031	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	GW48	Australia	1998	B	B1	IntE2L2	KC908090	GW
6	GW480	Australia	2008	B	B1	LCR	KC820240	GW
6	GW484	Australia	2008	B	B1	LCR	KC820241	GW
6	GW484	Australia	2008	B	B1	IntE2L2	KC908089	GW
6	GW487	Australia	2008	B	B1	LCR	KC820168	GW
6	GW487	Australia	2008	B	B1	IntE2L2	KC908032	GW
6	GW490	Australia	2008	B	B2	LCR	KC820242	GW
6	GW490	Australia	2008	B	B2	IntE2L2	KC908091	GW
6	GW491	Australia	2008	B	B2	LCR	KC820243	GW
6	GW491	Australia	2008	B	B2	IntE2L2	KC908092	GW
6	GW496	Australia	2008	B	B1	LCR	KC820147	GW
6	GW496	Australia	2008	B	B1	IntE2L2	KC908012	GW
6	GW499	Australia	2008	B	B1	LCR	KC820260	GW
6	GW499	Australia	2008	B	B1	IntE2L2	KC908096	GW
6	GW502	Australia	2008	B	B1	LCR	KC820261	GW
6	GW502	Australia	2008	B	B1	IntE2L2	KC908097	GW
6	GW505	Australia	2008	B	B1	LCR	KC820169	GW
6	GW505	Australia	2008	B	B1	IntE2L2	KC908033	GW
6	GW507	Australia	2008	B	B1	LCR	KC820262	GW
6	GW508	Australia	2008	B	B1	LCR	KC820263	GW
6	GW508	Australia	2008	B	B1	IntE2L2	KC908098	GW
6	GW510	Australia	2009	B	B2	LCR	KC820264	GW
6	GW510	Australia	2009	B	B2	IntE2L2	KC908099	GW
6	GW512	Australia	2009	B	B1	LCR	KC820265	GW
6	GW512	Australia	2009	B	B1	IntE2L2	KC908100	GW
6	GW519	Australia	2010	B	B1	IntE2L2	KC908093	GW
6	GW521	Australia	2010	B	B1	LCR	KC820170	GW
6	GW521	Australia	2010	B	B1	IntE2L2	KC908034	GW
6	GW525	Australia	2010	B	B1	LCR	KC820266	GW
6	GW525	Australia	2010	B	B1	IntE2L2	KC908101	GW
6	GW531	Australia	2010	B	B2	LCR	KC820267	GW
6	GW531	Australia	2010	B	B2	IntE2L2	KC908102	GW
6	GW537	Australia	2010	B	B2	LCR	KC820171	GW
6	GW537	Australia	2010	B	B2	IntE2L2	KC908035	GW
6	GW543	Australia	2010	B	B1	IntE2L2	KC908094	GW
6	GW545	Australia	2010	B	B1	LCR	KC820244	GW
6	GW554	Australia	2010	B	B1	LCR	KC820245	GW
6	GW560	Australia	2010	B	B1	LCR	KC820246	GW
6	GW571	Australia	2010	B	B1	LCR	KC820172	GW
6	GW571	Australia	2010	B	B1	IntE2L2	KC908036	GW
6	GW576	Australia	2010	B	B2	LCR	KC820247	GW
6	GW62	Australia	1999	B	B1	IntE2L2	KC908095	GW
6	GW702	Australia	2000	B	B1	LCR	KC820248	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	GW718	Australia	2005	B	B1	LCR	KC820249	GW
6	GW720	Australia	2005	B	B1	LCR	KC820250	GW
6	GW722	Australia	2005	B	B1	LCR	KC820251	GW
6	GW723	Australia	2005	B	B1	LCR	KC820252	GW
6	GW724	Australia	2005	B	B1	LCR	KC820253	GW
6	GW734	Australia	1999	B	B2	LCR	KC820254	GW
6	GW738	Australia	2006	B	B1	LCR	KC820255	GW
6	GW742	Australia	2007	B	B1	LCR	KC820256	GW
6	GW743	Australia	2008	B	B1	LCR	KC820257	GW
6	GW760	Australia	2008	B	B1	LCR	KC820173	GW
6	GW760	Australia	2008	B	B1	IntE2L2	KC908037	GW
6	RRP10	Colombia	2001	A		IntE2L2	KC907939	RRP
6	RRP103	Colombia	2007	B	B3	LCR	KC820268	RRP
6	RRP139	Colombia	2004	A		LCR	KC820270	RRP
6	RRP139'	Colombia	2005	A		LCR	KC820269	RRP
6	RRP139'	Colombia	2005	A		IntE2L2	KC907940	RRP
6	RRP141	Colombia	2000	B	B3	LCR	KC820271	RRP
6	RRP143	Colombia	2004	B	B3	LCR	KC820272	RRP
6	RRP143	Colombia	2004	B	B3	IntE2L2	KC907941	RRP
6	RRP145	Colombia	2001	A		IntE2L2	KC907942	RRP
6	RRP147	Colombia	2001	B	B3	LCR	KC820273	RRP
6	RRP147	Colombia	2001	B	B3	IntE2L2	KC907943	RRP
6	RRP153	Colombia	2009	B	B3	LCR	KC820274	RRP
6	RRP153	Colombia	2009	B	B3	IntE2L2	KC907944	RRP
6	RRP181	Colombia	2005	A		LCR	KC820275	RRP
6	RRP181	Colombia	2005	A		IntE2L2	KC907945	RRP
6	RRP26	Colombia	2000	A		LCR	KC820276	RRP
6	RRP26	Colombia	2000	A		IntE2L2	KC907946	RRP
6	RRP3	Colombia	2002	B	B3	IntE2L2	KC907947	RRP
6	RRP31	Colombia	1998	A		IntE2L2	#	RRP
6	RRP50	Colombia	2004	B	B3	LCR	KC820277	RRP
6	RRP55	Colombia	2003	A		LCR	KC820278	RRP
6	RRP55	Colombia	2003	A		IntE2L2	KC907948	RRP
6	RRP67	Colombia	2009	A		LCR	KC820279	RRP
6	RRP67	Colombia	2009	A		IntE2L2	KC907949	RRP
6	RRP70	Colombia	2003	A		IntE2L2	KC907950	RRP
6	RRP73	Colombia	2009	A		LCR	KC820280	RRP
6	RRP75	Colombia	2009	B	B2	LCR	KC820281	RRP
6	RRP75	Colombia	2009	B	B2	IntE2L2	KC907951	RRP
6	RRP8	Colombia	2003	B	B3	LCR	KC820282	RRP
6	RRP85	Colombia	1994	A		IntE2L2	#	RRP
6	RRP89	Colombia	1997	A		LCR	KC820283	RRP
6	RRP92	Colombia	2005	B	B3	IntE2L2	KC907952	RRP

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	RRP93	Colombia	2009	A		IntE2L2	#	RRP
11	GW102	Australia	2000	A	A2	LCR	#	GW
11	GW102	Australia	2000	A	A2	IntE2L2	KC907953	GW
11	GW10	Australia	1996	A	A2	LCR	KC907895	GW
11	GW143	Australia	2001	A	A2	LCR	#	GW
11	GW158	Australia	2001	A	A2	LCR	KC907896	GW
11	GW158	Australia	2001	A	A2	IntE2L2	KC907954	GW
11	GW159	Australia	2002	A	A2	LCR	KC907897	GW
11	GW159	Australia	2002	A	A2	IntE2L2	KC907955	GW
11	GW160	Australia	2002	A	A2	LCR	KC907898	GW
11	GW160	Australia	2002	A	A2	IntE2L2	KC907956	GW
11	GW16	Australia	1996	A	A2	LCR	KC907899	GW
11	GW16	Australia	1996	A	A2	IntE2L2	KC907957	GW
11	GW199	Australia	2002	A	A2	LCR	KC907900	GW
11	GW199	Australia	2002	A	A2	IntE2L2	KC907958	GW
11	GW221	Australia	2003	A	A2	LCR	KC907901	GW
11	GW221	Australia	2003	A	A2	IntE2L2	KC907959	GW
11	GW223	Australia	2003	A	A2	LCR	#	GW
11	GW301	Australia	2004	A	A2	LCR	KC907902	GW
11	GW301	Australia	2004	A	A2	IntE2L2	KC907960	GW
11	GW311	Australia	2004	A	A2	LCR	KC907903	GW
11	GW311	Australia	2004	A	A2	IntE2L2	KC907961	GW
11	GW367	Australia	2006	A	A2	LCR	KC907904	GW
11	GW367	Australia	2006	A	A2	IntE2L2	KC907962	GW
11	GW36	Australia	1997	A	A2	LCR	KC907905	GW
11	GW36	Australia	1997	A	A2	IntE2L2	KC907963	GW
11	GW384	Australia	2006	A	A2	LCR	KC907906	GW
11	GW384	Australia	2006	A	A2	IntE2L2	KC907964	GW
11	GW387	Australia	2006	A	A2	IntE2L2	KC907965	GW
11	GW422	Australia	2007	A	A2	LCR	KC907907	GW
11	GW422	Australia	2007	A	A2	IntE2L2	KC907966	GW
11	GW428	Australia	2007	A	A1	LCR	KC907908	GW
11	GW428	Australia	2007	A	A1	IntE2L2	KC907967	GW
11	GW466	Australia	2008	A	A2	LCR	KC907909	GW
11	GW466	Australia	2008	A	A2	IntE2L2	KC907968	GW
11	GW467	Australia	2008	A	A2	LCR	KC907910	GW
11	GW46	Australia	1998	A	A2	LCR	KC907911	GW
11	GW46	Australia	1998	A	A2	IntE2L2	KC907969	GW
11	GW4	Australia	1996	A	A2	LCR	KC907912	GW
11	GW4	Australia	1996	A	A2	IntE2L2	KC907970	GW
11	GW513	Australia	2009	A	A2	IntE2L2	KC907971	GW
11	GW514	Australia	2009	A	A2	LCR	#	GW
11	GW514	Australia	2009	A	A2	IntE2L2	KC907972	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
11	GW516	Australia	2009	A	A2	LCR	KC907913	GW
11	GW516	Australia	2009	A	A2	IntE2L2	KC907973	GW
11	GW518	Australia	2009	A	A2	LCR	KC907914	GW
11	GW533	Australia	2010	A	A2	LCR	KC907915	GW
11	GW533	Australia	2010	A	A2	IntE2L2	KC907974	GW
11	GW534	Australia	2010	A	A2	IntE2L2	KC907975	GW
11	GW535	Australia	2010	A	A2	LCR	#	GW
11	GW535	Australia	2010	A	A2	IntE2L2	KC907976	GW
11	GW548	Australia	2010	A	A2	LCR	KC907916	GW
11	GW548	Australia	2010	A	A2	IntE2L2	KC907977	GW
11	GW549	Australia	2010	A	A2	LCR	KC907917	GW
11	GW549	Australia	2010	A	A2	IntE2L2	KC907978	GW
11	GW552	Australia	2010	A	A2	LCR	KC907918	GW
11	GW552	Australia	2010	A	A2	IntE2L2	KC907979	GW
11	GW555	Australia	2010	A	A2	LCR	KC907919	GW
11	GW555	Australia	2010	A	A2	IntE2L2	KC907980	GW
11	GW557	Australia	2010	A	A2	IntE2L2	KC907981	GW
11	GW558	Australia	2010	A	A2	LCR	KC907920	GW
11	GW559	Australia	2010	A	A2	LCR	KC907921	GW
11	GW559	Australia	2010	A	A2	IntE2L2	KC907982	GW
11	GW564	Australia	2010	A	A2	LCR	#	GW
11	GW564	Australia	2010	A	A2	IntE2L2	KC907983	GW
11	GW565	Australia	2010	A	A2	LCR	KC907922	GW
11	GW565	Australia	2010	A	A2	IntE2L2	KC907984	GW
11	GW566	Australia	2010	A	A2	LCR	KC907923	GW
11	GW566	Australia	2010	A	A2	IntE2L2	KC907985	GW
11	GW5	Australia	1996	A	A2	LCR	KC907924	GW
11	GW5	Australia	1996	A	A2	IntE2L2	KC907986	GW
11	GW66	Australia	1999	A	A2	LCR	KC907925	GW
11	GW707	Australia	2003	A	A2	LCR	KC907926	GW
11	GW707	Australia	2003	A	A2	IntE2L2	KC907987	GW
11	GW712	Australia	2004	A	A2	LCR	KC907927	GW
11	GW712	Australia	2004	A	A2	IntE2L2	KC907988	GW
11	GW717	Australia	2005	A	A2	LCR	KC907928	GW
11	GW717	Australia	2005	A	A2	IntE2L2	KC907989	GW
11	GW71	Australia	1999	A	A2	LCR	KC907929	GW
11	GW71	Australia	1999	A	A2	IntE2L2	KC907990	GW
11	GW729	Australia	1999	A	A2	LCR	KC907930	GW
11	GW729	Australia	1999	A	A2	IntE2L2	KC907991	GW
11	GW72	Australia	1999	A	A2	IntE2L2	KC907992	GW
11	GW730	Australia	1999	A	A2	LCR	KC907931	GW
11	GW730	Australia	1999	A	A2	IntE2L2	KC907993	GW
11	GW740	Australia	2006	A	A1	LCR	KC907932	GW

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
11	GW740	Australia	2006	A	A1	IntE2L2	KC907994	GW
11	GW745	Australia	2008	A	A2	LCR	KC907933	GW
11	GW745	Australia	2008	A	A2	IntE2L2	KC907995	GW
11	GW78	Australia	1999	A	A2	LCR	#	GW
11	GW78	Australia	1999	A	A2	IntE2L2	KC907996	GW
11	GW79	Australia	1999	A	A2	LCR	KC907934	GW
11	GW79	Australia	1999	A	A2	IntE2L2	KC907997	GW
11	GW81	Australia	1999	A	A2	LCR	KC907935	GW
11	GW81	Australia	1999	A	A2	IntE2L2	KC907998	GW
11	GW93	Australia	2000	A	A2	LCR	KC907936	GW
11	GW93	Australia	2000	A	A2	IntE2L2	KC907999	GW
11	GW94	Australia	2000	A	A2	LCR	KC907937	GW
11	GW94	Australia	2000	A	A2	IntE2L2	KC908000	GW
11	GW9	Australia	1996	A	A2	LCR	KC907938	GW
11	RRP134	Colombia	2006	A	A2	IntE2L2	KC908001	RRP
11	RRP134	Colombia	2006	A	A2	LCR	#	RRP
11	RRP152	Colombia	2003	A	A2	LCR	KC907890	RRP
11	RRP152	Colombia	2003	A	A2	IntE2L2	KC908002	RRP
11	RRP159	Colombia	2001	A	A2	LCR	KC907891	RRP
11	RRP1	Colombia	2002	A	A2	IntE2L2	KC908003	RRP
11	RRP32	Colombia	2006	A	A1	LCR	KC907892	RRP
11	RRP32	Colombia	2006	A	A1	IntE2L2	KC908004	RRP
11	RRP34	Colombia	2009	A	A2	LCR	KC907893	RRP
11	RRP34	Colombia	2009	A	A2	IntE2L2	KC908005	RRP
11	RRP36	Colombia	2005	A	A2	LCR	KC907894	RRP
11	RRP36	Colombia	2005	A	A2	IntE2L2	KC908006	RRP
11	RRP51	Colombia	2007	A	A2	IntE2L2	KC908007	RRP
11	RRP96	Colombia	2009	A	A2	IntE2L2	KC908008	RRP
Sequences retrieved from GenBank								
HPV Type	**Sample / Isolate**	**Country**	**Year**	**Lineage**	**Sub-lineage**	**Region**	**NCBI#**	**Lesion**
6	PV1499	Sweden		A		Complete Genome	JN252322	LP
6	LP22	Slovenia		A		Complete Genome	HE599226	LP
6	PV2702	Sweden		A		Complete Genome	JN252323	LP
6	LPX29	Slovenia		A		Complete Genome	HE962031	LP
6	LPX26	Slovenia		A		Complete Genome	HE962032	LP
6	LP238	Slovenia		A		Complete Genome	HE962030	LP
6	CAC377	Slovenia		A		Complete Genome	FR751325	CA
6	LP130	Slovenia		A		Complete Genome	FR751323	LP

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	CAC251c	Slovenia		A		Complete Genome	FR751321	CA
6	LP5	Slovenia		A		Complete Genome	FR751324	LP
6	CAC26	Slovenia		A		Complete Genome	FR751322	CA
6	LP26	Slovenia		A		Complete Genome	FR751320	LP
6	LP182	Slovenia		A		Complete Genome	HE599243	LP
6	LP173	Slovenia		A		Complete Genome	HE599241	LP
6	LP15	Slovenia		A		Complete Genome	HE599239	LP
6	LP23	Slovenia		A		Complete Genome	HE599237	LP
6	LP143	Slovenia		A		Complete Genome	HE599229	LP
6	LP237	Slovenia		A		Complete Genome	HE599227	LP
6	LP244	Slovenia		A		Complete Genome	HE599246	LP
6	LP223	Slovenia		A		Complete Genome	HE599244	LP
6	LP240	Slovenia		A		Complete Genome	HE599240	LP
6	HPV-6f7rt	USA		B	B1	LCR	U61968	LP
6	RRP110	UK		B	B2	InterE2L2	JN169754	Unidentified
6	RRP115	UK		B	B2	InterE2L2	JN169755	Unidentified
6	RRP38	UK		B	B3	InterE2L2	JN169756	Unidentified
6	RRP48	UK		B	B1	InterE2L2	JN169757	Unidentified
6	GW36	UK		B	B1	E5A	JN169758	Unidentified
6	RRP56	UK		B	B3	E5A	JN169759	Unidentified
6	RRP45	UK		B	B3	E5A	JN169760	Unidentified
6	RRP111	UK		B	B3	E5A	JN169761	Unidentified
6	RRP44	UK		B	B3	E5A	JN169762	Unidentified
6	GW4	UK		B	B3	E5A	JN169763	Unidentified
6	JN573163	South Africa		B	B3	LCR	JN573163*	LP
6	JN573164	South Africa		B	B3	LCR	JN573164*	LP
6	JN573165	South Africa		B	B1	LCR	JN573165*	LP
6	JN573166	South Africa		B	B3	LCR	JN573166*	LP
6	JN573167	South Africa		B	B3	LCR	JN573167*	LP
6	JN573168	South Africa		B	B1	LCR	JN573168*	LP

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	JN573169	South Africa		B	B1	LCR	JN573169*	LP
6	JN573170	South Africa		B	B3	LCR	JN573170*	LP
6	JN573171	South Africa		B	B1	LCR	JN573171*	LP
6	JN573172	South Africa		B	B3	LCR	JN573172*	LP
6	JN573173	South Africa		B	B1	LCR	JN573173*	LP
6	JN573174	South Africa		B	B3	LCR	JN573174*	LP
6	HPV-6e1gt	USA		B	B1	LCR	U61962	CA
6	HPV-6e2gt	USA		B	B1	LCR	U61963	GW
6	HPV-6e3rt	USA		B	B1	LCR	U61964	LP
6	HPV-6e4rt	USA		B	B1	LCR	U61965	LP
6	HPV-6f5gt	USA		B	B1	LCR	U61966	CA
6	HPV-6f6gt	USA		B	B1	LCR	U61967	CA
6	PV1667	Sweden		B	B2	Complete Genome	JN252320	LP
6	PV1774	Sweden		B	B2	Complete Genome	JN252318	LP
6	PV1666	Sweden		B	B1	Complete Genome	JN252316	LP
6	PV2345	Sweden		B	B1	Complete Genome	JN252314	LP
6	PV1732	Sweden		B	B3	Complete Genome	JN252321	LP
6	PV2530	Sweden		B	B2	Complete Genome	JN252319	LP
6	PV1731	Sweden		B	B2	Complete Genome	JN252317	LP
6	PV529	Sweden		B	B2	Complete Genome	JN252315	LP
6	LP211	Slovenia		B	B1	Complete Genome	HE962029	LP
6	LPX34	Slovenia		B	B1	Complete Genome	HE962027	LP
6	LPX32	Slovenia		B	B1	Complete Genome	HE962028	LP
6	LPX44	Slovenia		B	B1	Complete Genome	HE962026	LP
6	CAC306	Slovenia		B	B1	Complete Genome	FR751337	CA
6	CAC96	Slovenia		B	B1	Complete Genome	FR751335	CA
6	LP176	Slovenia		B	B1	Complete Genome	FR751333	LP
6	LP137	Slovenia		B	B1	Complete Genome	FR751331	LP
6	LP11	Slovenia		B	B1	Complete Genome	FR751329	CA
6	CAC331	Slovenia		B	B1	Complete Genome	FR751327	CA

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
6	LP98(131)	Slovenia		B	B1	Complete Genome	FR751338	LP
6	CAC231	Slovenia		B	B1	Complete Genome	FR751336	CA
6	CAC11	Slovenia		B	B1	Complete Genome	FR751334	CA
6	LP96(175)	Slovenia		B	B1	Complete Genome	FR751332	LP
6	CAC23z	Slovenia		B	B1	Complete Genome	FR751330	CA
6	CAC301	Slovenia		B	B2	Complete Genome	FR751328	CA
6	CAC56	Slovenia		B	B3	Complete Genome	FR751326	CA
6	LP112	Slovenia		B	B1	Complete Genome	HE599245	LP
6	LP3	Slovenia		B	B1	Complete Genome	HE599235	LP
6	LP145	Slovenia		B	B1	Complete Genome	HE599233	LP
6	LP183	Slovenia		B	B1	Complete Genome	HE599231	LP
6	LP178	Slovenia		B	B1	Complete Genome	HE599242	LP
6	LP230	Slovenia		B	B1	Complete Genome	HE599238	LP
6	LP210	Slovenia		B	B1	Complete Genome	HE599236	LP
6	LP231	Slovenia		B	B1	Complete Genome	HE599234	LP
6	LP243	Slovenia		B	B1	Complete Genome	HE599232	LP
6	LP148	Slovenia		B	B1	Complete Genome	HE599230	LP
6	LP139	Slovenia		B	B1	Complete Genome	HE599228	LP
6	AF092932	Unidentified		B	B1	Complete Genome	AF092932	LP
11	A281	Slovenia		A	A1	E5A	FN870553	Anal sample
11	A86	Slovenia		A	A1	Complete genome	FN870021	Anal sample
11	CAC346	Slovenia		A	A1	Complete genome	FN870022	CA
11	CS123	Slovenia		A	A1	E5A	FN870541	Cervical sample
11	CS20	Slovenia		A	A1	Complete genome	FN907963	Cervical sample
11	HPV-11gt1	USA		A	A1	LCR	4097446	CA
11	HPV-11jrt1	USA		A	A1	LCR	4097447	LP
11	LZod45-11	China		A	A1	Complete genome	EU918768	Unidentified
11	1845353	Unidentified		A	A1	LCR	1845353	Unidentified
11	M14119	Germany		A	A1	Complete genome	M14119	LP

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
11	A101	Slovenia		A	A2	E5A	FN870530	Anal sample
11	A107	Slovenia		A	A2	E5A	FN870542	Anal sample
11	A122	Slovenia		A	A2	E5B	FN870594	Anal sample
11	A128	Slovenia		A	A2	E5A	FN870532	Anal sample
11	A129	Slovenia		A	A2	E5A	FN870533	Anal sample
11	A136	Slovenia		A	A2	E5A	FN870534	Anal sample
11	A140	Slovenia		A	A2	E5B	FN870598	Anal sample
11	A161	Slovenia		A	A2	E5B	FN870599	Anal sample
11	A171	Slovenia		A	A2	E5A	FN870537	Anal sample
11	A18	Slovenia		A	A2	E5A	FN870506	Anal sample
11	A187	Slovenia		A	A2	E5A	FN870543	Anal sample
11	A197	Slovenia		A	A2	E5A	FN870544	Anal sample
11	A200	Slovenia		A	A2	E5A	FN870545	Anal sample
11	A201	Slovenia		A	A2	E5A	FN870546	Anal sample
11	A202	Slovenia		A	A2	E5B	FN870610	Anal sample
11	A205	Slovenia		A	A2	E5A	FN870548	Anal sample
11	A218	Slovenia		A	A2	E5B	FN870612	Anal sample
11	A250	Slovenia		A	A2	E5A	FN870550	Anal sample
11	A253	Slovenia		A	A2	E5B	FN870614	Anal sample
11	A260	Slovenia		A	A2	Complete genome	FN907958	Anal sample
11	A297	Slovenia		A	A2	E5B	FN870617	Anal sample
11	A32	Slovenia		A	A2	E5A	FN870500	Anal sample
11	A320	Slovenia		A	A2	E5B	FN870618	Anal sample
11	A34	Slovenia		A	A2	E5A	FN870507	Anal sample
11	A345	Slovenia		A	A2	E5A	FN870556	Anal sample
11	A35	Slovenia		A	A2	E5A	FN870508	Anal sample
11	A383	Slovenia		A	A2	E5B	FN870621	Anal sample
11	A4	Slovenia		A	A2	E5A	FN870502	Anal sample
11	A409	Slovenia		A	A2	E5B	FN870622	Anal sample
11	A41	Slovenia		A	A2	E5A	FN870503	Anal sample
11	A418	Slovenia		A	A2	E5A	FN870560	Anal sample
11	A47	Slovenia		A	A2	Complete genome	FN907959	Anal sample
11	A48	Slovenia		A	A2	Complete genome	FN907961	Anal sample
11	A50	Slovenia		A	A2	E5A	FN870498	Anal sample
11	A51	Slovenia		A	A2	E5A	FN870501	Anal sample
11	A57	Slovenia		A	A2	E5A	FN870499	Anal sample
11	A63	Slovenia		A	A2	E5A	FN870527	Anal sample
11	A89	Slovenia		A	A2	E5A	FN870529	Anal sample
11	CAC11	Slovenia		A	A2	E5B	FN870573	CA
11	CAC246	Slovenia		A	A2	Complete genome	FN907960	CA
11	CAC256	Slovenia		A	A2	E5B	FN870576	CA

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
11	CAC266	Slovenia		A	A2	E5B	FN870577	CA
11	CAC321	Slovenia		A	A2	E5A	FN870515	CA
11	CAC336	Slovenia		A	A2	E5A	FN870516	CA
11	CAC372	Slovenia		A	A2	E5A	FN870518	CA
11	CAC86	Slovenia		A	A2	Complete genome	FN907962	CA
11	CS58	Slovenia		A	A2	E5A	FN870539	Cervical sample
11	CS93	Slovenia		A	A2	E5A	FN870603	Cervical sample
11	CU16	Thailand		A	A2	Complete genome	JQ773408	LP
11	CU17	Thailand		A	A2	IntE2L2/ LCR	JQ773409	Cervical sample
11	CU18	Thailand		A	A2	IntE2L2/ LCR	JQ773410	Cervical sample
11	CU19	Thailand		A	A2	IntE2L2/ LCR	JQ773411	LP
11	CU20	Thailand		A	A2	Complete genome	JQ773412	Lung
11	GUMC-AJ	China		A	A2	Complete genome	JN644141	LP
11	HPV-11art1	USA		A	A2	LCR	4097448	LP
11	JO-RRP_1	Hungary		A	A2	Complete genome	HE574701	LP
11	JO-RRP_2	Hungary		A	A2	Complete genome	HE574702	LP
11	JO-RRP_3	Hungary		A	A2	Complete genome	HE574703	LP
11	JO-RRP_4	Hungary		A	A2	Complete genome	HE574704	LP
11	JO-RRP_5	Hungary		A	A2	Complete genome	HE574705	LP
11	LP1	Slovenia		A	A2	IntE2L2/ LCR	HE611267	LP
11	LP12	Slovenia		A	A2	Complete genome	FN907957	LP
11	LP126	Slovenia		A	A2	Complete genome	HE611264	LP
11	LP13	Slovenia		A	A2	Complete genome	FN907964	LP
11	LP133	Slovenia		A	A2	IntE2L2/ LCR	HE611262	LP
11	LP138	Slovenia		A	A2	IntE2L2/ LCR	HE611258	LP
11	LP14	Slovenia		A	A2	IntE2L2/ LCR	HE611270	LP
11	LP146	Slovenia		A	A2	IntE2L2/ LCR	HE611265	LP
11	LP16	Slovenia		A	A2	Complete genome	HE611268	LP
11	LP170	Slovenia		A	A2	IntE2L2/ LCR	HE611273	CA

HPV Type	Sample / Isolate	Country	Year	Lineage	Sub-lineage	Region	NCBI#	Lesion
11	LP19	Slovenia		A	A2	Complete genome	HE611259	LP
11	LP2	Slovenia		A	A2	Complete genome	HE611271	LP
11	LP20	Slovenia		A	A2	Complete genome	HE611269	LP
11	LP220	Slovenia		A	A2	IntE2L2/ LCR	HE611263	LP
11	LP221	Slovenia		A	A2	IntE2L2/ LCR	HE611266	LP
11	LP229	Slovenia		A	A2	IntE2L2/ LCR	HE611274	CA
11	LP27	Slovenia		A	A2	IntE2L2/ LCR	HE611272	LP
11	LP6	Slovenia		A	A2	IntE2L2/ LCR	HE611260	LP
11	LP8	Slovenia		A	A2	IntE2L2/ LCR	HE611261	LP
11	M3	Slovenia		A	A2	E5A	FN870505	Anal sample
11	RRP50	Unidentified		A	A2	LCR	JN169764	LP
11	FR872717	Hungary		A	A2	Complete genome	FR872717	LP
11	AF029054	Australia		A	A2	LCR	AF029054	LP

Table S2. Primers used for amplification and sequencing

HPV6		
Primer	**Sequence (5'-3')**	**Starting Nucleotide***
URR_1F	CGCCTTACACACATAAGTAAT	7563
URR_1R	CAGGATATGATGCACTTGG	7748
URR_2F	GTGTTGCCTGTTAATCCTAT	7687
URR_2R	GCTTTTAAGAGTTGGCTACA	7881
E2L2_1F	ATTGTGGCTGCTATTAACAA	4055
E2L2_1R	TAAAGGCACATAACAACCAC	4207
E2L2_2F	TGGTTGTTATGTGCCTTTAT	4208
E2L2_2R	CAAAGCATACACTGTTACAAA	4417
HPV11		
Primer	**Sequence (5'-3')**	**Starting Nucleotide****
URR_1F	ATGTTGTTATGATGTTTGTGTGTTT	7387
URR_2R	TAAGTGTATGTAAGGGCAACCG	7705
E2L2_1F	AGGGCTGTACATGGTACTGAA	4251
E2L2_1R	CAGTGGCCTTGCATGTTTGA	4476

** In Reference sequence AF092932; ** In Reference sequence M14119*

Table S3. Values of accumulated frequencies at 95th and 10th percentile for each viral ORF analyzed

HPV6	95th PERCENTILE	10th PERCENTILE	MAD
CONCAT	0.0151819	0.000624	NA
LCR	0.020492	0,000001	0,2429014
E6	*0.011758*	0,000001	0,4329937
E7	0.006771	0,000001	0,5856697
E1	0.012322	0.000513	0,1324624
E2	0.021304	0,000002	0,5969294
IntE2L2	0.044041	0.001681	0,4808001
L2	0.0134831	0,000001	0,145485
L1	0.0082407	0.000666	0,1333648

HPV11	95th PERCENTILE	10th PERCENTILE	MAD
CONCAT	0.00356525	0.000379	NA
LCR	0.003944	0,000001	1,056302
E6	0.007043	0,000001	0,005063208
E7	0.003401	0	*
E1	0.0032	0.000529	0,5434211
E2	0.006447	0,000001	1,127627
IntE2L2	0.007291	0,000002	1,401046
L2	0.002233	0,000001	0,3839823
L1	0.00266	0,000001	0,5313555

95th percentile, 10th percentile and Median Absolute Deviation (MAD) values of the pair-wise distances for each viral ORF. "LCR": Long Control Region; "IntE2L2": Genomic region located between ORFs E2 and L2 of HPV." *": *Non analyzable due to low values*

Table S4. Likelihood Weights for assignment of HPV variants to the different clades.

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
HPV6		
JN573173	B1	1,00
JN573171	B1	0.959246
JN573169	B1	0.841844
	B2	0.052612
	B3	0.052621
	A	0.051615
JN573167	B3	0.965174
JN573165	B1	0.841844
	B2	0.052612
	B3	0.052621
	A	0.051615
JN573163	B3	0.965174
JN573174	B3	0.965174
JN573172	B3	0.681369
	B1	0.232859
	B2	0.036652
JN573170	B3	0.965174
JN573168	B1	0.841844
	B3	0.052621
	B2	0.052612
	A	0.051615
JN573166	B3	0.965174
JN573164	B3	0.965174
GWA284	B1	0.967320
GWA329	B1	0.967320
GWA034	B1	0.964985
GWA110	B1	0.963580
GWA118	A	0.951719
GWA244	B1	0.951115
GWA313	B2	0.999991
GWA369	B3	0.992139
GWA403	B1	0.963430
GWA408	B1	0.951174
GWA496	B1	0.953815
GWA23	B1	0.974251
GWA91	A	0.984737

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
GWA140	B1	0.962727
GWA174	B1	0.963569
GWA177	B1	0.960499
GWA224	B2	0.999993
GWA251	B1	0.963912
GWA258	B2	0.999993
GWA279	B2	0.999989
GWA319	B1	0.950263
GWA326	B2	0.999812
GWA331	B2	0.999993
GWA377	B2	0.999991
GWA389	B1	0.966013
GWA391	B2	0.999995
GWA412	B1	0.994421
GWA420	B2	0.999990
GWA440	B1	0.960516
GWA456	B1	0.953763
GWA477	B1	0.963893
GWA487	B1	0.960594
GWA505	B1	0.953437
GWA521	B1	0.963576
GWA537	B2	0.999812
GWA571	B1	0.963223
GWA760	B1	0.965034
GWA104	B2	0.999992
RRP110A	B2	0.996212
RRP115A	B2	0.996212
RRP38A	B3	0.427887
	B2	0.284847
	B1	0.142507
	A	0.142053
RRP48A	B1	0.963876
isolateGW36E5A	B1	0.996394
isolateRRP56E5A	B3	0.392220
	B2	0.338538
	B1	0.130333
	A	0.130016
isolateRRP45E5A	B3	0.427887
	B2	0.284847
	B1	0.142507

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
	A	0.142053
isolateRRP111E5A	B3	0.427888
	B2	0.427123
isolateRRP44E5A	B3	0.424883
	B2	0.423509
isolateGW4E5A	B2	0.996152
isolateHPV-6e1gt	B1	0.956139
isolateHPV-6e2gt	B1	0.956129
isolateHPV-6e3rt	B1	0.959121
isolateHPV-6e4rt	B1	0.956129
isolateHPV-6f5gt	A	0.966275
isolateHPV-6f6gt	B1	0.956129
isolateHPV-6f7rt	B1	0.956129
GW019	B1	0.967320
GW026	B2	0.965035
GW027	B2	0.999910
GW028	B2	0.999908
GW029	B1	0.951394
GW031	B1	0.967320
GW042	B1	0.951935
GW048	B1	0.967320
GW051	A	0.984332
GW055	B1	0.965031
GW063	B1	0.950339
GW070	A	0.995355
GW076	B1	0.967320
GW092	B2	0.995930
GW098	A	0.984332
GW108	B2	0.999999
GW113	B1	0.967320
GW115	B2	0.999909
GW117	B1	0.967320
GW120	A	0.999676
GW123	B1	0.964979
GW128	B1	0.965035
GW135	B1	0.952697
GW137	B2	0.999908
GW146	B3	0.989882
GW147	B2	0.999999
GW151	B1	0.967320

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
GW153	B1	0.964648
GW161	B1	0.967320
GW164	B1	0.967320
GW165	A	0.984332
GW168	B2	0.996707
GW172	B1	0.965033
GW179	B1	0.950202
GW203	B2	0.999999
GW206	B1	0.965034
GW216	B1	0.952624
GW248	B1	0.950987
GW264	B1	0.965062
GW271	B1	0.951775
GW280	B1	0.950372
GW282	B1	0.967320
GW287	B1	0.954743
GW289	B1	0.967524
GW295	B1	0.951657
GW33	B1	0.964828
GW362	B1	0.950179
GW363	B1	0.954788
GW365	B1	0.967320
GW366	B1	0.958339
GW370	B2	0.999998
GW380	B1	0.964255
GW383	B2	0.967320
GW392	B2	0.999998
GW397	B2	0.999908
GW39	B1	0.967320
GW400	B1	0.982353
GW404	B3	0.970347
GW406	B1	0.965034
GW411	B1	0.953547
GW415	B1	0.964614
GW419	B1	0.950466
GW429	B2	0.999954
GW432	B1	0.951818
GW437	B1	0.953521
GW441	B2	0.999999
GW444	B2	0.999998

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
GW445	B1	0.958739
GW447	B1	0.954161
GW455	B1	0.980487
GW468	B3	0.989883
GW469	B2	0.999999
GW480	B1	0.967320
GW484	B1	0.951285
GW48	B1	0.950151
GW490	B2	0.999997
GW491	B2	0.999998
GW519	B1	0.964985
GW543	B1	0.964984
GW545	B1	0.967320
GW554	B1	0.966663
GW560	B1	0.967320
GW576	B2	0.999553
GW62	B1	0.999810
GW702	B1	0.967320
GW718	B1	0.967320
GW720	B1	0.967320
GW722	B1	0.967320
GW723	B1	0.967320
GW724	B1	0.967320
GW734	B2	0.999911
GW738	B1	0.966978
GW742	B1	0.967320
GW743	B1	0.967320
GWA233-	A	0.999252
GWA239-	B1	0.967320
GWA499	B1	0.965034
GWA502	B1	0.965035
GWA507	B1	0.967320
GWA508	B1	0.951718
GWA510	B2	0.999999
GWA512	B1	0.951421
GWA525	B1	0.951454
GWA531	B2	0.999955
LP11	B1	0.999982
LP112	B1	0.989023
LP139	B1	0.999982

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
LP143	A	0.994763
LP145	B1	0.964970
LP15	A	0.994763
LP173	A	0.999988
LP178	B1	0.985144
LP183	B1	0.964970
LP211	B1	0.964970
LP22	A	0.999730
LP23	B1	0.991883
LP237	A	0.999730
LP238	A	0.999730
LP243	B1	0.985059
LP244	A	0.998432
LP3	B1	0.964970
LP98131	B1	0.989023
LPX26	A	0.997861
LPX29	A	0.999730
PRR103	B3	0.999619
PRR10	A	0.999659
PRR139'	A	0.984039
PRR139	A	0.984263
PRR141	B3	0.999496
PRR143	B3	0.997769
PRR145	A	0.998548
PRR147	B3	0.969398
PRR153	B3	0.973127
PRR181	A	0.983424
PRR26	A	0.993936
PRR31	A	0.952407
PRR3	B3	0.954063
PRR50	B3	0.999619
PRR55	A	0.984069
PRR67	A	0.984029
PRR70	A	0.998587
PRR73	A	0.984263
PRR75	B2	0.999945
PRR85	A	0.952407
PRR89	A	0.950553
PRR8	B3	0.999619
PRR92	B3	0.970225

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
PRR93	A	0.952408
HPV11		
1845353LCR627nt	A1	0.987177
AF029054	A2	0.795919
	A1	0.163289
JQ773410	A2	0.999998
JQ773409	A2	0.952082
JQ773411	A2	0.999998
GWA102	A2	0.969343
GWA10	A2	0.992862
GWA143	A2	0.721888
	A1	0.239928
GWA158	A2	0.953657
GWA159	A2	0.968313
GWA160	A2	0.969351
GWA16	A2	0.985486
GWA199	A2	0.968327
GWA221	A2	0.994733
GWA223	A2	0.970338
GWA301	A2	0.996908
GWA311	A2	0.745526
GWA367	A2	0.951618
GWA36	A2	0.996849
GWA384	A2	0.994733
GWA387	A2	0.777713
	A1	0.17777
GWA422	A2	0.996924
GWA428	A1	0.993357
GWA466	A2	0.9855
GWA467	A2	0.971868
GWA46	A2	0.994733
GWA4	A2	0.951656
GWA513	A2	0.777263
	A1	0.177668
GWA514	A2	0.996498
GWA516	A2	0.996924
GWA518	A2	0.970843
GWA533	A2	0.990223
GWA534	A2	0.777241
GWA535	A2	0.969343

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
GWA548	A2	0.99693
GWA549	A2	0.994733
GWA552	A2	0.994753
GWA555	A2	0.996924
GWA557	A2	0.777694
	A1	0.177762
GWA558	A2	0.993908
GWA559	A2	0.996924
GWA564	A2	0.951587
GWA565	A2	0.96582
GWA566	A2	0.996849
GWA5	A2	0.99694
GWA66	A2	0.992862
GWA707	A2	0.996924
GWA712	A2	0.996924
GWA717	A2	0.996924
GWA71	A2	0.985528
GWA729	A2	0.994733
GWA72	A2	0.777465
GWA730	A2	0.994733
GWA740	A1	0.981863
GWA745	A2	0.996924
GWA78	A2	0.966213
GWA79	A2	0.952296
GWA81	A2	0.996915
GWA93	A2	0.996849
GWA94	A2	0.954629
GWA9	A2	0.951662
HE611258	A2	0.952103
HE611260	A2	0.952103
HE611261	A2	0.952103
HE611262	A2	0.952103
HE611263	A2	0.952103
HE611265	A2	1,00
HE611266	A2	1,00
HE611267	A2	0.952103
HE611270	A2	0.993205
HE611272	A2	0.988569
HE611273	A2	0.988569
HE611274	A2	0.988569

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
isolateA101	A2	0.952075
isolateA107	A2	0.952075
isolateA122	A2	0.952075
isolateA128	A2	0.952075
isolateA129	A2	0.998652
isolateA136	A2	0.998653
isolateA140	A2	0.990953
isolateA161	A2	0.952075
isolateA171	A2	0.952075
isolateA18	A2	0.952075
isolateA187	A2	0.952075
isolateA197	A2	0.998653
isolateA200	A2	0.952075
isolateA201	A2	0.952075
isolateA202	A2	0.998653
isolateA205	A2	0.99288
isolateA218	A2	0.998652
isolateA250	A2	0.994629
isolateA253	A2	0.998653
isolateA281	A1	0.996718
isolateA32	A2	0.952075
isolateA320	A2	0.952075
isolateA34	A2	0.998653
isolateA345	A2	0.952075
isolateA35	A2	0.952075
isolateA383	A2	0.998652
isolateA4	A2	0.99237
isolateA409	A2	0.95119
isolateA41	A2	0.998653
isolateA418	A2	0.998653
isolateA50	A2	0.952075
isolateA51	A2	0.952075
isolateA57	A2	0.952075
isolateA63	A2	0.998653
isolateA89	A2	0.998653
isolateCAC11	A2	0.991425
isolateCAC256	A2	0.952075
isolateCAC266	A2	0.952075
isolateCAC321	A2	0.952075
isolateCAC336	A2	0.952075

SEQUENCE	ASSIGNED TO	POSTERIOR LIKELIHOOD WEIGHT
isolateCAC346	A2	0.952075
isolateCAC372	A2	0.952075
isolateCS123	A1	0.996687
isolateCS58	A2	0.952075
isolateCS93	A2	0.952075
isolateHPV-11art1	A2	0.990636
isolateHPV-11gt1	A1	0.995329
isolateHPV-11jrt1	A1	0.957821
isolateM3	A2	0.998652
isolateRRP50	A2	0.950399
RRP134	A2	0.968265
RRP152	A2	0.994799
RRP159	A2	0.953695
RRP1	A2	0.777713
RRP32	A1	0.956191
RRP34	A2	0.994688
RRP36	A2	0.994814
RRP51	A2	0.777713
	A1	0.17777
RRP96	A2	0.755499
isolateA297	A2	0.952075

Supplementary Text.

1) HPV reference trees generated for the analyses

i) HPV6 Reference Tree

(LP182_:0.00025336058259455691,(((((((PV529_:0.00037801202285943125,CAC301_:0.00063025155495666956)31:0.00000086753216491956,(PV1731_:0.00025218450131554487,PV1667_:0.00025228185921688929)21:0.00000086753216491956)44:0.00000086753216491956,(PV2530_:0.00025212451904382360,PV1774_:0.00037867103923089514)46:0.00012586549353599553)100:0.00186391644683312620,(CAC331_:0.00103472483959333700,(CAC96_:0.00075664916616337947,((AF092932.1_:0.00012602782014742850,((LP176_:0.00012574408809476635,CAC11_:0.00012574879406670588)8:0.00000086753216491956,((CAC231_:0.00037751616843511259,((LPX32_:0.00012572729955737680,(CAC23z_:0.00025179708714006611,LP137_:0.00012564814618706708)59:0.00012579679424533588)3:0.00000086753216491956,(((LP96175_:0.00037750824419838095,LP231_:0.00012571877522677427)6:0.00000086753216491956,(PV2345_:0.00050349446247458115,(LP210_:0.00000086753216491956,LP230_:0.00025136663398019451)87:0.00025146134885894100)12:0.00000086753216491956)1:0.00000086753216491956,(PV1666_:0.00025147138375741367,LP148_:0.00012574397592234217)6:0.00000086753216491956)0:0.00000086753216491956)4:0.00000086753216491956)50:0.00012574243593267606,LPX44_:0.00062947607827960298)16:0.00000086753216491956)8:0.00000086753216491956)48:0.00000086753216491956,(LPX34_:0.00037759450356455445,CAC306_:0.00025205990032002736)66:0.00012539987842561875)95:0.00037768572991613697)98:0.00035246683007799675)97:0.00083224820296583701)100:0.00125939099024752350,(PV1732_:0.00039038341489339113,CAC56_:0.00049252942060433490)100:0.00247363896800882780)100:0.01080505074726112100,((((PV1499_:0.00012581109597416937,LP130_:0.00012586739994228794)78:0.00012580308678247098,(CAC26_:0.00012574374607323587,LP26_:0.00012574507731849465)78:0.00012575243744685277)71:0.00000086753216491956,LP240_:0.00063293666245647462)77:0.00022761168557106659,CAC251c_:0.00059856258745886451)82:0.00031099743244559878)100:0.00126033704835719480,(LP5_:0.00000086753216491956,(CAC377_:0.00012571224441021729,PV2702_:0.00000086753216491956)72:0.00012572549184185005)45:0.00000086753216491956)64:0.00012570978057168370,LP223_:0.00000086753216491956);

ii) HP11 Reference Tree

(FN907961_:0.00012772820657329110,(FN907957_:0.00000145752489405818,(((((HE574702_:0.00038347132577645060,HE611259_:0.00000145752489405818)33:0.00000145752489405818,HE574701_:0.00025579079410906555)48:0.00012782944505154598,(((HE574703_:0.00000145752489405818,HE574704_:0.00051130959346642145)11:0.00000145752489405818,HE574705_:0.00025552441856018665)50:0.00012770750867035184,(FN907959_:0.00012786120044692712,(FN907958_:0.00063922329681825169,((((FN907963_:0.00000145752489405818,M14119_HPV-11_:0.00038371812846577005)71:0.00012765471478575925,(FN870021_:0.00025529158

589216513,EU918768_:0.00089414805646982794)33:0.000001457524894058 18)89:0.000
25505000544236747,FN870022_:0.00038386780541317466)100:0.001798148248048672
40,(FN907962_:0.00115480613804337100,(JQ773408_:0.00064051142493794667,(FR8727
17_:0.000001457524894058 18,JQ773412_:0.00012778779506702762)99:0.000384506257
72824181)82:0.00012657444985732588)77:0.00025575984704696375)41:0.0000014575
2489405818)74:0.00025582511102581771)62:0.00012788191696676176)42:0.00012776
647181931040)15:0.000001457524894058 18,FN907960_:0.00012772132459050038)1:0.0
000014575248940581 8,((HE611268_:0.00012771984476396198,HE611269_:0.000255677
89757664018)5:0.000001457524894058 18,(HE611271_:0.00012767627536085973,(FN90
7964_:0.00025527202889415255,HE611264_:0.000001457524894058 18)96:0.0003831643
7663546638)12:0.000001457524894058 18)2:0.000001457524894058 18)2:0.0000014575
2489405818)7:0.000001457524894058 18,JN644141_:0.00012772453259546194);

2) Sequences not submitted to GenBank, because total length is below 200 bp.

i) HPV6 IntE2L2

RRP31
TTTGTGGCTGCTATTAACAACCCCCTTGCAATTTTTCCTACTAACTCTACTTGTGTGTTACTGTCCC
GCATTGTATATACACTACTATATTGTTACCACACAGCAATGATGCTAACATGTCAATTTAATGATG
GAGATACCTGGCTGGGTTTGTGGTTGTTATGTGCCTTT
RRP85
TTGTGGCTGCTATTAACAACCCCCTTGCAATTTTTCCTACTAACTCTACTTGTGTGTTACTGTCCCG
CATTGTATATACACTACTATATTGTTACCACACAGCAATGATGCTAACATGTCAATTTAATGATGG
AGATACCTGGCTGGGTTTGTGGTTGTTATGTGCCTTTAA

RRP93
TTGTGGCTGCTATTAACAACCCCCTTGCAATTTTTCCTACTAACTCTACTTGTGTGTTACTGTCCCG
CATTGTATATACACTACTATATTGTTACCACACAGCAATGATGCTAACATGTCAATTTAATGATGG
AGATACCTGGCTGGGTTTGTGGTTGTTATGTGCCTTT

ii) HPV11 Long Control Region

RRP134
TTTGTGGATGTGTATGTATGTTTTTGTGCAATAAACAATTATTATGTGTGTTCTGTTACACCCAGT
GACTAAGTTGTGTTTGCACGCGCCGTTTGTGTTGCCTTCATATTATATTATACATATTTGTAATATA
CCTATACTATGTTACCCCCCCCCCCACTTGCAACCGTTTTCGGTTGCCCTTACATACACTTAA

GW102
TGTGTATGTATGTTTTTGTGCAATAAACAATTATTATGTGTGTTCTGTTACACCCAGTGACTAAGT
TGTGTTTGCACGCGCCGTTTGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTATACT
ATGTTACCCCCCCCCCCACTTGCAACCGTTTTCGGTTGCCCTTACATACACTT

GW143
TGTGTATGTTTTTGTGCAATAAACAATTATTATGTGTGTCCTGTTACACCCAGTGACTAAGTTGTG
TTTTGCACGCGCCGTTTGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTATACTATG
TTACCCCCCCCCCCACTTGCAACCGTTTTCGGTTGCCCTTACATACACTT

GW223

AAATGTGTATGTATGTTTTTGTGCAATAAACAATTATTATGTGTGTTCTGTTACACCCAGTGACTA
AGTTGTGTTTGCACGCGCCGTTTGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTAT
ACTATGTTACCCCCCCCCACTTGCAACCGTTTTCGGTTGCCCTT

GW514
TGTTGTTATGTATGTTTGTGTGTTTCGTGTGTGTATATATTTGTGGAATGTGTATGTATGTTTTTG
TGCAATAAACAATTATTATGTGTGTTTTGTTACACCCAGTGACTAAGTTGTGTTTGCACGCGCCGTT
TGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTATACTA

GW535
TGTGTATGTATGTTTTTGTGCAATAAACAATTATTATGTGTGTTCTGTTACACCCAGTGACTAAGT
TGTGTTTGCACGCGCCGTTTGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTATACT
ATGTTACCCCCCCCCCCACTTGCAACCGTTTTCGGTTGCCCTTACATACACTT

GW564
TTGTTGTTATGTATGTTTGTGTGTTTAGTGTGTGTATATATTTGTGGAATGTGTATGTATGTTTTT
GTGCAATAAACAATTATTATGTGTGTTTTGTTACACCCAGTGACTAAGTTGTGTTTGCACGCGCCGT
TTGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTATACTA

GW78
GGATGTGTATGTATGTTTTTGTGCAATAAACAATTATTATGTGTGTTCTGTTACACCCAGTGACTA
AGTTGTGTTTGCACGCGCCGTTTGTGTTGCCTTCATATTATATTATACATATTTGTAATATACCTAT
ACTATGTTACCCCCCCCCCCCACTTGCAACCGTTTTCGGTTGCCCTTACATACACTTAA

Cancer Medicine

Open Access

ORIGINAL RESEARCH

HPV16 variants distribution in invasive cancers of the cervix, vulva, vagina, penis, and anus

Sara Nicolás-Párraga[1,2,†], Carolina Gandini[1,2,†], Ville N. Pimenoff[1,2,†], Laia Alemany[1,2], Silvia de Sanjosé[1,2,3], F. Xavier Bosch[1,2] & Ignacio G. Bravo[1,4] on behalf of the RIS HPV TT and HPV VVAP study groups

[1]Infections and Cancer Unit, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), Barcelona, Spain
[2]Bellvitge Institute of Biomedical Research (IDIBELL), Barcelona, Spain
[3]CIBER in Epidemiology and Public Health (CIBERESP), Barcelona, Spain
[4]MIVEGEC, National Center for Scientific Research (CNRS), Montpellier, France

Keywords
Anogenital cancers, papillomavirus infection and cancer, viral diversity, viral evolution, virus–host interactions

Correspondence
Ignacio G. Bravo, National Center for Scientific Research (CNRS), Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle (MIVEGEC), UMR CNRS 5290, IRD 224, UM, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France. Tel: +33 467 41 5123; Fax: +33 467 41 63 30; E-mail: ignacio.bravo@ird.fr

Funding Information
This work was financially supported by the *Fundación Dexeus* for Women's Health grant to IGB, the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) to VNP, the Agència de Gestió d'Ajuts Universitaris i de Investigació, AGAUR, Generalitat de Catalunya (2014SGR1077 to XFB), and the Fondo de Investigaciones Sanitarias (FI12/00142 to XFB and SNP).

Received: 30 April 2016; Revised: 13 June 2016; Accepted: 27 July 2016

***Cancer Medicine* 2016; 5(10):2909–2919**

doi: 10.1002/cam4.870

[†]These authors contributed equally to this work.

Abstract

Human papillomavirus (HPV)16 is the most oncogenic human papillomavirus, responsible for most papillomavirus-induced anogenital cancers. We have explored by sequencing and phylogenetic analysis the viral variant lineages present in 692 HPV16-monoinfected invasive anogenital cancers from Europe, Asia, and Central/South America. We have assessed the contribution of geography and anatomy to the differential prevalence of HPV16 variants and to the non-synonymous *E6* T350G polymorphism. Most (68%) of the variance in the distribution of HPV16 variants was accounted for by the differential abundance of the different viral lineages. The most prevalent variant (above 70% prevalence) in all regions and in all locations was HPV16_A1-3, except in Asia, where HPV16_A4 predominated in anal cancers. The differential prevalence of variants as a function of geographical origin explained 9% of the variance, and the differential prevalence of variants as a function of anatomical location accounted for less than 3% of the variance. Despite containing similar repertoires of HPV16 variants, we confirm the worldwide trend of cervical cancers being diagnosed significantly earlier than other anogenital cancers (early fifties vs. early sixties). Frequencies for alleles in the HPV16 E6 T350G polymorphism were similar across anogenital cancers from the same geographical origin. Interestingly, anogenital cancers from Central/South America displayed higher 350G allele frequencies also within HPV16_A1-3 lineage compared with Europe. Our results demonstrate ample variation in HPV16 variants prevalence in anogenital cancers, which is partly explained by the geographical origin of the sample and only marginally explained by the anatomical location of the lesion, suggesting that tissue specialization is not essential evolutionary forces shaping HPV16 diversity in anogenital cancers.

Introduction

Certain human papillomaviruses (HPVs) are associated with certain human tumors. Based on the differential risk and association between infections and invasive cervical cancer (ICC), HPVs are classified into different risk groups [1]. A number of HPVs are considered carcinogenic for humans (group 1) or possibly/probably carcinogenic (groups 2a and 2b, respectively) and are commonly referred to as "high-risk HPVs" [1, 2]. ICC is worldwide the second most common cancer affecting women and responsible

for approximately 266,000 deaths per year (http://globocan.iarc.fr/Default.aspx). Persistent infection by oncogenic HPVs is considered a pre-requisite for the development of virtually all ICCs [3]. For this reason, the most extensive studies on HPVs have addressed cervical lesions and tumors [4]. A similar repertoire of HPVs may also be responsible for different fractions of other anogenital tumors [5], as viral DNA has been detected in malignant proliferations in the penis (33% prevalence) [6, 7], anus (88% prevalence) [7, 8], vulva (29% prevalence) [9], and vagina (74% prevalence) [10].

Oncogenic potential is not evenly distributed among oncogenic HPVs. Instead, HPV prevalence largely differs between types and between geographical regions, and the probability of progression from a clinically asymptomatic cervical infection to ICC is different for different HPVs [11, 12]. HPV16 is the most frequently detected HPV in all cervical infections, from normal cytology to ICC, in all world regions [11, 12]. HPV16 is also the most oncogenic HPV, responsible for 61% of all ICCs worldwide [4] and for even higher fractions of other HPV-associated anogenital carcinomas [6, 9, 10]. The biological reasons underlying the increased prevalence and oncogenicity of HPV16 compared with other closely related viruses, for example, the sister viruses HPV31 and HPV35, remain unclear [13].

Sequence diversity within HPV types is described in terms of viral variants [14]. The best classification for HPV16 variants has been proposed by Burk and coworkers, describing four lineages and a number of sublineages and applying an alphanumeric code, for example, HPV16_A4 [15]. Further, a large body of experimental research on the differential biological activities of HPV16 variants has focused on the E6 gene, especially on the T350G polymorphism, corresponding to the L83V amino acid substitution in the E6 oncoprotein. The initial literature described 350T as the "prototype," found in the "European" HPV16 variant, and the 350G allele as the "nonprototype," found in "non-European" HPV16 variants. However, the T350G polymorphism is found in different HPV16 variant lineages and is not a specific marker of any of them [16].

Papillomavirus variants are genetically very close, with above 98% nucleotide identity [14], but nonetheless HPV16 variants are suggested to differ in their oncogenic potential [17]. Particularly, the E6 T350G polymorphism has been associated with differential persistence and risk of progression to precancerous cervical lesions [17, 18].

The objective of this study was to characterize the viral component in a comprehensive set of invasive tumors of the cervix, vulva, vagina, penis, and anus, encompassing 35 countries within three continents. We aimed to analyze the differential prevalence of HPV16 variants as well as of the intensively studied T350G polymorphism as a function of the anatomical location of the lesion, the geographical origin of the samples, and the age at cancer diagnosis.

Materials and Methods

Samples

Samples analyzed in this study stem from a formalin-fixed paraffin-embedded (FFPE) sample repository from the Catalan Institute of Oncology (ICO), Barcelona, Spain, designed and constructed for the assessment of HPVs contribution to a number of anogenital human tumors [4, 6, 8–10]. All samples were tested for the presence of tumor tissue and for the presence of HPVs DNA using the SPF_{10}-$LiPA_{25}$ protocol, capable of genotyping HPV6, 11, 16, 18, 31, 33, 34, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 66, 68, 70, and 74 (version 1; Laboratory Biomedical Products, Rijswijk, the Netherlands). For this study, we selected 711 invasive squamous cell carcinomas showing exclusively the presence of HPV16 DNA after $LiPA_{25}$ genotyping. This data set consisted of samples from the cervix (n = 170), vulva (n = 128), vagina (n = 121), penis (n = 119), and anus (n = 172), spanning 35 different countries within three geographical regions: Europe, Central/South America, and Asia (Table S1).

Specimens were received anonymously and allocated a unique identification number upon reception, and the respective local and ICO ethic committees approved all the study protocols.

Identification and selection of the most informative regions

The most variable regions in the HPV16 genome were identified to maximize sequence diversity and phylogenetic signal in the targeted DNA fragments. We retrieved 109 HPV16 complete genome sequences from GenBank. Coding regions were aligned at the amino acid level with Muscle 3.7 [19] (http://www.ebi.ac.uk/Tools/msa/muscle/), while the upstream regulatory region (URR) was aligned at nucleotide level. Phylogenetic inference was performed at nucleotide level for each alignment, as well as for the concatenated full-length aligned genome, under a maximum likelihood framework using RAxML v7.2.8 [20] (http://www.exelixis-lab.org/) and the GTR+Γ_4 as substitution model. Robustness of tree individual nodes was assessed by bootstrap resampling analysis, as determined with the *-autoMRE* command [21]. Using the full-length genome data and the *–J MR_DROP* command [22], three rogue taxa were identified to show inconsistent positions during bootstrapping and were excluded from further analyses. The final alignment for the full-length genome comprised

7925 nucleotides and 548 distinct alignment patterns (Fig. S1). The well-resolved maximum likelihood phylogenetic trees obtained were employed to compute tree-based pairwise genetic distances (nucleotide substitutions per site) for each pair of taxa and for each genomic region analyzed, using the RAxML *-f x* command. Distances were then normalized with respect to the genetic distance between the corresponding taxa for the complete genome.

PCR and sequencing

DNA was extracted from four 5-μm paraffin slices by incubation overnight at 56°C with 250-μL proteinase K buffer (10 mg/mL proteinase K, 50 mmol/L Tris-HCl, pH 8.0) followed by incubation at 95°C for 8 min to inactivate proteinase K, and stored at −20°C until use. A 1:10 water dilution of this DNA solution was used for downstream processing. PCR primers were designed to target-specific HPV16 genome regions, so that well-described linage-specific polymorphisms were covered by the corresponding amplicons (Table S2). We also used primers previously designed by Larsson and coworkers [23] to span two positions in the *E6* gene that have been thoroughly analyzed in several studies (i.e., nt 131 and 350, reference sequence GenBank: NC_001526). All PCR reaction mixtures contained: 0.125 U/μL AmpliTaq Gold® DNA Polymerase (Life Technologies, Alcobendas, Spain), 2.0 mmol/L $MgCl_2$, 0.2 μmol/L deoxynucleotides triphosphate (Life Technologies, Alcobendas, Spain), 0.2 μmol/L forward and reverse primer (Biolegio , Nijmegen, The Netherlands), and 5 μL DNA solution. PCR conditions were 95°C for 10 min; 40 cycles of 30 sec at 94°C, 30 sec at 58°C, 30 sec at 72°C; plus 7-min final extension at 72°C. PCR products were Sanger-sequenced at Genoscreen (Lille, France) in both strands using four pairs of primers. (Table S2).

Phylogenetic analyses

Phylogenetic relationships of the amplified *E6*, *L2*, and LCR short fragments were placed in the global context of HPV16 genetic variability using an Evolutionary Placement Algorithm on RAxML v7.2.8 with the GTR+Γ_4 model [19, 20]. The algorithm provides likelihood weights for placing the partial sequences into the different nodes in the reference tree, in our case based on the pruned full-length genome alignment described above. Sequences obtained from our samples were incorporated into the reference alignment with MAFFT v7, and their phylogenetic placement was individually inferred with the *-f v* command in RAxML [21]. We integrated the results for all nodes and used 0.7 as a likelihood cutoff value to assign each sample into a specific variant lineage, namely A1-3, A4, B, C, and D (Table S3, Fig. S1). Using the 0.7 cutoff, 12 samples (1.7%) could only be classified as belonging to the A lineage and were subsequently classified as HPV16_A1-3 using a 0.6 cutoff value (Table S3).

Statistical analyses

A generalized linear model (GLM) with a Poisson distribution for count data and a log-link function was used to analyze the relationship between HPV16 variants prevalence with the two variables of interest: anatomical location and geographical origin. We explored as well the contribution of all double and triple interactions. Significance level was set at α value of 0.05. Analyses were performed using R in RStudio v0.98.939 (RStudio, Inc.). To corroborate the GLM results, differences in HPV16 variant distribution stratifying by anatomical location or by geographical origin were statistically assessed by means of Pearson's chi-square test and of Fisher's test, respectively. Prevalence ratios (PRs) of HPV16 variants among invasive anogenital cancers between Europe and Central/South America or Asia were estimated using Poisson multivariate regression model with robust variance. The different HPV16 variant lineages (i.e., A1–3, A4, and D) were used as dichotomous variables.

Distribution of the polymorphic site T350G within HPV16_A1 variants was assessed by Pearson's chi-square test when stratified by geographical origin and by Fisher's test when stratified by anatomical location. To assess the possible differential prevalence of the T|G alleles, we estimated the frequency of this polymorphism within HPV16_A1-3 variants for all anatomical locations for samples from Europe and Central/South America (Table 3). By focusing on the HPV16_A1-3 variants, we aimed to avoid the possible different epistatic interactions of the T|G alleles with the genetic background of each HPV16 variant, because the *E6* 350 position is also polymorphic T|G in HPV16_B, monomorphic T in HPV16_C, and monomorphic G in HPV16_D [16]. Asian cases were excluded from this analysis due to the small number of samples.

Cancer registry data show that cervical cancers are diagnosed earlier than other anogenital cancers associated with HPVs [6, 8–10]. Also, cancers caused by HPV16 are diagnosed earlier than cancers in the same anatomical location caused by other HPVs [4]. To disentangle the effects of virus genetics and anatomical location of the lesion on the age at diagnosis, we have followed a top-down approach, analyzing first age at cancer diagnosis for all HPV-related anogenital cancers available from our full clinical data set [4, 6, 8, 9], then for all cases exclusively linked to HPV16, and finally for all cases exclusively linked to HPV16_A1-3 (Fig. 2). For ages at tumor diagnosis, central values were estimated with the median, dispersion was estimated with the median absolute deviation, and differences were assessed

by Wilcoxon–Mann–Whitney test. Bonferroni correction for multiple comparisons was used when applicable.

Results

Choice of informative regions and sample set description

We identified, in decreasing order, *E4, E5,* LCR, *L2, E2,* and *E6* as the most informative regions in the HPV16 genome to perform phylogenetic inference (Fig. S2). PCRs were designed for each of these six genomic regions, and the LCR, *L2,* and *E6* targets rendered the best results in terms of amplicon quality and suitability for Sanger sequencing, as well as for the number of samples that tested positive. The final sample set comprised three continents (Europe, Central/South America, and Asia) and five anatomical sites (cervix, vulva, vagina, anus, and penis) (Table 1; Table S1). From 711 initially suitable amplicons, we were able to confidently classify 692 (97.3%) as belonging to HPV16_A1-3, A4, B, C, or D following an evolutionary placement algorithm (Table S3). Only nine samples belonged within the B or C lineages. Given the low numbers for both B and C lineages in our sample set, these sequences were not included in further analyses.

Geographical origin and anatomical location of the HPV16 variant distribution

The association between HPV16_A1-3, A4, and D variants (n = 683) with anatomical location and geographical origin was assessed using a GLM analysis. The model that fitted best our observations for the complete data set included the predictors "Geographical origin," "Anatomical location," and "Variant" (AIC = 225.88; Table S4). All predictors and their two-by-two interactions contributed significantly to the model (P < 0.0001 in all cases), but the triple interaction did not provide additional explanatory power (P = 0.36). The GLM analysis fitted very well our experimental data, as only <1.4% of all variance in HPV16 variant distribution remained unexplained by the model (Table S4). In our data set, 14.1% of the global variability arose from differential coverage of the three geographical regions (n = 342 for Europe, n = 261 for Central/South America, and n = 80 for Asia), and only 1.7% arose from differential coverage of the five anatomical origins analyzed (n = 163 for cervix, n = 121 for vulva, n = 114 for vagina, n = 115 for penis, and n = 170 for anus). Thus, the GLM approach allowed us to estimate and account for possible biases associated with design asymmetries in our data. We confirmed further the GLM results by estimating prevalence ratios for the different HPV16 variants stratifying by geography (Table 2) and by using a chi-square test after stratifying for geography and a Fisher's test after stratifying by anatomical location of the samples (Table S5).

We estimated that 68.2% of all variation in HPV16 variants abundance corresponded to actual differences in variant prevalence alone ($P < 2.2e^{-16}$; Table S4). Globally, HPV16_A1-3 was by far the most prevalent variant, with an overall prevalence of 95% in Europe, 86% in Central/South America, and 61% in Asia (F, Table 2). We

Table 1. Anatomical location and geographical distribution of amplified and classified samples.

Anatomical location	Europe		Central/South America		Asia		Total amplified	Total classified
	Amplified	Classified	Amplified	Classified	Amplified	Classified		
Cervix	72	70	71	69	27	26	170	165
Vulva	68	68	36	32	24	23	128	123
Vagina	61	60	51	48	10	9	122	117
Penis	74	73	42	40	3	2	119	115
Anal	79	79	72	72	21	21	172	172
Total	354	350	272	261	85	81	711	692

Table 2. Prevalence ratio (PR) of HPV16 variants between Europe and Central/South America or Asia.

Variant	Europe (%)	Central/South America (%)	Asia (%)	Europe vs. Central/South America		Europe vs. Asia	
				PR	95% CI	PR	95% CI
A1-3	324 (94.7)	225 (86.2)	49 (61.3)	Ref	—	Ref	—
A4	4 (1.2)	1 (0.4)	26 (32.5)	0.49	0.85–2.84	6.60	4.90–8.88
D	14 (4.1)	35 (13.4)	5 (6.2)	1.75	1.43–2.15	2.00	0.90–4.45
Wald's test				P < 0.001		P < 0.001	

quantified further that 9.0% of all variance in variant distribution was explained by differential association of viral lineages with geography ($P < 2.2e^{-16}$; Table S4). This variation corresponded to a significant 1.7-fold (95% CI: 1.4–2.1) increase of HPV16_D prevalence in Central/South America and to a significant 6.6-fold (95% CI: 4.9–8.9) increase of HPV16_A4 prevalence in Asia, in both cases compared with Europe (Table 2, see also Fig. 1). Finally, 2.8% of all variation in variant distribution corresponded to differential association of viral lineages with anatomical location ($P < 2.1e^{-05}$, Table S4, see also Fig. 1). This variation stemmed from the increased prevalence of HPV16_A4 in vagina and in anus in Asia, where this variant prevailed (Table S5, see also Fig. 1). Differences remained significant even after excluding data from Asia for vagina and penis, both locations with low number of cases (Table S5, see also Fig. 1).

HPV16 E6 gene T350G polymorphism

Prevalence for the 350G allele within HPV16_A1-3 ranged between 47% and 59% for Europe and between 59% and 90% for Central/South America. No differences between anatomical locations were observed within each geographical region (respectively, $P = 0.617$ and $P = 0.102$ for Europe and Central/South America). However, HPV16_A1-3 cases from Central/South America showed consistently higher 350G allele frequencies compared with Europe, especially for cervical ($P = 0.015$) and penile ($P < 0.0005$) cancers (Table 3).

Age at cancer diagnosis

Cervical cancers showed significantly younger ages at diagnosis compared with other anogenital cancers (early fifties vs. early sixties, $P < 0.0005$) regardless of the oncogenic HPV type or of the HPV16 variant driving the cancer (Fig. 2, see also Table S6). Notably, no significant differences were observed for age at cancer diagnosis among noncervical cancers (Fig. S3, see also Table S7).

Discussion

In this study, we have assessed the HPV16 variant diversity in a comprehensive set of invasive tumors of the cervix,

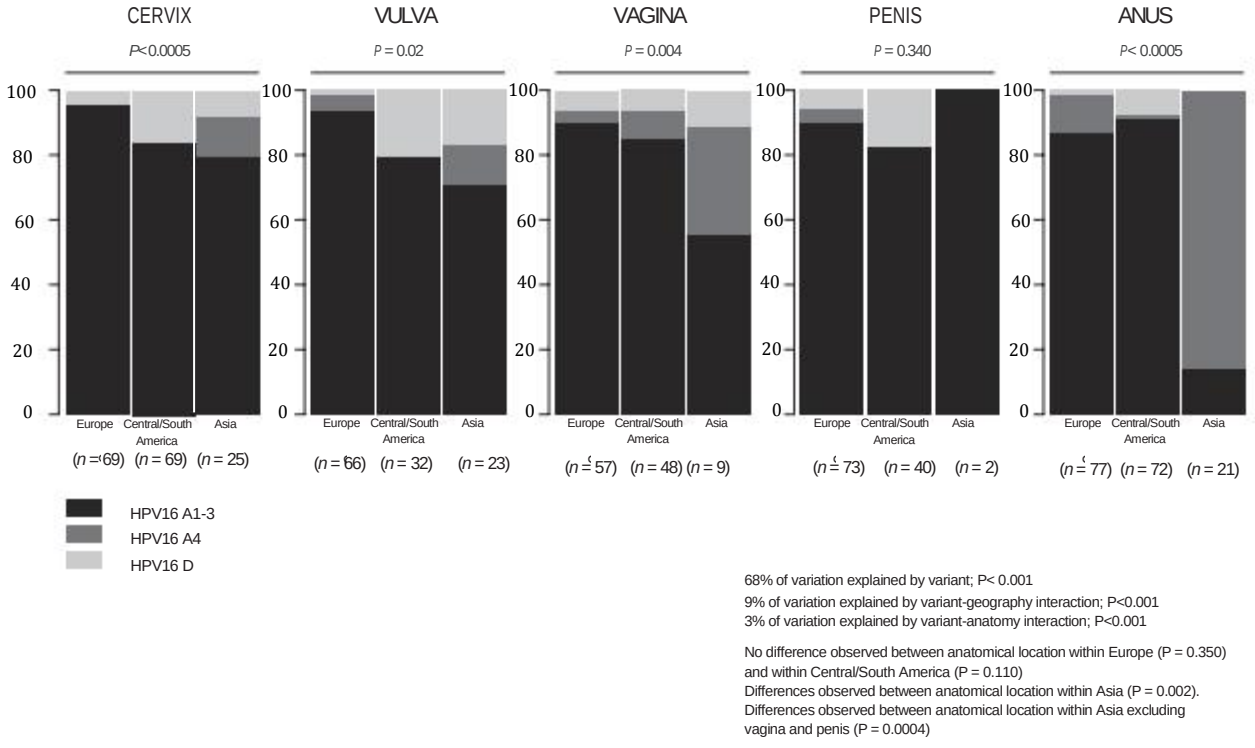


Figure 1. Distribution of HPV16_A1-3, A4 and D variants depending on geographical regions and anatomical location. For each combination of geography and anatomy, the number of samples is given in parentheses. Values for the contribution of differential variant prevalence (68%), for the contribution of geography (9%), and for the contribution of anatomy (3%) have been generated with a generalized linear model. For each anatomical location, the result of a chi-square test assessing homogeneity for variant prevalence values between the three geographical origins is provided (e.g., for vaginal cancers, the H0 hypothesis of the variant prevalence values being similar in Europe, Central/South America, and Asia is rejected with $P = 0.004$). For each geographical origin, the result of a chi-square test assessing homogeneity for variant prevalence values between the five anatomical locations is provided (e.g., for cancers from Central/South America, the null hypothesis of the variant prevalence values being similar in cervix, vulva, vagina, anus, and penis is accepted with $P = 0.074$). HPV16, Human papillomavirus type 16.

Table 3. HPV16_A1-3 variant distribution of the T350G polymorphic site for Europe and Central-South America and for anatomical location.

	Europe (*n* = 277)			Central/South America (*n* = 182)			
Anatomical location	*N*	350G	%	*N*	350G	%	*F* test[1]
Cervix	52	29	55.77	42	34	80.95	0.015
Vulva	49	29	59.18	17	10	58.82	1
Vagina	52	30	57.69	42	29	69.05	0.289
Penis	60	28	46.67	30	27	90.00	<0.0005
Anal	64	38	59.38	51	39	76.47	0.072
χ test[2]	*P* = 0.617			*P* = 0.102			

The number of samples (*N*), the samples with 350G allele (350G), and the percentage for the 350G allele frequencies are represented for each anatomical location for Europe and Central/South America. HPV16, Human papillomavirus type 16.

[1]Within each anatomical location, differences for the 350G allele frequency in the two geographical origins were assessed using Fisher's exact test.

[2]Within each geographical origin, differences for the 350G allele frequency in the different anatomical locations were assessed using chi-square test.

vulva, vagina, penis, and anus, analyzing the HPV16 variant distribution in 692 invasive squamous cancer samples from Europe, Central/South America, and Asia.

We have quantified for the first time the relative contributions of variant differential abundance, geographical origin, and anatomical location of the anogenital cancers to the observation of differential prevalence distribution of HPV16 variants. Our results show that there are no large differences between HPV16 lineage prevalence values among the anogenital cancers. The most prevalent viral lineage was by far HPV16_A1-3, independently of geographical origin and anatomical location of the samples, with the only exception of anal cancers in Asia, dominated by HPV16_A4, albeit based on small numbers. We have further estimated the contribution of geography and anatomical location to the observed differential HPV16 variant prevalence. The geographical origin of the cancer sample explains roughly 9% of all diversity in viral lineage distribution, and this contribution arises essentially from the increased prevalence of HPV16_A4 in samples from Asia and of HPV16_D in samples from Central/South America. In contrast, the anatomical location of the anogenital cancer explains only <3% of the observed diversity in viral lineages distribution. Indeed, we have not observed significant prevalence differences for HPV16 variants between anatomical locations of the anogenital cancers within Europe or Central/South America. In Asia, however, the higher contribution of HPV16_A4 variant in anogenital cancers exhibited a significant prevalence peak for anal cancers.

Variant distribution and diversity in HPV16 have mainly focused on the uterus cervix [16, 24, 25], but a sound description for viral lineages in other anogenital rscance sites was still wanting. Here, we have characterized the HPV16 variant component in a total of 692 anogenital invasive squamous cancers, including more than 550 cases from the vulva, vagina, penis, or anus. Our results confirm previous results with cervical samples and show further that the repertoire of viral HPV16 variants in anogenital cancers is largely the same regardless of the anatomical location. Consistent with our observations, two previous small studies in Northern Europe (HPV positive total *N* = 40; HPV16 positive *N* = 31) and in North America (HPV positive total *N* = 14; HPV16 positive *N* = 9) also reported an increased prevalence of HPV16_A1-3 variants in vulvar cancer (*N* = 29/31; *N* = 5/9 respectively) [26, 27]. Regarding vaginal cancer, the only previous study analyzing HPV16 variants (HPV positive total *N* = 37; HPV16 positive total *N* = 26) showed exclusively the presence of HPV16_A1-3 variants in European samples [26]. For anal cancer, a Canadian study (HPV positive total *N* = 96; HPV16 positive total *N* = 79) reported around 90% prevalence for HPV16_A variants [28]. Finally, and concerning viral diversity in HPV16-associated penile cancers, an Italian study (HPV positive total *N* = 19; HPV16 positive total *N* = 18) showed above 40% prevalence for both HPV16_A1-3 and D variants, along with above 10% minor nonnegligible contribution of HPV16_B variants [29]. However, a Mexican series of penile cancer samples (HPV positive total *N* = 67; HPV16 positive total *N* = 57) showed 92% prevalence of HPV16_A1-3 and 8% prevalence of HPV16_D [30]. In certain cases, the use in previous literature of imprecise naming schemes for HPV16 variants hampers a proper comparison. To avoid ambiguity, we have adhered here to the HPV16 variant terminology as standardized by Burk and coworkers [15] and strongly encourage further research on HPV variants to stick to it. Previous HPV16 variant nomenclatures included potentially misleading geographical references (e.g., "European") or ill-defined arbitrary classifications (e.g., "prototype" or "nonprototype"). The use of a geography-based nomenclature conveys a message of a close match between differential HPV16 variants prevalence and geography, which is not justified neither by the best previously available data [16] nor by our results presented here.

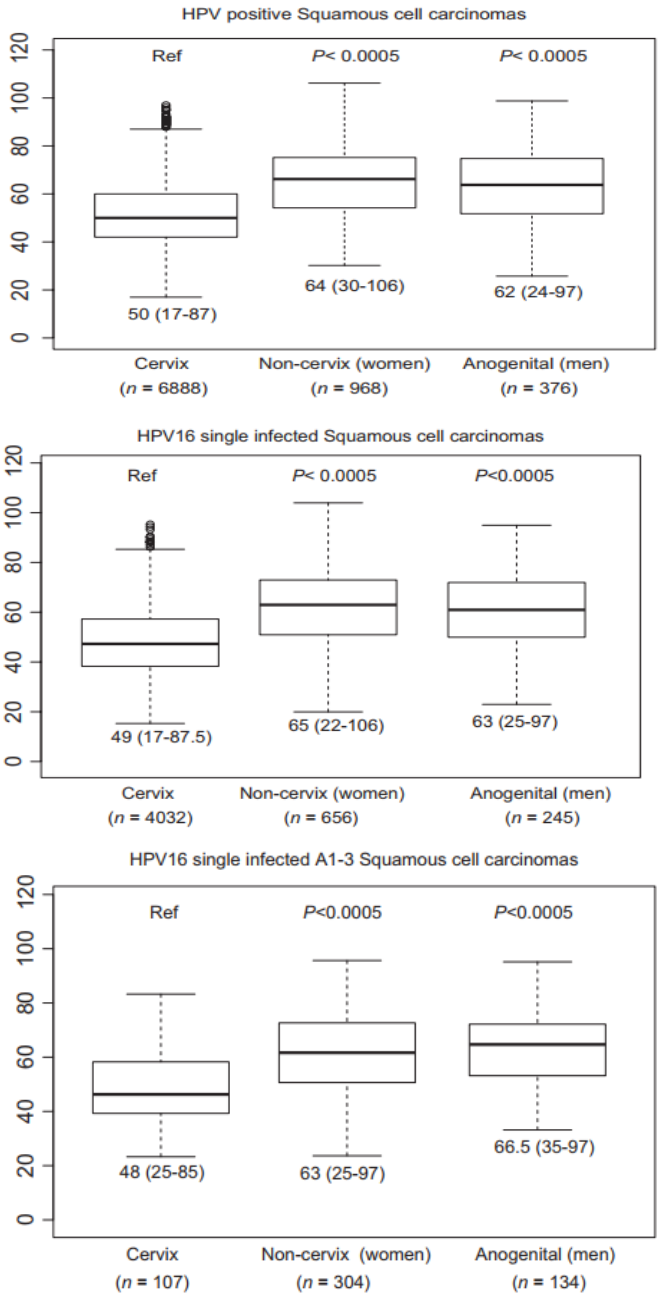


Figure 2. Age at tumor diagnosis for HPV-positive, HPV16 single infected, and HPV16_A1-3 single infected invasive SCC stratified by cervix, women anogenital noncervix (encompassing vagina, vulva, and anus), and men anogenital (encompassing anus and penis) samples. For each data set, the bar represents the median, the box encompasses the 25–75% percentiles, and the whiskers encompass the 95% percentiles. Numbers below each graph indicate the median and the range (1.5 × interquartile). Numbers in parentheses indicate the sample size for each location. Values for HPV-

Minor variations in the viral genome may be responsible for important changes regarding increased persistence or viral load [36, 37]. In addition, the adaptive host–pathogen interaction may further condition the differential probability for clearance or for eventual malignization of HPV16 infections [35, 36]. To tackle this connection between viral genotypic diversity and cancer risk, a long-studied candidate has been the T350G(L83V) single-nucleotide polymorphism in the HPV16 genome [38, 39]. In vitro studies have suggested an increased transformation potential for the 83V allele, especially for the HPV16_D lineage [38, 40], although these results may be linked to a specific host genetic background [39]. In European populations, prospective studies in cervical lesions as well as case–control studies have also communicated inconsistent results regarding the involvement of the T350G polymorphism in the persistence and progression to cancer [17, 18, 33, 41]. To address the question of the differential HPV16_*E6* 350G allele frequencies as a function of the geographical origin and the anatomical location of the cases, we focused exclusively on HPV16_A1-3 samples from Europe and Central/South America. We found that the 350G allele frequency did not significantly differ between anatomical locations for samples from the same geographical origin. In addition, our analysis revealed an increase in 350G allele frequency in samples from Central/South America compared with samples from Europe, consistent for all anatomical locations except for the vulva. This trend is in agreement with previous studies reporting an increased frequency of the 350G allele in Central/South America compared with European populations [31, 32, 33, 34], as well as with the minor contribution of this allele in vulvar and in vaginal lesions [23, 26].

Finally, we estimated the possible influence of the HPV16 variant on the age at cancer diagnosis. The rationale behind is threefold. First, cancer registry data show that cervical cancers are diagnosed earlier than other anogenital cancers associated with HPVs (www.hpvcentre.net). Second the studies from our group also show that cervical cancers caused by more aggressive HPVs, such as HPV16, HPV18, or HPV45, are diagnosed earlier than cervical cancers caused by other HPVs [4]. Third, the relative contribution of the different HPVs varies depending on the anatomical location of the cancer. Thus, the observed differences in age at diagnosis in HPV-related cancers of different anatomical origin could be linked to specific characteristics of the target tissue and/or to the different prevalence of the underlying viral agents. Making a coherent picture out of all available facts remains, however, a conundrum, because the contribution of HPV16 in noncervical cancers is higher than in cervical cancer: 61% in cervix [4], 62.9% in penis [6], 72.5% in vulva [9], and 75.8% in anus [8]. The only exception to this trend is vaginal cancers,

showing a 57% contribution of HPV16 [10]. One would thus expect that the increased contribution of HPV16 in noncervical cancers would result in earlier age at diagnosis when comparing HPV-related cancers among locations, but this is not the case. Our study design offered a unique opportunity to disentangle both alternatives and to test

repository enough samples from Europe, Central/South America, and Asia, as we did not have access to good quality samples from the African continent. Indeed, a thorough study on the evolution of any human pathogen should aim to sample the host–pathogen interaction there where the genetic diversity of the host is largest, that is,

Conflict of Interest

None declared.

References

1. IARC. 2012. IARC monographs on the evaluation of carcinogenic risks to humans - biological agents. 100B:100B.
2. Munoz, N., F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, et al. 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N. Engl. J. Med. 348:518–527.
3. Walboomers, J. M., M. V. Jacobs, M. M. Manos, F. X. Bosch, J. A. Kummer, K. V. Shah, et al. 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J. Pathol. 189:12–19.
4. de Sanjose, S., W. G. Quint, L. Alemany, D. T. Geraets, J. E. Klaustermeier, B. Lloveras, et al. 2010. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 11:1048–1056.
5. De Vuyst, H., G. M. Clifford, M. C. Nascimento, M. M. Madeleine, and S. Franceschi. 2009. Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: a meta-analysis. Int. J. Cancer 124:1626–1636.
6. Alemany, L., A. Cubilla, G. Halec, E. Kasamatsu, B. Quirós, E. Masferrer, et al. 2016. Role of human papillomavirus in penile carcinomas Worldwide. Eur. Urol.
7. Giuliano, A. R., G. Tortolero-Luna, E. Ferrer, A. N. Burchell, S. de Sanjose, S. K. Kjaer, et al. 2008. Epidemiology of human papillomavirus infection in men, cancers other than cervical and benign conditions. Vaccine 26(Suppl 1):K17–K28.
8. Alemany, L., M. Saunier, I. Alvarado-Cabrero, B. Quirós, J. Salmeron, H.-R. Shin, et al. 2015. Human papillomavirus DNA prevalence and type distribution in anal carcinomas worldwide. Int. J. Cancer 136:98–107.
9. de Sanjosé, S., L. Alemany, J. Ordi, S. Tous, M. Alejo, S. M. Bigby, et al. 2013. Worldwide human papillomavirus genotype attribution in over 2000 cases of intraepithelial and invasive lesions of the vulva. Eur. J. Cancer 49:3450–3461.
10. Alemany, L., M. Saunier, L. Tinoco, B. Quirós, I. Alvarado-Cabrero, M. Alejo, et al. 2014. Large contribution of human papillomavirus in vaginal neoplastic lesions: a worldwide study in 597 samples. Eur. J. Cancer 50:2846–2854.
11. Bruni, L., M. Diaz, X. Castellsague, E. Ferrer, F. X. Bosch, and S. de Sanjose. 2010. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. J. Infect. Dis. 202:1789–1799.
12. Bzhalava, D., P. Guan, S. Franceschi, J. Dillner, and G. Clifford. 2013. A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types. Virology 445:224–231.
13. Bravo, I. G., and M. Félez-Sánchez. 2015. Papillomaviruses: viral evolution, cancer and evolutionary medicine. Evol. Med. Public Health 2015:32–51.
14. de Villiers, E. M., C. Fauquet, T. R. Broker, H. U. Bernard, and H. Zur Hausen. 2004. Classification of papillomaviruses. Virology 324:17–27.
15. Burk, R. D., A. Harari, and Z. Chen. 2013. Human papillomavirus genome variants. Virology 445:232–243.
16. Cornet, I., T. Gheit, S. Franceschi, J. Vignat, R. D. Burk, B. S. Sylla, et al. 2012. Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR. J. Virol. 86:6855–6861.
17. Grodzki, M., G. Besson, C. Clavel, A. Arslan, S. Franceschi, P. Birembaut, et al. 2006. Increased risk for cervical disease progression of French women infected with the human papillomavirus type 16 E6-350G variant. Cancer Epidemiol. Biomarkers Prev. 15:820–822.
18. Gheit, T., I. Cornet, G. M. Clifford, T. Iftner, C. Munk, M. Tommasino, et al. 2011. Risks for persistence and progression by human papillomavirus type 16 variant lineages among a population-based sample of Danish women. Cancer Epidemiol. Biomarkers Prev. 20:1315–1321.
19. Stark, M., S. A. Berger, A. Stamatakis, and C. von Mering. 2010. MLTreeMap–accurate maximum likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. BMC Genom. 11:461.
20. Berger, S. A., and A. Stamatakis. 2011. Aligning short reads to reference alignments and trees. Bioinformatics 27:2068–2075.
21. Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
22. Pattengale, N. D., A. J. Aberer, K. M. Swenson, A. Stamatakis, and B. M. E. Moret. 2011. Uncovering hidden phylogenetic consensus in large data sets. IEEE/ACM Trans Comput Biol Bioinform [Internet] 8:902–911. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21301032 (accessed: September 30, 2015)
23. Larsson, G. L., G. Helenius, S. Andersson, F. Elgh, B. Sorbe, and M. G. Karlsson. 2012. Human papillomavirus (HPV) and HPV 16-variant distribution in vulvar squamous cell carcinoma in Sweden. Int. J. Gynecol. Cancer 22:1413–1419.
24. Yamada, T., C. M. Wheeler, A. L. Halpern, A. C. Stewart, A. Hildesheim, and S. A. Jenison. 1995. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence

analysis of the E6, L2 and L1 coding segments. J. Virol. 69:7743-7753.

25. Ho, L., S. Y. Chan, V. Chow, T. Chong, S. K. Tay, L. L. Villa, et al. 1991. Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and

36. Banister, C. E., A. R. Messersmith, B. Cai, L. B. Spiryda, S. H. Glover, L. Pirisi, et al. 2015. Disparity in the persistence of high-risk human papillomavirus genotypes between African American and European American women of college age. J. Infect. Dis. 211:100-108

Supporting Information

Additional supporting information may be found in the online version of this article:

Figure S1. Cumulative pair-wise distance frequencies for HPV16 genes and control region: Pair-wise distances (substitutions per site) are calculated for the full-genome, LCR, and all ORFs of reference sequences. Horizontal plain gray line represents the pair-wise distance 95th percentile. E2 - E4 (E2 minus E4) stands for the E2 gene nonoverlapping with the E4 gene.

Figure S2. Mid-point rooted HPV16 best-known maximum likelihood phylogenetic tree, constructed using 109 unique full-length genome sequences. HPV16 lineages are classified into four variants: A, B, C, and D. Bootstrap values above 700 are displayed closed to the corresponding node. GenBank accession numbers are given for all entries.

Figure S3. Age at tumor diagnosis for HPV positive, HPV16 single-infected band HPV16 A1-3 invasive squamous cell carcinomas stratified by anatomical location: Box plots represent the median, the 25% and 75% quantiles. Median and range (1.5 × interquantile) are represented in brackets below the box plots. Number of samples is represented in brackets below the anatomical location. HPV16-positive samples include all the types detectable through SPF10-LIPA25 protocol (version 1; Laboratory Biomedical Products, Rijswijk, the Netherlands).

Table S1. Sample distribution per anatomical location, geographical region, and country.

Table S2. Primer design.

Table S3. Likelihood weights for the attribution of each individual sequence to each (sub) variant, number of samples, and percentage.

Table S4. Generalized linear model (GLM) and analysis of deviance.

Table S5. HPV16 A1-3, A4 and D variant distribution by anatomical location within each geographical area.

Table S6. Age at tumor diagnosis for invasive squamous invasive carcinomas HPV positive, HPV negative, HPV16 single infected, and HPV16 A1-A2-A3 stratified by cervix, noncervix (women), and anogenital (men) samples.

Table S7. Age at tumor diagnosis for invasive squamous invasive carcinomas HPV positive, HPV16 single infected, and HPV16 A1-A2-A3 stratified by anatomical location.

Table S8. Collaborating centers at the RIS HPV TT and HPV VVAP study groups.

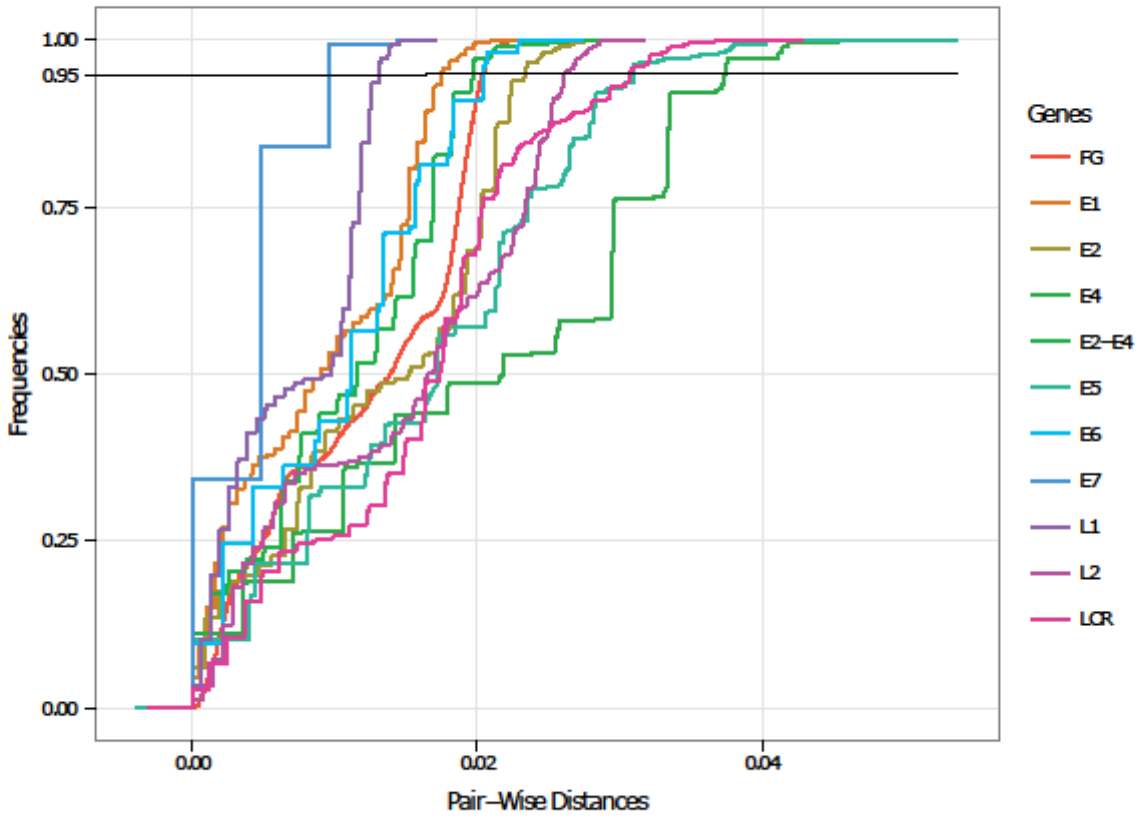


Figure S1: Cumulative pair-wise distance frequencies for HPV16 genes and control region: Pair-wise distances (substitutions per site) are calculated for the full-genome, LCR and all ORFs of reference sequences.Horitzontal plain grey line represent the pair-wise distance 95th percentile. E2-E4 (E2 minus E4) stands for the E2 gene non-overlapping with the E4 gene.

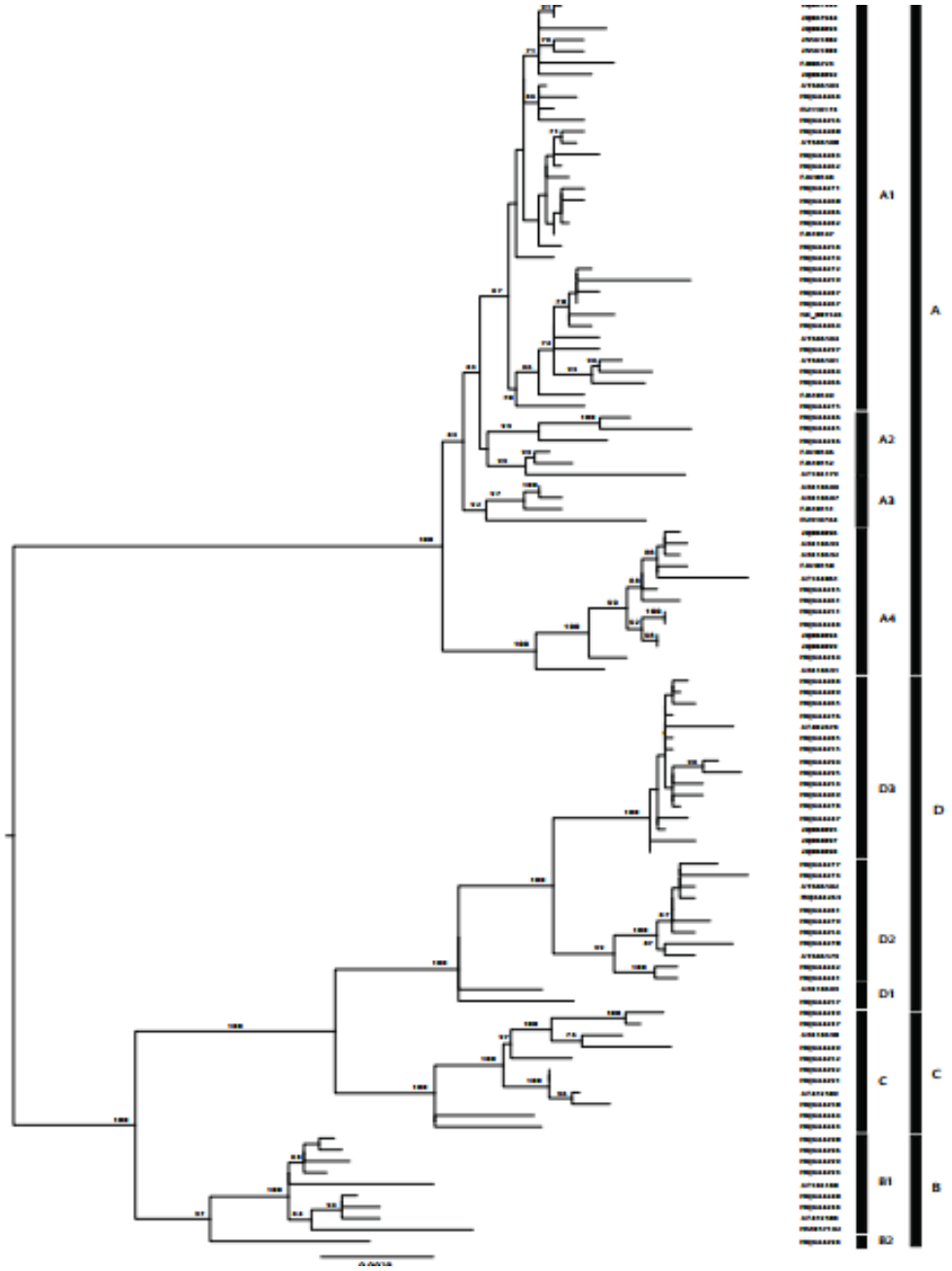
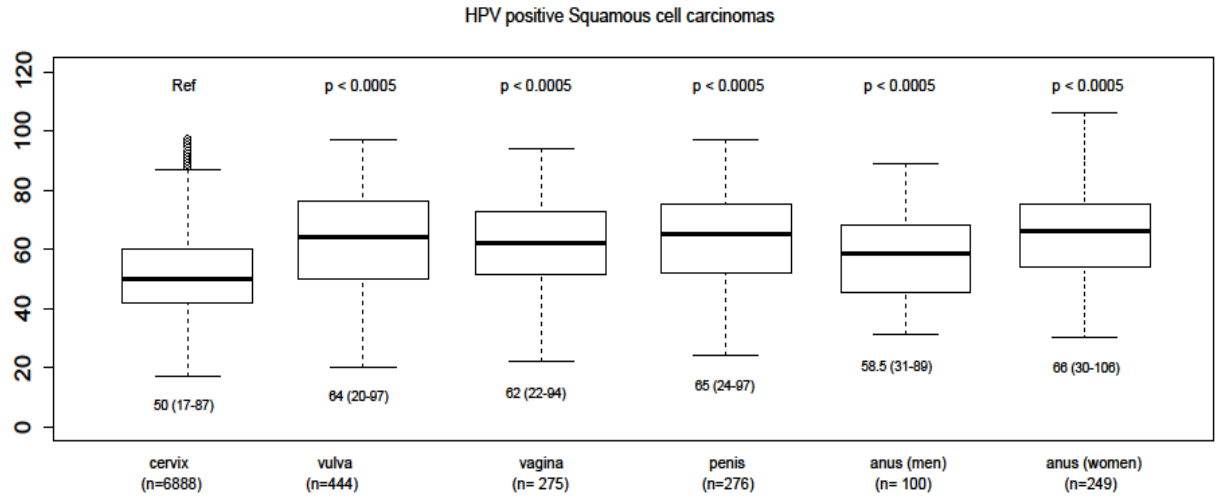
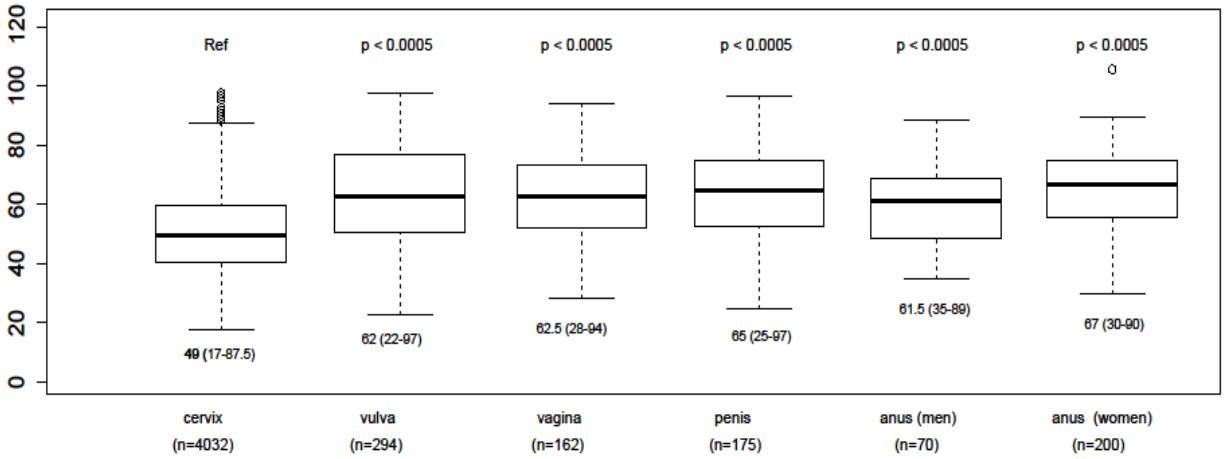


Figure S2. HPV16 Maximum likelihood phylogenetic best tree. Mid-point rooted HPV16 Best-known maximum likelihood phylogenetic tree, constructed using 109 unique full-length genome sequences. HPV16 lineages are classified into four variants: A, B, C and D. Bootstrap values above 70 are displayed closed to the corresponding branch. GenBank accession numbers are given for all entries.





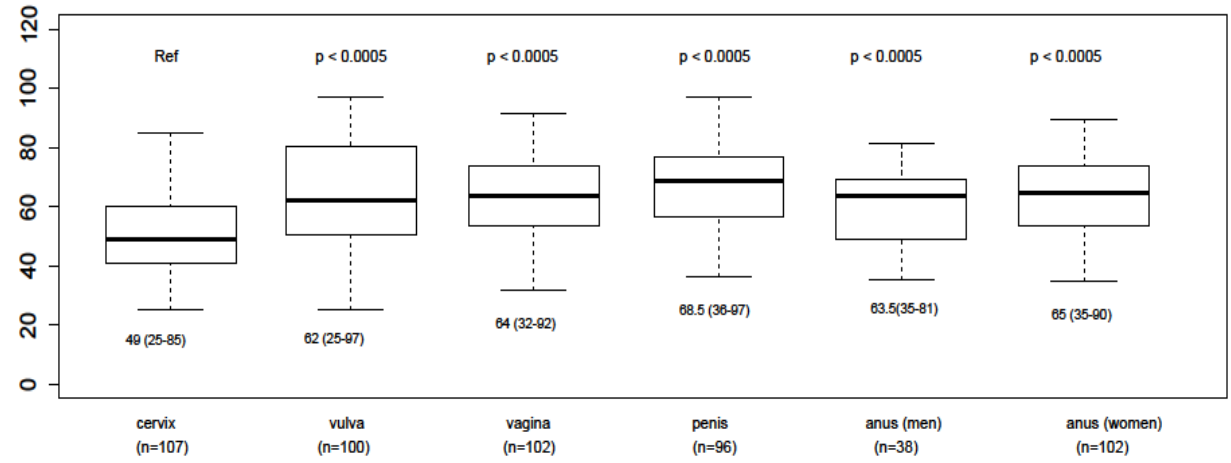


Figure S3: Age at tumour diagnosis for HPV positive, HPV16 single infected and HPV16 A1-3 invasive squamous cell carcinomas stratified by anatomical location: Boxplots represent the median, the 25% and 75% quantiles. Median and range (1.5 x Inter-quantile) are represented in brackets below the boxplots. Number of samples are represented in brackets below the anatomical location. HPV16 positive samples include all the types detectable through SPF10-LIPA25 protocol (version 1; Laboratory Biomedical Products, Rijswijk, Netherlands)

Table S1: Sample distribution *per* anatomical location, geographical region and country

Anatomical Location	Continent Category	Country	Number of Samples
Anal	Asia	India	1
		Korea	20
	Central/South America	Chile	2
		Colombia	25
		Ecuador	11
		Guatemala	1
		Mexico	26
		Paraguay	7
	Europe	Czech Republic	21
		France	2
		Germany	1
		Poland	4
		Portugal	20
		Slovenia	12
		Spain	19
Cervix	Asia	China	1
		India	5
		Kuwait	4
		Lebanon	5
		Taiwan	5
		Thailand	1
		Turkey	6
	Central/South America	Argentina	15
		Brazil	1
		Colombia	15
		Guatemala	11
		Mexico	5
		Paraguay	7
		Peru	7
		Venezuela	10

	Europe	France	4
		Greece	6
		Herzegovina	6
		Italy	13
		Netherlands	7
		Poland	17
		Portugal	5
		Spain	14
Penis	**Asia**	India	1
		Lebanon	1
		Korea	1
	Central/South America	Chile	3
		Colombia	12
		Ecuador	9
		Guatemala	2
		Honduras	3
		Mexico	4
		Paraguay	7
		Venezuela	2
	Europe	Czech Republic	14
		Poland	4
		Portugal	3
		Spain	36
		United Kingdom	17
Vagina	**Asia**	Bangladesh	1
		India	1
		Israel	1
		Korea	5
		Taiwan	1
		Turkey	1
	Central/South America	Argentina	1
		Chile	4
		Colombia	8
		Ecuador	14
		Guatemala	2

		Uruguay	1
		Mexico	22
	Europe	Austria	18
		Belarus	3
		Czech Republic	6
		France	13
		Germany	6
		Poland	5
		Spain	7
		United Kingdom	2
Vulva	**Asia**	India	6
		Korea	1
		Kuwait	2
		Lebanon	7
		Philippines	1
		Taiwan	2
		Turkey	5
	Central/South America	Argentina	2
		Brazil	2
		Chile	2
		Colombia	8
		Ecuador	5
		Guatemala	1
		Honduras	1
		Mexico	5
		Paraguay	5
		Uruguay	3
		Venezuela	2
	Europe	Austria	14
		Belarus	3
		Czech Republic	5
		France	3
		Germany	10
		Greece	2
		Italy	5

Poland	4
Portugal	5
Spain	8
United Kingdom	9
Total	**711**

Table S2: Primer design

Target Gene	Amplified region PCR product		Designed Primers	Specific SNPs	HPV16 (sub)variant
E6-1 [a]	nt 75-206 / 131 bp	F	5'-GACATTTTMTGCACCAAAAGAGA-3'	A83C G132C	B B
		R	5'-GCTTGCAGTACACACATTCTAATA-3'	G132T T178G	C A4
E6-2 [a]	nt 274-388 / 114 bp	F	5'-GAATCCATATGCTGTATGTGATAA-3'	T350G	A1
		R	5'-CGGTTTGTTGTATTGCTGTT-3'		
L2	nt 5106-5312 / 206 bp	F	5'-AGGCCAGCATTAACCTCTAGGCG-3'	A5117G T5138A C5138A C5144T C5230G A5258G	C A1 A1 A1 D C
		R	5'-AGTDGGTGAGGCTGCATGKGA-3'	T5285A A5286C A5294C T5309A	D D D D
LCR	nt 7712-7876 / 164 bp	F	5'- TGGCTTGTTTTAACTMMCCTAA -3'	T7712A A7728C T7741G T7779C G7824A A7828C	B A4 D A4 C A4
		R	5'- KKTGTAACCCAAAATCGGT-3'	A7835C A7837G G7840A C7873G	C C A4 A4

Table shows the target gene, the amplified region, the PCR product length, the designed primer sequence for each target gene and the lineage-specific polymorphism contained in each amplicon. F=Forward; R=Reverse. Reference sequence used for numbering: NC_001526. [a]E6 primer sequence retrieved from Larsson and colleagues (22).

Table S3: Likelihood weights for the attribution of each individual sequence to each (sub) variant, number of samples and percentage

(sub)variant	Likelihood Ascription	Number of Samples	%
A1-3	0.6-0.7	11	1,84
	0.70-0.89	99	16,56
	0.90-0.94	129	21,57
	> 0.95	359	60,03
	Sub-total	598	
A4	0.6-0.7	1	3,23
	0.70-0.89	2	6,45
	0.90-0.94	0	0,00
	> 0.95	28	90,32
	Sub-total	31	
B	0.6-0.7	0	0,00
	0.70-0.89	0	0,00
	0.90-0.94	2	50,00
	> 0.95	2	50,00
	Sub-total	4	
C	0.6-0.7	0	0,00
	0.70-0.89	0	0,00
	0.90-0.94	0	0,00
	> 0.95	5	100,00
	Sub-total	5	
D	0.6-0.7	0	0,00
	0.70-0.89	0	0,00
	0.90-0.94	1	1,85
	> 0.95	53	98,15
	Sub-total	54	
Unclassified samples	0.6-0.7	1	5,26
	0.70-0.89	8	42,11
	0.90-0.94	6	31,58
	> 0.95	4	21,05
	Sub-total	19	
Total		**711**	

Table S4: Generalized Linear Model (GLM) and analysis of deviance

	Df	Res. Dev	Df	Res. Dev	p-value	% of total variance
NULL			44	1289.29		
Anatomical location	4	22.15	40	1267.14	0.0001867	1.72
Geography	2	182.06	38	1085.08	< 2.2e-16	14.12
Variant	2	878.86	36	206.22	< 2.2e-16	68.16
Anatomical location: Geography	8	37.75	28	168.47	8,38E-03	2.93
Anatomical location: Variant	8	35.57	20	132.90	2,10E-02	2.76
Geography: Variant	4	115.48	16	17.42	< 2.2e-16	8.96
Anatomical location: Geography: Variant	16	17.42	0	0.00	0.3591931	-

Statistical mode approximation with the best fit for observed data for Europe, Central/ South America and Asia. Df= Degrees of freedom ; Res. Dev = Residual Deviance.

Table S5: HPV16 A1-3, A4 and D variant distribution by anatomical location within each geographical area

Continent	Variant	Cervix		Vulva		Vagina		Penis		Anus		Total	χ2 test
		n	%	n	%	n	%	n	%	n	%	n	
Europe	A1-3	66	95,65	62	93,93	54	94,73	68	93,150	74	96,10	324	0,350
	A4	0	0	2	3,03	0	0	0	0	2	2,53	4	
	D	3	4,35	2	3,03	3	5,26	5	6,84	1	1,29	14	
Sub-total		69		66		57		73		77			
Central/South America	A1-3	54	78,26	28	87,50	43	89,58	33	82,50	67	93,05	225	0,110
	A4	0	0	0	0	1	2,083	0	0	0	0	1	
	D	15	21,73	4	12,50	4	8,33	7	17,50	5	6,94	35	
Sub-total		69		32		48		40		72			
Asia	A1-3	20	80	17	73,91	6	66,66	2	100	4	19,04	49	0,002/ < 0.0005†
	A4	3	12	3	13,04	3	33,33	0	0	17	80,95	26	
	D	2	7,69	3	13,04	0	0	0	0	0	0	5	
Sub-total		25		23		9		2		21			
Total		163		121		114		115		170		683	
Fisher test		<0,0005/0,003		0,02/ 0,182		0,004 /0,562		0,340/0,114		<0,0005/0,059			

The contingency table shows HPV16 variants distribution for the 683 samples analysed, according to geographical region and anatomical location. Differences in variant prevalence between anatomical sites within a given geographical region are given for each row (χ2 test). Differences in variant prevalence between geographical regions, within a given anatomical location are given for each column (Fisher test). Abbreviations: A1-3= HPV16_A1, HPV16_A2 and HPV16_A3; A4= HPV16_A4; D=HPV16_D. †p-value for the χ2 test b excluding vaginal and penile samples from Asia, both present in low numbers.

Table S6 : Age at tumour diagnosis for invasive squamous invasive carcinomas HPV positive, HPV negative, HPV16 single infected and HPV16 A1-A2-A3 stratified by cervix, non-cervix (women) and anogenital (men) samples

Age of tumour diagnosis							
SCC	**Anatomical location**	**Median ± MAD**	**Quartile 1**	**Quartile 3**	**Range†**	**number**	**Wilcox. p-value**
HPV positive	**cervix**	50±13	42	60	(17-87)	6888	Ref
	non-cervix (women)	64±17	51	75	(30-106)	968	p<0.0005
	anogenital (men)	62±18	50	73	(24-97)	376	p<0.0005
HPV16 single infected	**cervix**	49±13	40	59	(17-87.5)	4032	Ref
	non-cervix (women)	65±16	53	75	(22-106)	656	p<0.0005
	anogenital (men)	63±16	52	74	(25-97)	245	p<0.0005
HPV16 A1-3	**cervix**	48±13	41	60	(25-85)	107	Ref
	non-cervix (women)	63±16	52	74	(25-97)	304	p<0.0005
	anogenital (men)	66.5±14	55	74	(35-97)	134	p<0.0005

Table represent the median , the 25% and 75% quatiles, range and number of samples. Median differences between cervix (Reference) and the other anogenital locations are performed by Wilcoxon Mann-Whitney test. HPV positive samples include all the types detectable though SPF_{10}-$LiPA_{25}$ protocol (version 1; Laboratory Biomedical Products, Rijswijk, Netherlands) † Range, 1.5x Inter-quartile.

Table S7 : Age at tumour diagnosis for invasive squamous invasive carcinomas HPV positive, HPV16 single infected and HPV16 A1-A2-A3 stratified by anatomical location

Age of tumour diagnosis							
SCC	**Antomical location**	**Median ± MAD**	**Quartile 1**	**Quartile 3**	**Range†**	**number**	**Wilcox. p-value**
HPV positive	**Cervix**	50±13	42.00	60.00	(17-87)	6888	Ref
	Vulva	64±19	50.00	76.00	(20-97)	444	p<0.0005
	Vagina	62±16	51.5	72.5	(22-94)	275	p<0.0005
	Penis	65±18	52.00	75.00	(24-97)	276	p<0.0005
	Anus (men)	58.5±14	45.75	68	(31-89)	100	p<0.0005
	Anus (women)	66±15	54	75	(30-106)	249	p<0.0005
HPV16 single infected	**Cervix**	49±13	40.00	59.00	(17-87.5)	4033	Ref
	Vulva	62±19	50.00	76.00	(22-97)	294	p<0.0005
	Vagina	62.5±16	52.25	72.75	(28-94)	162	p<0.0005
	Penis	65±16	53.00	75.00	(25-97)	175	p<0.0005
	Anus (men)	61.5±16	49.00	68.75	(35-89)	70	p<0.0005
	Anus (women)	67±14	56.00	75.00	(30-90)	200	p<0.0005
HPV16 A1-3	**Cervix**	49±13	41.00	60.00	(25-85)	107	Ref
	Vulva	62±21	50.75	80.00	(25-97)	100	p<0.0005
	Vagina	64±15	54.00	74.00	(32-92)	102	p<0.0005
	Penis	68.5±16	56.75	76.25	(36-97)	96	p<0.0005
	Anus (men)	63.5±13	49.00	68.75	(35-81)	38	p<0.0005
	Anus (women)	65±13	54.00	74.00	(35-90)	102	p<0.0005

Table represent the median , the 25% and 75% quatiles, range and number of samples. Median differences between cervix (Reference) and the other anogenital locations are performed by Wilcoxon Mann-Whitney test. HPV positive samples include all the types detectable though SPF_{10}-$LiPA_{25}$ protocol (version 1; Laboratory Biomedical Products, Rijswijk, Netherlands). †Range, 1.5xInter-quartile

Table S8: **Collaborating centres at the RIS HPV TT and HPV VVAP study groups**

1.RIS HPV TT study group:

1. Algeria: Doudja Hammouda (National Institute of Health, Registre des Tumeurs d'Alger), Anissa Bouhadef (C.H.U Hussein-Dey- Hospital Nafissa Hamoud).

2. Argentina: Silvio Alejandro Tatti, Susana Vighi (Hospital de Clínicas José de San Martín, Universidad de Buenos Aires).

3. Australia: Suzanne M. Garland, Sepher Tabrizi (The Royal Women's Hospital, The University of Melbourne).

4. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Faruk Ahmed, Mohammad Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU).

5. Bosnia Herzegovina: Ermina Iljazovic (University Clinical Center Tuzla BiH).

6. Brazil: Marcus Aurelho de Lima, Sérgio Henrique (Laboratório de Anatomia Patológica e Citologia Ltda, Associação de Combate ao Câncer do Brasil Central, Hospital Dr. Hélio Angotti, PATMED).

7. Chile: Rodrigo Prado (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile), Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Complejo Hospitalario San José).

8. China: You Lin Qiao, Chen Wen, Wu Su Hui (Cancer Institute, Chinese Academy of Medical Sciences and Peking Union Medical College).

9. Colombia: Hector Jaime Posso (Liga Contra el Cáncer de Bogotá), Luis Eduardo Bravo, Tito Collazos, Luz Stella García (Hospital Universitario del Valle), Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón (Instituto Nacional de Cancerología), Gloria I. Sánchez (Universidad de Antioquía).

10. Croatia: Magdalena Grce, Ivan Sabol (Rudjer Boskovic Institutem), Sonja Dzebro, Mara Dominis (Clinical Hospital Merkur).

11. Czech Republic: Ivo Steiner (Faculty of Medicine and Faculty Hospital, Hradec Kralove), Vladimir Vonka (Institute of Hematology and Blood Transfusion).

12. France: Christine Clavel (CHU Reims, Laboratoire Pol Bouin, Hôpital Maison Blanche), Massimo Tommasino (International Agency for Research on Cancer).

13. Greece: Maria Tzardi (Medical School of University of Crete), Theodoros Agorastos (Aristotle University of Thessaloniki).

14. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva - CIESAR, Hospital General San Juan de Dios), Walter Guerra (Instituto Nacional del Cáncer), Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer).

15. Honduras: Annabelle Ferrera (Escuela de Microbiología, Universidad Nacional

Autónoma de Honduras), Ricardo Bulnes (Hospital General San Felipe).

16. India: Asha Jain (Cancer Prevention and Relief Society Raipur), PSA Sarma (BSP Hospital), Sushil K Giri (Regional Cancer Center); Maheep Bhalla (JLN Hospital & Research Center), Bharat Patel (Lab One).

17. Italy: Luciano Mariani, Ferdinando Marandino (Regina Elena Cancer Institute).

18. Israel: Jacob Bornstein, Alejandro Livoff, Hector Itzhac Cohen (Western Galilee Hospital).

19. Japan: Toshiyuki Sasagawa (Kanazawa University Hospital), Shintaro Terahata (Tonami General Hospital), Kazuhisa Ishi (Juntendo University, Urayasu Hospital).

20. South Korea: Hai-Rim Shin, Jin-Kyoung Oh (National Cancer Center), Jung-II Suh (National Medical Center), and Seo-Hee Rha (Dong-A University).

21. Kuwait: Waleed Fahad Al-Jassar (Faculty of Medicine, Kuwait University), Rema'a Al-Safi (Maternity Hospital).

22. Lebanon: Muhieddine Seoud (The American University of Beirut Medical Center).

23. México: Patricia Alonso de Ruiz, Gustavo Lastra, Alma Karina Olivares Montano (Hospital General de México, Facultad de Medicina, Universidad Nacional Autónoma de México).

24. Mozambique: Clara Menéndez, Cleofé Romagosa, Carla Carrillo (Barcelona Center for International Health Research, Hospital Clínic/Universitat de Barcelona, and Manhiça Health Research Center).

25. Nigeria: Adekunbiola Banjo, Rose Anorlu, Fatimah B. Abdulkareem, A.O Daramola, CC Anunobi (Lagos University Teaching Hospital Idi-Araba).

26. Paraguay: Elena Kasamatsu, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción).

27. Perú: Carlos Santos, Eduardo Cáceres, Henry Gómez, Juvenal Sánchez, Carlos S. Vallejos, (Instituto Nacional de Enfermedades Neoplásicas).

28. The Philippines: Efren J. Domingo, María Julieta V. Germar, Jericho Thaddeus P. Luna, Carolyn Zalameda-Castro, Arnold M. Fernandez, Roslyn Balacuit (University of the Philippines College of Medicine); Cecilia Ladines Llave, Jean Anne Toral (Cervical Cancer Prevention Center – CECAP, Cancer Institute (UP-CM-PGH)).

29. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin).

30. Portugal: Eugenia Cruz (Centro Regional de Oncologia Coimbra, Instituto Português de Oncologia); Manuela Lacerda, Manuel Sobrinho-Simoes (Institute of Molecular Pathology and Immunology of the University of Porto); Ana Felix (Instituto Portugues de Oncologia de Lisboa Francisco Gentil).

31. Spain: Enrique Lerma (Hospital de la Santa Creu i Sant Pau); Enrique Poblet (Hospital General Universitário de Albacete); Lluís Eleuteri Pons (Hospital de Tortosa Verge de La Cinta); Antonio Llombart-Bosch, Morelva Toro de Méndez (Facultad de Medicina, Universidad de Valencia); Belen Lloveras (Hospital del Mar); Ana Puras Gil (Hospital Virgen del Camino); Miguel Andújar (Complejo Hospitalario Universitario Insular Marteno-Infantil); Jaume Ordi (CRESIB - Hospital Clínic); Adela Pelayo (Clinica San Carlos); Julio Velasco, Cristina Pérez (Hospital San Agustín, & IUOPA (Oncologic and Universitary Institute of Principality of Astúrias)); Maria Alejo (Hospital General de l'Hospitalet); Ignacio G. Bravo (CSISP - Centre for Public Health Research, Centro Superior de Investigación en Salud Pública, Conselleria de Sanidad (Generalitat Valenciana)); Laia Alemany, F. Xavier Bosch, Vanesa Camón, Gabriel Capellà, Cristina Caupena, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Mireia Diaz, Ana Esteban, Rebeca Font, Jose M. Godínez, Nuria Guimerà, Yolanda Florencia, Helena Frayle, Mercedes Hurtado, Joellen Klaustermeier, Anna Merchán, Carles Miralles, Nuria Monfulleda, Nubia Muñoz, Bea Quirós Cristina Rajo, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia (ICO) - Catalan Institute of Oncology); August Vidal (Hospital Universitari de Bellvitge).

32. Taiwan: Chou Cheng-Yang (National Cheng Kung University Medical College, Taiwan Association of Gynecologic Oncologists); Tang-Yuan Chu (Buddhist Tzuchi Genral Hospital); Kuo-Feng Huang (Chi Mei Medical Center); Cheng Wen-Fang (National Taiwan University Hospital); Chih- Ming HO (Gynecologic Cancer Center, Cathay General Hospital).

33. Thailand: Saibua C. Bunnag Chichareon, Kobkul Tungsinmunkong, Jintamard Suwanjarat (Prince of Songkla University).

34. The Netherlands: Chris J.L.M Meijer, Peter J.F Snijders (Vrije Universiteit Medical Center); Wim G.V. Quint, Jean-Paul Brunsveld, Anco C. Molijn, Daan T. Geraets (DDL Diagnostic Laboratory).

35. Turkey: Alp Usubutun (Medical School, Hacettepe University).

36. Uganda: Michael Odida (Uganda Makerere University); Elisabete Weiderpass (Karolinska Institutet, Sweden; The Norwegian Cancer Registry, Norway; Samfunded Folkhalsan, Finland).

37. United States of America: Esther Oliva (Massachusetts General Hospital); Thomas C. Wright (New York Presbyterian Hospital, Columbia University Medical Center).

38. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

2.HPV VVAP study group for vulvar site, updated September 2012:

1. Argentina: Myriam Perrotta, Ana Jaen, Kevin Davies, Henry Kitchener, Godfrey Wilson (Hospital Italiano de Buenos Aires);

2. Australia: Suzanne M. Garland, Sepher N. Tabrizi (The Royal Women's Hospital, The University of Melbourne); Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

3. Austria: Elmar Armin Joura (University Hospital and Medical School);

4. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

5. Belarus: Halina Viarheichyk, Sitnikov Valeriy (Gomel State Medical University); Achynovich Searhei (Gomel Regional Clinical Oncological Hospital);

6. Bosnia Herzegovina: Ermina Iljazovic (University Clinical Center Tuzla BiH);

7. Brazil: Paula Maldonado, Gutemberg Leão Almeida, Isabel Val, Renata Fonseca, Roberto José Lima, Marcia Mannarino, Yara Furtado (Instituto de Ginecologia da Universidad Federal do Rio de Janeiro);

8. Chile: Rodrigo Prado, Carla Molina, rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

9. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón (Instituto Nacional de Cancerología); Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (Faculty Hospital Hradec Kralove);

10. Ecuador: Leopoldo Tinoco, Hospital Oncológico de Quito, Quito, Ecuador;

11. France: Christine Clavel, Philippe Birembaut, Veronique Dalstein (CHU de Reims, Laboratoire Pol Bouin/ INSERM UMR-S 903, Reims); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

12. Germany: Monika Hampl, Pof. Baldus (University Hospital of Duesseldorf); Karl Ulrich Petry, Alexander Luyten (Klinikum Wolfsburg); Michael Pawlita, Gordana Halec (Department Genome Changes and Carcinogenesis. Heildelberg);

13. Greece: Theodoros Agorastos (Aristotle University of Thessaloniki);

14. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer);

15. Honduras: Annabelle Ferrera (Escuela de Microbiología, Universidad Nacional Autónoma de Honduras);

16. India: Neerja Bhatla (Institute of Medical Science New Dehli);

17. Israel: Jacob Bornstein, Alejandro Livoff, Hector Itzhac Cohen (Western Galilee

Hospital- Nahariya);

18. Italy: Luciano Mariani, Amina Vocaturo, Maria Benevolo, Fernando Marandino, Francesca Rollo (Regina Elena Cancer Istitute);

19. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asian Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

20. Kuwait: Waleed Al-Jassar (Faculty of Medicine, Kuwait University), Rema'a Al-Safi (Maternity Hospital);

21. Lebanon: Muhieddine Seoud (The American University of Beirut Medical Center); Mali: Bakarou Kamate, Cathy Ndiaye (Hospital National DU Point G);

22. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Rubén López-Revilla, Claudia Magaña-León (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

23. Mozambique: Carla Carrilho (Eduardo Mondlane University);

24. New Zealand: Susan M Bigby, RW Jones, KL Fong, D Rowan, J Baranyai, L Eva (Middlemore Hospital);

25. Nigeria: A.A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba); Sani Malami, Ali Bala Umar (Faculty of Medicine, Bayero University);

26. Paraguay: Elena Kasamatsu, Antonio Leoploldo Cubilla, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción);

27. Philippines: Celia Ladines Llave, Jean Anne Toral (Cervical Cancer Prevention Center-CECAP, Cancer Institute (UP-CM-PGH)); Efren j Domingo, Maria Julieta V. Germar, Jerico Thaddeus, P. Luna, Arnold M. Fernandez, Carolyn Zalameda Castro, Roslyn Balacuit (University of the Philippine College of Medicine General Hospital);

28. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin); Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

29. Portugal: Eugenia Cruz (Centro Regional de Oncologia Coimbra, Instituto Português de Oncologia); Ana Félix, Jorge Manuel Soares (Instituto Portugues de Oncologia de Lisboa Francisco Gentil);

30. Senegal: Cathy Ndiaye, Nafissatou Ndiaye Ba, Victorino Mendes (HOGGY stands for Hôpital Général de Grand Yoff; DANTEC - Hôpital A. Le Dantec; FAC - Faculté de Médecine - Université Cheikh A. Diop);

31. Spain: Maria Alejo (Hospital General d'Hospitalet); Belén Lloveras (Hospital del Mar); Laia Alemany, F. Xavier Bosch, Ignacio Bravo, Vanesa Camón, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Ana Esteban, Jose M. Godínez, Yolanda Florencia, Joellen Klaustermeier, Nubia Muñoz, Beatriz Quirós, Maëlle Saunier, Cristina Rajo, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia – Catalan Institute of Oncology), August Vidal, Enric Condom (Hospital Universitari de Bellvitge), Jaume Ordi (Hospital Clínic), Julio Velasco, Cristina Pérez (Hospital San Agustín);

32. Taiwan: Chou Cheng-Yang (National Cheng Kung University Medical College,Taiwan Association of Gynecologic Oncologists); Tang-Yuan Chu (Buddhist Tzuchi Genral Hospital); Kuo-Feng Huang (Chi Mei Medical Center); Cheng Wen-Fang (National Taiwan University Hospital); Chih- Ming HO (Gynecologic Cancer Center, Cathay General Hospital);

33. The Netherlands: Wim Quint, , Anco C. Molijn, Daan T. Geraets, Nuria Guimera (DDL Diagnostic Laboratory); (Chris J.L.M Meijer (Vrije Universiteit Medical Center);

34. Turkey: Alp Usubutun (Hacettepe University); UK: Henry Kitchener (School of Medicine, University of Manchester); Robyn Davies (Manchester Royal Infirmary); Paul Cross (Queen Elizabeth Hospital, Sheriff Hill);

35. Uruguay: Adela Rosa Sica, Benedicta Caserta, Mabel Cedeira, Daniel Mazal, Guillermo Rodríguez (Laboratorio de Anatomía patológica del hospital de la Mujer, Montevideo);

36. USA: Wendy Cozen, Marc T. Goodman, Brenda Y. Hernández, Charles Lynch, Daniel B. Olson, Freda R. Selk (Cancer Center, Hawaii-Iowa);

37. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

3.HPV VVAP study group for vagina site, updated December 2013:

1. Argentina: Myriam Perrotta, Ana Jaen, Kevin Davies (Hospital Italiano de Buenos Aires);

2. Australia: Suzanne M. Garland, Sepehr N. Tabrizi (The Royal Women's Hospital, The University of Melbourne); Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

3. Austria: Elmar Armin Joura (University Hospital and Medical School); Josefine Stani and Reinhard Horvat, MD (Medical University and General Hospital Vienna, Austria);

4. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

5. Belarus: Halina Viarheichyk, Sitnikov Valeriy (Gomel State Medical University); Achynovich Searhei (Gomel Regional Clinical Oncological Hospital);

6. Brazil: Paula Maldonado, Gutemberg Leão Almeida, Isabel Val, Renata Fonseca, Roberto José Lima, Marcia Mannarino, Yara Furtado (Instituto de Ginecologia da Universidad Federal do Rio de Janeiro);

7. Chile: Rodrigo Prado, Carla Molina, Rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

8. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón, Nubia Muñoz (Instituto Nacional de Cancerología);

9. Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (The Fingerland Department of Pathology, Charles University in Prague, Faculty of Medicine and University Hospital Hradec Kralove);

10. Ecuador: Leopoldo Tinoco (Hospital Oncológico de Quito, Quito, Ecuador);

11. France: Christine Clavel, Philippe Birembaut, Véronique Dalstein (CHU de Reims, Laboratoire Pol Bouin / INSERM UMR-S 903, Reims, France); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

12. Germany: Karl Ulrich Petry, Alexander Luyten (Klinikum Wolfsburg); Michael Pawlita, Gordana Halec, Dana Holzinger (Department Genome Changes and Carcinogenesis, Heildelberg);

13. Greece: Theodoros Agorastos (Aristotle University of Thessaloniki);

14. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer; Instituto de Cancerologia Dr. Bernardo del Valle S);

15. India: Asha Jain (Cancer Prevention and Relief Society Raipur); Sushil K Giri (Regional Cancer Center, Cuttack); Maheep Bhalla (JLN Hospital & Research Center,BSP, Bhilai); Bharat Patel (Lab One Raipur); PSA Sarma (BSP Hospital);

16. Israel: Jacob Bornstein, Alejandro Livoff, Hector Itzhac Cohen (Western Galilee Hospital- Nahariya);

17. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asian Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

18. Kuwait: Dr. Waleed Al- Jassar. Faculty of Medicine, Kuwait University, Dr. Rema'a Al-Safi. Maternity Hospital, Kuwait;

19. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Rubén López-Revilla, Claudia Magaña-León (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

20. Mozambique: Carla Carrilho (Eduardo Mondlane University);

21. Nigeria: Adekunbiola A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba);

22. Paraguay: Elena Kasamatsu, Antonio Leoploldo Cubilla, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Instituto de Patología e Investigación);

23. Philippines: Celia Ladines Llave, Jean Anne Toral (Cervical Cancer Prevention Center, Cancer Institute); Efren Javier Domingo, Jericho Thaddeus P. Luna, Maria Julieta V. Germar, Arnold M. Fernandez, Carolyn Zalameda Castro, Roslyn Balacuit (University of the Philippine College of Medicine General Hospital);

24. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin); Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

25. Spain: Belén Lloveras (Hospital del Mar); August Vidal, Enric Condom (Hospital Universitari de Bellvitge); Jaume Ordi (Hospital Clínic); Julio Velasco Alonso, Cristina Pérez (Hospital San Agustín); Maria Alejo (Hospital General de l'Hospitalet, Barcelona); Laia Alemany, Francesc Xavier Bosch, Ignacio G. Bravo, Vanesa Camón, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Ion Espuña, Anna Esteban, José M. Godínez, Yolanda Florencia, Klaustermeier, Natividad Patón, Beatriz Quirós, Cristina Rajo, Maëlle Saunier, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia-Catalan Institute of Oncology);

26. Taiwan: Chou Cheng-Yang (National Cheng Kung University Medical College,Taiwan Association of Gynecologic Oncologists); Tang-Yuan Chu (Buddhist Tzuchi Genral Hospital); Kuo-Feng Huang (Chi Mei Medical Center); Cheng Wen-Fang (National Taiwan University Hospital); Chih-Ming HO (Gynecologic Cancer Center, Cathay General Hospital);

27. The Netherlands: Wim G.V. Quint, Anco C. Molijn, Daan T. Geraets, Núria Guimerà (DDL Diagnostic Laboratory); Chris J.L.M Meijer (Vrije Universiteit Medical Center);

28. Turkey: Alp Usubutun (Hacettepe University);

29. UK: Henry Kitchener, Godfrey Wilson (School of Medicine, University of Manchester); Paul Cross (Queen Elizabeth Hospital, Sheriff Hill);

30. Uruguay: Adela Rosa Sica, Benedicta Caserta, Mabel Cedeira, Daniel Mazal, Guillermo Rodríguez (Laboratorio de Anatomía Patológica del Hospital de la Mujer, Montevideo);

31. USA: Marc T. Goodman, Wendy Cozen (Cedars Sinai Medical Center, Los Angeles, California); Marc T. Goodman, Brenda Y. Hernández (Cancer Center, Hawaii); Charles F. Lynch, Daniel B. Olson, Freda R. Selk (Iowa);

32. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García

Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

33. The advisory committee members are: Chris J Meijer, Massimo Tommasino, Michael Pawlita, Wim Quint and Nubia Muñoz

4.**HPV VVAP study group for penile site, updated November 2014:**

1. Australia: Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

2. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

3. Chile: Rodrigo Prado, Carla Molina, rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

4. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón (Instituto Nacional de Cancerología);

5. Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (Faculty Hospital Hradec Kralove);

6. Ecuador: Leopoldo Tinoco (Hospital Oncológico Solca-Quito);

7. France: Christine Clavel, Philippe Birembaut, Veronique Dalstein (CHU de Reims, Laboratoire Pol Bouin/ INSERM UMR-S 903, REIMS); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

8. Germany: Michael Pawlita, Gordana Halec (Department Genome Changes and Carcinogenesis. Im Neuenheimer Feld 242. D-69120 Heildelberg);

9. Greece: Maria Tzardi (Medical School of University of Crete);

10. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer);

11. Honduras: Annabelle Ferrera (Escuela de Microbiología, Universidad Nacional Autónoma de Honduras);

12. India: Asha Jain (Cancer Prevention and Relief Society Raipur); Sushil K Giri (Regional Cancer Center, Cuttack); Maheep Bhalla (JLN Hospital & Research Center,BSP, Bhilai); Bharat Patel (Lab One Raipur); PSA Sarma (BSP Hospital); Ravi

Mehrotra, Mamta Singh (M.L.N Medical College, Allahabad);

13. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asan Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

14. Lebanon: Muhieddine Seoud (The American University of Beirut Medical Center);

15. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Claudia Magaña-León, Rubén López-Revilla (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

16. Mozambique: Carla Carrilho (Eduardo Mondlane University);

17. Nigeria: A.A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba); Sani Malami, Ali Bala Umar (Faculty of Medicine, Bayero University);

18. Paraguay: Antonio Leopoldo Cubilla, Elena Kasamatsu, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción);

19. Philippines: Efren j Domingo, Maria Julieta V. Germar, Jerico Thaddeus, P. Luna, Arnold M. Fernandez, Carolyn Zalameda Castro, Roslyn Balacuit (University of the Philippine College of Medicine General Hospital);

20. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin); Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

21. Portugal: Eugenia Cruz (Centro Regional de Oncologia Coimbra, Instituto Português de Oncologia); Ana Felix, Jorge Manuel Soares (Instituto Portugues de Oncologia de Lisboa Francisco Gentil);

22. Senegal: Cathy Ndiaye, Nafissatou Ndiaye Ba, Victorino Mendes (HOGGY stands for Hôpital Général de Grand Yoff ; DANTEC - Hôpital A. Le Dantec; FAC - Faculté de Médecine - Université Cheikh A. Diop);

23. Spain: Enrique Poblet (Hospital General Universitario de Albacete); August Vidal, Enric Condom (Hospital Universitari de Bellvitge); Lluís Eleuteri Pons Ferré, Patrícia Escrivà Beltri, Marylene Lejeune (Hospital de Tortosa Verge de La Cinta); Belén Lloveras, Emili Masferrer (Hospital del Mar); Julio Velasco Alonso, Cristina Pérez (Hospital San Agustin); Maria Alejo (Hospital General de L'Hospitalet); Laia Alemany, Francesc Xavier Bosch, Ignacio G. Bravo, Vanesa Camón, Gabriel Capellà, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Anna Esteban, José M. Godínez, Yolanda Florencia, Joellen Klaustermeier, Núbia Muñoz, Beatriz Quirós, Cristina Rajo, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia-Catalan Institute of Oncology);

24. The Netherlands: Wim G.V. Quint, Anco C. Molijn, Daan T. Geraets, Núria Guimerà (DDL Diagnostic Laboratory); Chris J.L.M Meijer (Vrije Universiteit Medical Center);

25. UK: Ray Lonsdale (Norfolk & Norwich University Hospital NHS Foundation Trust);

26. USA: Wendy Cozen, Marc T. Goodman, Brenda Y. Hernández, Charles Lynch, Daniel B. Olson, Freda R. Selk (Cancer Center, Hawaii-Iowa);

27. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

28. The advisory committee members are: Chris J Meijer, Massimo Tommasino, Michael Pawlita, Wim Quint and Nubia Muñoz

5.**HPV VVAP study group for anal site, updated June 2013:**

1. Australia: Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

2. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

3. Bosnia Herzegovina: Ermina Iljazovic (University Clinical Center Tuzla BiH);

4. Chile: Rodrigo Prado, Carla Molina, rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

5. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón, Nubia Muñoz (Instituto Nacional de Cancerología);

6. Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (Faculty Hospital Hradec Kralove);

7. Ecuador: Leopoldo Tinoco (Hospital Oncológico Solca-Quito);

8. France: Christine Clavel, Philippe Birembaut, Veronique Dalstein (CHU de Reims, Laboratoire Pol Bouin/ INSERM UMR-S 903, REIMS); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

9. Germany: Karl Ulrich Petry, Alexander Luyten (Klinikum Wolfsburg); Michael Pawlita, Gordana Halec, Dana Holzinger (Department Genome Changes and Carcinogenesis, Heildelberg);

10. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional

del Cáncer);

11. Honduras: Annabelle Ferrera (Universidad Autónoma de Honduras); Odessa Henríquez and Silvia Portillo (Instituto Nacional Cardiopulmonar en Tegucigalpa).

12. India: Asha Jain (Cancer Prevention and Relief Society Raipur); Sushil K Giri (Regional Cancer Center, Cuttack); Maheep Bhalla (JLN Hospital & Research Center,BSP, Bhilai); Bharat Patel (Lab One Raipur); PSA Sarma (BSP Hospital); Ravi Mehrotra, Mamta Singh (M.L.N Medical College, Allahabad);

13. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asan Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

14. Mali: Bakarou Kamate, Cathy Ndiaye (Hospital National DU Point G);

15. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Rubén López-Revilla, Claudia Magaña-León (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

16. Nigeria: Adekunbiola A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba); Sani Malami, Ali Bala Umar (Faculty of Medicine, Bayero University);

17. Paraguay: Elena Kasamatsu, Antonio Leopoldo Cubilla, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Instituto de Patología e Investigación);

18. Poland: Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

19. Portugal: Ana Felix (Instituto Portugues de Oncologia de Lisboa Francisco Gentil);

20. Senegal: Cathy Ndiaye, Nafissatou Ndiaye Ba, Victorino Mendes (HOGGY stands for Hôpital Général de Grand Yoff ; DANTEC - Hôpital A. Le Dantec; FAC - Faculté de Médecine - Université Cheikh A. Diop);

21. Slovenia: Mario Poljak, Boris Pospihalj, Pavle Košorok (Institute of Microbiology and Immunology, Ljubljana);

22. Spain: Maria Alejo (Hospital General de l'Hospitalet); Mar Iglesias, Belén Lloveras, David Parés (Hospital del Mar); Laia Alemany, Francesc Xavier Bosch, Ignacio G. Bravo, Vanesa Camón, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Ion Espuña, Anna Esteban, José M. Godínez, Yolanda Florencia, Joellen Klaustermeier, Nubia Muñoz, Nati Patón, Beatriz Quirós, Cristina Rajo, Maëlle Saunier, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia-Catalan Institute of Oncology, Barcelona); August Vidal, Enric Condom (Hospital Universitari de Bellvitge, Barcelona); Julio Velasco Alonso, Cristina Pérez (Hospital San Agustín);

23. The Netherlands: Wim G.V. Quint, Anco C. Molijn, Daan T. Geraets, Núria Guimerà (DDL Diagnostic Laboratory); Chris J.L.M Meijer (Vrije Universiteit Medical Center);

24. UK: Henry Kitchener, Godfrey Wilson (School of Medicine, University of Manchester);

25. USA: Wendy Cozen (Los Angeles, CA); Marc T. Goodman, Brenda Y. Hernández (Hawaii); Charles Lynch, Daniel B. Olson, Freda R. Selk (Iowa); Edyta C. Pirog (New York Hospital - Cornell Medical Centre).

26. The advisory committee members are: Chris J Meijer, Massimo Tommasino, Michael Pawlita, Wim Quint, and Nubia Muñoz

IJC
International Journal of Cancer

Differential HPV16 variant distribution in squamous cell carcinoma, adenocarcinoma and adenosquamous cell carcinoma

S. Nicolas-Parraga [1,2], L. Alemany [1,2], S. de Sanjose [1,2,3], F.X. Bosch 1,2, I.G. Bravo [1,4] and on behalf of the RIS HPV TT and HPV VVAP study groups

[1] Infections and Cancer Unit, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), Barcelona, Spain
[2] Bellvitge Institute of Biomedical Research (IDIBELL), Barcelona, Spain
[3] CIBER in Epidemiology and Public Health (CIBERESP), Barcelona, Spain
[4] UMR MIVEGEC, National Center for Scientific Research (CNRS), Montpellier, France

Human Papillomavirus 16 (HPV16) causes 70% of invasive cervical cancers (ICC) worldwide. Interaction between HPV16 genetic diversity, host genetics and target tissue largely determine the chances to trigger carcinogenesis. We have analyzed the differential prevalence of viral variants in 233 HPV16-monoinfected squamous (SCC), glandular (ADC) and mixed (ADSC) ICCs from four continents, assessing the contribution of geographical origin and cancer histology. We have further quantified the contribution of viral variants and cancer histology to differences in age at tumor diagnosis. The model fitted to the data explained 97% of the total variance: the largest explanatory factors were differential abundance among HPV16 variants (78%) and their interaction with cancer histology (9.2%) and geography (10.1%). HPV16_A1-3 variants were more prevalent in SCC while HPV16_D variants were increased in glandular ICCs. We confirm further a non-random geographical structure of the viral variants distribution. ADCs were diagnosed at younger ages than SCCs, independently of the viral variant triggering carcinogenesis. HPV16 variants are differentially associated with histological ICCs types, and ADCs are systematically diagnosed in younger women. Our results have implications for the implementation of cervical cancer screening algorithms, to ensure proper early detection of elusive ADCs.

Introduction

Invasive cervical cancer (ICC) is the second most common cancer affecting women,[1] being responsible for approximately 266,000 deaths per year worldwide (http://globocan.iarc.fr/Default.aspx). Around 88% of the global burden occurs in developing countries: approximately 53.000 in Africa, almost 32.000 in Central-South America and Caribbean and ca. 160.000 in Asia.[1] Persistent infections by oncogenic Human Papillomaviruses (HPVs) are the etiologic cause of virtually all cervical cancers.[2] This well-established connection between HPVs infection and disease is observed for the most prevalent histological presentations of ICC, namely squamous cell carcinomas (SCC), adenocarcinomas (ADC) and adenosquamous carcinomas (ADSC)[2] (https://www.iarc.fr/en/publications/pdfs-online/pat-gen/bb4/bb4-chap5.pdf).

Cervical SCC is an epithelial invasive cancer that affects the squamous cells covering the outer surface of the cervix, i.e. the ectocervix. SCCs most often arise at the squamo-columnar junction between the non-keratinized stratified

Key words: papillomavirus infection and cancer, squamous cell carcinoma, adenocarcinoma, adenosquamous cell carcinoma, histological type, virus-host interactions, viral diversity

Abbreviations: ADC: adenocarcinoma; ADSC: adenosquamous cell carcinoma; AIC: Akaike Information Criterion; AS: Asia; CSA: Central-South America; EUR: Europe; FFPE: formalin-fixed paraffin embedded; GLM: generalized linear model; HPV: human papillomavirus; ICC: invasive cervical carcinoma; LCR: long control region; SCC: squamous cell carcinoma; URR: up-stream regulatory region
Additional Supporting Information may be found in the online version of this article.

Grant sponsor: This work was financially supported by a Fundacion Dexeus for Women's Health grant to IGB, the Age`ncia de Gestio d'Ajuts Universitaris i de Investigacio, AGAUR, Generalitat de Catalunya (2014SGR1077 to XFB), and the Fondo de Investigaciones Sanitarias (FI12/00142 to XFB and SNP). FXB has received institutional funding support and has received honoraria and/or consultation feed from GlaxoSmithKline, Sanofi Pasteur MSD and MSD. SdS has received institutional funding support from Merck. Funders had no role in study design, data collection and analysis, and publication. The authors want to thank all participants at the RIS HPV TT and HPV VVAP collaborating centres listed in Table S14.
DOI: 10.1002/ijc.30636
History: Received 26 July 2016; Accepted 20 Jan 2017; Online 10 Feb 2017
Correspondence to: Ignacio G. Bravo, National Center for Scientific Research (CNRS), Maladies Infectieuses et Vecteurs: Ecologie, Genetique, Evolution et Contr^ole (MIVEGEC), UMR CNRS 5290, IRD 224, UM, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France, E-mail: ignacio.bravo@ird.fr; Tel. 133 467 41 5123

What's new?

Interaction between Human Papillomavirus 16 genetic diversity, host genetics and target tissue largely determines the odds of HPV16 triggering invasive cervical cancers (ICCs), but the mechanisms remain unclear. Our study assessed HPV16 variant diversity in three ICC histological types in European, Central-South American, Asian and African samples. Different viral variants displayed different prevalence depending on geographical origin and histological cancer type. Genuine differences in HPV16 lineage prevalence explained more than 70% of all variance in the viral lineage distribution, with the interaction of geographical origin and histological cancer type with HPV16 variants together accounting for 20% of the data variance.

squamous epithelium of the ectocervix and the non-ciliated simple columnar epithelium of the endocervix.[3] Instead, most cervical ADCs originate mostly from endocervix glandular precursor lesions.[4] Finally, cervical ADSC is a mixed histological type amalgamating malignant glandular and squamous components consisting of intermingled ADC and SCC.[5] Given the mixed nature of ADSC, there has historically been some controversy with this diagnosis. It was considered as a subtype of ADC, but it has been classified as an independent entity, as the ADSC histological presentation is a clinicpathological factor that influences prognosis.[6] After radical hysterectomy, both ADC and ADSC present a poorer prognosis than SCC,[7] with nearly 10–20% difference in 5-year overall survival rates.[8]

The most common presentation of ICC is SCC, accounting for 80–85% of all ICC cases,[4] compared to 10–15% of ADC and 2–3% ADSC.[9,10] However, the epidemiology of ICC seems to be changing in the last years. Public health interventions and efforts in cervical cancer screening have proven to be an effective approach to reducing the cervical cancer burden through early detection of precursor lesions.[11–13] The differential anatomical location of the precursor lesions of SCC (essentially the ectocervix) and the ADC (essentially the endocervix) could be partly responsible for the increased success at early detection of SCC compared to ADC,[14,15] as the endocervix is more likely to be improperly sampled during routine screening sampling. Indeed, cervical screening has lead to a decrease in SCC incidence mainly in high income countries such United States, Canada, New South Wales, most European countries and in some Asian countries.[16,17] Certain exceptions to this trend are remarkable, as it is the case of Ireland.[18–20] But the overall trend seems to be the opposite for ADC and for ADSC, which show an increment of both relative and absolute incidence in certain developed countries,[14,21,22] especially among young adult women.[5,20,23,24] The forces driving this increase in ADC and ADSC detection remain nevertheless unclear.

Not all HPVs are equally associated with the different histological presentation of ICCs. A clear trend of differential HPV prevalence is obvious between SCC and glandular ICC (i.e., ADC and ADSC);[25,26] HPV16 is associated with 55–59.3% of SCC cases and with 33–36.3% of ADC cases, while HPV18 is associated with 12–13% SCCs and 37–56% ADCs.[25–27] Globally, SCCs are closely related to HPV16 and its close relatives (HPV31, 35 and 52, members of Alphapapillomavirus species 9) whereas ADCs and ADSCs are more closely to HPV18 and its close relatives (HPV39, 45 and 59, members of the Alphapapillomavirus species 7).[27–30] Thus, oncogenic HPVs are differentially associated with the various histological presentations of ICC. This specificity is reported at the level of type and at the level of variant.[2,31–34] Indeed, within HPV16, the HPV16_A1-3 variants may show an increased prevalence in SCCs, while HPV16_D and to a lesser extent HPV16_A4, B and C variants might be more prevalent in ADCs.[2,33,35]

In this study, we explore the association between the differential prevalence of HPV16 lineages in SCC, ADSC and ADC from Europe, Central-South America, Asia and Africa.

Materials and Methods

Samples

Cervical samples analyzed in our study stem from a Formalin Fixed Paraffin Embedded (FFPE) repository from the Catalan Institute of Oncology (ICO), Barcelona, Spain.[36] All samples were tested for the presence of tumor tissue as well as for the presence of PV DNA using the SPF$_{10}$-LiPA$_{25}$ protocol (version 1; Laboratory Biomedical Products, Rijswijk, the Netherlands). For the purpose of our study, only HPV16-monoinfected samples were selected from three different histological ICC types: SCC, ADC and ADSC. The selection of ADC and ADSC samples was geographically paired to the SCC samples. A total 118 samples were selected for SCC, 120 for ADC and 53 for ADSC from Europe, Central-South America, Asia and Africa (Table 1).

PCR and sequencing

Briefly, four 5 mm paraffin sections were systematically obtained from each block (sandwich method). The first and last sections were used for histopathological assessment, and the second and third sections were used for analysis of HPVs DNA.[37] A blank paraffin section was cut and processed in-between specimens to control for carryover contaminations in addition to routine controls. DNA was released by incubation overnight at 56°C with 250 mL proteinase K buffer (10 mg/mL proteinase K, 50 mM Tris-HCl, pH 8.0) followed by incubation at 95°C for 8 minutes to inactivate proteinase K, and stored at −20°C until use. A 1:10 water dilution of this DNA solution was used for downstream processes.

Table 1. Histological cancer type and geographical distribution of amplified and classified samples

	EUR-CSA-AS-AF SAMPLES			
	Initial	Amplified	Classified[1]	Unclassified[2]
SCC	118	111	109	2
ADSC	53	32	29	3
ADC	120	97	95	2
Total	291	240	233	7

The table shows the number of initial, amplified, classified and unclassified samples according to histological cancer type.
Abbreviations: ADC, adenocarcinoma; SCC, Squamous Cell Carcinoma; ADSC, adenosquamous cell carcinoma
[1] Samples classified in HPV16_A1-3, A4, B, C and D variants.
[2] Samples that are classified basal to a particular HPV16 variant cluster (i.e., basal to HPV16_A1-3 and A4 variants) and samples not classified with likelihood values below 0.6 within any HPV16 variant cluster.

The upstream regulatory region (URR), and the E6 and L2 HPV16 genes were chosen as amplification targets (Table S1). All PCR reaction and conditions were performed as previously described.[38] All PCR products were Sanger-sequenced in both strands at Genoscreen (Lille, France). For those samples in which the target was difficult to amplify, PCR conditions were adjusted as follows: 95°C for 10 min; 40 cycles of 30 s at 94°C, 50 s at 56°C, 30 s at 72°C; plus 7 min final extension at 72°C.

Phylogenetic analyses

Phylogenetic relationships of the E6, L2 and URR sequences generated from the samples in the global context of HPV16 genetic variability were inferred using an Evolutionary Placement Algorithm on RAxML_v7.2.8 with the GTR+C_4 model,[39,40] as previously described.[38] The reference tree was constructed using 109 HPV16 full-genome sequences alignment (Fig. S1). Sequences retrieved from our samples were incorporated into the reference alignment with MAFFT_v7 and their phylogenetic placement was individually inferred with the -f v command in RAxML.[41] The results were integrated for all nodes within a variant lineage, and the threshold for assigning each sequence to a specific variant lineage was set to 0.60.

Statistical analyses

A Generalized Linear Model (GLM) with a Poisson distribution and a log link function was used to analyze the relationships between HPV16 variant prevalence and the two variables of interest: histological cancer type and sample geographical origin, as well as with the interaction of both variables. HPV16 variant distribution was statistically analyzed by means of Fisher's test and Prevalence Ratios (PR) were calculated. PRs of HPV16 variants among histological cancers between Europe and Central-South America or Asia were estimated using Poisson multivariate regression model with robust variance. The different HPV16 variant lineages (i.e., HPV16_A1-3, A4, B, C and D) were used as dichotomous variables.

An analysis of association between age at tumor diagnosis and histological cancer type and sample geographical region was performed through a two-way ANOVA and Wilcoxon Mann–Whitney test. All analyses were performed using R in RStudio v0.98.939 (RStudio, Inc. https://www.rstudio.com/products/RStudio/).

Results

Dataset construction, study design and data collection bias, and explanatory power

From the initial 118 SCC, 120 ADC and 53 ADSC we were able to amplify 111 SCC, 97 ADC and 32 ADSC, covering 28 different countries (Table 1, Table S2). Sequences were subsequently classified as belonging to HPV16_A1-3, A4, B , C and D variants. The final dataset included 109 SCC, 95 ADC and 29 ADSC, (Table1, Table S3). We assessed the impact of cancer histology and geographical origin on the differential prevalence of HPV16 variants, by applying a GLM, initially performed including all histologies and all geographies (Europe, Central-South America, Asia and Africa) (Table S4). The model reached a good fit to the data, capturing above 96% of the variance in the original data (Table S4). As our work did not include samples from North America, we performed two additional models incorporating data from HPV16 SCC and ADC isolates from United States communicated by Mirabello and colleagues.[35] Both GLMs, the one including all histologies (SCC, ADSC and ADC) (Table S5) and the one including the two cancer presentations shared with Mirabello and colleagues (SCC and ADC) (Table S6)[35] fitted also well the data (Tables S5 and S6). A more homogeneous variance distribution was observed in the model that included only SCC and ADC cases (Table S6). Despite our efforts for a balanced representation of all three histologies and all four geographical origins studied in this work, the low number of ADSC and African samples may have been responsible for the spurious explanatory power of the factors Histology and Geography in the global analyses (9.8% and 9.6% respectively in Table S4). We confirmed thus the overall results by performing all analyses after excluding the underrepresented levels "ADSC" as histological cancer type and "Africa" as geographical region (Table 2). This model analysis using our cleaner, best data showed that the dataset was well balanced for both histology (accounting only for 0.31% of the variance, $p = 0.275$) and geography (accounting only for 0.69% of the variance, $p = 0.267$) (Table 2). The model performed very well on these refined data, as it was able to fit >97% of the data variance (Table 2).

Prevalence of HPV16 variants depends largely on variant biology, and additionally on cancer histology and on the geographical origin of the sample

Our data reflected the different prevalence of HPV16 variants in distinct histological cervical cancer types and geographical regions. Globally we observed the highest prevalence values for HPV16_A1-3 in SCC (from 76.9% to 97% for different

Table 2. Generalized Linear Model (GLM) for the main two histologies (squamous cell carcinoma and adenocarcinoma) and the best represented geographic origins (Europe, Central-South-America and Asia)

	Df	Res. Dev.	Df	Res. Dev.	% exp. Dev.	p-Value
NULL			29	382.41		
Variant	4	298.3	25	84.15	78	<0.0001
Histology	1	1.192	24	82.96	0.31	0.2750
Geography	2	2.635	22	80.32	0.69	0.2677
Variant:histology	4	35.193	18	46.13	9.20	<0.0001
Variant:geography	8	38.598	10	6.53	10.1	<0.0001
Histology:geography	2	3.789	8	2.74	1	0.1504
Variant:histology:geography	8	2.742	0	0	0.7	0.9495

Abbreviations: Df, degrees of freedom; Res. Dev., residual deviance, % exp.dev., percentage of data deviance explained by the corresponding factor or factor combination.
Data should be read as follows (using "Variant" as an example): the factor "Variant" has five levels (HPV16_A1-3, A4, B, C and D) and thus contributes with four degrees of freedom; it explains in the model 298.3 units of deviance, i.e. 78.0% of the whole deviance in the original data; the probability of a factor to explain at random this proportion of the data deviance is below 0.0001.

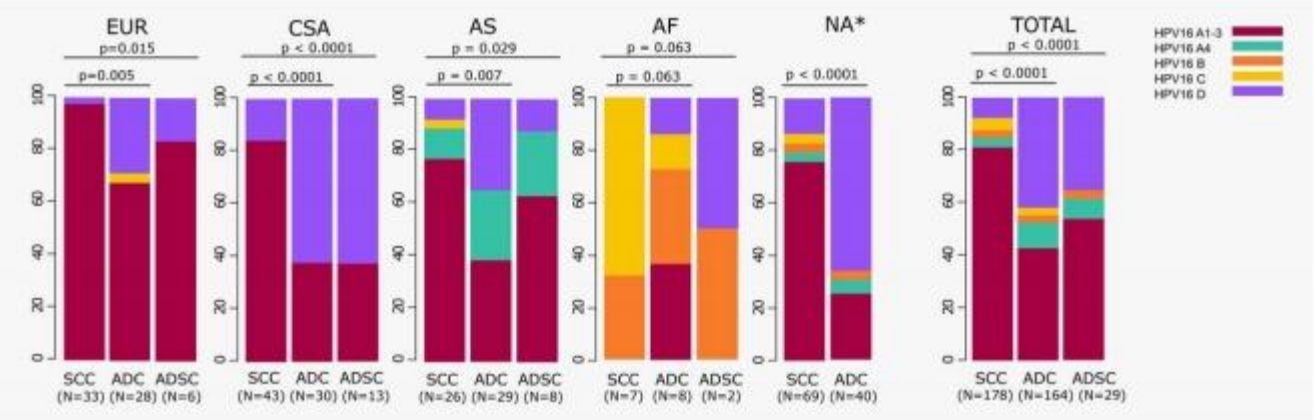


Figure 1. Distribution of HPV16_A1-3, A4, B, C and D variants depending on geographical regions and histological cancer type. For each combination of geography and histology the number of samples is given in parentheses. For each geographical origin, the result of Fisher's test assessing homogeneity for variant prevalence values between the three cancer histologies is provided (e.g., for CSA the H0 hypothesis of the variant prevalence values being similar in SCC, ADSC and ADC is rejected with p value below 0.0001). Abbreviations: A1-3, HPV16_A1, HPV16_A2 and HPV16_A3; A4, HPV16_A4; B, HPV16_B; C, HPV16_C; D, HPV16_D; SCC, squamous cell carcinoma; ADSC, adenosquamous cell carcinoma; ADC, adenocarcinoma; EUR, Europe; CSA, Central-South America; AS, Asia; AF, Africa. *Data for North America were extracted from Mirabello et al. 2016. [Color figure can be viewed at wileyonlinelibrary.com]

geographical regions) and large variations in HPV16_D variant prevalence as a function of geography in ADC (from 28.6 to 63.3% for different geographical regions) and ADSC (from 12.5 to 61.5% for different geographical regions) (Table S7). The data showed increased prevalence of HPV16_A1-3 in Europe (from 67.9% to 97% for all histologies), of HPV16_D in Central-South America (61.5–63%, mainly for ADC and ADSC histologies), of HPV16_A4 in Asia (from 11.5 to 27.6% for all histologies) and of HPV16_B and C variants mostly restricted to Africa (from 28% to 66.7% and from 12.3 to 37.5% for SCC and ADC) (Fig. 1; Table S7). Figures 1 and 2 and Table S7 include data from Mirabello and colleagues[35] to incorporate North America in the analyses. For these samples we observed an increased HPV16_A1-3 variants in SCC (75.4%) and an enhanced prevalence of HPV16_D variants for ADC (67.5%), largely dominated by HPV16_D3 (Table S8).

Different HPV16 variants displayed major differences in prevalence, and that such differences explain the largest fraction of the total variance (49.7% for the complete data, $p < 0.0001$, Table S4; 78% for the most reliable data, $p < 0.0001$, Table 2). The HPV16_A1-3 variant was overall the most prevalent lineage, with a global prevalence of 63.1% (Table S7). However, important differences in variant prevalence depended on the geographical origin of the samples and on the histological presentation of the cervical cancer. The factors Variant and Geography, and their interaction accounted together for 20.2% of the total variation in the complete dataset (Table S4) and for 10.1% of the total variation in the filtered dataset (Table 2). The HPV16_A1-3

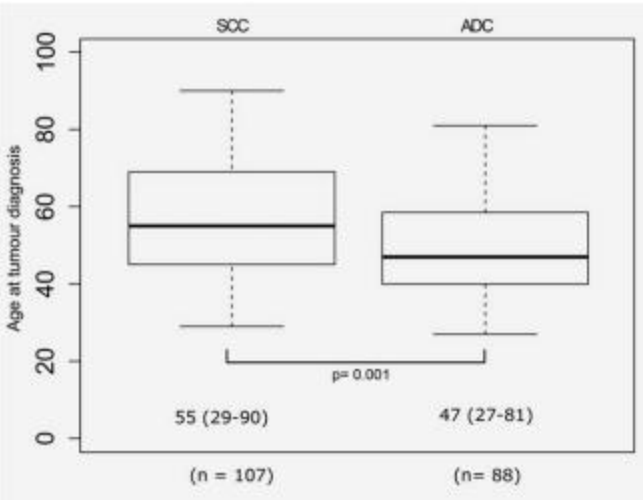


Figure 2. Age at tumor diagnosis for HPV16 single infected squamous cell carcinoma (SCC) and adenocarcinoma (ADC): For each dataset, the bar represents the median, the box encompasses the 25–75% percentiles. Numbers below each graph indicate the median and the range (1.5 3 Inter-quartile). Numbers in parentheses at the bottom indicate sample size for each location.

variant showed an evident decreasing trend in prevalence in the different continents: 83.6% in Europe, 61.2% in Central-South-America, 57.1 in Asia, and 17.6 in Africa (Table S7). The interaction Variant*Histology accounted for 7.3% of the total variation in the complete dataset (Table S4) and for 9.2% of the total variation in the filtered dataset (Table 2). The decreasing trend for the HPV16_A1-3 variant in different cancer histologies was also obvious: it accounted for 80.7% of all SCCs, 51.7% of all ADSCs and 46.3% of all ADCs (Table S7). Results obtained with the GLMs were validated using a Fisher's test after stratifying by cancer histology and by geographic origin. These tests further confirmed the significant difference in prevalence distribution of HPV16 variants within the same cancer type between geographical regions (for SCC, $p = 0.013$; and for ADC, $p < 0.0001$) (Table 3, Fig. 1), as well as the different prevalence of HPV16 variants within the same geographic region between histologic presentations (for Europe, $p = 0.005$; for Central and South America, $p < 0.0001$ and for Asia, $p = 0.007$) (Table 3). Fisher's test for the complete dataset (including ADSC in histology and Africa in geography) confirmed that variant prevalence was

Table 3. HPV16_A1-3, A4 B, C and D variant distribution analysis by the two main histologies (SCC and ADC) within the best represented geographic origins (Europe, Central-South America and Asia).

		EUR		CSA		AS				
Histology	Variants	n	%	n	%	n	%	Total n	Total %	Fisher test
SCC	A1-3	32	97	36	83.7	20	76.9	88	86.3	0.013
	A4	0	0	0	0	3	11.5	3	2.9	
	B	0	0	0	0	0	0	0	0	
	C	0	0	0	0	1	3.8	1	1	
	D	1	3	7	16.3	2	7.7	10	9.8	
	Sub-total	33	100	43	100	26	100	102	100	
ADC	A1-3	19	67.9	11	36.6	11	37.9	41	47.1	<0.0001
	A4	0	0	0	0	8	27.6	8	9.2	
	B	0	0	0	0	0	0	0	0	
	C	1	3.6	0	0	0	0	1	1.1	
	D	8	28.6	19	63.3	10	34.5	37	42.5	
	Sub-total	28	100	30	100	29	100	87	100	
Total	A1-3	51	83.6	47	64.4	31	56.4	129	68.3	<0.0001
	A4	0	0	0	0	11	20	11	5,8	
	B	0	0	0	0	0	0	0	0	
	C	1	1,6	0	0	1	1.8	2	1.1	
	D	9	14.8	26	35.6	12	21.8	47	24.9	
	Total	61	100	73	100	55	100	189	100	
	Fisher test	0.005		<0.0001		0.007		<0.0001		

The contingency table shows HPV16 variants distribution for the 189 samples analyzed, according to geographical region and anatomical location. Differences in variant prevalence between anatomical sites within a given geographical region are given through Fisher's test values (columns). Differences in variant prevalence between geographical regions, within an anatomical location are given through Fisher's test values (rows).
Abbreviations: A1-3, HPV16_A1, HPV16_A2 and HPV16_A3; A4, HPV16_A4; B, HPV16_B; C, HPV16_C; D, HPV16_D; SCC, squamous cell carcinoma; ADC, adenocarcinoma; EUR, Europe; CSA, Central-South America, AS, Asia.
Data should be read as follows for Fisher's test: (Using "ADC" as an example): H_0 hypothesis of the variant prevalence values being similar for EUR, CSA and AS is rejected with p value below 0.0001; (Using "AS" as an example): H_0 hypothesis of the variant prevalence values being similar for SCC and ADC is rejected with p value 0.007.

Table 4. Prevalence ratio (PR) of HPV16 variants by the two main histologies (SCC and ADC) for the best represented geographic origins (Europe, Central-South-America and Asia)

	Variants	SCC ^{Ref} (n var)	ADC (n var)	PR	PR Wald-test	95% CI	Fisher-test
EUR	A1-3	32	19	Ref	Ref	Ref	0.004
	A4/B/C/D	1 (-/-/-/1)	9 (-/-/1/8)	2.42	0.002	1.60-3.65	
CSA	A1-3	36	11	Ref	Ref	Ref	<0.0001
	A4/B/C/D	7 (-/-/-/7)	19 (-/-/-/19)	3.12	<0.0001	1.77-5.51	
AS	A1-3	20	11	Ref	Ref	Ref	0.006
	A4/B/C/D	6 (3/-/1/2)	18 (8/-/-/10)	2.11	0.004	1.25-3.58	

PR for each stratum is accompanied by Wald's test result and score confidence intervals (95%CI) and by the Fisher's test for the null hypothesis that the variant prevalence values are similar for SCC and ADC.
Abbreviations: Ref, reference histology; ADC, adenocarcinoma, SCC, squamous cell carcinoma; ADSC, adenosquamous cell carcinoma; EUR, Europe, CSA, Central-South America; AS, Asia, AF, Africa; A1-3, HPV16_A1-A3 variants; A4, HPV16_A4 variants; D, HPV16_D variants; PR, prevalence ratio.
Data should be read as follows (using CSA as an example): PR shows 3.12 (95% 1.77–5.51) times higher prevalence of HPV16_D variants in ADC than in SCC, the Reference group (i.e., 11 over 36 compared to 19 over 7). The Wald's test shows that this PR value is significantly different from one. The Fisher's test shows that the probability of obtaining this shift in PR by chance given the sample sizes of the two groups being compared is lower than 0.0001.

different between squamous and glandular ICCs (SCC, $p < 0.0001$; ADC, $p < 0.0001$) as well as for samples from the same geographical regions (Europe, $p = 0.015$; Central-South America, $p < 0.0001$; Asia, $p = 0.029$) (Table S5 and Fig. 1).

Cervical cancers were associated with different HPV16 variants depending on the squamous or glandular nature of the lesions as well as on the geographical origin of the samples (Fig. 1, Table S7). The mixed presentation ADSC displayed somehow intermediate features between SCC and ADC with regards to the viral lineages present. Compared to SCC the decrease in HPV16_A1-3 in both ADC and ADSC was accompanied by an increase in HPV16_D and of HPV16_A4 variants, depending on geography. Specifically, we observed an increase of HPV16_D in Central-South America (16.3% and 63.3% for SCC and ADC respectively), a unique presence of HPV16_A4 in Asia (11.5% and 27.6% for SCC and ADC respectively), a low frequency of HPV16_B and HPV16_C outside Africa (one HPV16_C in Asia (1.6%); one HPV16_C in Europe (1.5%) and four HPV16_B and one HPV16_C in North America (4.6%)), and a decreased presence of A and D variants in Africa (overall 29%), although sample size in Africa is smaller than in other geographical regions (Table S7, Fig. 1). The estimated ratios between prevalence values for HPV16 variants after stratifying by histology and geography confirmed the trend of the significant decrease in prevalence of HPV16_A1-3 and the increase of non-HPV16_A1-3 variants in SCC compared to ADC in Asia (2.11 fold increase, $p = 0.006$), Central-South America (3.12 fold increase, $p < 0.0001$) and Europe (2.42 fold increase, $p = 0.004$) (Table 4). Similar results were obtained when the full dataset included the data from the less represented ADSC and African samples (Table S9). No values for Africa could be calculated as PR are estimated with integer data.

ADC and ADSC are diagnosed in younger patients

Age at diagnosis and prognosis has been shown to differ between squamous and glandular cervical cancers. Indeed in our dataset, we confirmed that ADCs are diagnosed in significantly younger women than SCCs (respectively 47 ± 13.3 and 55 ± 16.3 years of age at diagnosis, median and median absolute deviation; $p = 0.001$, Wilcoxon Mann–Whitney test) (Fig. 2, Table S10). Similar results were obtained either applying a GLM (Table S11) or a three-way ANOVA (Table S12). We have further tried to assess whether the differences in prevalence of viral variants in different histologic presentations of cervical cancer were also associated with differences in age at cancer diagnosis. Our dataset provided with statistical power for analyzing only the two more frequent variants, with contrasting results (Table S13): while ADCs were diagnosed significantly earlier than SCCs for HPV16_A1-3 (56 ± 19.2 vs. 46.5 ± 13.3; $n = 124$; $p = 0.004$) we did not detect differences in age at diagnosis between SCCs and ADCs for HPV16_D (46 ±9.6 vs. 47.5 ±10.3; $n = 46$; $p = 0.862$). This differential behavior of the variable age at diagnosis was consistent with the explanatory power for the factor Variant and for the interaction Variant*Histology found in the GLM results (Table S11).

Discussion

In our study, we have assessed the HPV16 variant diversity in a comprehensive set ($n = 240$) of HPV16-monoinfected cervical ADCs, ADSCs and SCCs, in samples originating from Europe, Central-South America, Asia and Africa. We show that different viral variants display different prevalence depending on the geographical origin of the samples and on the histologic cancer type. The main novelty of our study is that we have been able to quantify for the first time the relative contribution of each factor to the uneven HPV16 variant prevalence. With a balanced dataset, we observe that genuine differences in prevalence between HPV16 lineages explain >70% of data variance, while the geographical origin and histological cancer type interaction with HPV16 variants combined account roughly for around 20% of all variance in viral lineage distribution. The main strength of our study is

the epidemiologic design, as we have restricted ourselves to well-characterized invasive cancer cases, analyzing the hitherto largest collection of HPV16-monoinfected SCCs, ADSCs and ADCs so far.

Genetic variation within HPV16 has been widely studied, with an interest in SCC,[38,42–45] as this histological type remains the most prevalent ICC.[16,20] A number of studies had addressed other cancer histologies but had focused on data from a restricted geographic origin.[2,34,35,46,47] Globally, our results confirm and expand previous reports.[34,35] We communicate an increased prevalence of HPV16_D variants in ADC and ADSC compared to SCC, that had been reported in studies using samples from United States (38% of 21 ADCs compared to 3% of 37 SCCs;[2] 41.7% of 24 ADCs compared to 2.4% of 42 SCCs;[33] 67.5% of 40 ADCs compared to 15.9% of 69 SCCs[35]) and from Spain (85.7% from 7 glandular pathologies compared to 28.6% from 7 SCCs[32]). We further describe an increased prevalence of HPV16_A1-3 variants in SCC compared to ADC or ADSC, as previously reported in two American studies (86.8% prevalence in 38 SCCs compared to 57.1% prevalence in 21 ADCs;[2] and 75.4% prevalence in 69 SCCs compared to 25% prevalence in 40 ADCs[35]), and in other geographically more extended works (60% of 98 SCC compared to 42% in ADC).[48] In addition, we describe an increment of HPV16_A4 variants in glandular cancer types, 28% for ADC and 25% for ADSC, as reported in other studies including African, Central-South American and Asian isolates (18% of 50 ADC).[48] Regarding variation in HPV16 lineage prevalence depending on the geographical origin of the samples, our results largely confirm the best data available[44,49] showing a large dominance of HPV16_A1-3 variants in Europe, the virtually exclusive presence of HPV16_B and C variants in Africa, the increased prevalence of HPV16_A4 variants in Asia and the enrichment of HPV16 D variants in the Americas.

Our results showing a differential association between HPV16 variant lineages and the histological presentation of the cervical cancer open interesting research prospects. Independently of the geographic origin of the samples, we observe a sharp decrease in prevalence of the HPV16_A1-3 variant in cancers with a glandular component in Europe, Central-South America and Asia whereas we observe a globally increased prevalence of HPV16_D variants (Table 4). Molecular differences between viral variants in the virus-host interaction may underlie these differences in prevalence. Indeed, specific polymorphisms in the regulatory region of HPV16_D variants may facilitate regulation of viral gene expression as response to progesterone and estrogen hormones, which are produced in large amounts in endocervical columnar epithelia where ADC and ADSC occur.[2,50] Some authors have identified polymorphisms in HPV16_D variants glucocorticoid response elements (GREs)[50–52] that confer facilitated activation of promoter p97, leading to an enhanced E6-E7 transcription activity.[50–53] An alternative hypothesis would be that the cellular targets for malignization associated preferentially to HPV16_A1-3 variants are rarer in glandular epithelia. The existence of particular cell types associated with the development of ICC is well documented.[54] The scarcity in the glandular epithelia of such cell types, more prone to transformation by HPV16_A1-3, could thus explain simultaneously the lower prevalence of HPV16_A1-3 in ADC and ADSC and also the overall lower incidence of ADC and ADSC compared to SCC, globally some six to eight times lower.

A number of previous studies suggested that cervical ADCs are diagnosed in younger women than cervical SCCs.[20,32,55] However, other large studies did not find differences in age between glandular and squamous ICCs.[35] Because distinct HPVs are differentially associated with either cancer presentation[25,26,29] and because more aggressive HPVs such as HPV16, 18 or 45 cause cancers in younger ages than other HPVs,[36] differences in age at diagnosis could be associated with different factors. Our study design, focused exclusively on HPV16 monoinfections and with a paired sample choice between glandular and squamous ICCs, offered a unique opportunity to pinpoint the source of the proposed differences in age at diagnosis between ADC and SCC. Our results confirm that HPV16-associated ADCs are diagnosed significantly earlier than HPV16-associated SCCs (late forties compared to early fifties). In our dataset, differences in age at diagnosis between squamous and glandular cancer forms essentially arise essentially from two factors: first, for ICCs associated with the more prevalent HPV16_A1-3 variants, glandular cancers are diagnosed earlier than squamous cancers (late forties compared to early fifties); and second, although ADCs and SCCs associated with HPV16_D variants do not display differences in age at diagnosis (late forties in both cases), the increased prevalence of HPV16_D in ADCs contributes further the younger presentation of glandular ICCs. Our results contrast with Mirabello and colleagues, who did not identified an age pattern.[35] The differences between these findings may arise from the different age definitions used: Mirabello and coworkers reported "age at enrollment" in the screening program in which the samples were generated, which could largely predate the age at cancer diagnosis, while we have analyzed actual age at cancer diagnosis.

Besides differences in age at diagnosis, early stage ADCs and ADSCs display a poorer prognosis compared to SCCs.[56,57] Other factors, such as a differential efficacy of screening procedures, have also been directly linked with the distinct patterns of age at tumor diagnosis observed among the different histological presentations of ICC.[20,23] Indeed, standard screening procedures perform very well at detecting precursor squamous lesions, and in recent years, the rising implementation of cervical cancer screening programs has achieved an important decrease in SCC incidence.[23] But exfoliation cytology may be less efficient at capturing the early cytopathologic signs of ADC because it tends to occur in the endocervical canal.[16,58,59] Since the detection of HPV genetic

material in cervical samples using standardized screening techniques seems to be more sensitive than the cytological identification of precursor lesions[60] the early detection of glandular precursor lesions may benefit from a tailored, more detailed report targeting viral genotypes differentially enriched in ADCs compared to SCCs. Such differential targeting could address types with higher prevalence in ADCs, being HPV18 the most cogent example, with 3.2 world prevalence increase in ADCs compared to SCCs[36] (respectively 36.2% vs. 11.2%; http://www.hpvcentre.net/). Our data here presented, as well as another large study[35] suggest that HPV16_D, and possibly more specifically HPV16_D2/D3 sublineages and particularly D2, display increased prevalence and could have an enhanced risk in glandular ICCs. Integrating this knowledge of type-specific or even variant-specific differential risk into future screening algorithms may help ensure proper early detection of elusive ADCs.

Despite the large sample size and the rigorous molecular classification of viral variants, our study suffers from a number of limitations. We have been able to cover with good depth only three large geographical regions, while the African continent was underrepresented and North America and Oceania were not included. Also we did not have access to the genetic background of the patients nor to data on self-reported ethnicity, which could have helped disentangle relationships between viral variants and human populations. Notwithstanding, our study provides the hitherto largest sample of well-characterized HPV16-monoinfected ICCs. Furthermore, ADSC is a rare condition, so that we had to work with a small sample set, and certain analyses were thus re-assessed without ADSC data to yield more robust results. However, compared to other studies that lump ADC and ADSC, our work classified separately ADC and ADSC, as they are different histological cancer types (https://www.iarc.fr/en/publications/pdfs-online/pat-gen/bb4/bb4-chap5.pdf). Finally, our work is not a case control study and we therefore cannot provide any data regarding differential cancer risk for HPV16 variants.

We conclude that differences in HPV16 variant prevalence values are largely explained by genuine lineage-specific differences in viral fitness and/or oncogenicity, and additionally shaped by the interaction between viral variant with cervical cancer histology and with the geographical origin of the sample. We confirm that cancer histology presentation strongly conditions age at cancer diagnosis, especially for HPV16_A1-3 variants. Our results highlight the need for understanding the differential interaction between viral genetics and host genetic background, even at very shallow levels of virus diversity. Particular histochemistry and structure within the epithelia create different niches that allow for particular interactions between viruses and cells, with substantial variation in the chances for malignization. Our knowledge of such cell-type specific cellular environment and its impact on the virus life cycle remains very limited, but it probably holds the key to understand the connection between the large diversity of HPVs genotypes and the plurality of clinical manifestations of the associated infections. Finally, the enrichment of certain viral variants in ICC with glandular component opens a way for improved screening algorithms aiming at early detection of ADC and ADSC, which tend to be diagnosed in younger women and to bear a poorer prognosis.

Authors' Contributions

Conceived the project: IGB. Generated HPV sequence data and performed analyses: SNP. Sample collection and epidemiological design: FXB, LA, SDS. Drafted the article: SNP, IGB. All authors contributed to, read and approved the last version of the article.

Bibliography

1. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC CancerBase. No. 11 [Internet]. Lyon Fr Int Agency Res Cancer 2013.
2. Burk RD, Terai M, Gravitt PE, et al. Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix. Cancer Res 2003;63:7215–20.
3. Spriggs AI, Boddington MM. Progression and regression of cervical lesions. Review of smears from women followed without initial biopsy or treatment. J Clin Pathol 1980;33:517–22.
4. Kosary CL. FIGO stage, histology, histologic grade, age and race as prognostic factors in determining survival for cancers of the female gynecological system: an analysis of 1973-87 SEER cases of cancers of the endometrium, cervix, ovary, vulva, and vagina. Semin Surg Oncol 1994;10:31–46.
5. Baek M-H, Park J-Y, Kim D, et al. Comparison of adenocarcinoma and adenosquamous carcinoma in patients with early-stage cervical cancer after radical surgery. Gynecol Oncol 2014;135:462–7.
6. Yan M, Zhang YN, He JH, et al. Prognosis analysis of 83 cases of cervical adenosquamous carcinoma. Ai Zheng 2008;27:956–61.
7. Lai CH, Hsueh S, Hong JH, et al. Are adenocarcinomas and adenosquamous carcinomas different from squamous carcinomas in stage IB and II cervical cancer patients undergoing primary radical surgery? Int J Gynecol Cancer 1999;9:28–36.
8. Davy MLJ, Dodd TJ, Luke CG, et al. Cervical cancer: effect of glandular cell type on prognosis, treatment, and survival. Obstet Gynecol 2003;101:38–45.
9. Quinn MA, Benedet JL, Odicino F, et al. Carcinoma of the cervix uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. Int J Gynaecol Obstet 2006;95:S43–S103.
10. Alfsen GC, Reed W, Abeler VM. Reproducibility of classification in non-squamous cell carcinomas of the uterine cervix. Gynecol Oncol 2003;90:282–9.
11. Arbyn M, Verdoodt F, Snijders PJF, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. Lancet Oncol 2014;15:172–83.
12. Arbyn M, Anttila A, Jordan J, et al. European guidelines for quality assurance in cervical cancer screening. Second edition-summary document. Ann Oncol 2010;21:448–58.
13. Ronco G, Dillner J, Elfstr€om KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014;383:524–32.
14. Gien LT, Beauchemin MC, Thomas G. Adenocarcinoma: a unique cervical cancer. Gynecol Oncol 2010;116:140–6.
15. Di Bonito L, Bergeron C. Cytological screening of endocervical adenocarcinoma. Ann Pathol 2012;32:e8–e14, 394–400.
16. Vizcaino AP, Moreno V, Bosch FX, et al. International trends in incidence of cervical cancer: II. Squamous-cell carcinoma. Int J Cancer 2000;86:429–35.
17. Castanon A, Landy R, Sasieni PD. Is cervical screening preventing adenocarcinoma and adenosquamous carcinoma of the cervix? Int J Cancer 2016;139:1040–5.
18. Fujiwara K, Monk B, Devouassoux-Shisheboran M. Adenocarcinoma of the uterine cervix: why is it different? Curr Oncol Rep 2014;16:416.

19. Mathew A, George PS. Trends in incidence and mortality rates of squamous cell carcinoma and adenocarcinoma of cervix - worldwide. Asian Pac J Cancer Prev 2009;10:645–50.
20. Vinh-Hung V, Bourgain C, Vlastos G, et al. Prognostic value of histopathology and trends in cervical cancer: a SEER population study. BMC Cancer 2007;7:164.
21. Bray F, Carstensen B, Møller H, et al. Incidence trends of adenocarcinoma of the cervix in 13 European countries. Cancer Epidemiol Biomarkers Prev 2005;14:2191–9.
22. Sasieni P, Adams J. Changing rates of adenocarcinoma and adenosquamous carcinoma of the cervix in England. Lancet 2001;357:1490–3.
23. Castellsague X, Diaz M, de Sanjose S, et al. Worldwide human papillomavirus etiology of cervical adenocarcinoma and its cofactors: implications for screening and prevention. J Natl Cancer Inst 2006;98:303–15.
24. Smith HO, Tiffany MF, Qualls CR, et al. The rising incidence of adenocarcinoma relative to squamous cell carcinoma of the uterine cervix in the United States - a 24-year population-based study. Gynecol Oncol 2000;78:97–105.
25. Smith JS, Lindsay L, Hoots B, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. Int J Cancer 2007;121:621–32.
26. Li N, Franceschi S, Howell-Jones R, et al. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: variation by geographical region, histological type and year of publication. Int J Cancer 2011;128:927–35.
27. Bosch FX, Manos MM, Mun~oz N, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J Natl Cancer Inst 1995;87:796–802.
28. Altekruse SF, Lacey JV, Jr. Brinton LA, et al. Comparison of human papillomavirus genotypes, sexual, and reproductive risk factors of cervical adenocarcinoma and squamous cell carcinoma: Northeastern United States. Am J Obs Gynecol 2003;188:657–63.
29. Bosch FX, Lorincz A, Mun~oz N, et al. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol 2002;55:244–65.
30. Lacey JV, Frisch M, Brinton LA, et al. Associations between smoking and adenocarcinomas and squamous cell carcinomas of the uterine cervix (United States). Cancer Causes Control 2001;12: 153–61.
31. Berumen J, Ordon~ez RM, Lazcano E, et al. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. J Natl Cancer Inst 2001;93:1325–30.
32. Perez S, Cid A, In~arrea A, Pato M, et al. Prevalence of HPV 16 and HPV 18 lineages in Galicia, Spain. PLoS One 2014;9:e104678.
33. Quint KD, de Koning MNC, van Doorn LJ, et al. HPV genotyping and HPV16 variant analysis in glandular and squamous neoplastic lesions of the uterine cervix. Gynecol Oncol 2010;117:297–301.
34. Tornesello ML, Losito S, Benincasa G, et al. Human papillomavirus (HPV) genotypes and HPV16 variants and risk of adenocarcinoma and squamous cell carcinoma of the cervix. Gynecol Oncol 2011;121:32–42.
35. Mirabello L, Yeager M, Cullen M, et al. HPV16 sublineage associations with histology-specific cancer risk using HPV whole-genome sequences in 3200 women. J Natl Cancer Inst 2016;108.
36. de Sanjose S, Quint WG, Alemany L, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol 2010;11: 1048–56.
37. Guimera N, Lloveras B, Lindeman J, et al. The occasional role of low-risk human papillomaviruses 6, 11, 42, 44, and 70 in anogenital carcinoma defined by laser capture microdissection/PCR methodology: results from a global study. Am J Surg Pathol 2013;37:1299–310.
38. Nicolas-Parraga S, Gandini C, Pimenoff VN, et al. HPV16 variants distribution in invasive cancers of the cervix, vulva, vagina, penis, and anus. Cancer Med 2016;5:2909–19.
39. Stark M, Berger SA, Stamatakis A, et al. MLTreeMap - accurate maximum likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. BMC Genomics 2010;11:461.
40. Berger SA, Stamatakis A. Aligning short reads to reference alignments and trees. Bioinformatics 2011;27:2068–75.
41. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–90.
42. Chopjitt P, Ekalaksananan T, Pientong C, et al. Prevalence of human papillomavirus type 16 and its variants in abnormal squamous cervical cells in Northeast Thailand. Int J Infect Dis 2009;13: 212–9.
43. Cornet I, Gheit T, Iannacone M. HPV16 genetic variation and the development of cervical cancer worldwide. Br J Cancer 2013;108:240–4.
44. Cornet I, Gheit T, Franceschi S. Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR. J Virol 2012; 86:6855–61.
45. Hu X, Pang T, Guo Z, et al. HPV16 E6 gene variations in invasive cervical squamous cell carcinoma and cancer in situ from Russian patients. Br J Cancer 2001;84:791–5.
46. Lizano M, De la Cruz-Hernandez E, Carrillo-Garcıa A, et al. Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population. Gynecol Oncol 2006;102:230–5.
47. Qmichou Z, Khyatti M, Berraho M, et al. Analysis of mutations in the E6 oncogene of human papillomavirus 16 in cervical cancer isolates from Moroccan women. BMC Infect Dis 2013;13:378.
48. Sanchez GI, Kleter B, Gheit T, et al. Clinical evaluation of polymerase chain reaction reverse hybridization assay for detection and identification of human papillomavirus type 16 variants. J Clin Virol 2011;51:165–9.
49. Pimenoff VN, Mendes de Oliveira C, Bravo IG. Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. Mol Biol Evol 2016;34:4–19.
50. Chan WK, Klock G, Bernard HU. Progesterone and glucocorticoid response elements occur in the long control regions of several human papillomaviruses involved in anogenital neoplasia. J Virol 1989;63:3261–9.
51. Kammer C, Warthorst U, Torrez-Martinez N, et al. Sequence analysis of the long control region of human papillomavirus type 16 variants and functional consequences for P97 promoter activity. J Gen Virol 2000;81:1975–81.
52. Lace MJ, Isacson C, Anson JR, et al. Upstream regulatory region alterations found in human papillomavirus type 16 (HPV-16) isolates from cervical carcinomas increase transcription, ori function, and HPV immortalization capacity in culture. J Virol 2009;83:7457–66.
53. Sichero L, Franco EL, Villa LL. Different P105 promoter activities among natural variants of human papillomavirus type 18. J Infect Dis 2005; 191:739–42.
54. Herfs M, Yamamoto Y, Laury A, et al. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. Proc Natl Acad Sci U S A 2012;109:10516–21.
55. dos Reis R, Frumovitz M, Milam MR, et al. Adenosquamous carcinoma versus adenocarcinoma in early-stage cervical cancer patients undergoing radical hysterectomy: an outcomes analysis. Gynecol Oncol 2007;107:458–63.
56. Eifel PJ, Burke TW, Morris M, et al. Adenocarcinoma as an independent risk factor for disease recurrence in patients with stage IB cervical carcinoma. Gynecol Oncol 1995;59:38–44.
57. Hopkins MP, Morley GW. A comparison of adenocarcinoma and squamous cell carcinoma of the cervix. Obstet Gynecol 1991;77:912–7.
58. Paloma Vizcaino A, Moreno V, Xavier Bosch F, et al. International trends in the incidence of cervical cancer: I. Adenocarcinoma and adenosquamous cell carcinomas. Int J Cancer 1998;75:536–45.
59. Sasieni P, Castanon A, Cuzick J. Screening and adenocarcinoma of the cervix. Int J Cancer 2009; 125:525–9.
60. Schiffman M, Castle PE. The promise of global cervical-cancer prevention. N Engl J Med 2005; 353:2101–4.

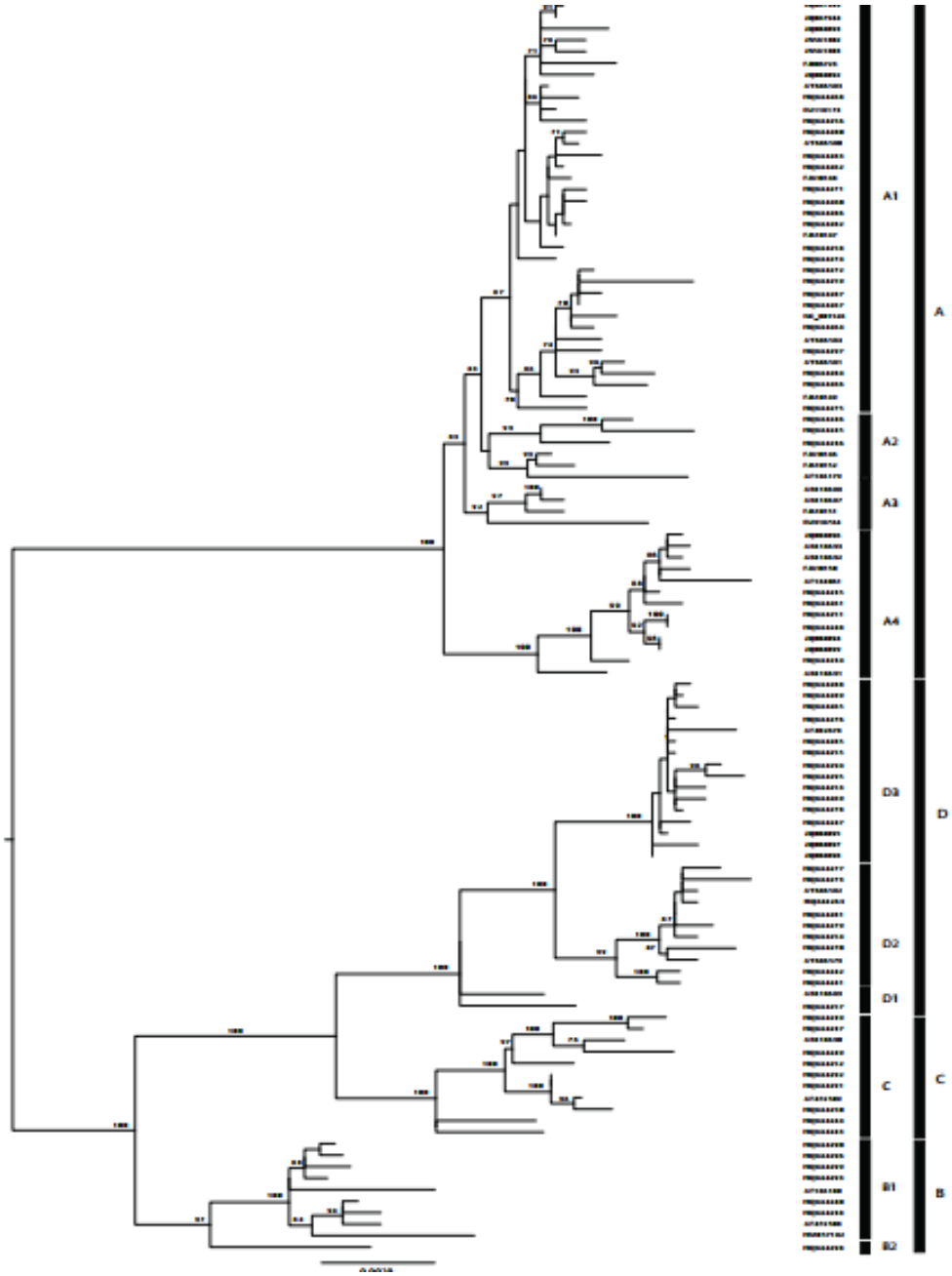


Figure S1. HPV16 Maximum likelihood phylogenetic best tree. Mid-point rooted HPV16 Best-known maximum likelihood phylogenetic tree, constructed using 109 unique full-length genome sequences. HPV16 lineages are classified into four variants: A, B, C and D. Bootstrap values above 70 are displayed closed to the corresponding branch. GenBank accession numbers are given for all entries.

Table S1: Primer design

Target Gene	Amplified region PCR product		Designed Primers	Specific SNPs	HPV16 (sub)variant
E6-1 [a]	nt 75-206 / 131 bp	F	5'-GACATTTTMTGCACCAAAAGAGA-3'	A83C G132C G132T T178G	B B C A4
		R	5'-GCTTGCAGTACACACATTCTAATA-3'		
E6-2 [a]	nt 274-388 / 114 bp	F	5'-GAATCCATATGCTGTATGTGATAA-3'	T350G	A1
		R	5'-CGGTTTGTTGTATTGCTGTT-3'		
L2	nt 5106-5312 / 206 bp	F	5'-AGGCCAGCATTAACCTCTAGGCG-3'	A5117G T5138A C5138A C5144T C5230G A5258G T5285A A5286C A5294C T5309A	C A1 A1 A1 D C D D D D
		R	5'-AGTDGGTGAGGCTGCATGKGA-3'		
LCR	nt 7712-7876 / 164 bp	F	5'- TGGCTTGTTTTAACTMMCCTAA -3'	T7712A A7728C T7741G T7779C G7824A A7828C A7835C A7837G G7840A C7873G	B A4 D A4 C A4 C C A4 A4
		R	5'- KKTGTAACCCAAAATCGGT-3'		

Table shows the target gene, the amplified region, the PCR product length, the designed primer sequence for each target gene and the lineage-specific polymorphism contained in each amplicon. F=Forward; R=Reverse. Reference sequence used for numbering: NC_001526. [a]E6 primer sequence retrieved from Larsson and colleagues (22).

Table S2: Sample distribution *per* histological cancer type, geographical region and country

			Initial	Amplified
SCC	Asia	Thailand	1	1
		Turkey	6	6
		Taiwan	5	5
		China	8	1
		India	5	5
		Kuwait	4	4
		Lebanon	5	5
	Central-South America	Mexico	2	2
		Paraguay	4	4
		Peru	5	5
		Venezuela	7	7
		Colombia	10	10
		Argentina	6	6
		Guatemala	9	9
	Europe	Netherlands	5	5
		Poland	6	6
		Portugal	1	1
		Spain	3	3
		Bosnia-Herzegovina	5	5
		France	4	4
		Italy	10	10
	Africa	Nigeria	7	7
ADSC	Asia	Turkey	3	3
		Taiwan	2	2
		India	1	1
		Kuwait	2	2
		Korea	2	0
	Central/South America	Mexico	9	6
		Paraguay	3	3
		Brazil	2	1
		Chile	2	1
		Colombia	2	1
		Argentina	8	3
		Guatemala	1	1
	Europe	Poland	2	1
		Portugal	1	1
		Spain	6	4
		Czech Republic	1	0
		Bosnia-Herzegovina	4	0
	Africa	Nigeria	2	2

		Philippines	1	1
		Taiwan	10	9
	Asia	China	3	1
		India	2	2
		Lebanon	5	3
		Turkey	13	13
		Argentina	4	1
		Brazil	9	7
		Colombia	10	7
ADC	**Central/South America**	Guatemala	3	3
		Mexico	13	9
		Paraguay	2	1
		Chile	2	2
		Italy	6	5
	Europe	Poland	2	2
		Portugal	3	1
		Spain	23	21
	Africa	Nigeria	3	2
		Uganda	6	6
			291	**240**

Table S3: EPA Likelihood weights for the attribution of each individual sequence to each variant, number of samples and percentage

Evolutionary Placement Algorithm (EUR-CSA-AS-AF)							
		SCC		ADSC		AD	
Variant	Likelihood Ascription	Number	%	Number	%	Number	%
A1-3	0.6-0.7	2	2.3	0	0	10	22.7
	0.70-0.89	19	21.6	0	0	7	15.9
	0.90-0.94	15	17	7	46.7	5	11.4
	> 0.95	52	59.1	8	53.3	22	50
	Sub-total	88	100	15	100	44	100
A4	0.6-0.7	0	0	0	0	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	0	0	1	50	0	0
	> 0.95	3	100	1	50	8	100
	Sub-total	3	100	2	100	8	100
B	0.6-0.7	0	0	0	0	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	2	100	0	0	1	33.3
	> 0.95	0	0	1	100	2	66.7
	Sub-total	2	100	1	100	3	100
C	0.6-0.7	0	0	0	0	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	0	0	0	0	0	0
	> 0.95	6	100	0	0	2	100
	Sub-total	6	100	0	0	2	100
D	0.6-0.7	0	0	0	0	0	0
	0.70-0.89	0	0	0	0	1	2.6
	0.90-0.94	1	10	0	0	3	7.9
	> 0.95	9	90	11	100	34	89.5
	Sub-total	10	100	11	100	38	100
Unclassified	0.5-0.6	0	0	2^{HPV_A1-3}	66.7	1^{A}	50
	0.6-0.7	0	0	1^{A}	33.3	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	1^{A}	50	0	0	0	0
	> 0.95	1^{A}	50	0	0	1^{UC}	50
	Sub-total	2	100	3	100	2	100
TOTAL		**111**		**32**		**97**	

Un-classified samples contain isolates not classified within a particular HPV16 variant group with LH weights > 0.6. The superindex indicates the HPV16 linage the sample is basally classified. Using ^{A} Samples classified basal to HPV16_A1-3 and A4 variants; $^{HPV\_A1-3}$ Samples classified basal to HPV16_A1-3; ^{UC} Samples not classified with LH weights >0.6 within any HPV16 variant group.

Table S4: Generalized Linear Model (GLM) and analysis of deviance for the three histologies (SCC, ADSC and ADC) for all the geographic origins (EUR, CSA, AS and AF)

	Df	*Dev.Res*	*Df*	*Dev.Res*	*% exp.dev*	*p-value*
NULL			59	562.29		
Variant	4	279.61	55	282.68	49.7	< 0.0001
Histology	2	55.02	53	227.66	9.8	< 0.0001
Geography	3	53.77	50	173.89	9.6	< 0.0001
Variant : Histology	8	40.93	42	132.96	7.3	< 0.0001
Variant : Geography	12	113.82	30	29.14	20.2	< 0.0001
Histology : Geography	6	7.77	24	11.38	1.4	0.2557
Variant : Histology : Geography	24	11.38	0	0	2.02	0.9861

Abbreviations: *Df*= Degrees of freedom ; *Res. Dev* = Residual Deviance, *% exp.dev* = percentage of data deviance explained by the corresponding factor or factor combination.

Data should be read as follows (using “variant” as an example): the factor “Variant” has five levels (HPV16_A1-1, A4, B, C and D) and thus contributes with four degrees of freedom; it explains in the model 279.6 units of deviance, i.e. 49.7% of the whole deviance in the original data; the probability of a factor to explain at random this proportion of the data deviance is below 0.0001.

Table S5: Generalized Linear Model (GLM) and analysis of deviance for three histologies (SCC , ADSC and ADC) for five geographical origins including North American (NA) samples from Mirabello *et al.*, 2016 (EUR, CSA, AS , AF and NA)

	Df	*Dev.Res*	*Df*	*Dev.Res*	*% exp.dev*	*p-value*
NULL			74	855.46		
Variant	4	412.37	70	443.10	48.2	< 0.0001
Histology	2	124.88	68	318.22	14.5	< 0.0001
Geography	4	80.50	64	237.72	9.41	< 0.0001
Variant : Histology	8	66.20	56	171.52	7.7	< 0.0001
Variant : Geography	16	117.88	40	53.64	13.7	< 0.0001
Histology : Geography	8	41.43	32	12.20	4.84	< 0.0001
Variant : Histology : Geography	32	12.20	0	0	1.43	0.9994

Abbreviations: *Df*= Degrees of freedom ; *Res. Dev* = Residual Deviance, *% exp.dev* = percentage of data deviance explained by the corresponding factor or factor combination.

Data should be read as follows (using “variant” as an example): the factor “Variant” has five levels (HPV16_A1-1, A4, B, C and D) and thus contributes with four degrees of freedom; it explains in the model 412.37 units of deviance, i.e. 48.2% of the whole deviance in the original data; the probability of a factor to explain at random this proportion of the data deviance is below 0.0001.

Table S6: Generalized Linear Model (GLM) and analysis of deviance for two histologies (SCC and ADC) for five geographical origins including North American (NA) samples from Mirabello *et al.*, 2016 (EUR, CSA, AS , AF and NA)

	Df	*Dev.Res*	*Df*	*Dev.Res*	*% exp.dev*	*p-value*
NULL			49	660.79		
Variant	4	381.03	44	279.76	57.6	< 0.0001
Histology	1	5.93	44	273.84	0.89	0.015
Geography	4	83.08	40	190.76	12.6	< 0.0001
Variant : Histology	4	62.72	36	128.03	9.49	< 0.0001
Variant : Geography	16	103.88	20	24.15	15.72	< 0.0001
Histology : Geography	4	15.57	16	8.59	2.35	0.003
Variant : Histology : Geography	16	8.59	0	0	1.29	0.292

Abbreviations: *Df*= Degrees of freedom ; *Res. Dev* = Residual Deviance, *% exp.dev* = percentage of data deviance explained by the corresponding factor or factor combination.

Data should be read as follows (using “variant” as an example): the factor “Variant” has five levels (HPV16_A1-1, A4, B, C and D) and thus contributes with four degrees of freedom; it explains in the model 381.03 units of deviance, i.e. 57.6% of the whole deviance in the original data; the probability of a factor to explain at random this proportion of the data deviance is below 0.0001.

Table S7: HPV16 A1-3, A4 B, C and D variant distribution analysis by the three histologies (SCC, ADSC and ADC) for all the geographic origins (EUR, CSA, AS and AF)

		EUR		CSA		AS		AF		NA†		Total (our study)		
Histology	**Variants**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**Chisq test**
SCC	**A1-3**	32	97	36	83.7	20	76.9	0	0	52	75.4	88	80.7	< 0.0001 / 0.07
	A4	0	0	0	0	3	11.5	0	0	2	2.9	3	2.8	
	B	0	0	0	0	0	0	2	28.6	2	2.9	2	1.8	
	C	0	0	0	0	1	3.8	5	71.4	2	2.9	6	5.5	
	D	1	3	7	16.3	2	7.7	0	0	11	15.9	10	9.2	
	Sub-total	33	100	43	100	26	100	7	100	69	100	109	100	
ADSC	**A1-3**	5	83.3	5	38.5	5	62.5	0	0	-	-	15	51.7	0.014/ 0.03
	A4	0	0	0	0	2	25	0	0	-	-	2	6.9	
	B	0	0	0	0	0	0	1	50	-	-	1	3.4	
	C	0	0	0	0	0	0	0	0	-	-	0	0	
	D	1	16.7	8	61.5	1	12.5	1	50	-	-	11	37.9	
	Sub-total	6	100	13	100	8	100	2	0	-	-	29	100	
ADC	**A1-3**	19	67.9	11	36.6	11	37.9	3	37.5	10	25	44	46.3	<0.0001 / <0.0001
	A4	0	0	0	0	8	27.6	0	0	2	5	8	8.4	
	B	0	0	0	0	0	0	3	37.5	1	2.5	3	3.2	
	C	1	3.6	0	0	0	0	1	12.5	0	0	2	2.1	
	D	8	28.6	19	63.3	10	34.5	1	12.5	27	67.5	38	40	
	Sub-total	28	100	30	100	29	100	8	100	40	100	95	100	
Total	**A1-3**	56	83.6	52	60.5	36	57.1	3	17.6	62	56.9	147	63.1	<0.0001 / <0.0001
	A4	0	0	0	0	13	20.6	0	0	4	3.7	13	5.6	
	B	0	0	0	0	0	0	6	35.3	3	2.7	6	2.6	
	C	1	1.5	0	0	1	1.6	6	35.3	2	1.8	8	3.4	
	D	10	14.9	34	39.5	13	20.6	2	11.8	38	34.9	59	25.3	
	Total	67	100	86	100	63	100	17	100	109	100	233	100	
	Fisher test	0.015 / 0.005		<0.0001 / <0.0001		0.029/0.007		0.063/ 0.063		< 0.0001		< 0.0001 / < 0.0001		

The contingency table shows HPV16 variants distribution for the 233 samples analysed, according to geographical region and anatomical location. Differences in variant prevalence between anatomical sites within a given geographical region are given through Fisher test values (columns). Differences in variant prevalence between geographical regions, within a anatomical location are given through Fisher test values (rows). Abbreviations: † Data for ADC and SCC samples from Mirabello *et al.*, 2016; *A1-3*= HPV16_A1, HPV16_A2 and HPV16_A3*; A4*= HPV16_A4*; B*= HPV16_B; *C*=HPV16_C*;D*=HPV16_D*; SCC*=Squamous cell carcinoma*; ADSC*= Adenosquamous cell carcinoma*; ADC*=Adenocarcinoma; *EUR*=Europe*; CSA*=Central-South America *and AS*=Asia*, AF*=Africa*.*

Data should be read as follows for Chi-square test: (Using “ADC” as an example): H_0 hypothesis of the variant prevalence values being similar for EUR, CSA, AS ,NA and AF is rejected with p-value below 0.0001 / H_0 hypothesis of the variant prevalence values being similar for EUR, CSA, AS and AF is rejected with p-value below 0.0001. Data should be read as follows for the Fisher test (Using “CSA” as an example): H_0 hypothesis of the variant prevalence values being similar for SCC, ADSC and ADC is rejected with p-value below 0.0001 / (Using “CSA” as an example): H_0 hypothesis of the variant prevalence values being similar for SCC and ADC is rejected with p-value below 0.0001.

Table S8: EPA Likelihood weights for the attribution of each HPV16 D variant sequence to each sub-variant (HPV16_D1, D2 and D3), number of samples and percentage

Evolutionary Placement Algorithm (EUR-CSA-AS-AF)							
		SCC		ADSC		AD	
Variant	Likelihood Ascription	Number	%	Number	%	Number	%
D1	0.6-0.7	0	0	0	0	0	0
	0.70-0.89	1	50	1	100	1	33,3
	0.90-0.94	0	0	0	0,0	0	0,0
	> 0.95	1	50	0	0,0	2	66,7
	Sub-total	2	100	1	100	3	100
D2	0.6-0.7	0	0	0	0	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	0	0	0	0	3	42,86
	> 0.95	0	0	0	0	4	57,14
	Sub-total	0	100	0	100	7	100
D3	0.6-0.7	0	0	0	0,0	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	0	0	0	0	0	0,0
	> 0.95	7	100	6	100,0	28	100
	Sub-total	7	100	6	100	28	100
D[(a)]	0.6-0.7	0	0	1	0	0	0
	0.70-0.89	0	0	0	0	0	0
	0.90-0.94	1	0	1	0	0	0
	> 0.95	0	0	2	0	0	0
	Sub-total	1	0	4	0	0	0
TOTAL		**10**		**11**		**38**	

D[(a)] samples contain isolates not classified within a particular HPV16 D variant group with LH weights > 0.6.

Table S9: Prevalence ratios (PR) for HPV16 variants by the three histologies (SCC, ADSC and ADC) for all the geographic origins (EUR, CSA, AS and AF).

	Variants	SCC [Ref]	ADSC	PR	PR Wald-test	95% CI	Fisher-test	ADC (%)	PR	PR Wald-test	95% CI	Fisher-test
EUR	**A1-3 (Ref)**	32	5				0.287	19				0.004
	A4/B/C/D	1 (-/-/-/1)	1 (-/-/-/1)	3.70	0.163	0.74-18.47		9 (-/-/1/8)	1.34	0.715	0.32-5.71	
CSA	**A1-3 (Ref)**	36	5				0.003	11				< 0.0001
	A4/B/C/D	7 (-/-/-/7)	8 (-/-/-/8)	4.37	0.001	1.69-11.29		19 (-/-/-/19)	3.12	< 0.0001	1.77-5.51	
AS	**A1-3 (Ref)**	20	5				0.648	11				0.006
	A4/B/C/D	6 (3/-/1/2)	3(2/-/-/1)	1.67	0.418	0.50-5.60		18 (8/-/-/10)	2.11	0.003	1.25-3.58	
AF	**A1-3 (Ref)**	0	0				1	3				0.2
	A4/B/C/D	7 (-/-/2/5)	2(-/1/-/1)	-	-	-		5 (-/3/1/1)	-	-	-	

Abbreviations: *Ref*= Reference histology, *ADC*=Adenocarcinoma, *SCC*=Squamous Cell Carcinoma and *ADSC*= Adenosquamous cell carcinoma. *EUR*=Europe, *CSA*=Central-South America, *AS*=Asia, *AF*=Africa. *HPV16_A1-3*= HPV16 A1, A2 and A3 variants; *HPV16_A4*= HPV16 A4 variants; *HPV16_D*= HPV16 D variants; *PR*=Prevalence Ratio. No PR values could be calculated for the African samples because of the lack of power due to low numbers. Data should be read as follows: "Using CSA as an example": PR show 4.37 times higher prevalence of HPV16_A4/B/C/D variants in ADSC than in SCC (used as Reference group, Ref) and 3.12 times higher prevalence of HPV16_A4/B/C/D variants in ADC than in SCC (Ref). PR for each strata is accompanied with Wald test p-value and score confidence intervals (95%CI). Regarding Fisher test data should read as follows: "Using CSA as an example": H0 hypothesis of the variant prevalence values being similar for SCC and ADC within CSA is rejected with $p<0.0001$.

Table S10: Age at tumour diagnosis by the two main histologies (SCC and ADC):

Histology	Median ± MAD	Q1 (25%)	Q3 (75%)	Range†	N	Wilcox. p-value
SCC	55±16.3	45	69	(29-90)	107	Ref
ADC	47 ±13.34	40.5	59.25	(27-81)	88	0.001

Abbreviations: *ADC*=Adenocarcinoma, *SCC*=Squamous Cell Carcinoma, MAD=Median Absolute Deviation; Ref= Reference group (SCC).Table represents the median ± median absolute deviation, the 25% and 75% quartiles, range († Range, 1.5x Inter-quartile) and number of samples. Median differences between SCC (Reference) and ADC are performed by Wilcoxon Mann-Whitney test.

Data should read as follows: H_0 hypothesis of the median age (Median ± MAD) being similar for SCC and ADC is rejected with *p-value* below 0.01.

Table S11: Generalized Linear Model (GLM) and analysis of deviance by the two main histologies (SCC and ADC) for the best represented geographic origins (EUR, CSA and AS) for age at tumor diagnosis.

	Df	*Dev.Res*	*Df*	*Dev.Res*	*% exp. Variance*	*p-value*
NULL			180	700.70		
Variant	3	10.34	177	690.35	1.47	0.015
Histology	1	24.52	176	665.38	3.49	< 0.0001
Geography	2	4.93	174	660.89	0.7	0.084
Variant : Histology	3	12.02	171	648.87	1.72	0.007
Variant : Geography	2	4.01	169	644.85	0.57	0.134
Histology : Geography	2	3.31	167	641.54	0.47	0.190
Variant : Histology : Geography	2	23.24	165	618.30	3.32	< 0.0001

Abbreviations: *Df*= Degrees of freedom; *Res. Dev* = Residual Deviance, *% exp.dev* = percentage of data deviance explained by the corresponding factor or factor combination.

Data should be read as follows (using “Histology” as an example): the factor “Histology” has two levels (SCC and ADC) and thus contributes with one degrees of freedom; it explains in the model 24.52 units of deviance, i.e. 3.49% of the whole deviance in the original data; the probability of a factor to explain at random this proportion of the data deviance is below 0.0001.

Table S12: Three-way ANOVA comparing mean age at tumour diagnosis for HPV16 variant (A1-3, A4, B, C and D), cancer histology (SCC and ADC) and geographical origin (Europe, Central-South-America, Asia)

	Df	*Sum Sq*	*Mean*	*Mean Sq*	*p-value*
Age vs. Histology	1	1736	1736.3	8.663	0.00371
Age vs. Variant	3	109	36.4	0.182	0.90860
Age vs. Geography	2	266	133.0	0.663	0.51645
Age vs. Histology:Variant	3	640	213.2	1.064	0.36615
Age vs. Histology:Geography	2	68	33.8	0.169	0.84503
Age vs. Variant:Geography	2	311	155.5	0.776	0.46194
Age vs. Histology:Variant:Geography	2	1157	578.6	2.887	0.05856
Residuals	165	33070	200.4		

Abbreviations: *Df*= Degrees of freedom*; Sum Sq* = Sum of squares*; Mean Sq*= Mean of squares. Three-way ANOVA assess the contribution of multiple factors onto the age at tumour diagnosis variance.

Data should be read as follows: “Using Age vs. Histology”: ANOVA shows that *cancer histology* is the factor with the largest effect on the variable “age at cancer diagnosis” with significance below 0.005 (0.00371).

Table S13: Median and Median Absolute Deviation (MAD) for SCC and ADC stratified for variants (HPV16 A1-3; A4; B; C and D)

	SCC	N	ADC	N	Total N	Wilcox-test
A1-3	56 ± 19.2	84	46.5 ± 13.3	40	124	0.004
A4	50±7.4	3	47 ± 17	8	11	0.918
B	59.5 ± 15.6	2	52.5±11.1	2	4	0.667
C	56.5 ± 3.7	6	63	1	7	0.517
D	46 ± 9.6	10	47.5 ± 10.3	36	46	0.862

Abbreviations: *ADC*=Adenocarcinoma, *SCC*=Squamous Cell Carcinoma, *N*= Number of samples; *A1-3*=HPV16_A1, HPV16_A2, HPV16_A3, *A4*=HPV16_A4, *B*=HPV16_B, *C*=HPV16_C, *D*=HPV16_D

Table represents the median ± median absolute deviation, number of isolates and total number of isolates with age data available. Median differences between SCC (Reference) and ADC are performed by Wilcoxon Mann-Whitney test. Data should read as follows: H0 hypothesis of the median age being similar for SCC and ADC for A1-3 is rejected with $p<0.005$.

Table S14: **Collaborating centres at the RIS HPV TT and HPV VVAP study groups**

1.RIS HPV TT study group:

1. Algeria: Doudja Hammouda (National Institute of Health, Registre des Tumeurs d'Alger), Anissa Bouhadef (C.H.U Hussein-Dey- Hospital Nafissa Hamoud).

2. Argentina: Silvio Alejandro Tatti, Susana Vighi (Hospital de Clínicas José de San Martín, Universidad de Buenos Aires).

3. Australia: Suzanne M. Garland, Sepher Tabrizi (The Royal Women's Hospital, The University of Melbourne).

4. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Faruk Ahmed, Mohammad Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU).

5. Bosnia Herzegovina: Ermina Iljazovic (University Clinical Center Tuzla BiH).

6. Brazil: Marcus Aurelho de Lima, Sérgio Henrique (Laboratório de Anatomia Patológica e Citologia Ltda, Associação de Combate ao Câncer do Brasil Central, Hospital Dr. Hélio Angotti, PATMED).

7. Chile: Rodrigo Prado (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile), Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Complejo Hospitalario San José).

8. China: You Lin Qiao, Chen Wen, Wu Su Hui (Cancer Institute, Chinese Academy of Medical Sciences and Peking Union Medical College).

9. Colombia: Hector Jaime Posso (Liga Contra el Cáncer de Bogotá), Luis Eduardo Bravo, Tito Collazos, Luz Stella García (Hospital Universitario del Valle), Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón (Instituto Nacional de Cancerología), Gloria I. Sánchez (Universidad de Antioquía).

10. Croatia: Magdalena Grce, Ivan Sabol (Rudjer Boskovic Institutem), Sonja Dzebro, Mara Dominis (Clinical Hospital Merkur).

11. Czech Republic: Ivo Steiner (Faculty of Medicine and Faculty Hospital, Hradec Kralove), Vladimir Vonka (Institute of Hematology and Blood Transfusion).

12. France: Christine Clavel (CHU Reims, Laboratoire Pol Bouin, Hôpital Maison Blanche), Massimo Tommasino (International Agency for Research on Cancer).

13. Greece: Maria Tzardi (Medical School of University of Crete), Theodoros Agorastos (Aristotle University of Thessaloniki).

14. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva - CIESAR, Hospital General San Juan de Dios), Walter Guerra (Instituto Nacional del Cáncer), Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer).

15. Honduras: Annabelle Ferrera (Escuela de Microbiología, Universidad Nacional Autónoma de Honduras), Ricardo Bulnes (Hospital General San Felipe).

16. India: Asha Jain (Cancer Prevention and Relief Society Raipur), PSA Sarma (BSP Hospital), Sushil K Giri (Regional Cancer Center); Maheep Bhalla (JLN Hospital & Research Center), Bharat Patel (Lab One).

17. Italy: Luciano Mariani, Ferdinando Marandino (Regina Elena Cancer Institute).

18. Israel: Jacob Bornstein, Alejandro Livoff, Hector Itzhac Cohen (Western Galilee Hospital).

19. Japan: Toshiyuki Sasagawa (Kanazawa University Hospital), Shintaro Terahata (Tonami General Hospital), Kazuhisa Ishi (Juntendo University, Urayasu Hospital).

20. South Korea: Hai-Rim Shin, Jin-Kyoung Oh (National Cancer Center), Jung-II Suh (National Medical Center), and Seo-Hee Rha (Dong-A University).

21. Kuwait: Waleed Fahad Al-Jassar (Faculty of Medicine, Kuwait University), Rema'a Al-Safi (Maternity Hospital).

22. Lebanon: Muhieddine Seoud (The American University of Beirut Medical Center).

23. México: Patricia Alonso de Ruiz, Gustavo Lastra, Alma Karina Olivares Montano (Hospital General de México, Facultad de Medicina, Universidad Nacional Autónoma de México).

24. Mozambique: Clara Menéndez, Cleofé Romagosa, Carla Carrillo (Barcelona Center for International Health Research, Hospital Clínic/Universitat de Barcelona, and Manhiça Health Research Center).

25. Nigeria: Adekunbiola Banjo, Rose Anorlu, Fatimah B. Abdulkareem, A.O Daramola, CC Anunobi (Lagos University Teaching Hospital Idi-Araba).

26. Paraguay: Elena Kasamatsu, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción).

27. Perú: Carlos Santos, Eduardo Cáceres, Henry Gómez, Juvenal Sánchez, Carlos S. Vallejos, (Instituto Nacional de Enfermedades Neoplásicas).

28. The Philippines: Efren J. Domingo, María Julieta V. Germar, Jericho Thaddeus P. Luna, Carolyn Zalameda-Castro, Arnold M. Fernandez, Roslyn Balacuit (University of the Philippines College of Medicine); Cecilia Ladines Llave, Jean Anne Toral (Cervical Cancer Prevention Center – CECAP, Cancer Institute (UP-CM-PGH)).

29. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin).

30. Portugal: Eugenia Cruz (Centro Regional de Oncologia Coimbra, Instituto Português de Oncologia); Manuela Lacerda, Manuel Sobrinho-Simoes (Institute of Molecular Pathology and Immunology of the University of Porto); Ana Felix (Instituto Portugues de Oncologia de Lisboa Francisco Gentil).

31. Spain: Enrique Lerma (Hospital de la Santa Creu i Sant Pau); Enrique Poblet (Hospital General Universitário de Albacete); Lluís Eleuteri Pons (Hospital de Tortosa Verge de La Cinta); Antonio Llombart-Bosch, Morelva Toro de Méndez (Facultad de Medicina, Universidad de Valencia); Belen Lloveras (Hospital del Mar); Ana Puras Gil (Hospital Virgen del Camino); Miguel Andújar (Complejo Hospitalario Universitario Insular Marteno-Infantil); Jaume Ordi (CRESIB - Hospital Clínic); Adela Pelayo (Clinica San Carlos); Julio Velasco, Cristina Pérez (Hospital San Agustín, & IUOPA (Oncologic and Universitary Institute of Principality of Astúrias)); Maria Alejo (Hospital General de l'Hospitalet); Ignacio G. Bravo (CSISP - Centre for Public Health Research, Centro Superior de Investigación en Salud Pública, Conselleria de Sanidad (Generalitat Valenciana)); Laia Alemany, F. Xavier Bosch, Vanesa Camón, Gabriel Capellà, Cristina Caupena, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Mireia Diaz, Ana Esteban, Rebeca Font, Jose M. Godínez, Nuria Guimerà, Yolanda Florencia, Helena Frayle, Mercedes Hurtado, Joellen Klaustermeier, Anna Merchán, Carles Miralles, Nuria Monfulleda, Nubia Muñoz, Bea Quirós Cristina Rajo, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia (ICO) - Catalan Institute of Oncology); August Vidal (Hospital Universitari de Bellvitge).

32. Taiwan: Chou Cheng-Yang (National Cheng Kung University Medical College, Taiwan Association of Gynecologic Oncologists); Tang-Yuan Chu (Buddhist Tzuchi Genral Hospital); Kuo-Feng Huang (Chi Mei Medical Center); Cheng Wen-Fang (National Taiwan University Hospital); Chih- Ming HO (Gynecologic Cancer Center, Cathay General Hospital).

33. Thailand: Saibua C. Bunnag Chichareon, Kobkul Tungsinmunkong, Jintamard Suwanjarat (Prince of Songkla University).

34. The Netherlands: Chris J.L.M Meijer, Peter J.F Snijders (Vrije Universiteit Medical Center); Wim G.V. Quint, Jean-Paul Brunsveld, Anco C. Molijn, Daan T. Geraets (DDL Diagnostic Laboratory).

35. Turkey: Alp Usubutun (Medical School, Hacettepe University).

36. Uganda: Michael Odida (Uganda Makerere University); Elisabete Weiderpass (Karolinska Institutet, Sweden; The Norwegian Cancer Registry, Norway; Samfunded Folkhalsan, Finland).

37. United States of America: Esther Oliva (Massachusetts General Hospital); Thomas C.

Wright (New York Presbyterian Hospital, Columbia University Medical Center).

38. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

2.HPV VVAP study group for vulvar site, updated September 2012:

1. Argentina: Myriam Perrotta, Ana Jaen, Kevin Davies, Henry Kitchener, Godfrey Wilson (Hospital Italiano de Buenos Aires);

2. Australia: Suzanne M. Garland, Sepher N. Tabrizi (The Royal Women's Hospital, The University of Melbourne); Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

3. Austria: Elmar Armin Joura (University Hospital and Medical School);

4. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

5. Belarus: Halina Viarheichyk, Sitnikov Valeriy (Gomel State Medical University); Achynovich Searhei (Gomel Regional Clinical Oncological Hospital);

6. Bosnia Herzegovina: Ermina Iljazovic (University Clinical Center Tuzla BiH);

7. Brazil: Paula Maldonado, Gutemberg Leão Almeida, Isabel Val, Renata Fonseca, Roberto José Lima, Marcia Mannarino, Yara Furtado (Instituto de Ginecologia da Universidad Federal do Rio de Janeiro);

8. Chile: Rodrigo Prado, Carla Molina, rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

9. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón (Instituto Nacional de Cancerología); Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (Faculty Hospital Hradec Kralove);

10. Ecuador: Leopoldo Tinoco, Hospital Oncológico de Quito, Quito, Ecuador;

11. France: Christine Clavel, Philippe Birembaut, Veronique Dalstein (CHU de Reims, Laboratoire Pol Bouin/ INSERM UMR-S 903, Reims); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

12. Germany: Monika Hampl, Pof. Baldus (University Hospital of Duesseldorf); Karl Ulrich Petry, Alexander Luyten (Klinikum Wolfsburg); Michael Pawlita, Gordana Halec (Department Genome Changes and Carcinogenesis. Heildelberg);

13. Greece: Theodoros Agorastos (Aristotle University of Thessaloniki);

14. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer);

15. Honduras: Annabelle Ferrera (Escuela de Microbiología, Universidad Nacional Autónoma de Honduras);

16. India: Neerja Bhatla (Institute of Medical Science New Dehli);

17. Israel: Jacob Bornstein, Alejandro Livoff, Hector Itzhac Cohen (Western Galilee Hospital- Nahariya);

18. Italy: Luciano Mariani, Amina Vocaturo, Maria Benevolo, Fernando Marandino, Francesca Rollo (Regina Elena Cancer Istitute);

19. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asian Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

20. Kuwait: Waleed Al-Jassar (Faculty of Medicine, Kuwait University), Rema'a Al-Safi (Maternity Hospital);

21. Lebanon: Muhieddine Seoud (The American University of Beirut Medical Center); Mali: Bakarou Kamate, Cathy Ndiaye (Hospital National DU Point G);

22. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Rubén López-Revilla, Claudia Magaña-León (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

23. Mozambique: Carla Carrilho (Eduardo Mondlane University);

24. New Zealand: Susan M Bigby, RW Jones, KL Fong, D Rowan, J Baranyai, L Eva (Middlemore Hospital);

25. Nigeria: A.A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba); Sani Malami, Ali Bala Umar (Faculty of Medicine, Bayero University);

26. Paraguay: Elena Kasamatsu, Antonio Leoploldo Cubilla, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción);

27. Philippines: Celia Ladines Llave, Jean Anne Toral (Cervical Cancer Prevention Center-CECAP, Cancer Institute (UP-CM-PGH)); Efren j Domingo, Maria Julieta V. Germar, Jerico Thaddeus, P. Luna, Arnold M. Fernandez, Carolyn Zalameda Castro, Roslyn Balacuit (University of the Philippine College of Medicine General Hospital);

28. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin); Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

29. Portugal: Eugenia Cruz (Centro Regional de Oncologia Coimbra, Instituto Português de Oncologia); Ana Félix, Jorge Manuel Soares (Instituto Portugues de Oncologia de Lisboa Francisco Gentil);

30. Senegal: Cathy Ndiaye, Nafissatou Ndiaye Ba, Victorino Mendes (HOGGY stands for Hôpital Général de Grand Yoff; DANTEC - Hôpital A. Le Dantec; FAC - Faculté de Médecine - Université Cheikh A. Diop);

31. Spain: Maria Alejo (Hospital General d'Hospitalet); Belén Lloveras (Hospital del Mar); Laia Alemany, F. Xavier Bosch, Ignacio Bravo, Vanesa Camón, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Ana Esteban, Jose M. Godínez, Yolanda Florencia, Joellen Klaustermeier, Nubia Muñoz, Beatriz Quirós, Maëlle Saunier, Cristina Rajo, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia – Catalan Institute of Oncology), August Vidal, Enric Condom (Hospital Universitari de Bellvitge), Jaume Ordi (Hospital Clínic), Julio Velasco, Cristina Pérez (Hospital San Agustín);

32. Taiwan: Chou Cheng-Yang (National Cheng Kung University Medical College,Taiwan Association of Gynecologic Oncologists); Tang-Yuan Chu (Buddhist Tzuchi Genral Hospital); Kuo-Feng Huang (Chi Mei Medical Center); Cheng Wen-Fang (National Taiwan University Hospital); Chih- Ming HO (Gynecologic Cancer Center, Cathay General Hospital);

33. The Netherlands: Wim Quint, , Anco C. Molijn, Daan T. Geraets, Nuria Guimera (DDL Diagnostic Laboratory); (Chris J.L.M Meijer (Vrije Universiteit Medical Center);

34. Turkey: Alp Usubutun (Hacettepe University); UK: Henry Kitchener (School of Medicine, University of Manchester); Robyn Davies (Manchester Royal Infirmary); Paul Cross (Queen Elizabeth Hospital, Sheriff Hill);

35. Uruguay: Adela Rosa Sica, Benedicta Caserta, Mabel Cedeira, Daniel Mazal, Guillermo Rodríguez (Laboratorio de Anatomía patológica del hospital de la Mujer, Montevideo);

36. USA: Wendy Cozen, Marc T. Goodman, Brenda Y. Hernández, Charles Lynch, Daniel B. Olson, Freda R. Selk (Cancer Center, Hawaii-Iowa);

37. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

3.HPV VVAP study group for vagina site, updated December 2013:

1. Argentina: Myriam Perrotta, Ana Jaen, Kevin Davies (Hospital Italiano de Buenos Aires);

2. Australia: Suzanne M. Garland, Sepehr N. Tabrizi (The Royal Women's Hospital, The University of Melbourne); Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

3. Austria: Elmar Armin Joura (University Hospital and Medical School); Josefine Stani and Reinhard Horvat, MD (Medical University and General Hospital Vienna, Austria);

4. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

5. Belarus: Halina Viarheichyk, Sitnikov Valeriy (Gomel State Medical University); Achynovich Searhei (Gomel Regional Clinical Oncological Hospital);

6. Brazil: Paula Maldonado, Gutemberg Leão Almeida, Isabel Val, Renata Fonseca, Roberto José Lima, Marcia Mannarino, Yara Furtado (Instituto de Ginecologia da Universidad Federal do Rio de Janeiro);

7. Chile: Rodrigo Prado, Carla Molina, Rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

8. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón, Nubia Muñoz (Instituto Nacional de Cancerología);

9. Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (The Fingerland Department of Pathology, Charles University in Prague, Faculty of Medicine and University Hospital Hradec Kralove);

10. Ecuador: Leopoldo Tinoco (Hospital Oncológico de Quito, Quito, Ecuador);

11. France: Christine Clavel, Philippe Birembaut, Véronique Dalstein (CHU de Reims, Laboratoire Pol Bouin / INSERM UMR-S 903, Reims, France); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

12. Germany: Karl Ulrich Petry, Alexander Luyten (Klinikum Wolfsburg); Michael Pawlita, Gordana Halec, Dana Holzinger (Department Genome Changes and Carcinogenesis, Heildelberg);

13. Greece: Theodoros Agorastos (Aristotle University of Thessaloniki);

14. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer; Instituto de Cancerologia Dr. Bernardo del Valle S);

15. India: Asha Jain (Cancer Prevention and Relief Society Raipur); Sushil K Giri (Regional Cancer Center, Cuttack); Maheep Bhalla (JLN Hospital & Research Center,BSP, Bhilai); Bharat Patel (Lab One Raipur); PSA Sarma (BSP Hospital);

16. Israel: Jacob Bornstein, Alejandro Livoff, Hector Itzhac Cohen (Western Galilee Hospital- Nahariya);

17. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asian Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

18. Kuwait: Dr. Waleed Al- Jassar. Faculty of Medicine, Kuwait University, Dr. Rema'a Al-Safi. Maternity Hospital, Kuwait;

19. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Rubén López-Revilla, Claudia Magaña-León (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

20. Mozambique: Carla Carrilho (Eduardo Mondlane University);

21. Nigeria: Adekunbiola A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba);

22. Paraguay: Elena Kasamatsu, Antonio Leoploldo Cubilla, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Instituto de Patología e Investigación);

23. Philippines: Celia Ladines Llave, Jean Anne Toral (Cervical Cancer Prevention Center, Cancer Institute); Efren Javier Domingo, Jericho Thaddeus P. Luna, Maria Julieta V. Germar, Arnold M. Fernandez, Carolyn Zalameda Castro, Roslyn Balacuit (University of the Philippine College of Medicine General Hospital);

24. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin); Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-

Prykan (Jagiellonian University Medical College);

25. Spain: Belén Lloveras (Hospital del Mar); August Vidal, Enric Condom (Hospital Universitari de Bellvitge); Jaume Ordi (Hospital Clínic); Julio Velasco Alonso, Cristina Pérez (Hospital San Agustín); Maria Alejo (Hospital General de l'Hospitalet, Barcelona); Laia Alemany, Francesc Xavier Bosch, Ignacio G. Bravo, Vanesa Camón, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Ion Espuña, Anna Esteban, José M. Godínez, Yolanda Florencia, Klaustermeier, Natividad Patón, Beatriz Quirós, Cristina Rajo, Maëlle Saunier, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia-Catalan Institute of Oncology);

26. Taiwan: Chou Cheng-Yang (National Cheng Kung University Medical College,Taiwan Association of Gynecologic Oncologists); Tang-Yuan Chu (Buddhist Tzuchi Genral Hospital); Kuo-Feng Huang (Chi Mei Medical Center); Cheng Wen-Fang (National Taiwan University Hospital); Chih-Ming HO (Gynecologic Cancer Center, Cathay General Hospital);

27. The Netherlands: Wim G.V. Quint, Anco C. Molijn, Daan T. Geraets, Núria Guimerà (DDL Diagnostic Laboratory); Chris J.L.M Meijer (Vrije Universiteit Medical Center);

28. Turkey: Alp Usubutun (Hacettepe University);

29. UK: Henry Kitchener, Godfrey Wilson (School of Medicine, University of Manchester); Paul Cross (Queen Elizabeth Hospital, Sheriff Hill);

30. Uruguay: Adela Rosa Sica, Benedicta Caserta, Mabel Cedeira, Daniel Mazal, Guillermo Rodríguez (Laboratorio de Anatomía Patológica del Hospital de la Mujer, Montevideo);

31. USA: Marc T. Goodman, Wendy Cozen (Cedars Sinai Medical Center, Los Angeles, California); Marc T. Goodman, Brenda Y. Hernández (Cancer Center, Hawaii); Charles F. Lynch, Daniel B. Olson, Freda R. Selk (Iowa);

32. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

33. The advisory committee members are: Chris J Meijer, Massimo Tommasino, Michael Pawlita, Wim Quint and Nubia Muñoz

4.**HPV VVAP study group for penile site, updated November 2014:**

1. Australia: Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

2. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

3. Chile: Rodrigo Prado, Carla Molina, rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

4. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón (Instituto Nacional de Cancerología);

5. Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (Faculty Hospital Hradec Kralove);

6. Ecuador: Leopoldo Tinoco (Hospital Oncológico Solca-Quito);

7. France: Christine Clavel, Philippe Birembaut, Veronique Dalstein (CHU de Reims, Laboratoire Pol Bouin/ INSERM UMR-S 903, REIMS); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

8. Germany: Michael Pawlita, Gordana Halec (Department Genome Changes and Carcinogenesis. Im Neuenheimer Feld 242. D-69120 Heildelberg);

9. Greece: Maria Tzardi (Medical School of University of Crete);

10. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer);

11. Honduras: Annabelle Ferrera (Escuela de Microbiología, Universidad Nacional Autónoma de Honduras);

12. India: Asha Jain (Cancer Prevention and Relief Society Raipur); Sushil K Giri (Regional Cancer Center, Cuttack); Maheep Bhalla (JLN Hospital & Research Center,BSP, Bhilai); Bharat Patel (Lab One Raipur); PSA Sarma (BSP Hospital); Ravi Mehrotra, Mamta Singh (M.L.N Medical College, Allahabad);

13. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asan Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

14. Lebanon: Muhieddine Seoud (The American University of Beirut Medical Center);

15. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Claudia Magaña-León, Rubén López-Revilla (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

16. Mozambique: Carla Carrilho (Eduardo Mondlane University);

17. Nigeria: A.A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba); Sani Malami, Ali Bala Umar (Faculty of Medicine, Bayero University);

18. Paraguay: Antonio Leopoldo Cubilla, Elena Kasamatsu, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción);

19. Philippines: Efren j Domingo, Maria Julieta V. Germar, Jerico Thaddeus, P. Luna, Arnold M. Fernandez, Carolyn Zalameda Castro, Roslyn Balacuit (University of the Philippine College of Medicine General Hospital);

20. Poland: Andrzej Marcin Nowakowski (Medical University of Lublin); Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

21. Portugal: Eugenia Cruz (Centro Regional de Oncologia Coimbra, Instituto Português de Oncologia); Ana Felix, Jorge Manuel Soares (Instituto Portugues de Oncologia de Lisboa Francisco Gentil);

22. Senegal: Cathy Ndiaye, Nafissatou Ndiaye Ba, Victorino Mendes (HOGGY stands for Hôpital Général de Grand Yoff ; DANTEC - Hôpital A. Le Dantec; FAC - Faculté de Médecine - Université Cheikh A. Diop);

23. Spain: Enrique Poblet (Hospital General Universitario de Albacete); August Vidal, Enric Condom (Hospital Universitari de Bellvitge); Lluís Eleuteri Pons Ferré, Patrícia Escrivà Beltri, Marylene Lejeune (Hospital de Tortosa Verge de La Cinta); Belén Lloveras, Emili Masferrer (Hospital del Mar); Julio Velasco Alonso, Cristina Pérez (Hospital San Agustin); Maria Alejo (Hospital General de L'Hospitalet); Laia Alemany, Francesc Xavier Bosch, Ignacio G. Bravo, Vanesa Camón, Gabriel Capellà, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Anna Esteban, José M. Godínez, Yolanda Florencia, Joellen Klaustermeier, Núbia Muñoz, Beatriz Quirós, Cristina Rajo, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia-Catalan Institute of Oncology);

24. The Netherlands: Wim G.V. Quint, Anco C. Molijn, Daan T. Geraets, Núria Guimerà (DDL Diagnostic Laboratory); Chris J.L.M Meijer (Vrije Universiteit Medical Center);

25. UK: Ray Lonsdale (Norfolk & Norwich University Hospital NHS Foundation Trust);

26. USA: Wendy Cozen, Marc T. Goodman, Brenda Y. Hernández, Charles Lynch, Daniel B. Olson, Freda R. Selk (Cancer Center, Hawaii-Iowa);

27. Venezuela: Enrique López Loyo (Sociedad Venezolana de Patología); Victoria García Barriola, Mirian Naranjo de Gómez, Adayza Figueredo, Janira Navarro (Universidad Central de Venezuela).

28. The advisory committee members are: Chris J Meijer, Massimo Tommasino, Michael Pawlita, Wim Quint and Nubia Muñoz

5.**HPV VVAP study group for anal site, updated June 2013:**

1. Australia: Gerard Vincent Wain, Catherine Jane Kennedy, Yoke-Eng Chiew (Gynaecological Oncology, Westmead Hospital); Raghwa Sharma (Department of Tissue Pathology and Diagnostic Oncology, University of Sydney and University of Western Sydney Westmead Hospital);

2. Bangladesh: Ashrafun Nessa, AJE Nahar Rahman, Mohammed Kamal (Bangabandhu Sheikh Mujib Medical University – BSMMU); Faruk Ahmed (Dhaka Medical College Hospital);

3. Bosnia Herzegovina: Ermina Iljazovic (University Clinical Center Tuzla BiH);

4. Chile: Rodrigo Prado, Carla Molina, rosa Muñoz (Centro de Oncología Preventiva, Facultad de Medicina, Universidad de Chile); Ximena Rodriguez, Marisol Guerrero,Virginia Leiva, Elsa Olave, Claudia Ramis, Viviana Toro (Hospital de San José);

5. Colombia: Raúl Murillo, Gustavo Adolfo Hernández Suárez, Carlos Eduardo Pinzón, Nubia Muñoz (Instituto Nacional de Cancerología);

6. Czech Republic: Václav Mandys (3rd Faculty of Medicine and University Hospital King's Wineyards); Jan Laco (Faculty Hospital Hradec Kralove);

7. Ecuador: Leopoldo Tinoco (Hospital Oncológico Solca-Quito);

8. France: Christine Clavel, Philippe Birembaut, Veronique Dalstein (CHU de Reims, Laboratoire Pol Bouin/ INSERM UMR-S 903, REIMS); Christine Bergeron (Laboratoire Cerba, Department de Pathology, Cergy Pontoise); Massimo Tommasino (International Agency for Research on Cancer);

9. Germany: Karl Ulrich Petry, Alexander Luyten (Klinikum Wolfsburg); Michael Pawlita, Gordana Halec, Dana Holzinger (Department Genome Changes and Carcinogenesis, Heildelberg);

10. Guatemala: Luis Estuardo Lombardi, Edgar Kestler, Obdulia Salic, Sergio Marroquin, Victor Argueta (Centro de Investigación Epidemiológica en Salud Sexual y Reproductiva-CIESAR, Hospital General San Juan de Dios); Walter Guerra (Instituto Nacional del Cáncer); Hesler Morales (Instituto Guatemalteco de Seguridad Social, Instituto Nacional del Cáncer);

11. Honduras: Annabelle Ferrera (Universidad Autónoma de Honduras); Odessa Henríquez and Silvia Portillo (Instituto Nacional Cardiopulmonar en Tegucigalpa).

12. India: Asha Jain (Cancer Prevention and Relief Society Raipur); Sushil K Giri (Regional Cancer Center, Cuttack); Maheep Bhalla (JLN Hospital & Research Center,BSP, Bhilai); Bharat Patel (Lab One Raipur); PSA Sarma (BSP Hospital); Ravi Mehrotra, Mamta Singh (M.L.N Medical College, Allahabad);

13. Korea-South: Hai-Rim Shin, Jin-Kyung Oh (National Cancer Center); Shin Gwang Kang (Asan Medical Center); Dong-chul Kim (Kangnam St. Mary's Hospital);

14. Mali: Bakarou Kamate, Cathy Ndiaye (Hospital National DU Point G);

15. Mexico: Isabel Alvarado-Cabrero (Instituto Mexicano del Seguro Social); Rubén López-Revilla, Claudia Magaña-León (Instituto Potosino de Investigación Científica y Tecnológica, AC); Cuauhtémoc Oros (Hospital Central Ignacio Morones Prieto, San Luis Potosí);

16. Nigeria: Adekunbiola A.F. Banjo, F.B. Abdulkareem, A.O. Daramola, C.C. Anunobi, R.U. Anorlu (Lagos University Teaching Hospital Idi-Araba); Sani Malami, Ali Bala Umar (Faculty of Medicine, Bayero University);

17. Paraguay: Elena Kasamatsu, Antonio Leopoldo Cubilla, Francisco Perrota (Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Instituto de Patología e Investigación);

18. Poland: Robert Jach, Jolanta Orlowska-Heitzman, Monika Kabzinska-Turek, Paulina Przybylska, Marzena kula-Prykan (Jagiellonian University Medical College);

19. Portugal: Ana Felix (Instituto Portugues de Oncologia de Lisboa Francisco Gentil);

20. Senegal: Cathy Ndiaye, Nafissatou Ndiaye Ba, Victorino Mendes (HOGGY stands for Hôpital Général de Grand Yoff ; DANTEC - Hôpital A. Le Dantec; FAC - Faculté de Médecine - Université Cheikh A. Diop);

21. Slovenia: Mario Poljak, Boris Pospihalj, Pavle Košorok (Institute of Microbiology and Immunology, Ljubljana);

22. Spain: Maria Alejo (Hospital General de l'Hospitalet); Mar Iglesias, Belén Lloveras, David Parés (Hospital del Mar); Laia Alemany, Francesc Xavier Bosch, Ignacio G. Bravo, Vanesa Camón, Xavier Castellsagué, Omar Clavero, Silvia de Sanjosé, Ion Espuña, Anna Esteban, José M. Godínez, Yolanda Florencia, Joellen Klaustermeier, Nubia Muñoz, Nati Patón, Beatriz Quirós, Cristina Rajo, Maëlle Saunier, Sara Tous, Marleny Vergara (IDIBELL, Institut Català d'Oncologia-Catalan Institute of Oncology, Barcelona); August Vidal, Enric Condom (Hospital Universitari de Bellvitge, Barcelona); Julio Velasco Alonso, Cristina Pérez (Hospital San Agustín);

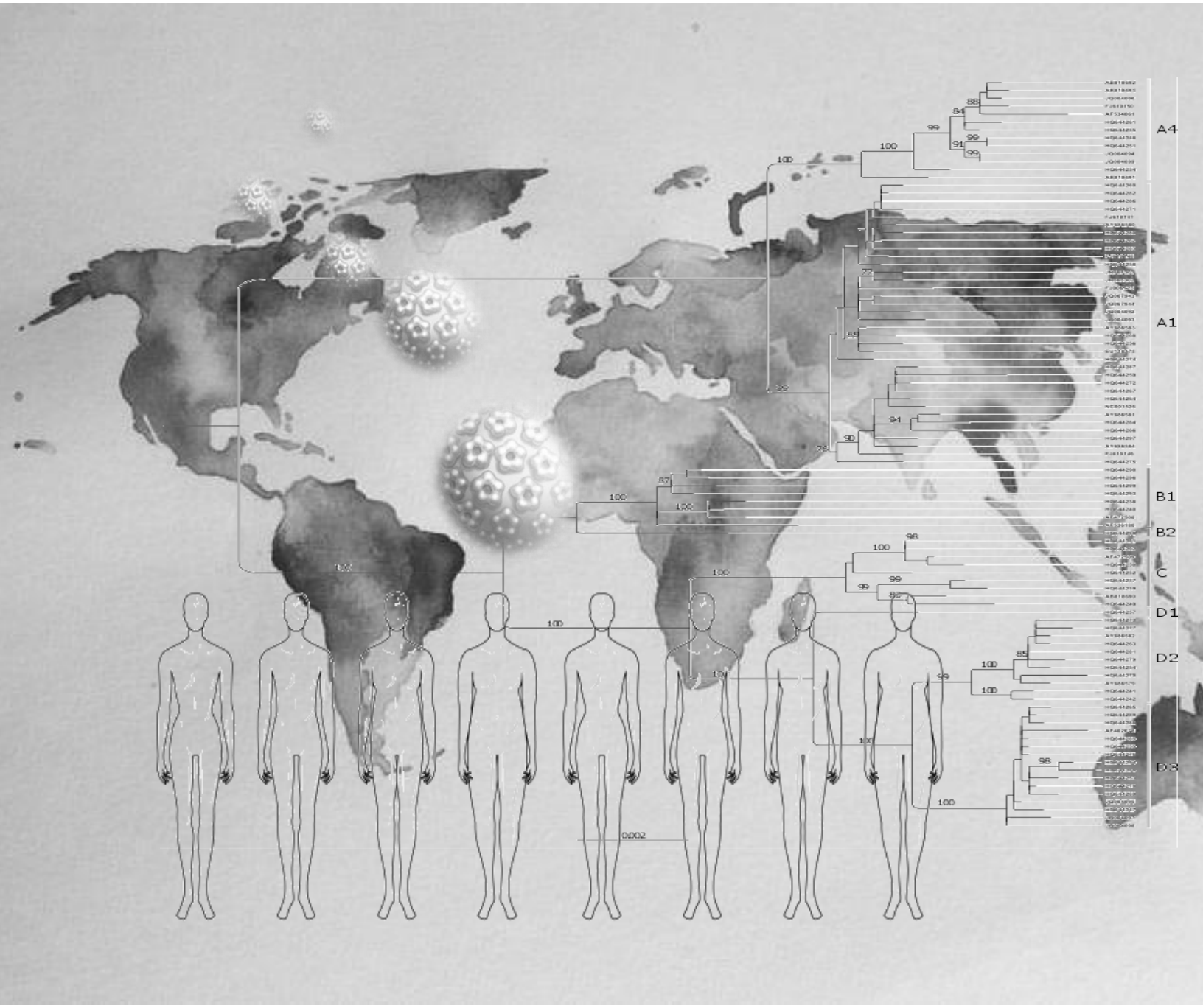
23. The Netherlands: Wim G.V. Quint, Anco C. Molijn, Daan T. Geraets, Núria Guimerà (DDL Diagnostic Laboratory); Chris J.L.M Meijer (Vrije Universiteit Medical Center);

24. UK: Henry Kitchener, Godfrey Wilson (School of Medicine, University of Manchester);

25. USA: Wendy Cozen (Los Angeles, CA); Marc T. Goodman, Brenda Y. Hernández (Hawaii); Charles Lynch, Daniel B. Olson, Freda R. Selk (Iowa); Edyta C. Pirog (New York Hospital - Cornell Medical Centre).

26. The advisory committee members are: Chris J Meijer, Massimo Tommasino, Michael Pawlita, Wim Quint, and Nubia Muñoz

4. RESULTS RESUME



4.1 Choice of informative regions and sample set description

In the three Manuscripts we initially studied the distribution of the pairwise genetic distances for the different genomic regions among full genomes of HPV6, HPV11 and HPV16. We observed that the most variable regions for HPV6 and HPV11 were the Inter*E2L2*, *E2* and LCR. We identified *E6* as one of the most variable regions solely for HPV6. We discarded *E2* due to the scarce number of sequences in databases. We finally selected Inter*E2L2* and LCR to assess our analyses for Manuscript 1. For HPV16, we identified, from the most to the least nucleotide variable ORF, the following regions: *E4*, *E5*, LCR, *L2*, *E2*, *E6, E1, L1 and E7* We designed amplicons and set up PCRs for all of them. We finally selected LCR, *L2*, and *E6* targets as they rendered the best results in terms of amplification and Sanger sequencing. *E6* was selected additionally due to be one of the most studied regions and because it spanned the T350G polymorphic site, one of the targets of our study.

4.2 Dataset construction: initial and amplified samples

For Manuscript 1, we worked with two sequence sources: sequences generated from the amplification and sequencing of GWs and RRPs repository samples and sequences obtained from GenBank database. Regarding GWs repository, we generated 142 sequences out of the 143 original samples from HPV6 and 56 amplified sequences out of the 64 original samples from HPV11. From RRPs repository we obtained 24 sequences out of the 41 original samples from HPV6 and from HPV11 we got 9 sequences out of the original 11 samples (Table 8).

	Anatomical lesion	**Initial (n)**	**Amplified (n)**			
			LCR and *IntE2L2*	**LCR**	***IntE2L2***	**Total**
HPV6	**RRPs**	41	13	3	8	24
	GWs	143	89	42	11	142
	Total	184	102	45	19	166
HPV11	**RRPs**	11	5	1	3	9
	GWs	64	45	6	5	56
	Total	75	50	7	8	65

Table 8: Sample contribution to the study for HPV6 and for HPV11: Initial and amplified samples according to the targeted regions: LCR, IntE2L2 or both

From GenBank, we recovered 87 sequences from HPV6 and 94 sequences from HPV11 (Table 9) The total number of sequences raised to 253 for HPV6 and 159 for HPV11 and encompassed ten countries (Australia, China, Colombia, Germany, Hungary, Slovenia, South Africa, Sweden, Thailand and USA) and covered five continents. For the final analyses nonetheless, ten sequences were excluded due to the lack geographical information. Thus, the final dataset contained in 243 sequences for HPV6 and 149 sequences for HPV11 (Table 9).

HPV6	**RRPs**	24
	GWs	142
	GenBank sequences	87
	Total	253†
HPV11	**RRPs**	9
	GWs	56
	GenBank sequences	94
	Total	159†

Table 9: Number of alignments for HPV6 and for HPV11 †For the final analyses ten sequences we excluded due to the lack of geographical information.

For Manuscript 2 we worked with an amplified dataset of 711 isolates. Our samples were originated from five different invasive anogenital carcinomas produced by HPV16 monoinfection: ICC (N=170), IVuC (N=128), IVaC (N=122), IPeC (N= 119) and IAnC (N=172) and covered three different geographical regions: Europe (N=354), Central/South America (N=272), and Asia (N=85) (Table 10).

Invasive anogenital cancer	**Amplified samples**			
	Europe	**Central/South America**	**Asia**	**Total**
Cervix (ICC)	72	71	27	170
Vulva (IVuC)	68	36	24	128
Vagina (IVaC)	61	51	10	122
Penis (IPeC)	74	42	3	119
Anus (IAnC)	79	72	21	172
Total	354	272	85	711

Table 10: Anatomical location and geographical distribution of amplified samples.

In Manuscript 3, we worked with an initial dataset of 118 SCC, 120 ADC and 53 ADSC, a total dataset of 291 samples, all of them HPV16 monoinfections. From these, we were able to amplify and sequence the viral component in 111 SCC, 97 ADC and 32 ADSC. Thus our final dataset contained 240 samples covering 28 different countries (Table 11).

Anatomical Location	Initial	Amplified
SCC	118	111
ADSC	53	32
ADC	120	97
Total	291	240

Table 11: The table shows the number of initial and amplified samples according to histological cancer type.Abbreviations: ADC, adenocarcinoma; SCC, Squamous Cell Carcinoma; ADSC, adenosquamous cell carcinoma.

4.3 HPV variant classification, study design and data collection bias, and explanatory power

In Manuscript 1, for HPV6, differences on variant distribution were observed among the two types of lesion (chi-squared test; p <0.01). Moreover, the contribution of subclades within HPV6 B variants was different among GWs and in RRPs (p <0.01). While in GWs almost two-thirds of sequences belonged to B variant, specially to B1 subclade; in RRP we found an increased contribution of A variants and within B variants, an increased presence of B3 (Table 12) .

HPV6 variants	GWs		RRPs	
	Nº samples	%	Nº samples	%
A	10(7)	6.33%	32 (14)	37.65%
B	148 (135)	93.67%	53 (10)	62.35%
B1	108(97)	68.35%	30(-)	35.30%
B2	34 (33)	21.52%	6(1)	7.05%
B3	6(5)	3.80%	17(9)	20.00%
TOTAL	158(142)	100.00%	85(24)	100.00%

Table 12: The comparison of the distribution of HPV6 variants among both pathologies, genital warts (GWs) and Recurrent respiratory Papillomatosis (RRPs), shows a statistically significant difference (chi-squared test p <0.01).Numbers without brackets and percentages represent the global set of sequences belonging to each clade and sublade. The numbers in brackets show the number of sequences belonging to each clade coming from the GWs and RRPs sample repositories. Ten partial sequences were not identified at the level of lesion and are not included in this analysis. "GWs": Genital Warts; "RRPs": Recurrent Respiratory Papillomatosis.

For HPV11, no significant difference in variant distribution depending on the different type of lesion was observed through Pearson's Chi- Squared test (chi-squared test; p = 0.493).

	GWs		RRPs	
Clade	**Nº samples**	**%**	**Nº samples**	**%**
A1	6(2)	5.50%	3(1)	7.5%
A2	103(54)	94.50%	37(8)	93.5%
TOTAL	109(56)	100.00%	40(9)	100.00%

Table 13: The comparison of the distribution of HPV11 variants among both pathologies, genital warts (GWs) and Recurrent respiratory Papillomatosis (RRPs), shows no statistically significant difference (chi-squared test p =0.493). Numbers without brackets and percentages represent the global set of sequences belonging to each clade and sublade. The numbers in brackets show the number of sequences belonging to each clade coming from the GWs and RRPs sample repositories. Eleven samples were not included in the analysis. They were classified as "unidentified", "lung", "cervical sample". "GWs": Genital Warts; "RRPs": Recurrent Respiratory Papillomatosis.

In Manuscript 2 we could classify into HPV16 variants 692 samples (97.3%) from the 711 amplified isolates in the five anogenital cancer sites for Europe, Central-South America and Asia (Table 14).

Anatomical Location	**EUR**		**CSA**		**AS**		**Total amplified**	**Total classif.**
	Amplified	**Classified**	**Amplified**	**Classified**	**Amplified**	**Classified**		
CERVIX	72	70	71	69	27	26	170	165
VULVA	68	68	36	32	24	23	128	123
VAGINA	61	60	51	48	10	9	122	117
PENIS	74	73	42	40	3	2	119	115
ANAL	79	79	72	72	21	21	172	172
Total	354	350	272	261	85	81	711	692

Table 14: Anatomical location and geographical distribution of amplified and classified samples Abbreviations: EUR=Europe; CSA=Central-South America; AS=Asia, Total classif.= Total of samples classified.

In Manuscript 3 we were able to classify into HPV16 variants 233 from the 240 amplified sequences. The final dataset included 109 SCC, 95 ADC and 29 ADSC, (Table 15).

	EUR-AS-CSA-AF SAMPLES			
	Initial	**Amplified**	**Classified[a]**	**Unclassified[b]**
SCC	118	111	109	2
ADSC	53	32	29	3
ADC	120	97	95	2
Total	291	240	233	7

Table 15: The table shows the number of initial, amplified, classified and unclassified samples according to histological cancer type. Abbreviations: ADC, adenocarcinoma; SCC, Squamous Cell Carcinoma; ADSC, adenosquamous cell carcinoma.; [a]Samples classified in HPV16_A1-3, A4, B, C and D variants. [b]Samples that are classified basal to a particular HPV16 variant cluster (i.e., basal to HPV16_A1-3 and A4 variants) and samples not classified with likelihood values below 0.6 within any HPV16 variant cluster.

Distinctly from Manuscript 1, for Manuscripts 2 and 3 we initially studied through a Generalized Linear Model (GLM) the association among HPV16 variants with the variables studied in each work (For manuscript 2 the variables (or predictors) were "Geographical origin," "Anatomical location," and "Variant" and for Manuscript 3 the variables (or predictors) were "Geographical origin," "Histological cancer type," and "Variant". In Manuscript 2, all the variables and their two by two interaction with HPV16 variants contributed significantly to the model ($p < 0.0001$ in all cases). Nonetheless, the triple interaction did not provide additional explanatory power ($p = 0.359$). (Manuscript 2, Table S4). Importantly, the model explained almost the total of variance in HPV16 variant distribution (98.6 %): Solely 14.1% of the total data variance arose from differential coverage of the three geographical regions ($n = 342$ for Europe, $n = 261$ for Central/South America, and $n = 80$ for Asia), and 1.7% come from differential coverage of the five anogenital sites analyzed ($n = 163$ for cervix, $n = 121$ for vulva, $n = 114$ for vagina, $n = 115$ for penis, and $n = 170$ for anus).

For Manuscript 3, the GLM showed good fit to the data, capturing above 96% of the variance (Manuscript 3, Table S4). Two additional GLMs (including data originating from North America (Mirabello et al. 2016) and considering 2 or 3 histological levels: (i) Only SCC and ADC and (ii) SCC, ADSC and ADC, fitted also well with data, explaining both 98% of the total variance (Manuscript 3, Supp. Tables 5 and 6). The explanatory power of the factors *Histology* and *Geography* in the global analyses (9.8% and 9.6% respectively in Table S4) was actually linked to imbalances in the access to samples, *i.e.* underrepresentation of certain histologies and of certain geographical origins. We therefore designed a balanced GLM excluding the

"variable levels" that presented low numbers: ADSC histology and African geography. Our fourth GLM presented thus the cleaner and more balanced data set including two best-represented histologies (SCC and ADC) and the three best-represented geographies (Europe, Central/South America and Asia) (Table 2). This model showed the best fit to the data as it could explain almost 98% of the data variance mainly from the two by two interactions and from the different HPV16 variant distribution, showing no significance for the "geography" (accounting only for 0.69% of the variance, p=0.267) and "histology" (accounting only for 0.31% of the variance, p=0.275) variables alone, what showed no bias provided by these predictors selection (Table 2)

The two by two interaction results provided by GLM assessment were confirmed through counting data statistics and prevalence ratios. The Chi-square test was used after stratifying data by geography and the Fisher's test was applied after stratifying data by anatomical location of the samples. The same statistical evaluation was also performed in Manuscript 1 after stratifying by pathological outcome. In Manuscript 2 we observed that 68.2% of all variation in HPV16 variants abundance originated mainly from differences in variant prevalence alone ($p < 0.001$; Manuscript 2, Table S4). Indeed, we observed that HPV16_A1-3 was by far the most prevalent variant, with an overall prevalence of 95% in Europe, 86% in Central/South America, and 61% in Asia (Manuscript 2, Figure 1 and Table 2). We quantified further that 9% of all variance in variant distribution was explained by differential association of HPV16 variants with geography ($p < 0.001$; Manuscript 2. Table S4). Prevalence ratios showed a significant 1.7-fold (95% CI: 1.4–2.1) increase prevalence of HPV16_D in Central/South America and a significant 6.6-fold (95% CI: 4.9–8.9) increased prevalence of HPV16_A4 in Asia compared, in both cases, with Europe (Manuscript 2, Table 2,Table S5 and Figure 1). Finally, we observed that 2.8% of all variation in variant distribution corresponded to differential association of HPV16 variants with anatomical location ($p < 0.001$, Manuscript 2, Table 2,Table S5 and Figure 1). The variation observed originated mainly from the increased prevalence of HPV16_A4 in vagina and in anus in Asia (Manuscript 2, Table S5 and Figure 1). We evaluated the significance of our results, excluding those regions less represented and we confirmed that differences remained significant even after excluding data from Asia for vagina and penis (Manuscript 2, Table S5 and Figure 1). Correction with Bonferroni for Multiple comparisons was applied when needed.

Similarly, in our Manuscript 3, the GLM showed that the data variance was explained mainly by the predictor "*Variant*", suggesting that HPV16 variants different prevalence explained the largest fraction of the total variance (78% for the most reliable data, $p < 0.0001$; (Manuscript 3,Table 2) and 48.2% for the complete data, $p < 0.0001$ (Manuscript 3,Table S5)). Indeed, HPV16_A1-3 was overall the most prevalent lineage, with a global prevalence of 68.3% in the most reliable data and of 63.1% for the complete data (Manuscript 3,Table 3 and Table S7 respectively). Furthermore, variant prevalence depended on the geographical origin. The interaction "*Variants*Geography*" explained 10% of the variance in the filtered data (Manuscript 3,Table 2) and 20% in the complete data (Manuscript 3,Table S4) In our results, HPV16_A1-3 variant showed a decreasing trend in prevalence in the different continents: 83.6% in Europe, 60.5% in Central/SouthAmerica, 57.1 in Asia, and 17.6 in Africa (Manuscript 3,Table S7). Furthermore, for both, filtered and complete data, we observed an increased HPV16_A4 prevalence in Asia (from 11.5 to 27.6 % for all histologies) and a HPV16_B and C variants restricted distribution to Africa (from 28.6% to 71.4% and from 12.3 to 37.5% for SCC and ADC) (Manuscript 3, Table S7). Statistics showed differences of HPV16 variants distribution within the same cancer histological presentation among geographical regions: for filtered data: for SCC ($p = 0.013$) and for ADC ($p < 0.0001$) (Manuscript 3,Table 3) and for the complete data: for SCC ($p < 0.0001$ / 0.07); for ADSC (p=0.014/ 0.03) and for ADC (<0.0001 / <0.0001) (Manuscript 3,Table S7). Furthermore, the GLM showed that variant prevalence depended on the histological presentation of the cervical cancer. The interaction "*Variant*Cancer Histology*" explained 9% of the variance in the filtered data and 7% in the total data. In our results, we detected a decreasing trend for the HPV16_A1-3 variant in different cancer histologies: it accounted for 80.7% of all SCCs, 51.7% of all ADSCs and 46.3% of all ADCs (Manuscript 3,Table S7). Parallel to this trend, we detected an increased pattern of HPV16_D variants: it accounted for 9.2% of all SCCs, 37.9% of all ADSCs and 40% of all ADCs (Manuscript 3,Table S7). For North American data (from Mirabello's and colleagues)(Mirabello et al. 2016) we observed also an increased prevalence of HPV16_A1-3 in SCC (75.4%) and an enhanced prevalence of HPV16_D for ADC (67.5%) (Manuscript 3, Figure 1 and Table S7) largely dominated by HPV16_D3. Finally, the estimated ratios between prevalence values for HPV16 variants after stratifying by histology and geography confirmed the trend of the significant decrease in prevalence of HPV16_A1-3 and the increase of non-HPV16_A1-3 variants in ADC compared with SCC in Asia (2.11 fold increase, p=0.006), Central-South America (3.12 fold increase, $p < 0.0001$) and Europe (2.42 fold increase,

p=0.004) (Manuscript 3,Table 4). Similar results were obtained when the full dataset included the data from the less represented ADSC and African samples (Manuscript 3,Table S9). No values for Africa were given as PR are calculated with integer data (≥ 1).

4.4 HPV16 *E6* gene T350G polymorphism

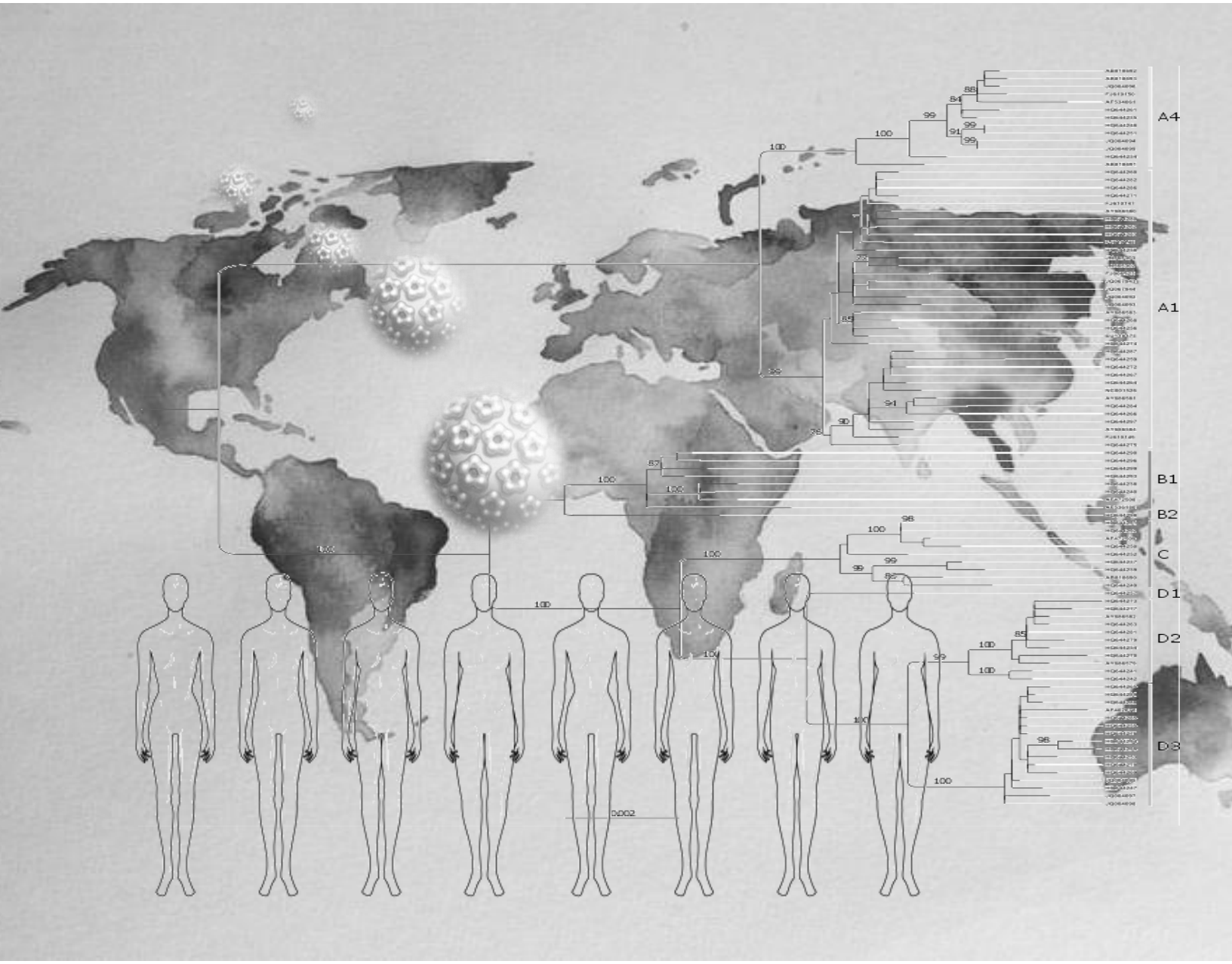
Focusing on a particular HPV16 variant (*i.e.* HPV16_A1-3 variant) we show different frequency of the T350G polymorphism depending on the geographical region: HPV16_A1-3 cases from Central/South America showed consistently higher 350G allele frequencies compared with Europe, especially for cervical (p = 0.015) and penile cancers (p < 0.0005) (Table 3, Manuscript 2). Prevalence for the 350G allele within HPV16_A1-3 among anatomical location ranged between 47% and 59% for Europe and between 59% and 90% for Central/South America (Table 3, Manuscript 2). We further showed no differences between anatomical locations within each geographical region (p = 0.617 and P = 0.102 for Europe and Central/South America respectively).

4.5 Age at cancer diagnosis: patterns detected for cervical cancer and its glandular histological presentations ADC and ADSC

Among different anogenital cancers, independently of the oncogenic HPV type or of the HPV16 variant driving the cancer, we show differences in age at tumour diagnosis: cervical cancers showed significantly younger ages at diagnosis compared with other anogenital cancers (early fifties vs. early sixties, p < 0.0005) (Manuscript 2, Figure 2 andTable S6). Furthermore, we did not detect differences in age at cancer diagnosis among noncervical cancers (Manuscript 2, Figure S3, and Table S7). Regarding the different histological presentations of ICC, we confirmed that ADCs were diagnosed in younger women than SCCs (respectively 47±13.3 and 55±16.3 years of age at diagnosis, median and median absolute deviation; p = 0.001, Wilcoxon Mann-Whitney test) (Manuscript 3, Figure 2 and Table S10). Similar results were obtained either applying a GLM (Manuscript3, Table S11) or a three way ANOVA (Manuscript 3, Table S12). At variant level we showed that while ADCs were diagnosed significantly earlier than SCCs for HPV16_A1-3 (56±19.2 vs 46.5±13.3; N=124; *p*=0.004) we did not detect differences in age at diagnosis between SCCs and ADCs for HPV16_D (46±9.6 vs 47.5±10.3; N=46*; p*=0.862) (Manuscript 3, Table S13). This differential behavior of the variable age at diagnosis was consistent with the explanatory power for the

factor *Variant* and for the interaction *Variant*Histology* found in the GLM results (Manuscript 3, Table S11).

5. DISCUSSION



The main objective of this thesis was to identify the genetic variants of HPV6, HPV11 and HPV16 DNA sequences retrieved from different pathologies named GWs, RRPs and anogenital cancers and further identify HPV16 linages in the most prevalent histological presentations of cervical cancer. The second objective was to analyze the evolutionary relationships of the HPV variants detected according to the context of those available in the databases by means of phylogenetic reconstructions. Subsequent aims comprised: the analyses of the possible differential distribution of HPV variants in function of the pathological outcome (HPV variant lesion-dependent distribution) and the possible differential distribution of HPV variants in function of the sample geographical origin (HPV variant geographical-dependent distribution). Furthermore, the research pretended to analyze possible correlations regarding age at tumor diagnosis and possible associations of the variables studied (geography and anatomy) with the long-studied and most oncogenic-related polymorphism: T350G located in HPV16 *E6* ORF.

5.1 Strenghts of identifying HPV single infected pathological outcomes

To provide clean estimates of HPV6, HPV11 and HPV16 variant lineage prevalence in different pathologies we decided to work with a well-defined dataset of clinical samples associated only to monoinfection (HPV6, HPV11 and HPV16 single infected outcomes). Through this selection we could restrict our assessments to a collection of cases with well-defined pathological outcome without the possible added bias introduced by co-infections or multi-infections, which could blur our observations. To reach this initial objective, we worked with repository samples that were initially tested for HPV DNA detection by SPF10-DEIA-LiPA protocol (version 1; Laboratory Biomedical Products Rijswijk, The Nederlands) (as described in the introduction and in the Manuscripts). To avoid possible ill-defined genotyping, in the protocols used in the original publications characterising the complete ICO's sample set, for each of the different anatomical locations, all samples with HPV detected pattern were identified and submitted to further amplification and sequencing (Alemany et al. 2014, 2015, 2016; de Sanjose et al. 2010; de Sanjosé et al. 2013).

5.2 Selection of the most informative HPV genomic regions and phylogenetic classification at level of variant. How our methodology conforms to the previously used classification procedures and the strengths and limitations that it presents

According to the ICTV, nucleotide identity of the *L1* gene has been the criteria for PV classification (De Villiers et al. 2004). However, the PVs committee does not define standards below species phylogenetic level (Chen et al. 2015). A single ORF or region, such *L1* gene, does not always contain enough sequence information for unambiguously distinguish between closely related HPV variants, and it may not to be sufficient for variant classification (Chen et al. 2015). Thus, approaches alike using the complete genome sequence to classify variants have been suggested (Chen et al. 2015) and some authors provide HPV variant classification based on the application of high-throughput full genome sequencing that provides deeper genome coverage(Cullen et al. 2015; Mirabello et al. 2016). However, for most research laboratories, including ours, these technologies remain expensive for large studies, such as those presented here. Furthermore, in pathology routine, virtually all biopsies and surgical samples are formalin fixed paraffin embedded (FFPE) and constitute a valuable source of samples over histopathology departments worldwide giving to researchers, the possibility to perform large histological and molecular retrospective studies (Ademà et al. 2014; Kokkat et al. 2013; Morshed et al. 2010). Our repository samples indeed, are all FFPE isolates that have allowed us to work with a well-archived, extensive, easy to handle and relatively inexpensive samples. However, the degradation of DNA due to the fixatives has prevented us from amplifying full genome sequence due to the fragmented nature of our FFPE-samples produced mainly by the cross-links between amino groups in DNA (Dietrich et al. 2013). Alternatively to NGS, Sanger methodology has been widely used to amplify distinct HPV genomic regions as it presents accuracy in sequencing middle and long fragments of viral DNA and most importantly, because it is cheaper. Although great number of published works use Sanger sequencing, it has been described, nonetheless, incongruence of classification when different genomic fragments are used (Chen et al. 2015) what suggests the necessity of standardized procedures at this research level. We decided to amplify HPV genomic regions of 150-200bp to full-cover amplified product obtaining. The same procedure was followed by other authors, suggesting that the short fragment amplification increases the PCR efficiency (Chen et al. 2015)

HPV variant current literature performs phylogenetic classification assessment using a non-defined gene repertoire (selection of distinct genomic regions), resulting in some cases, in a misleading classification when different target amplicons are selected (Combrinck et al. 2012; Cullen et al. 2015; Heinzel et al. 1995). There is therefore, no consensus about how to conduct an exhaustive variant classification using ORF selection. However, viral variants within HPVs are described in terms of sequence diversity and fine taxonomic classification provides a powerful epidemiological tool(Bernard et al. 2010). However, not all genomic regions are equally apt to classify HPV linages (Chen et al. 2015; Lavezzo et al. 2016; Pande et al. 2008), and only stable polymorphisms specifically associated to a variant could serve as diagnostic sites in research works in which complete HPV genome cannot be amplified (Burk et al. 2011, 2013; Harari et al. 2014) So, the first step of each Manuscript included in this thesis, was to assess the HPV genomic intra-type variability in order to select the most informative region and exhaustively classify through the amplification and sequencing of short fragments the HPV6, HPV11 and HPV16 variants according with the existent phylogenetic background of each virus. For this we performed a pair-wise nucleotide distance assessment of the full complete genome sequences available in the databases. We tried to select our target regions covering two conditions: on one hand, that the fragments provide enough phylogenetic information, what means: enough nucleotide diversity (high number of variable positions) to classify our isolates correctly; and on the other hand, that the selected genome regions, were some of the most used in the literature (*e.g.* HPV16 *E6*), to allow better comparisons with previous published data regarding the same HPV variants. In our works, we analyzed further, the intra-type variability for each of the types studied. The variation rate for each ORF/region was consistent with that presented by Ying Liu and colleagues (Liu et al. 2017). We further performed a visual inspection of the candidate gene selected in order to choose a short region with a well-defined diversity profile. Our targeted regions furthermore, should include flanking regions with low rate of nucleotide polymorphisms in order to design our amplification primers. For HPV16, PCRs were designed for each of the genomic regions, nonetheless, LCR, *L2* and *E6* targets rendered the best results in terms of amplicon quality and suitability for Sanger sequencing, as well as for the number of samples that tested positive. Additionally, *E6* was selected as target region because it is one of the HPV16 most chosen genes in HPV16 literature (Andersson et al. 2000; Asadurian et al. 2007; Grodzki et al. 2006; Zehbe et al. 2003).

5.2.1 HPV variant classification and nomenclature systems

Not only the selection of the genetic targets, but also the nomenclature used provides misleading variant information. HPV16 variants have been referred with delusive nomenclatures based on geographical references (*e.g.* "European") or on ill-defined arbitrary classifications (*e.g.* "prototype" or "non-prototype") that affect as well, HPV6 and HPV11 types (Cornet et al. 2012; Heinzel et al. 1995) In the case of "prototype" and "non-prototype" terms, it may suggest an ancestry, however, this nomenclature is not an indicative of an ancestral sequence and consequently, this approach provide tricky classification (Sichero and Villa 2006). In the case of HPV16, the usage of a geography-based nomenclature conveys a message of a close match between differential HPV16 variants prevalence and geography, which is again not justified by the best available data (Cornet et al. 2012). Historically, for HPV16, *E6* was used as target to phylogenetically classify variants (Andersson et al. 2000; Zehbe et al. 2009). In most cases, this ORF selection was due to the fact that SNPs within the gene had been related to an increase in persistence and risk of progression in cervical lesions produced by HPV16 (Togtema et al. 2015; Zehbe et al. 1998). This is classically the case of the polymorphic site T350G (L83V) (Togtema et al. 2015; Zehbe et al. 1998). Based mainly on T350G SNP or based on the combination of SNPs within *E6*, it is common to find in the literature linages referred as "*E6* variants" (Cornet et al. 2013; Larsson et al. 2012, 2013) and the subsequent classification in "European prototype *E6*" or "European T350G *E6* variants" (Togtema et al. 2015). However this nomenclature is also delusory as T350G polymorphism is not "European specific" and can be also found in African and NA/AA lineages (Cornet et al. 2012). In our Manuscript 2, we have explored the T350G SNP to confirm previous literature communications: T350G is an specific trait of a determinate HPV16 variant (Cornet et al. 2012). We have shown that the T350G status is not a synapomorphy of any HPV16 lineage, and consequently no HPV16 lineage can be fully and solely described based on the T350G status. We suggest that any analysis of the differential oncogenic potential of position *E6*-350 that does not correct for the variant genotype effect is indeed, most likely flawed. We present this finding as an example of the current biased nomenclature system used to define HPV variants and we suggest that studying HPV variants requires clear-cut and unambiguous definition of viral lineages.

Given the coexistence of several, often confounding variant nomenclature systems, the efforts of some authors such Burk and coworkers in delineating boundaries for taxonomical description of *AlphaPVs* variants have been remarkable (Burk et al. 2011). We subscribe to the

need of using normalized variants taxonomy, as only this consensus will afford inter-study comparison and consistency, and we have adhered to the proposed alphanumeric variant terminology (Burk et al. 2013).

5.2.2 HPV6, HPV11 and HPV16 different variability and selection of the most informative regions

For each HPV type studied, we created multiple sequence alignment using all available full-length genome sequences. Our values show different grades of variability between viruses: for HPV11 full-genome alignment, only thirty-nine (39) out of the 7878 nucleotides to present variability (0.5%), resulting in 77 alignment patterns; for HPV6, the number of variable positions in the full-genome alignment was 207 out of 8047 (2.6%), rendering 172 different alignment patterns. Genetic diversity was around four times higher for HPV6 than for HPV11 (p <0.01, Wilcoxon Mann-Whitney test). These findings agreed with previous literature (Burk et al. 2011; Matos et al. 2013). Indeed, Matos and colleagues observed an increased nucleotide variability for HPV6 isolates compared with HPV11 in a RRP sample set (Matos et al. 2013). They observed around 25% higher amino acid variation rate for HPV6 than for HPV11 sequences (Matos et al. 2013). For HPV16, the final alignment for the full-length genome comprised 7925 nucleotides, 735 variable positions and 548 distinct alignment patterns. The genetic variability observed is in greement with previous reports (Burk et al. 2013; Shang et al. 2011; Vrtačnik Bokal et al. 2010).

Our findings further evidenced that certain genomic regions exhibit greater heterogeneity than others, what agreed with previous published data (Harari et al. 2014; Liu et al. 2017; van der Weele, Meijer, and King 2017). Although not as clear correlation with malignant potential as in HPV16 variants, some authors suggest that HPV6 variability within LCR (especially rearrangements) might be linked to an enhanced HPV6 malignant potential, related with Buschke-Löwenstein tumours rather than in condylomata acuminate (Rübben et al. 1992). For HPV16, Ying Liu and colleagues suggest that T cells recognition of *E6* and *E7* epitopes might be greatly affected by this sequence genetic variability, which implies that HPV variant-specific epitopes differentially activate T cells (Liu et al. 2017). Sun and colleagues, further report that HPV16 intratype variability could harbour in distinct capacity of HPV16 variants to escape from immune recognition by the host immune system providing an increased fitness (Orlando et al. 2013; Sun et al. 2012) and may be responsible for important changes regarding increased persistence or viral load (Amador-Molina et al. 2013; Banister et al. 2015). Although the significance of nucleotide changes that lead to amino acid shifts is not properly known,

and some studies indicate that they are not specific to any related pathology, (Danielewski et al. 2013) amino acid alterations cannot be ruled out to play a role in disease progression without further investigation (Xi et al. 2017).

At the type level, which is considered as the reference taxonomic level for the PVs community (Bernard et al. 2010), even two types as close as HPV16 and HPV33 show differential prevalence in asymptomatic cervical infections and in cervical cancer (Bruni et al. 2010; de Sanjose et al. 2010). A similar picture of phenotypic diversity could arise when considering HPV variants (Chen et al. 2014; Cornet et al. 2012, 2013; Xi et al. 2013). Nonetheless, future larger studies on HPV viral diversity in asymptomatic, productive, benign, premalignant and malignant infections including invasive cancers are warranted.

5.2.3 Strengths on describing HPV intratype genetic variability according to the phylogenetic background comparing the evolutionary relationships between the detected DNA and those available in the databases

Recently, new methodologies to place short sequences into reference phylogenies have been proposed (Berger and Stamatakis 2011). These procedures have been designed to face problems such assigning a phylogenetic position to the vast amount of short reads obtained in using NGS procedures and one good example of such new methods is the evolutionary placement algorithm (EPA) (Berger and Stamatakis 2011; Stark et al. 2010) implemented in the RAxML software, which identifies the optimal insertion position of a sequence onto a well-resolved tree with the maximum accuracy, providing Likelihood weights for the placement of the partial sequences into the different nodes (Berger and Stamatakis 2011; Stark et al. 2010). The EPA methodology has been previously tested and applied to the phylogenetic placement of short PV sequences (Mengual-Chuliá et al. 2012). Importantly, the selection of likelihood-based methods provides alternative results to the well-known BLAST search and can further be translated under a phylogenetic background. Our methodology (used in the three Manuscripts) works with strictly bifurcating phylogenetic reference trees (RefTrees) as scaffold, in which the short sequences obtained through PCR and Sanger sequencing are placed in their most likely position, without modifying the topology of the RefTree, maintaining fixed the input developed to cope topology of the reference phylogeny (Stark et al., 2010). To be more exhaustive, in some cases, we edited the reference alignment using Gblocks, just because this approach eliminates poorly aligned positions and divergent regions of the DNA or protein alignment so that it becomes more suitable for phylogenetic analysis,

(Castresana 2000). In our manuscripts, the phylogenetic relationships reconstructed for the RefTrees rendered high-support values, and matched previous descriptions in the literature other reference phylogenies (Burk et al. 2011; Pimenoff et al. 2016). For each partial sequence, we integrated the likelihood weights for all nodes using 0.6, as a cut-off value, to confidently assign each sample into a particular HPV variant lineages and to not fall in misleading phylogenetic classifications. In our works, almost the totality of our samples were classified with LH weights > 0.7. Using the ML-based EPA approach to place short sequences into a well-resolved tree provides an evident gain in the output, which contains information not only about the best hit to the new short sequence but also about the node(s), to which it is more likely to belong (Berger and Stamatakis 2011).

The strong points of the methodology applied are that the scaffold tree is computed with a large amount of genomic information, and therefore with confidence. Compared to other approaches, it is less time-consuming, decreasing the required time from hours/days to seconds. Additionally, the (different) insertion position(s) are identified and evaluated through the assigned likelihood. EPA approach left us to compare and discuss better our results with the previous data published.

5.3 Genetic variants present different distribution among anatomical lesion or cancer presentation

5.3.1 Quantification of the relative contributions of variant differential abundance, geographical origin, and anatomical location/histological presentation to differential prevalence of HPV16 linages. Strengths of our statistical models

To provide *bonafide* results in HPV variant distribution among distinct anatomical sites, we selected statistical tests that could better fit our categorical data presented in contingency tables and Bonferroni-corrected in all cases for multiple comparison (Manuscript 2 Supplementary Material 5 or Manuscript 3 Table 3). For HPV16, in order to not reduce our statistical power, subvariants were lumped at variant level (*i.e.* HPV16_D1, D2 and D3 subvariants were lumped at HPV16_D). We have quantified for the first time the relative contributions of variant overall differential abundance, geographical origin, and anatomical location/histological presentation to the observation of HPV16 lineage differential prevalence, using a Generalized Linear Model (GLM). We selected this approach because it allowed the

generalization of ordinary linear regression to fit response variables presenting distribution models other than a normal distribution, and because it rendered a more robust fit by estimating individual variance values for each measurement via a link function (Mccullagh and Nelder 1972; Piegorsch and Bailer 2005). Additionally, the GLM allowed us to quantify the explanatory power of each of the factor considered in the response variable (Mccullagh and Nelder 1972). This methodology provides thus with important additional information compared with other classic approaches such as prevalence ratios (Piegorsch and Bailer 2005). Indeed we could quantify that the anatomical location of the cancer explained only 3% of the HPV16 variant differential prevalence (Manuscript 2) whereas the different ICC histological presentation explained around 10% of the HPV16 variant differential prevalence (Manuscript 3).

Most of the current literature describes HPV variant distribution in one single anatomical location, analyzing either variant geographical distribution or variant association with distinct interacting variables (*e.g* sex, genetic background, age at HPV onset or duration of disease) (Komloš et al. 2013; Larsson et al. 2012, 2013; Li et al. 2011; Mounts and Kashima 1984; Tornesello et al. 2008; Xi et al. 1998). Other authors analyze different anatomical locations but nonetheless, they remain at the taxonomic level of type instead of variants (Garland et al. 2009; Komloš et al. 2012; de Sanjosé et al. 2014). For example, Komlos and colleagues assess HPV6 and HPV11 distribution in a large sample set of GWs and RRPs, nonetheless, they do not analyse below type level (Komloš et al. 2012). Regarding HPV16 genetic diversity, most literature focuses on the uterus cervix (Cornet et al. 2012), with a lack of HPV16 lineages description in other anogenital sites. In order to cope with this biased literature, the presented manuscripts try to provide all together information about LR and HR-HPV6, 11 and 16 diversity among different anatomical sites and among different geographical regions.

Although without large sample sizes, some studies describe differential prevalence of HPV variants among distinct pathologies and point towards the same direction as our results. Danielewski and coworkers performed a comparative analysis of HPV6 and HPV11 genetic variability to determine whether different clinical manifestations (RRP, genital warts, cervical lesions and anal cancer) were differentially associated to distinct variant prevalences. For HPV6, they observed a higher presence of HPV6_B1 variants in genital warts compared to other anogenital lesions and compared to RRP. In contrast, for HPV11, they observed that almost all isolates were HPV11_A2 independently of the lesion studied (*e.g.* RRP, genital warts, cervical lesions and anal cancer). Kristina Komlos and colleagues also described the

presence of HPV6_B1 in 17 out of 18 patients presenting concurrent multiple GWs This work shows that the prevalence of HPV6_B1 in GWs is almost eight times higher than that of HPV6_A .Recently, Flores-Díaz and colleagues have published the only two case-control studies analyzing HPV6 and HPV11 variants distribution in GWs and RRP (Flores-Díaz et al. 2017a, 2017b). They communicate a higher prevalence of HPV6_B1 for GWs in their case group compared to the control group, while for HPV6_B3 they observe a similar prevalence in both cases and controls (Flores-Díaz et al. 2017b). For HPV11, the same authors detect HPV11_A2 as the most prevalent in GWs and in the control group, with no particular increased risk for GW development (Flores-Díaz et al. 2017a). Flores-Díaz and colleagues findings agree with previous published works (Heinzel et al. 1995; Jelen et al. 2014; Matos et al. 2013; Maver et al. 2011) and with our results. Although our work is not a case-control study, the *ca.* 2-fold increased ratio of HPV6_B1-positive GWs compared to positive healthy genital tissue presented by Flores-Díaz and colleagues may suggest differences on disease development depending on HPV6 variants status. HPV6 variants differential success, however, cannot be evaluated alone on the distribution of viral variants in pathological outcome, as we have assessed, but rather on the description of HPV6 lineages in asymptomatic infections, which hold the largest fraction of this viral population, that could provide more information about HPV variant transmission. Furthermore, checking ratios of HPV6 variants in healthy/disease population may disentangle whether HPV6 variants, and specially B variants present an associated risk of developing GWs.

HPV16 initial infection progression to cervical cancer has been linked to sequence properties (Berumen et al. 2001; Freitas et al. 2014; Grodzki et al. 2006; Jackson et al. 2016; Schiffman et al. 2010; Villa et al. 2000; van der Weele et al. 2017; Xi et al. 1997, 2007; Zuna et al. 2009) . Recently, van der Weele and colleagues published a study analyzing HPV16 variant prevalence in persistent infections compared with clearing infections. They observed that HPV16_A4, C, and D seem to clear preferentially compared to HPV16_A1 and A2 (van der Weele et al. 2017), concordant with the increased prevalence of HPV16_A1-3 detected in our ICC data. The HPV16 variant distribution described in our works agrees further with Zuna and coworkers follow-up study in HPV16 cervical cancers (Zuna et al. 2011). They observed that 85.2% of the HPV16-positive cervical cancers (N=155) harbored HPV16_A1-3 (Zuna et al. 2011) Furthermore, they showed a worse prognosis for women infected with HPV16_A1-3 compared with those infected with other HPV16 variants. While 31% of women with cancers with HPV16_A1-3 variants died from cervical cancer during follow-up, only 4.4% non-

HPV16_A1-3 cases died of cancer (Zuna et al. 2011). These results showed that viral genetic background influences clinical patterns and/or prognosis in fully evolved cervical cancers. For the rest of anatomical locations, the same scenario is described: in vulvar cancer, local studies in Sweden and in North America described an increased prevalence of HPV16_A1-3 (N=29/31; N=5/9 respectively) (de Koning et al. 2008; Larsson et al. 2013). Other research describes only the presence of HPV16_A1-3 in European vaginal samples (Larsson et al. 2013). A local study in Canada showed that 90% of HPV16 variants in anal cancers were HPV16_A (Ouhoummane et al. 2013). In penile cancers, an Italian study showed above 40% prevalence for both HPV16_A1-3 and D variants(Tornesello et al. 2008). Finally, a Mexican study showed 92% prevalence of HPV16_A1-3 and 8% prevalence of HPV16_D in penis cancers (López-Romero et al. 2013). Nonetheless, the high prevalence of HPV16_A1-3 in almost all anogenital sites is not observed by other authors (Freitas et al. 2014; Sichero et al. 2007).Our results agree partially with the data published: we show that there are no large differences between HPV16 lineage prevalence values among the anogenital cancers and that the most prevalent viral lineage is by far HPV16_A1-3, independently of the anatomical location of the samples. The only exception observed in our data was the sharply increased HPV16_A4 variants for anal cancers and the slightly but also evident increased presence of the same A4 variants in vagina for Asian isolates. Nonetheless, this small difference of HPV16 variant distribution among anogenital sites should be considered carefully as it is based on small sample size (N=21 and N=9, respectively)and because for anal location, data are not stratified by gender variable. When the analyses were performed after gender stratification (data not shown), the sample size obtained was reduced and the statistical power of the assessment was largely decreased, and we finally refrained from such stratification.

HPV16 distribution among different histological cancer types has been less studied. Generally, epidemiological data are based on SCC (Chopjitt et al. 2009; Hu et al. 2001) as these are the most prevalent HPV-associated cancers (Vinh-Hung et al. 2007; Vizcaino et al. 2000). A number of local studies however have addressed this question (Burk et al. 2003; Lizano 2006; Qmichou et al. 2013; Tornesello et al. 2011). Thus, with the third article (Manuscript 3) we wanted to analyse the HPV16 variant distribution in a large sample size selection of isolates from the three most prevalent ICC histological cancer types: SCC, ADC and ADSC. Our observations agree with the previous literature, describing an increased prevalence of HPV16_D variants mainly in glandular ICC presentations compared to the increased presence of HPV16 A1-3 in SCC (Burk et al. 2003; Mirabello et al. 2016; Quint et al. 2010). Burk and

colleagues, for example show an increased presence of HPV16 D variants in ADC and ADSC (38% of N=21) compared with SCC (3% of N=57) (Burk et al. 2003; Quint et al. 2010). Mirabello and colleagues, in a recent meta-analysis, show the same results: increased prevalence of HPV16_D in glandular ICC presentations, whereas they detect HPV16_A1 and A2 to be more present in SCC. Almost equally, in our work we observe that HPV16_D, and more specifically HPV16_D2 and D3 (particularly HPV16_D2), show an enhanced prevalence in ADC and might display enhanced risk in glandular ICCs. There is some evidence that HPV16 entry and processing in glandular compared with squamous cells is variant-specific (Quint et al. 2010). The possibility that variants might differ in their carcinogenic potential depending on the cell type infected may be an alternative hypothesis. Thus, the basis of the relationship between specific HPV16 variants and cancer histological subtypes warrants further study.

5.4 Host-virus interactions are the driving forces leading differential niche colonization and malignisation among HPV

Interaction between host genetic background (*i.e.* host immune surveillance system, specific-tissue hormone production or local infected cellular tissues microenviroinment) and viral genetic background (*i.e* intratype variability or specific SNPs) (Jackson et al. 2016; Xi et al. 2017) may lead to uneven HPV variants prevalence between anogenital locations, histological presentation or geographical region (Burk et al. 2003; Doorbar 2006; Fujiwara et al. 1997; Herfs et al. 2012; Mittal, Pater, and Pater 1993)

5.4.1 Particular polymorphisms within HPV16 E6 gene may increase carcinogenic potential

For HPV16, the T350G polymorphism resulting in a L83V change in the *E6* oncoprotein has been a long-studied SNP, as it has been differentially associated with an enhanced cell transformation and carcinogenic potential (Grodzki et al. 2006; Jackson et al. 2016; Zehbe et al. 1998, 2001). Furthermore, several authors suggested that coding changes in *E6,* and specially L83V shift, provided strong mechanistic and functional consequences for infection and contributed to marked differences in differential cancer risk for HPV16 variants (Berumen et al. 2001; Freitas et al. 2014; Jackson et al. 2016; Schiffman et al. 2010; Villa et al. 2000; Xi et al. 1997, 2007; Zuna et al. 2009). In our research we did not perform any functional assessment, but we studied the T350G polymorphism distribution among different anatomical locations and among geographical regions. It is common to find in the literature variants referred to as "*E6* variants" based on the combination of SNPs within this gene

(Larsson et al. 2012, 2013). However, T350G status is not a synapomorphic trait of any specific HPV16 lineage, and consequently no HPV16 lineage can be uniquely determined on the T350G status: position 350 is polymorphic T/G in the A1-3 and D clades, monomorphic G in the A4 clade and monomorphic T in the C clade. This distribution was in agreement with other works (Cornet et al. 2012). For this reason, we took a mixed approach, classifying samples using phylogenetic relationships and in addition, we studied samples according to SNPs variations inside the *E6* gene.

To address the question of the differential T350G prevalence as a function of the geographical origin and the anatomical location, in our research, we focused exclusively on HPV16_A1-3 as we had enough sample size and because it is the largely predominant linage in almost all geographic regions and anatomical sites.Within HPV16_A1-3 variants, we found that the T350G allele frequencies were not different among anatomical locations for the same geographical origin. However, our assessment revealed an enhanced 350G allele frequency in isolates from Central-South America compared with Europe. This pattern was observed for all anatomical locations with exception of vulva. Our results agree with previous studies reporting an increased frequency of the 350G allele in Central-South America compared with European populations (Cornet et al. 2013; Freitas et al. 2014; Villa et al. 2000; Zuna et al. 2011), as well as with the reduced contribution of this allele in vulvar and in vaginal lesions (Larsson et al. 2012; Ouhoummane et al. 2013). Indeed, Cornet and colleagues showed that the 350G allele was more common among cases from South/Central America than among controls for HPV16 A1-3 variants, what suggested a certain relationship with cervical cancer (Cornet et al. 2013)

5.4.2 Contribution of sexual hormones in HPV16 variant distinct carcinogenic capacity

Strong evidence of sexual hormone contribution in HPV16 variant cervical carcinogenesis among distinct histological presentations has been provided (Chan, Klock, and Bernard 1989; Sanborn, Held, and Kuo 1976). Studies with transgenic mice expressing HPV16 *E6* and *E7* showed distinct capacity of tumour development depending on sexual hormones (Chung, Franceschi, and Lambert 2010). These experimental conditions are intended to mimic those in premenopausal women exposed to continuous estrogenic stimulation such as from oral contraceptives or as a consequence of pregnancy (Chung et al. 2010). In these transgenic mice, cancers frequently arose in the transformation zone of the cervix where columnar epithelium converts to squamous epithelium (Egawa et al. 2015; Yang et al. 2015).

Endocervical columnar cells, where ADC and ADSC occur, express progesterone and estrogen receptors that are scarcely or absent in the basal cells of the squamous epithelium (Burk et al. 2003; Chan et al. 1989). The presence of this receptors may affect the Glucocorticoid Responsive Elements (GREs) and estrogen-response elements present in the HPV16 URR facilitating the carcinogenic process (Chan et al. 1989). In HPV16_D, Burk and colleagues and other authors show that particular polymorphisms located in the LCR GREs may probably confer HPV16_D lineage an increased response to sexual hormones (Burk et al. 2003; Fujiwara et al. 1997) affecting the D linage viral machinery and probably leading the tumor progression (Burk et al. 2003; Chan et al. 1989). Lace and coworkers and other researchers described that HPV16_D variants show higher transcriptional activity than HPV16_A1-3 from promoter regions that contain Glucocorticoids and estrogen responsive elements. (Lace et al. 2009; Sichero, Franco, and Villa 2005). In our work, we observe an increased prevalence of HPV16_D in ADC and ADSC where GREs are more prevalent compared to SCC, almost absent
At functional level, Kammer and colleagues showed enhanced *p97* promoter activity for HPV16_D (ca. 3 fold increased) driving to an augmented *E6-E7* oncogene activity compared to HPV16_A1-3, A4 , B and C, which presented a similar activity of the promoter (Kammer et al. 2000). Thus, SNPs located in glucocorticoid responsive elements within the regulatory region (Burk et al. 2003) could be related to the uneven early promoter activation among HPV16 variants (specifically HPV16_D), suggesting a synergic interplay between sexual hormones and HPV infection progression in humans (Chung et al. 2010). In Manuscript 3 we observe an increased prevalence of HPV16_D in ADC and ADSC compared to SCC, in agreement with the hormone context described and with previous published works (Mirabello et al. 2016). Mirabello and colleagues describe HPV16_D2 and D3 linked to a substantially increased risk in ADC (odds ratios >100 for D2) (Mirabello et al. 2016). Our descriptive results only show an enhanced prevalence of HPV16_D variants in adenocarcinoma composed principally by glandular cellular component.

5.4.3 Host genetic background may contribute in HPV16 uneven capacity of cancer developement

It has been described that host immune surveillance system may occasionally provide advantageous conditions for HPV16 variants within a particular cancer niche (Vartanian et al. 2008; Wang et al. 2014) . For example, it has been described that the APOBEC3 internal mutators introduce nucleotide modification in the HPV genome, leading directional C to T changes in the *E2* and the URR of the HPV16 genome which, within a particular infected cell

type, may originate an HPV16 variant mixture (Vartanian et al. 2008; Wang et al. 2014). Variants mixture may facilitate the oncogenic progression of HPV16 infected lesions (Kukimoto and Muramatsu 2015). Kukimoto and colleagues for example, describe the presence of HPV16 *E1* variants mixture in single clinical specimens and suggest that mixed variants may facilitate lesion oncogenic progression as *E1* functions are normally lost or silenced during cervical cancer progression (Kukimoto et al. 2013). Furthermore, it has been proposed that HLA locus also play a role in HPV induced lesions, influencing the clinical outcome(Madeleine et al. 2008; Marangon et al. 2013) and some HLA alleles in conjunction with specific HPV16 variants, may predispose to the development of cervical cancer (Hildesheim and Wang 2002). However, this association may vary depending on the studied population (Hildesheim and Wang 2002). Zehbe and colleagues, for example, showed that women infected with the HPV16 variants containing the T350G polymorphism within *E6* that presented HLA-B*44, B*51 and B*57 alleles had four to five times greater risk for cervical cancer than those bearing the HLA-B*15 allele, which presented a protective effect (Zehbe et al. 2003).

HPV16 in interaction with host genetic background could have developed a number of key adaptations, that allow its persistence in infected epithelial cells, even in the face of an active immune system (Christensen 2016). Identification of the key changes in the HPV16 genome and clarification of their role in cervical carcinogenesis are central research topics and are likely to provide novel strategies for the diagnosis and clinical management of HPV16 infected lesions in the future. Finally, although prophylactic vaccines targeting HPV16 are expected to be effective against all HPV16 variant, careful monitoring of the changes in the distributions of HPV16 variants may be necessary to fully assess the effectiveness of these vaccines in the post-vaccination era (Galani and Christodoulou 2009; Kabekkodu et al. 2015; Paavonen et al. 2009; Roden et al. 1997; Zhang et al. 2016)

5.5 HPV16 variants present different distribution among geographical region

It has been proposed that the differential risk of suffering cervical cancer produced by HPV16 variants infection may differ depending on the genetic background of the patient (Lopera et al. 2014). Unfortunately, the samples we used were not collected with an informed consent that may have allowed the analysis on the human genetic component. Thus, for Manuscripts 2 and 3, we resorted therefore to geography as a surrogate to genetic background. We followed the

geographical units used by the United Nations, although we definitely aware that certain region (extreme parts) within the same continent would most likely display large genetic distances. Previous literature suggests that geography-specific lineages are not common for HPV6 or HPV11 as prevalence values are similar among different geographical regions throughout the world (Danielewski et al. 2013). Other studies including more than 15 countries, further confirmed the absence of geographical clustering of HPV6 variants (Jelen et al. 2014). For Manuscript 1, our initial aim was to describe the different distribution of HPV6 and HPV11 variants among distinct pathologies, while in our second and third Manuscripts the main aims contemplated the HPV16 variant distribution depending on geographical region, supported by bibliographical data. Indeed HPV16 presents a phylogeographic match, indicating a correlation between HPV16 variant distribution and geographical origin and ethnicity (Cornet et al. 2012; Jelen et al. 2014; Pimenoff et al. 2016; Yamada et al. 1997). After out-of-Africa migration of modern human ancestors, sexual transmission between human populations introduced HPV16A into modern human ancestor populations. It is hypothesized that differential coevolution of HPV16 lineages with different but closely related ancestral human populations and subsequent host-switch events in parallel with introgression of archaic alleles into the genomes of modern human ancestors may be largely responsible for the present-day differential prevalence and association with cancers for HPV16 variants (Pimenoff et al. 2016). Indeed, the prevalence of HPV16_D for example, in some aboriginal regions such New Guinea or in some Asian Indians and Amerindian ethnics suggests that the HPV16 distribution was previously to the arrival of people with distinct ethnicity (*i.e.* caucasic-European) who conform the current population (Yamada et al. 1997). In our manuscripts, for all invasive anogenital cancers and cervical cancer histologies, we show further a geography-dependent distribution of HPV16 variants largely based on the dominance of HPV16 A1-3 variants in Europe, the virtually exclusive presence of HPV16 B and C variants in Africa, the increased prevalence of HPV16 A4 variants in Asia and the enrichment of HPV16 D variants in the Americas. The observed distribution indeed, reflects an association between HPV16 variants and geography in anogenital cancer cases. Different factors such founder effects (*i.e.* small proportion of people infected with HPV16_D variant that formed the native American population) or coevolution of HPVs with human populations could also influence the observed distribution (Xi et al. 2006). Furthermore, human migration fluxes among time and viral fitness are proposed further to be possible explanation factors for the observed HPV16 variants among different geographical regions (Lehoux, D'Abramo, and

Archambault 2009; Orlando et al. 2013; Stanley 2012; Xi et al. 2006) Other authors such Bontkes and coworkers or de Araujo Souza and colleagues have proposed the importance of certain polymorphisms in the MHC, suggesting the immune-surveillance system as other factor that affects HPV16 variants infection and persistence among geographical regions (de Araujo Souza et al. 2008; Bontkes et al. 2000).

5.6 HPV16 cervical cancers and adenocarcinoma histological presentation are diagnosed at younger ages.

Available data shows that cervical cancers are diagnosed earlier than other anogenital cancers associated with HPVs (https://www.cdc.gov/cancer/hpv/statistics/age.htm). (Larsson et al. 2012, 2013; Ouhoummane et al. 2013; de Sanjose et al. 2010; Tornesello et al. 2008). Regarding ICC histology, ADC and ADSC are diagnosed earlier than SCC (Vinh-Hung et al. 2007). Other studies described a nearly ten years difference with early mid-forties for glandular carcinomas and early-mid fifties for SCC (Pérez, Cid, Iñarrea, Pato, Lamas, Couso, Gil, Alvarez, et al. 2014; dos Reis et al. 2007). From the viral perspective, specific HPVs are differentially associated with either certain cancers or cancer presentations (Bosch et al. 2002; Li et al. 2011; Smith et al. 2007). For HPV16, the contribution in noncervical cancers is higher than in cervical cancer: 61% in cervix (de Sanjose et al. 2010), 62.9% in penis (Alemany et al. 2016), 72.5% in vulva (de Sanjosé et al. 2013), and 75.8% in anus (Alemany et al. 2015). Thus, the increased contribution of HPV16 in noncervical cancers could result in earlier age at diagnosis of cervical cancers when comparing HPV-related cancers among locations. In order to avoid this bias, we performed the assessment with an strategy: (1) To compare the age at tumor diagnosis considering all the $LIPA_{25}$ detected types and (2) To limit our study cohort to solely HPV16 single infected. Furthermore, previously to the age at tumour diagnosis assessment, we corroborated that our HPV16 variants present small/minor differences in prevalence among distinct anatomical sites, what prevent us to obtain a biased age at tumour diagnosis due to the different presence of the HPV variants among tissues. We considered that the selection of single HPV16 infections was one of the main strength of our work because it provided the opportunity to pinpoint the source of the proposed differences in age at diagnosis between anatomical locations (in Manuscript 2) and for ADC and SCC (In Manuscript 3) for a unique oncogenic HPV. In our results we confirmed the worldwide trend of cervical cancers to be diagnosed significantly earlier than other anogenital cancers (early fifties vs. early sixties), in agreement with other authors (Viens et al. 2016). We further

compared age at tumour diagnosis between ADC and ADSC compared to SCC (mid-forties vs. mid-fifties). For IC histological presentations, our data showed that ADC were diagnosed earlier (mid-forties) than SCC (mid-fifties) only for HPV16_A1-3. We did not observe additional patterns on age at cancer diagnosis for other HPV16 variants. Our results agree with other studies suggesting the same age trend (Pérez, Cid, Iñarrea, Pato, Lamas, Couso, Gil, Alvarez, et al. 2014; dos Reis et al. 2007; Vinh-Hung et al. 2007). Our findings however, did not agree with Alfaro and colleagues work in which they described an increased risk for ADC development at younger ages (≤ 49 years old) associated with D2 variants (Alfaro et al. 2016). Such disagreement could arise from the low statistical power within D variants, in which our sample size was reduced (N=10 for SCC and N=36 for ADC). Nonetheless, data regarding HPV16 variant age associations in distinct IC histological presentation are scarce, preventing HPV16 variants to be considered as risk factor for early diagnosis of certain IC histological presentations. Indeed, more research on this area may elucidate strong age at tumour diagnosis patterns for certain histological presentations of IC and other anogenital carcinomas.

Cytological screening has decreased the incidence of SCC in a wide range of developed countries (Castellsagué et al. 2006; Ronco et al. 2014). However, it has not been effective in decreasing the incidence of cervical ADC (Bergström, Sparén, and Adami 1999). This could be the cause of glandular cancer types increasing in incidence in women aged between their twenties-forties(Vinh-Hung et al. 2007). As proposed by other authors (Burk et al. 2003), and previously discussed, we additionally suggest that adenocarcinomas early age of detection could be related with factors mainly present in glandular epithelia, such as progesterone and estrogen receptors present in endocervical cells and the possible interactions of hormons with HPV16 GREs located in the LCR (Burk et al. 2003). According to this background, early use of oral contraceptives could be considered as another risk factor for developing adenocarcinoma of the cervix uteri (Thomas and Ray 1996; Vizcaino et al. 2000). Our results however, contrast with the recent published data of Mirabello and colleagues, who did not observe trends in age (Mirabello et al. 2016). However, this disagreement may be methodological: Mirabello and coworkers reported “age at enrollment” which could largely predate the age at cancer diagnosis, data analysed in our work.

The younger age at ADC presentation may indicate a shorter time of progression, which reduces the opportunity to prevent these tumours through screening (Sasieni, Castanon, and Cuzick 2009). Since a large proportion of ADC are HPV-positive, it is possible that HPV tests

have a stronger impact on the prevention of ADC than cytological screening, especially in young females, a trend observed in our data for HPV16 type (Visioli et al. 2004). Boulet and colleagues indicate that a HPV test is more sensitive and equally specific as a cytological examination for the detection of women at risk of developing cervical cancer (Boulet et al. 2008). In a large prospective study, in which it has been assessed exposure to HPV before the date of diagnosis of the ADC cases, it has been found that infection with HPV16 and 18, detected in cytologically normal smears up to at least 14 years before ADC diagnosis is associated with substantially increased risks of subsequent development of invasive ADC and its precursor AIS at younger ages, 36 median age for AIS and 43 median age for ADC (Dahlström et al. 2010). The strong link between infection with HPV 16 or 18 and malignant transformation of the glandular epithelium of the cervix suggests that screening for these HPV types might be a useful tool for improving the prevention and/or early detection of adenocarcinomas of the cervix, which has proven to be difficult through regular cytological screening (Sasieni et al. 2009). Consequently, a deeper understanding of the etiology of cervical adenocarcinoma, and better preventive efforts are urgently called for.

5.7 Strengths and weaknesses

Despite the large sample size and the molecular identification of viral variants, the works compiled in this thesis, nonetheless, presented number of limitations:

Manuscript 1 for example, contained a large number of HPV6 and HPV11 partial sequences originated from our samples that only covered two main geographical regions, Europe and Oceania. We had to resort to GenBank available sequences to increase our geographical representation, but certain regions, such as Africa, Asia and South America, remained underrepresented. Therefore, it could be argued that the observed differences were rather due to a geographical bias for the origin of the samples investigated. However, previous research did not identify geographical origin as an important component of viral diversity for HPV6 and HPV11(Danielewski et al. 2013; Jelen et al. 2014) and local studies reported that the whole repertoire of HPV6 variants could be found in isolates originating from a single country (Kocjan *et al.*, 2009). Other recent research on RRPs isolates from South America, Europe and South Africa further confirmed this absence of geographical structure on the HPV6 sequence variability (Matos et al. 2013). For HPV11 nonetheless, the absence of differences in variant distribution among the two studied pathological outcomes could be due to small sample size and lack of statistical power.

In Manuscript 2, we selected only HPV16-monoinfected SCC, restricting our coverage only to Europe, Central-South America, and Asia with enough sample representation. This selection prevented us from providing sound numbers from other continents such Africa. Indeed, studies that assess evolution and host-pathogen interaction should include Africa as it is the continent where the genetic diversity in host is larger (The 1000 Genomes Project Consortium 2012). Furthermore, the lack of coverage for Africa strongly impacted HPV16_B and C variant representation, almost absent in the rest of continents.

A proper individual analysis of the genetic background of each patient would be extremely useful in the all presented Manuscripts. It would have allowed us to quantitatively handle each individual's ancestry and to introduce the human genetic background as a better factor for explaining HPV16 variants distribution. Nevertheless, when no alternative was available, as in our case, geography remained a useful proxy for human genetic background even in large regions with very admixed populations, as shown by Homburger and coworkers for South America (Homburger et al. 2015).

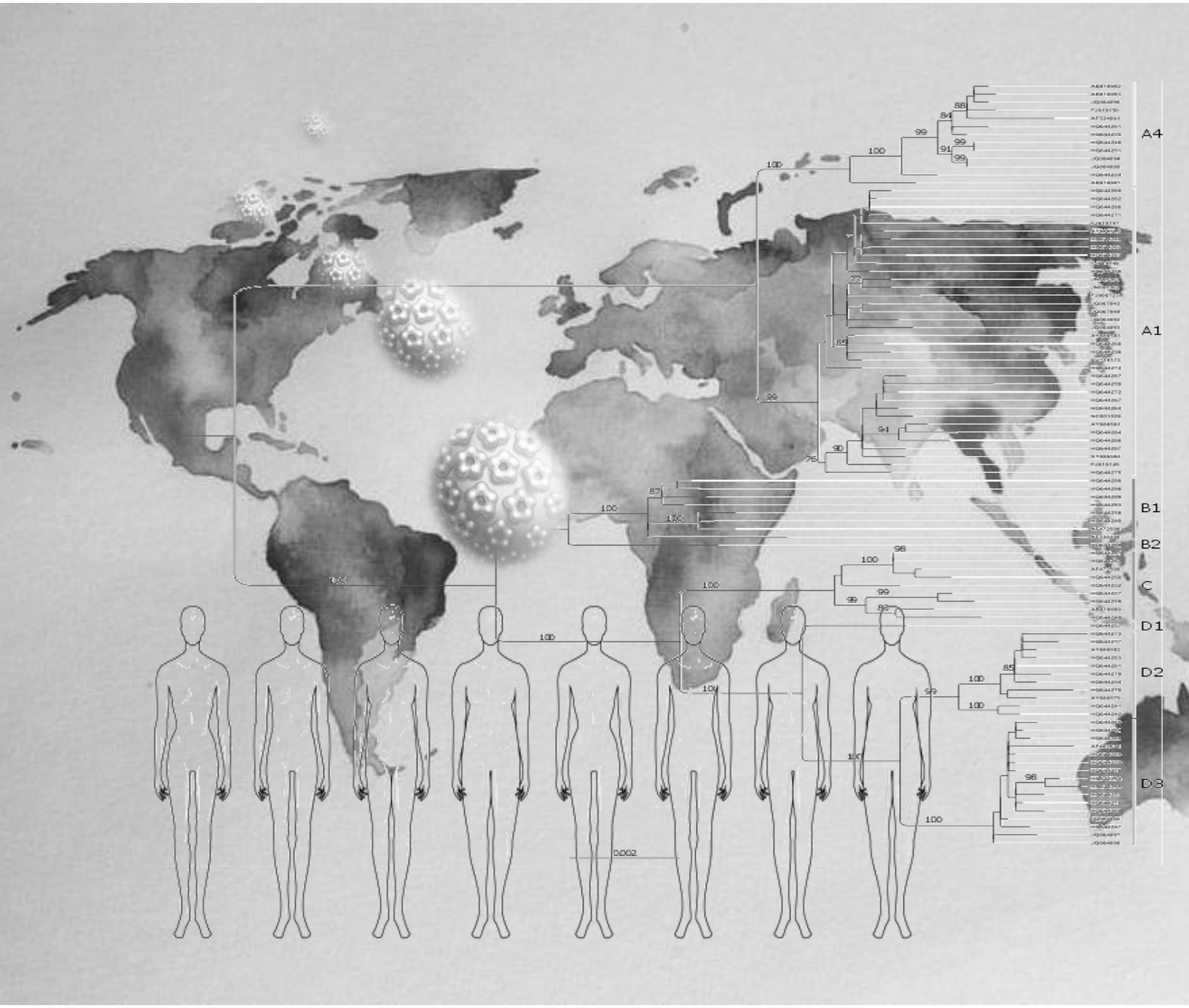
In Manuscript 2, we made the choice of studying a well-defined entity (*i.e.* squamous cell carcinomas associated only to HPV16 infections), and as a trade-off we could not find in our repositories enough samples from North America, Oceania or Africa (only present in small number in Manuscript 3). Furthermore, for the same manuscript, HPV16 variant oncogenic potential was proxied through the long-studied candidate T350G located in the *E6* oncogene.

In Manuscript 3, although we provided data for Africa, we only covered with good depth Europe, Central-South America, and Asia. Furthermore, the fact that adenosquamous cell carcinoma is considered a histological rare type, prevented us to work with a well-represented sample set, reducing it to small numbers compared to other two histological cancer types (SCC and ADC). For this reason, certain analyses were assessed twice.

Finally, the works compiled in this thesis were of descriptive nature, no case-control studies have been provided. Unfortunately, we could not stablish a differencial risk between variants for the variables analyzed (anatomy, geography and histology). Nonetheless, our works present a well-accurate description of the different HPV variants prevalence according to the interacting factors studied and suggest strong association of certain variants with different pathological outcomes and geographical regions.

As strengths, the presented works represent a complete analysis at worldwide level characterizing the viral component in the to our knowledge, most comprehensive sets of HPV6, HPV11 and HPV16 single infected pathological outcomes. Additionally, all works have been performed with a well-characterized study population: only restricting our analyses to single infections what has prevented to work with an ill-defined populations (infected with more than one HPV type). This curated selection, allowed us to define trends such younger ages at cancer diagnosis in the cervix, compared to other anogenital sites or in adenocarcinomas compared to squamous cell carcinomas. Furthermore, our third Manuscript has tackled adenosquamous cell carcinoma as an independent identity, providing new interesting data for this glandular histological cancer type, nowadays, scarce in current literature.

6. IMPLICATIONS



LR and HR-HPV intra-type genetic diversity and variant uneven distribution among pathological outcomes such HPV6 and HPV11 variants in GWs and RRPs and HPV16 variants in ICs, especially for certain histological presentations (i.e. SCC and ADC) open different perspectives and implications on public health framework. The current state of knowledge show not only different distribution of HPV variants according to lesion/cancer or to geographical world regions, but further presents substantial differences in the oncogenic potential of distinct HR-HPV linages, such HPV 16 variants, that are more pronounced than previously anticipated (Bernard 2013; Von Knebel Doeberitz 2016). Nonetheless, it is still unclear which are the mechanisms that contribute to these differences (Bernard 2013; Von Knebel Doeberitz 2016). The strongly increased risk of HPV 16_D2 variant in ADC in situ and ADC, suggests that this particular variant-specific increased oncogenicity is specifically linked to cells of glandular nature (Mirabello et al. 2016) but further, may be presumably linked to a plethora of viral and host genetics and immunologic factors. The data provide may impact on the design of refined HPV-typing tools for early detection of HPV based cervical cancer and screening algorithms in view that, there are still clinical difficulties associated with the identification of glandular endocervical lesions and cancers by current cytological(Mirabello et al. 2016; Quint et al. 2010; Smith et al. 2000). Knowing more about the associated risks of certain variants might help on global diagnostic as we could better understand the natural history of HPV infections (Schiffman et al. 2010) and the biological and clinical ramifications(Guan et al. 2012; Schiffman et al. 2010).

The works presented herein focus on those HPV types more prevalent in each of their related pathologies and cancers. Until the day, the best documented HPVs in the context of viral variants are HPV16 and HPV18 (De Boer et al. 2005; Lizano 2006; Pérez, Cid, Iñarrea, Pato, Lamas, Couso, Gil, Álvarez, et al. 2014) Nonetheless, there is also evidence that variants of other HR-HPVs (e.g. HPV31, HPV33 HPV58, HPV45) present further differential involvement in IC (Godínez et al. 2013; Xin et al. 2001). For HPV31, it has been reported that infections with A and B variants are associated with a significant increase in risk of CIN2 and CIN3 (Mirabello et al. 2018; Xi et al. 2012). For HPV33 it has been reported that non-A1 variants are associated more frequently with CIN1 and CIN2 while the A1 variants are associated with CIN3 and ICCs diagnoses (Xin et al. 2001). According to the variants increased risk context and together with the fact that some of these types are increasing its prevalence in high grade lesions and cancers (Liu et al. 2015) showing low rates of infection clearance (*e.g.* HPV31 and HPV33) (Bulkmans et al. 2007), it is warranted more research on variants among these types

in order to delineate which infections require treatment versus those that naturally regress, having subsequently, implications on the financial burden associated with cancer treatment globally. Indeed, HPV variants and subvariants assessments need to be extended not only at genitalia sites but also in still deeply to explore anatomical locations, such head and neck locations (HNC) (i.e oropharynx, tonsils, base of the tongue) (Combes and Franceschi 2018). High frequency of HPV16 or HPV33 in tonsil tumours confirms the hypothesis on the relevance of more research development in these anatomic site and show the relevance of determining the tissue susceptibility to infection and subsequent malignant events (Hassani et al. 2015; Snijders et al. 1992). Indeed, HPV16 account for the majority of HPV positive cases with a hot-spot in the tonsil area, described to be *ca.* 83% of the cases (Badaracco et al. 2007; Combes and Franceschi 2018). Although there is data on HPV associated ADC at the base of the tongue, it does not tackle HPV variant distribution (Hanna et al. 2013) which could be an interesting topic in order to establish correlations among tissue characteristics at variant level. Equally to cervical and other non-cervical anogenital SCC, other studies reveal an increased prevalence of HPV16 A1-3 variants in squamous HNC (Agrawal et al. 2008; Barbieri et al. 2014; Blakaj et al. 2012; Du et al. 2012) and show an increased prevalence of the polymorphic site T350G, compared with SCC (Boscolo-Rizzo et al. 2009; Combes and Franceschi 2018; Hassani et al. 2015). Nonetheless, other works analyzing oropharyngeal squamous cell carcinomas (OPSCCs) also show an increased prevalence of HPV16 B and C variants. Thus, the scarce literature on this anatomical locations and the fact that it has been further described an increased prevalence of B and C variants, nearly absent in our data and almost solely present in samples originating from Africa, emphasizes the necessity of further studies on this field.

Polymorphic sites present in both early (e.g. E1 and E2) and late genes (L1 and L2) support the fact that either viral entry (L2) or oncogene regulation (E2 and possibly E1) may play a role in the cellular differentiation pattern of the respective host cell and may provide a facilitated cellular environment for determinate HPV variants (i.e HPV16 A1-3 in SCC and HPV16_D in ADC) (Lee et al. 2008). For example, there is evidence that HPV16 E7 ORF is virtually invariable in ICC whereas it presents high variability in controls (Mirabello et al. 2017). Additionally, HPV16 E7 sequence is described to be less variable than other HR-HPVs (e.g. HPV31) (Mirabello et al. 2018). Although to be confirmed, a strict conservation of E7 sequence could represent a highly promising specific biomarker. All these data suggest again

that genetic variability within HPVs and pathological outcome present strong associations that should be considered in much deeper way.

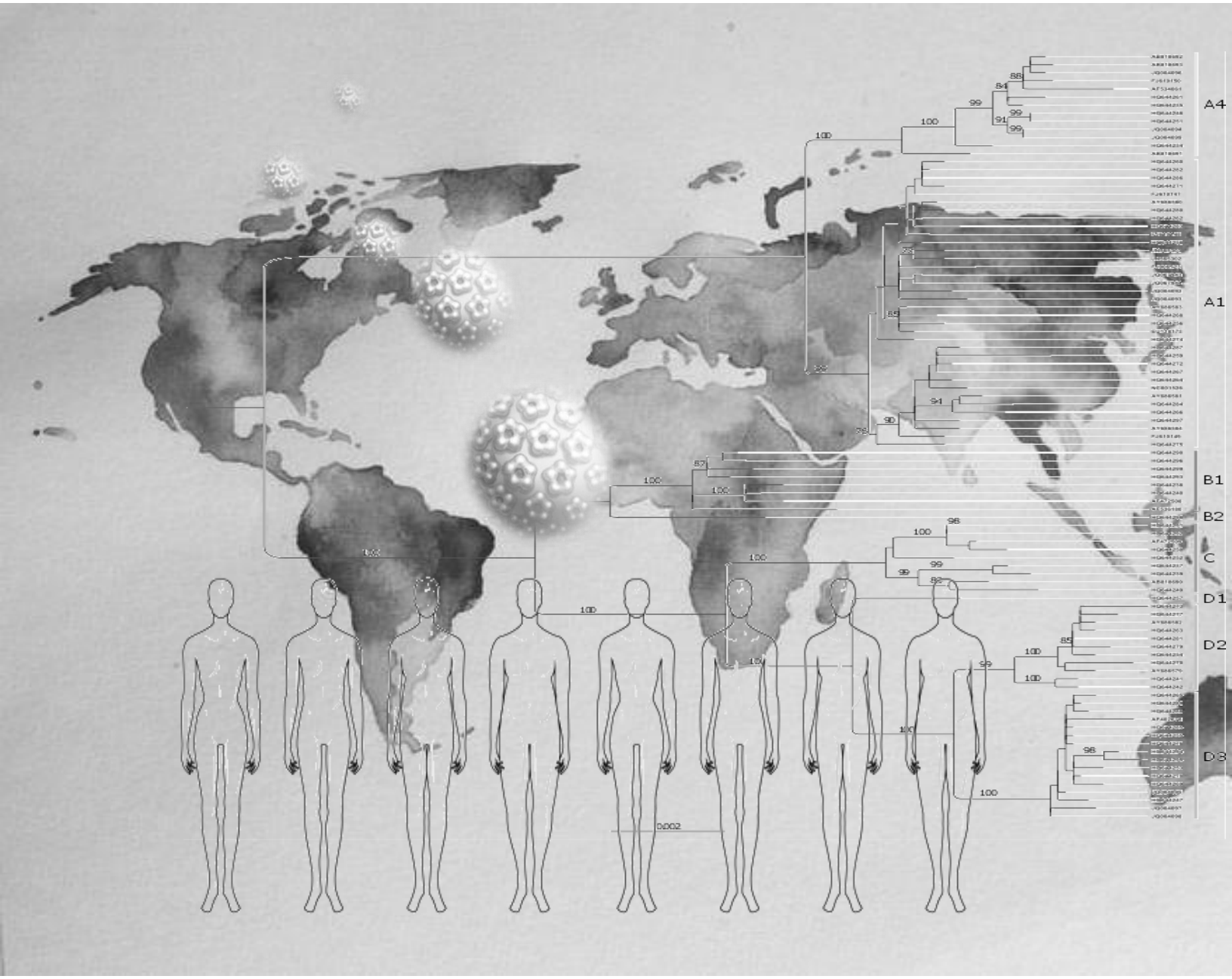
HPV variants seem further to have some impact in therapeutics. Functional studies have evidenced that HPV-16 variants respond differently to radiotherapy (Moreno-Acosta et al. 2017) presenting diverse radioresistance response (Moreno-Acosta et al. 2017). Regarding prophylaxis, the marketed HPV vaccines are composed of major capsid L1 Virus Like Particles (VLP) of HPV16/18 or HPV6/11/16/18 and VLP multivalent-based vaccine containing nine HPV Types (HPV 6/11/16/18/31/33/45/52/58) (Ma et al. 2011). The antibody response is mainly generated against conformational epitopes present on the outer VLP surface that are responsible for neutralizing antibody production (Stanley, Lowy, and Frazer 2006). The current state of knowledge clearly point to a complex interaction of minor differences of the viral genome and its host cell that may strongly influence the carcinogenic activity in a given cellular differentiation context (Mirabello et al. 2016). Complete sequencing of variant capsid genes followed by the development of immunologic test systems should be taken into account to achieve improved broad-spectrum vaccines. In a not so far future, an additional challenge will be then, to determine whether HPVs variants are relevant to vaccine strategies against HPVs.

While molecular sequence analyses from over the world show a high degree of sequence conservation between HPV 6 and 11 isolates (Ahmed, Bissett, and Beddows 2013; Danielewski et al. 2013) revealing that the genomic diversity of these two LR-HPVs has a minimal effect on vaccine antibody formation (Ahmed et al., 2013); for HR-HPVs, there is evidence that single amino acid substitutions within the L1 and L2 capsid genes may be important in the viral escape from neutralizing antibodies possibly leading scape tumors in vaccinated patients (Gurgel et al. 2015; Ryding et al. 2007). Indeed, there is evidence of nucleotide differences that result in amino acid changes for HPV16 L1 and that some of them map closely to the principal neutralization epitopes of monoclonal antibodies V5 and E70 (Chen et al. 2000), which might suggest that the HPV16 variants were established to escape neutralization (Fleury, Touzé, and Coursaget 2014; Gurgel et al. 2015; Ryding et al. 2007) For example, monoclonal E70 is able to neutralize HPV16_A1-3 variants but not HPV16_C variants pseudotype virions (Roden et al. 1997). Thus, mutation “hot spots,” which generally lie on the surface exposed residues of L1 pentamers, are likely to be the areas where neutralizing epitopes lie, but even these regions are structurally constrained. (Chen et al. 2000). Currently,

only young women are vaccinated but vaccination of men is going on in some countries since it can prevent penile, anal, and anogenital warts; and it could prevent oropharynx cancers and the transmission of HPV to their sexual partners. Providing insight into the selective pressures that may contribute to the phylogenetic diversity between HPV16 variants, determining the degree to which immunization with L1 VLPs of one variant may induce antibodies that cross-neutralize other HPV16 variants has important implications for the design of vaccine efficacy trials and for vaccine valency.

All this clinical context presented emphasize the huge importance and implications of HPV variants in public health context, spanning different clinical fields that involve gynecological screening, cancer development, cancer therapies, vaccine design and follow-up of vaccinated women in order to avoid possible not expected pathological outsiders. All the data provided and its application as far as possible, may help clinicians to identify the exact oncogenic potential of different HPVs in each patient separately. These results might impact prediction of HPV infections prognosis and treatment selection improving the personalized patient management.

7. CONCLUSIONS



For manuscript 1:

- HPV6 variants present different distribution among the two most common associated pathological outcomes, GWs and RRPs, with HPV6_B1 variants being more prevalent in GWs than in RRPs
- HPV11 variants are not differently distributed among GWs and RRPs showing for both pathologies an increased prevalence of HPV11_A2 variants

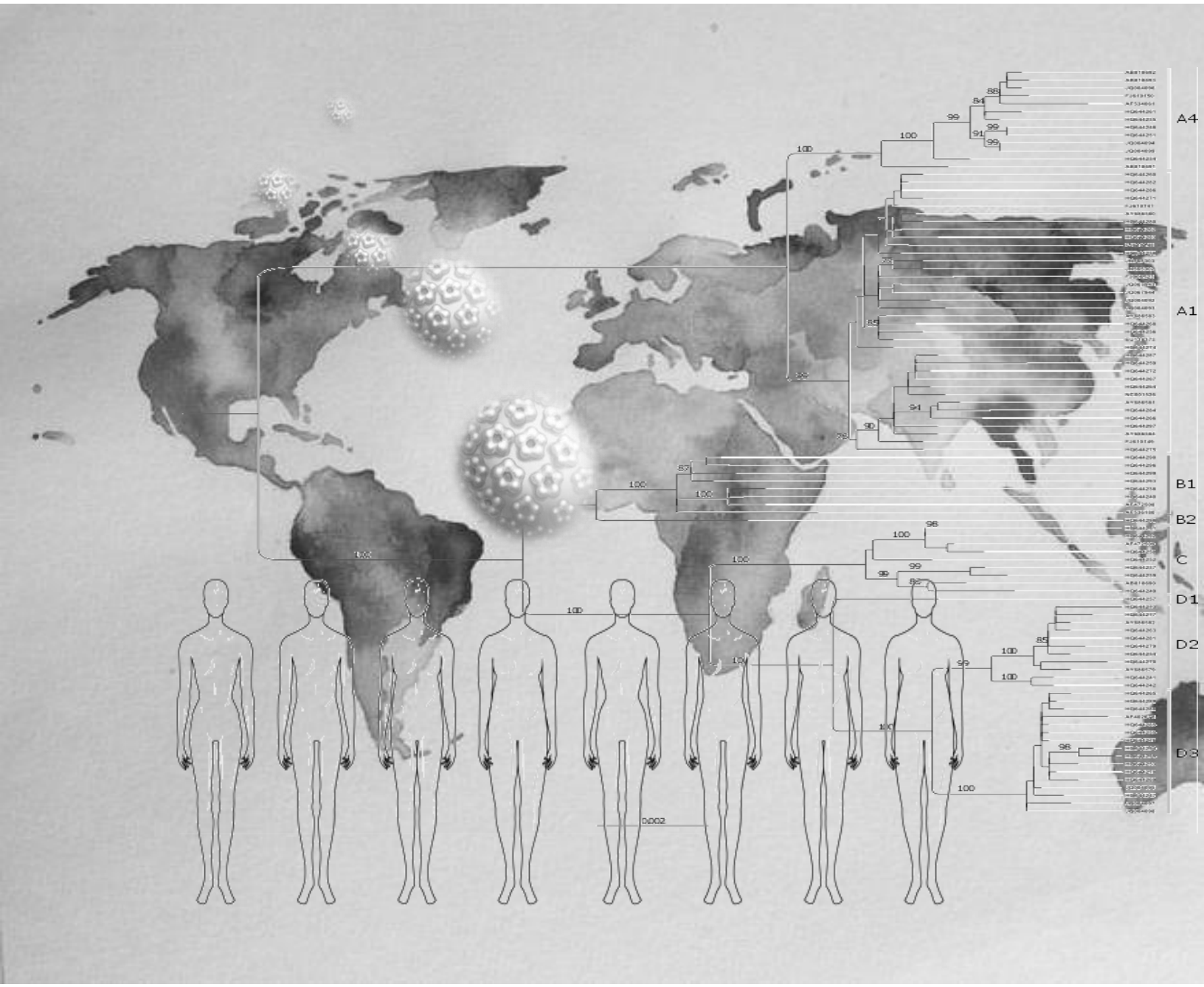
For manuscript 2:

- Anogenital invasive cancers anatomical location and geographical origin shape the differential prevalence of HPV16 variants.
- The most prevalent viral lineage is by far HPV16_A1-3, independently of geographical origin and anatomical location of the samples with the only exception of anal cancers in Asia, dominated by HPV16_A4 variants and the increased prevalence of HPV16_D variants in samples from Central/South America.
- The *E6* gene long-studied candidate T350G polymorphism presents a geographical dependent distribution within HPV16 A1-3 variants, showing an increased 350G allelic frequency in samples from Central/South America compared with samples from Europe
- Cervical cancers are diagnosed earlier (diagnosed in the early fifties) than non-cervical cancers (diagnosed in the early sixties). The differences in age at diagnosis remain unchanged when all HPVs positive cancers are considered, when only HPV16 monoinfections are considered, and even after focusing on cancers associated to the most prevalent viral lineage, HPV16 A1-3 variants
- We suggest different biological interplay between viruses and epithelia that could originate from genuine lineage-specific differences in viral fitness, in the potential to produce lesions or in HPV variant oncogenicity

For manuscript 3:

- HPV16 variants display different prevalence depending on the geographical origin of the samples and on the histologic cancer type.
- There is a large dominance of HPV16_A1-3 in Europe, a virtually exclusive presence of HPV16_B and C in Africa, an increased prevalence of HPV16_A4 in Asia and an enrichment of HPV16 D in the Americas.
- There is a higher prevalence of HPV16_A1-3 in SCC compared to ADC and ADSC where HPV16_D is more prevalent
- Cancer histology presentation conditions age at cancer diagnosis: HPV16-associated ADCs are diagnosed significantly earlier (late forties) than HPV16-associated SCCs (early fifties)

8. ABBREVIATIONS



A

ADC: *Adenocarcinoma*

ADSC: *Adenosquamous cell carcinoma*

AnIC: *Anal Invasive Carcinoma*

AF: *Africa*

AS: *Asia*

C

CIN: *Cervical Intraepithelial Neoplasia*

CSA: *Central-South America*

D

DEIA: *DNA Enzyme Immunoassay*

DNA: *Deoxyribonucleic Acid*

E

E: *Early gene*

EPA: *Evolutionary Placement Algorithm*

EUR: *Europe*

EV: *Epidermoplasia verruciformis*

F

FFPE*: Formalin Fixed Paraffin Embedded*

FG: *Full genome*

G

GRE: *Glucocorticoid Response Element*

GLM: *Generalized Linear Model*

GW*: Genital Wart*

H

HLA: *Human leukocyte antigen*

HPV: *Human Papillomavirus*

***HR-HPV**: High Risk Human Papillomavirus*

I

***IARC**: International Agency for Research on Cancer*

***IC:** Invasive carcinoma*

***ICC:** Invasive Cervical Carcinoma*

***ICO:** Institut Català d'Oncologia*

ICTV: International Committee on Taxonomy of Viruses

***IntE2L2**: Intergenic region between E2 and L2*

L

L: Late gene

LCR: Long Control Region

LH: Likelihood

***LIPA**: Linear Probe Assay*

***LP**: Laringeal Papillomatosis*

***LR-HPV**: Low Risk Human Papillomavirus*

M

***MAD**: Median Absolute Deviation*

***ML:** Maximum Likelihood*

***MHC:** Major Histocompatibility Complex*

***MSA**: Multiple Sequence Alignment*

N

***NA:** North America*

***NGS**: Next Generation Sequencing*

O

***ORF:** Open Reading Frame*

P

PCR: *Polymerase Chain Reaction*

PeIC: *Penile Invasive carcinoma*

PV: *Papillomavirus*

Q

QS: *Query sequences*

R

RA: *Reference Alignment*

RAxML: *Randomized Axelerated Maximum Likelihood*

RRP: *Recurrent Respiratory Papillomatosis*

RT: *Reference Tree*

S

SCC: *Squamous cell carcinoma*

SCJ: *Squamous columnar junction*

SNP: *Single Nucleotide Polymorphism*

SPF: *Short PCR Fragment*

T

TZ:*Transition zone*

U

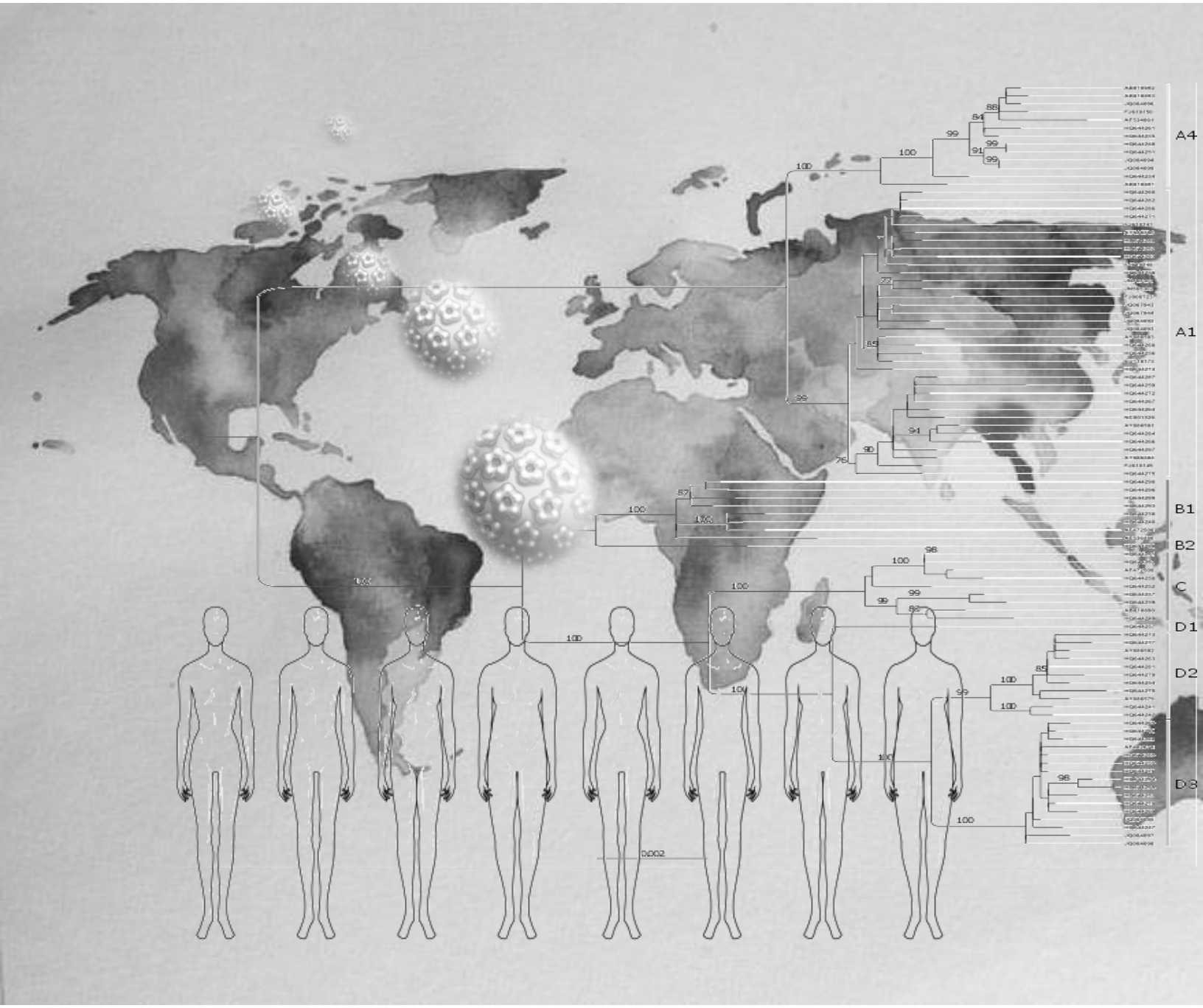
URR: *Upstream Regulatory Region*

V

VaIC: *Vaginal Invasive Carcinoma*

VuIC: *Vulvar Invasive Carcinoma*

9. BIBLIOGRAPHY



Agrawal, Yuri et al. 2008. "Oral Human Papillomavirus Infection before and after Treatment for Human Papillomavirus 16-Positive and Human Papillomavirus 16-Negative Head and Neck Squamous Cell Carcinoma." *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 14(21):7143–50.

Ahmed, Amina I., Sara L. Bissett, and Simon Beddows. 2013. "Amino Acid Sequence Diversity of the Major Human Papillomavirus Capsid Protein: Implications for Current and next Generation Vaccines." *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 18:151–59.

Alberti, Alberto et al. 2010. "Ovis Aries Papillomavirus 3: A Prototype of a Novel Genus in the Family Papillomaviridae Associated with Ovine Squamous Cell Carcinoma." *Virology* 407(2):352–59.

Alemany et al. 2014. "Large Contribution of Human Papillomavirus in Vaginal Neoplastic Lesions: A Worldwide Study in 597 Samples." *European Journal of Cancer* 50(16):2846–54.

Alemany, Laia et al. 2014. "Time Trends of Human Papillomavirus Types in Invasive Cervical Cancer, from 1940 to 2007." *International Journal of Cancer. Journal International Du Cancer* 135:88–95.

Alemany, Laia et al. 2015. "Human Papillomavirus DNA Prevalence and Type Distribution in Anal Carcinomas Worldwide." *International Journal of Cancer. Journal International Du Cancer* 136(1):98–107.

Alemany, Laia et al. 2016. "Role of Human Papillomavirus in Penile Carcinomas Worldwide." *European Urology*.

Alfaro, Ana et al. 2016. "Different Association of Human Papillomavirus 16 Variants with Early and Late Presentation of Cervical Cancer." *PloS One* 11(12):e0169315.

Amador-Molina, Alfredo, José Luis González-Montoya, Alejandro García-Carrancá, Alejandro Mohar, and Marcela Lizano. 2013. "Intratypic Changes of the E1 Gene and the Long Control Region Affect Ori Function of Human Papillomavirus Type 18 Variants." *Journal of General Virology* 94:393–402.

Andersson, S. et al. 2000. "Uneven Distribution of HPV 16 E6 Prototype and Variant (L83V) Oncoprotein in Cervical Neoplastic Lesions." *British Journal of Cancer* 83(3):307–10.

Anic, Gabriella M. and Anna R. Giuliano. 2011. "Genital HPV Infection and Related Lesions in Men." *Preventive Medicine* 53(SUPPL. 1).

Ansink, A. 1996. "Vulvar Squamous Cell Carcinoma." *Semin Dermatol* 15:51–59.

de Araujo Souza, P. S. et al. 2008. "Interaction between Polymorphisms of the Human Leukocyte Antigen and HPV-16 Variants on the Risk of Invasive Cervical Cancer ." *BMC Cancer* 8 :246.

Asadurian, Y. et al. 2007. "Activities of Human Papillomavirus 16 E6 Natural Variants in Human Keratinocytes." *J Med Virol* 79:1751–60.

Aubin, François et al. 2008. "Human Papillomavirus Genotype Distribution in External Acuminata Condylomata: A Large French National Study (EDiTH IV)." *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America* 47(5):610–15.

Baandrup, Louise et al. 2013. "Significant Decrease in the Incidence of Genital Warts in Young Danish Women after Implementation of a National Human Papillomavirus Vaccination Program." *Sexually Transmitted Diseases* 40(2):130–35.

Badaracco, Gianna et al. 2007. "Molecular Analyses and Prognostic Relevance of HPV in Head and Neck Tumours." *Oncology Reports* 17(4):931–39.

Baek, Min-Hyun et al. 2014. "Comparison of Adenocarcinoma and Adenosquamous Carcinoma in Patients with Early-Stage Cervical Cancer after Radical Surgery." *Gynecologic Oncology* 135(3):462–67.

Ball, Siolian L. R. et al. 2011. "Analyses of Human Papillomavirus Genotypes and Viral Loads in Anogenital Warts." *Journal of Medical Virology* 83(8):1345–50.
Banister, Carolyn E. et al. 2015. "Disparity in the Persistence of High-Risk Human Papillomavirus Genotypes between African American and European American Women of College Age." *The Journal of Infectious Diseases* 211(1):100–108.
Barbieri, Daniela et al. 2014. "Detection of HPV16 African Variants and Quantitative Analysis of Viral DNA Methylation in Oropharyngeal Squamous Cell Carcinomas." *Journal of Clinical Virology* 60(3):243–49.
Berger, Simon A. and Alexandros Stamatakis. 2011. "Aligning Short Reads to Reference Alignments and Trees." *Bioinformatics (Oxford, England)* 27(15):2068–75.
Bergström, R., P. Sparén, and H. O. Adami. 1999. "Trends in Cancer of the Cervix Uteri in Sweden Following Cytological Screening." *British Journal of Cancer* 81(1):159–66.
Bernard, H. U. 2013. "Taxonomy and Phylogeny of Papillomaviruses: An Overview and Recent Developments." *Infect Genet Evol* 18:357–61.
Bernard, Hans Ulrich. 1994. "Coevolution of Papillomaviruses with Human Populations." *Trends in Microbiology* 2(4):140–43.
Bernard, Hans Ulrich. 2002. "Gene Expression of Genital Human Papillomaviruses and Considerations on Potential Antiviral Approaches." *Antiviral Therapy* 7(4):219–37.
Bernard, Hans Ulrich et al. 2010. "Classification of Papillomaviruses (PVs) Based on 189 PV Types and Proposal of Taxonomic Amendments." *Virology* 401(1):70–79.
Bernard, Hans Ulrich, Itzel E. Calleja-Macias, and S. Terence Dunn. 2006. "Genome Variation of Human Papillomavirus Types: Phylogenetic and Medical Implications." *International Journal of Cancer* 118:1071–76.
Berumen, J. et al. 2001. "Asian-American Variants of Human Papillomavirus 16 and Risk for Cervical Cancer: A Case-Control Study." *Journal of the National Cancer Institute* 93(17):1325–30.
Bezerra, Artur L. R. et al. 2001. "Human Papillomavirus as a Prognostic Factor in Carcinoma of the Penis: Analysis of 82 Patients Treated with Amputation and Bilateral Lymphadenectomy." *Cancer* 91(12):2315–21.
Blakaj, D. et al. 2012. "Characterization of Human Papillomavirus 16 Variants in Headand-Neck Squamous Cell Carcinoma Patients." *International Journal of Radiation Oncology Biology Physics* 84(3):S711–12.
Blomberg, Maria, Soren Friis, Christian Munk, Andrea Bautz, and Susanne K. Kjaer. 2012. "Genital Warts and Risk of Cancer: A Danish Study of Nearly 50000 Patients with Genital Warts." *Journal of Infectious Diseases* 205(10):1544–53.
De Boer, Marjon A. et al. 2005. "Human Papillomavirus Type 18 Variants: Histopathology and E6/E7 Polymorphisms in Three Countries." *International Journal of Cancer. Journal International Du Cancer* 114(3):422–25.
Bogaert, Lies et al. 2012. "EcPV2 DNA in Equine Genital Squamous Cell Carcinomas and Normal Genital Mucosa." *Veterinary Microbiology* 158(1-2):33–41.
Bontkes, H. J. et al. 2000. "Human Papillomavirus Type 16 E6/E7-Specific Cytotoxic T Lymphocytes in Women with Cervical Neoplasia." *International Journal of Cancer. Journal International Du Cancer* 88(1):92–98.
Bosch, F. X. et al. 1992. "Risk Factors for Cervical Cancer in Colombia and Spain." *Int J Cancer* 52(5):750–58.
Bosch, F. X., A. Lorincz, N. Muñoz, C. J. L. M. Meijer, and K. V Shah. 2002. "The Causal Relation between Human Papillomavirus and Cervical Cancer." *Journal of Clinical Pathology* 55(4):244–65.
Boscolo-Rizzo, Paolo et al. 2009. "HPV-16 E6 L83V Variant in Squamous Cell Carcinomas of the Upper Aerodigestive Tract." *Journal of Cancer Research and Clinical Oncology*

135(4):559–66.
Boshart, M. et al. 1984. “A New Type of Papillomavirus DNA, Its Presence in Genital Cancer Biopsies and in Cell Lines Derived from Cervical Cancer.” *The EMBO Journal* 3(5):1151–57.
Boulet, Gaëlle a V, Caroline a J. Horvath, Sarah Berghmans, and Johannes Bogers. 2008. “Human Papillomavirus in Cervical Cancer Screening: Important Role as Biomarker.” *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 17(4):810–17.
Braaten, Kari P. and Marc R. Laufer. 2008. “Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine.” *Reviews in Obstetrics & Gynecology* 1(1):2–10.
Bravo, I. G., S. de Sanjosé, and M. Gottschling. 2010. “The Clinical Importance of Understanding the Evolution of Papillomaviruses.” *Trends in Microbiology* 18(10):432–38.
Brentjens, Mathijs H., Kimberly A. Yeung-Yue, Patricia C. Lee, and Stephen K. Tyring. 2002. “Human Papillomavirus: A Review.” *Dermatologic Clinics* 20(2):315–31.
Bruni, Laia et al. 2010. “Cervical Human Papillomavirus Prevalence in 5 Continents: Meta-Analysis of 1 Million Women with Normal Cytological Findings.” *The Journal of Infectious Diseases* 202(12):1789–99.
Buchinsky, F. J. et al. 2008. “Age of Child, More than HPV Type, Is Associated with Clinical Course in Recurrent Respiratory Papillomatosis.” *PLoS One* 3(5):e2263.
Bulkmans, N. W. J. et al. 2007. “High-Risk HPV Type-Specific Clearance Rates in Cervical Screening.” *British Journal of Cancer* 96(9):1419–24.
Burk, R. D. et al. 2011. “Classification and Nomenclature System for Human Alphapapillomavirus Variants: General Features, Nucleotide Landmarks and Assignment of HPV6 and HPV11 Isolates to Variant Lineages.” *Acta Dermatovenerologica Alpina, Pannonica et Adriatica* 20(3):113–23.
Burk, Robert D. et al. 2003. “Distribution of Human Papillomavirus Types 16 and 18 Variants in Squamous Cell Carcinomas and Adenocarcinomas of the Cervix.” *Cancer Research* 63(21):7215–20.
Burk, Robert D., Ariana Harari, and Zigui Chen. 2013. “Human Papillomavirus Genome Variants.” *Virology* 445(1-2):232–43.
Bzhalava, Davit, Peng Guan, Silvia Franceschi, Joakim Dillner, and Gary Clifford. 2013. “A Systematic Review of the Prevalence of Mucosal and Cutaneous Human Papillomavirus Types.” *Virology* 445(1-2):224–31.
Castellsagué, Xavier et al. 2006. “Worldwide Human Papillomavirus Etiology of Cervical Adenocarcinoma and Its Cofactors: Implications for Screening and Prevention.” *Journal of the National Cancer Institute* 98(5):303–15.
Castellsagué, Xavier et al. 2016. “HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients.” *Journal of the National Cancer Institute* 108(6):djv403.
Castresana, Jose. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution* 17(4):540–52.
Chan, Paul K. S., María Alejandra Picconi, Tak Hong Cheung, Lucia Giovannelli, and Jong Sup Park. 2012. “Laboratory and Clinical Aspects of Human Papillomavirus Testing.” *Critical Reviews in Clinical Laboratory Sciences* 49(4):117–36.
Chan, W. K., G. Klock, and H. U. Bernard. 1989. “Progesterone and Glucocorticoid Response Elements Occur in the Long Control Regions of Several Human Papillomaviruses Involved in Anogenital Neoplasia.” *Journal of Virology* 63(8):3261–69.
Chansaenroj, J., A. Theamboonlers, P. Junyangdikul, P. Supiyaphan, and Y. Poovorawan. 2012. “Whole Genome Analysis of Human Papillomavirus Genotype 11 from Cervix, Larynx and

Lung." *Asian Pac J Cancer Prev* 13(6):2619–23.
Chen, Alyce a et al. 2014. "Human Papillomavirus 33 Worldwide Genetic Variation and Associated Risk of Cervical Cancer." *Virology* 448:356–62.
Chen, X. S., R. L. Garcea, I. Goldberg, G. Casini, and S. C. Harrison. 2000. "Structure of Small Virus-like Particles Assembled from the L1 Protein of Human Papillomavirus 16." *Molecular Cell* 5(3):557–67.
Chen, Z. et al. 2011. "Evolution and Taxonomic Classification of Human Papillomavirus 16 (HPV16)-Related Variant Genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67." *PLoS One* 6(5):e20183.
Chen, Z., R. DeSalle, M. Schiffman, R. Herrero, and R. D. Burk. 2009. "Evolutionary Dynamics of Variant Genomes of Human Papillomavirus Types 18, 45, and 97." *J Virol* 83(3):1443–55.
Chen, Zigui, Luciana Bueno de Freitas, and Robert D. Burk. 2015. "Evolution and Classification of Oncogenic Human Papillomavirus Types and Variants Associated with Cervical Cancer." Pp. 3–26 in *Cervical Cancer: Methods and Protocols*, vol. 1249.
Chopjitt, Peechanika et al. 2009. "Prevalence of Human Papillomavirus Type 16 and Its Variants in Abnormal Squamous Cervical Cells in Northeast Thailand." *International Journal of Infectious Diseases : IJID : Official Publication of the International Society for Infectious Diseases* 13(2):212–19.
Chow, Louise T., Thomas R. Broker, and Bettie M. Steinberg. 2010. "The Natural History of Human Papillomavirus Infections of the Mucosal Epithelia." *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 118(6-7):422–49.
Christensen, Neil D. 2016. "HPV Disease Transmission Protection and Control." *Microbial Cell* 3(9):475–89.
Chung, Sang-Hyuk, Silvia Franceschi, and Paul F. Lambert. 2010. "Estrogen and ERalpha: Culprits in Cervical Cancer?" *Trends in Endocrinology and Metabolism: TEM* 21(8):504–11.
Combes, Jean Damien and Silvia Franceschi. 2018. "Human Papillomavirus Genome Variants and Head and Neck Cancers: A Perspective." *Infectious Agents and Cancer* 13(1).
Combrinck, C. E., R. Y. Seedat, C. Randall, Y. Roodt, and F. J. Burt. 2012. "Novel HPV-6 Variants of Human Papillomavirus Causing Recurrent Respiratory Papillomatosis in Southern Africa." *Epidemiology and Infection* 140(6):1095–1101.
Cornet, I., T. Gheit, and MR Iannacone. 2013. "HPV16 Genetic Variation and the Development of Cervical Cancer Worldwide." *British Journal of Cancer* 108(1):240–44.
Cornet, Iris, Tarik Gheit, and Silvia Franceschi. 2012. "Human Papillomavirus Type 16 Genetic Variants: Phylogeny and Classification Based on E6 and LCR." *Journal of Virology* 86(12):6855–61.
Crawford, L. V. 1965. "A Study of Human Papilloma Virus DNA." *Journal of Molecular Biology* 13(2):362–72.
Cromme, F. V, C. J. Meijer, et al. 1993. "Analysis of MHC Class I and II Expression in Relation to Presence of HPV Genotypes in Premalignant and Malignant Cervical Lesions." *Br J Cancer* 67(6):1372–80.
Cromme, F. V, P. J. Snijders, et al. 1993. "MHC Class I Expression in HPV 16 Positive Cervical Carcinomas Is Post-Transcriptionally Controlled and Independent from c-Myc Overexpression." *Oncogene* 8(11):2969–75.
Crum, C. P. 2000. "Contemporary Theories of Cervical Carcinogenesis: The Virus, the Host, and the Stem Cell." *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 13:243–51.
Crum, C. P., H. Ikenberg, R. M. Richart, and L. Gissman. 1984. "Human Papillomavirus Type 16 and Early Cervical Neoplasia." *The New England Journal of Medicine* 310(14):880–83.
Cullen, Michael et al. 2015. "Deep Sequencing of HPV16 Genomes: A New High-Throughput

Tool for Exploring the Carcinogenicity and Natural History of HPV16 Infection." *Papillomavirus Research* 1:3–11.
Dahlström, Lisen Arnheim et al. 2010. "Prospective Study of Human Papillomavirus and Risk of Cervical Adenocarcinoma." *International Journal of Cancer. Journal International Du Cancer* 127(8):1923–30.
Danielewski, Jennifer A., Suzanne M. Garland, Jenny McCloskey, Richard J. Hillman, and Sepehr N. Tabrizi. 2013. "Human Papillomavirus Type 6 and 11 Genetic Variants Found in 71 Oral and Anogenital Epithelial Samples from Australia." *PLoS ONE* 8(5).
Davy, M. L., T. J. Dodd, C. G. Luke, and D. M. Roder. 2003. "Cervical Cancer: Effect of Glandular Cell Type on Prognosis, Treatment, and Survival." *Obstetrics and Gynecology* 101(1):38–45.
Delius, H., M. A. Van Ranst, A. B. Jenson, H. zur Hausen, and J. P. Sundberg. 1994. "Canine Oral Papillomavirus Genomic Sequence: A Unique 1.5-Kb Intervening Sequence between the E2 and L2 Open Reading Frames." *Virology* 204:447–52.
Dietrich, Dimo et al. 2013. "Improved PCR Performance Using Template DNA from Formalin-Fixed and Paraffin-Embedded Tissues by Overcoming PCR Inhibition." *PLoS ONE* 8(10).
Dillon, S. et al. 2007. "Resolution of Cervical Dysplasia Is Associated with T-Cell Proliferative Responses to Human Papillomavirus Type 16 E2." *J Gen Virol* 88(Pt 3):803–13.
DiMaio, Daniel. 2015. "Nuns, Warts, Viruses, and Cancer." *The Yale Journal of Biology and Medicine* 88(2):127–29.
Donne, A. J., L. Hampson, J. J. Homer, and I. N. Hampson. 2010. "The Role of HPV Type in Recurrent Respiratory Papillomatosis." *International Journal of Pediatric Otorhinolaryngology* 74(1):7–14.
Doorbar, J. et al. 2012. "The Biology and Life-Cycle of Human Papillomaviruses." *Vaccine* 30 Suppl 5:F55–70.
Doorbar, J., D. Campbell, R. J. Grand, and P. H. Gallimore. 1986. "Identification of the Human Papilloma Virus-1a E4 Gene Products." *The EMBO Journal* 5(2):355–62.
Doorbar, John. 2005. "The Papillomavirus Life Cycle." *Journal of Clinical Virology* 32(SUPPL.).
Doorbar, John. 2006. "Molecular Biology of Human Papillomavirus Infection and Cervical Cancer." *Clinical Science (London, England : 1979)* 110(5):525–41.
Doorbar, John. 2013. "Latent Papillomavirus Infections and Their Regulation." *Current Opinion in Virology* 3(4):416–21.
Doorbar, John. 2016. "Model Systems of Human Papillomavirus-Associated Disease." *Journal of Pathology* 238(2):166–79.
Doorbar, John. 2017. "Host Control of Human Papillomavirus Infection and Disease." *Best Practice & Research. Clinical Obstetrics & Gynaecology*.
Doorbar, John, Nagayasu Egawa, Heather Griffin, Christian Kranjec, and Isao Murakami. 2015. "Human Papillomavirus Molecular Biology and Disease Association." *Reviews in Medical Virology* 25(S1):2–23.
Du, Juan et al. 2012. "Human Papillomavirus (HPV) 16 E6 Variants in Tonsillar Cancer in Comparison to Those in Cervical Cancer in Stockholm, Sweden." *PLoS ONE* 7(4).
Dürst, M., L. Gissmann, H. Ikenberg, and H. zur Hausen. 1983. "A Papillomavirus DNA from a Cervical Carcinoma and Its Prevalence in Cancer Biopsy Samples from Different Geographic Regions." *Proceedings of the National Academy of Sciences of the United States of America* 80(12):3812–15.
Egawa, Nagayasu and John Doorbar. 2017. "The Low-Risk Papillomaviruses." *Virus Research* 231:119–27.
Egawa, Nagayasu, Kiyofumi Egawa, Heather Griffin, and John Doorbar. 2015. "Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia." *Viruses* 7(7):3863–90.

Ferenczy, Alex and Eduardo Franco. 2002. "Persistent Human Papillomavirus Infection and Cervical Neoplasia." *The Lancet Oncology* 3(1):11–16.
Ferlay, J. et al. 2013. "GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase. No. 11 [Internet]." *Lyon, France: International Agency for Research on Cancer.* 11.
Fleury, Maxime J. J., Antoine Touzé, and Pierre Coursaget. 2014. "Human Papillomavirus Type 16 Pseudovirions with Few Point Mutations in L1 Major Capsid Protein FG Loop Could Escape Actual or Future Vaccination for Potential Use in Gene Therapy." *Molecular Biotechnology* 56(5):479–86.
Flores-Díaz, Ema et al. 2017a. "HPV-11 Variability, Persistence and Progression to Genital Warts in Men: The HIM Study." *Journal of General Virology* 98(9):2339–42.
Flores-Díaz, Ema et al. 2017b. "HPV-6 Molecular Variants Association with the Development of Genital Warts in Men: The HIM Study." *Journal of Infectious Diseases* 215(4):559–65.
Forman, David et al. 2012. "Global Burden of Human Papillomavirus and Related Diseases." *Vaccine* 30 Suppl 5:F12–23.
Freitas, Luciana Bueno et al. 2014. "Human Papillomavirus 16 Non-European Variants Are Preferentially Associated with High-Grade Cervical Lesions." *PloS One* 9(7):e100746.
Fujiwara, H., G. Tortolero-Luna, M. F. Mitchell, J. P. Koulos, and T. C. Wright. 1997. "Adenocarcinoma of the Cervix. Expression and Clinical Significance of Estrogen and Progesterone Receptors." *Cancer* 79(3):505–12.
Gabbott, M. et al. 1997. "Human Papillomavirus and Host Variables as Predictors of Clinical Course in Patients with Juvenile-Onset Recurrent Respiratory Papillomatosis." *Journal of Clinical Microbiology* 35(12):3098–3103.
Galani, E. and C. Christodoulou. 2009. "Human Papilloma Viruses and Cancer in the Post-Vaccine Era." *Clinical Microbiology and Infection* 15(11):977–81.
Gáll, Tamás, Andrea Kis, Eniko Fehér, Lajos Gergely, and Krisztina Szarka. 2011. "Virological Failure of Intralesional Cidofovir Therapy in Recurrent Respiratory Papillomatosis Is Not Associated with Genetic or Epigenetic Changes of HPV11: Complete Genome Comparison of Sequential Isolates." *Antiviral Research* 92(2):356–58.
García-Pérez, Raquel et al. 2014. "Novel Papillomaviruses in Free-Ranging Iberian Bats: No Virus-Host Co-Evolution, No Strict Host Specificity, and Hints for Recombination." *Genome Biology and Evolution* 6(1):94–104.
Garland, Suzanne M. et al. 2009. "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." *The Journal of Infectious Diseases* 199(6):805–14.
Gewin, Lindy, Hadley Myers, Tohru Kiyono, and Denise A. Galloway. 2004. "Identification of a Novel Telomerase Repressor That Interacts with the Human Papillomavirus Type-16 E6/E6-AP Complex." *Genes and Development* 18(18):2269–82.
Giri, I., O. Danos, and M. Yaniv. 1985. "Genomic Structure of the Cottontail Rabbit (Shope) Papillomavirus." *Proceedings of the National Academy of Sciences of the United States of America* 82(6):1580–84.
Gissmann, L. et al. 1983. "Human Papillomavirus Types 6 and 11 DNA Sequences in Genital and Laryngeal Papillomas and in Some Cervical Cancers." *Proceedings of the National Academy of Sciences of the United States of America* 80(2):560–63.
Gissmann, L., V. Diehl, H. J. Schultz-Coulon, and H. zur Hausen. 1982. "Molecular Cloning and Characterization of Human Papilloma Virus DNA Derived from a Laryngeal Papilloma." *Journal of Virology* 44(1):393–400.
Gissmann, L. and H. zur Hausen. 1980. "Partial Characterization of Viral DNA from Human Genital Warts (Condylomata Acuminata)." *International Journal of Cancer* 25(5):605–9.
Giuliano, Anna R. et al. 2008. "Epidemiology of Human Papillomavirus Infection in Men,

Cancers Other than Cervical and Benign Conditions." *Vaccine* 26 Suppl 1:K17–28.
Godínez, J. M. et al. 2013. "Differential Presence of Papillomavirus Variants in Cervical Cancer: An Analysis for HPV33, HPV45 and HPV58." *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 13:96–104.
Gottschling, Marc et al. 2011. "Quantifying the Phylodynamic Forces Driving Papillomavirus Evolution." *Molecular Biology and Evolution* 28(7):2101–13.
Grodzki, Martha et al. 2006. "Increased Risk for Cervical Disease Progression of French Women Infected with the Human Papillomavirus Type 16 E6-350G Variant." *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 15(4):820–22.
Guan, Peng et al. 2012. "Human Papillomavirus Types in 115,789 HPV-Positive Women: A Meta-Analysis from Cervical Infection to Cancer." *International Journal of Cancer* 131(10):2349–59.
Guess, J. C. and D. J. McCance. 2005. "Decreased Migration of Langerhans Precursor-Like Cells in Response to Human Keratinocytes Expressing Human Papillomavirus Type 16 E6/E7 Is Related to Reduced Macrophage Inflammatory Protein-3 Production." *Journal of Virology* 79(23):14852–62.
Gurgel, Ana Pavla Almeida Diniz et al. 2015. "Prevalence of Human Papillomavirus Variants and Genetic Diversity in the L1 Gene and Long Control Region of HPV16, HPV31, and HPV58 Found in North-East Brazil." *BioMed Research International* 2015:130828.
Handsfield, H. H. 1997. "Clinical Presentation and Natural Course of Anogenital Warts." *American Journal of Medicine* 102(5A):16–20.
Hanna, John, Julie D. R. Reimann, Robert I. Haddad, and Jeffrey F. Krane. 2013. "Human Papillomavirus-Associated Adenocarcinoma of the Base of the Tongue." *Human Pathology* 44(8):1516–23.
Harari, Ariana, Zigui Chen, and Robert D. Burk. 2014. "Human Papillomavirus Genomics: Past, Present and Future." *Current Problems in Dermatology* 45:1–18.
Harrison, Christopher et al. 2014. "Decreased Management of Genital Warts in Young Women in Australian General Practice Post Introduction of National HPV Vaccination Program: Results from a Nationally Representative Cross-Sectional General Practice Study." *PLoS ONE* 9(9).
Hassani, Soroush et al. 2015. "Molecular Pathogenesis of Human Papillomavirus Type 16 in Tonsillar Squamous Cell Carcinoma." *Anticancer Research* 35(12):6633–38.
zur Hausen, H. 1977. "Human Papillomaviruses and Their Possible Role in Squamous Cell Carcinomas." *Current Topics in Microbiology and Immunology* 78:1–30.
zur Hausen, H., L. Gissmann, W. Steiner, W. Dippold, and I. Dreger. 1975. "Human Papilloma Viruses and Cancer." *Bibliotheca Haematologica* (43):569–71.
zur Hausen, Harald. 2009. "Papillomaviruses in the Causation of Human Cancers - a Brief Historical Account." *Virology* 384(2):260–65.
Heinzel, P. A. et al. 1995. "Variation of Human Papillomavirus Type 6 (HPV-6) and HPV-11 Genomes Sampled throughout the World." *J Clin Microbiol* 33(7):1746–54.
Herbst, Lawrence H. et al. 2009. "Genomic Characterization of Two Novel Reptilian Papillomaviruses, Chelonia Mydas Papillomavirus 1 and Caretta Caretta Papillomavirus 1." *Virology* 383(1):131–35.
Herfs, Michael et al. 2012. "A Discrete Population of Squamocolumnar Junction Cells Implicated in the Pathogenesis of Cervical Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 109(26):10516–21.
Hernandez, Jonathan et al. 2011. "HPV L1 Capsid Protein Detection and Progression of Anal Squamous Neoplasia." *American Journal of Clinical Pathology* 135(3):436–41.

Hibma, Merilyn H. 2012. "The Immune Response to Papillomavirus During Infection Persistence and Regression." *The Open Virology Journal* 6(1):241–48.
Hildesheim, Allan and Sophia S. Wang. 2002. "Host and Viral Genetics and Risk of Cervical Cancer: A Review." Pp. 229–40 in *Virus Research*, vol. 89.
Ho, L. et al. 1993. "The Genetic Drift of Human Papillomavirus Type 16 Is a Means of Reconstructing Prehistoric Viral Spread and the Movement of Ancient Human Populations." *Journal of Virology* 67(11):6413–23.
Hobbs, Christopher G. L. and Martin a Birchall. 2004. "Human Papillomavirus Infection in the Etiology of Laryngeal Carcinoma." *Current Opinion in Otolaryngology & Head and Neck Surgery* 12(2):88–92.
Homburger, Julian R. et al. 2015. "Genomic Insights into the Ancestry and Demographic History of South America." *PLoS Genetics* 11(12).
Hu, X. et al. 2001. "HPV16 E6 Gene Variations in Invasive Cervical Squamous Cell Carcinoma and Cancer in Situ from Russian Patients." *British Journal of Cancer* 84:791–95.
IARC. 2012. "IARC Monographs on the Evaluation of Carcinogenic Risks to Humans - Biological Agents." *Volume 100B*.
Ikenberg, H., L. Gissmann, G. Gross, E. I. Grussendorf-Conen, and H. zur Hausen. 1983. "Human Papillomavirus Type-16-Related DNA in Genital Bowen's Disease and in Bowenoid Papulosis." *International Journal of Cancer. Journal International Du Cancer* 32(5):563–65.
Jablonska, S. et al. 1978. "Immunological Studies in Epidermodysplasia Verruciformis." *Bulletin Du Cancer* 65(2):183–90.
Jablonska, S., J. Dabrowski, and K. Jakubowicz. 1972. "Epidermodysplasia Verruciformis as a Model in Studies on the Role of Papovaviruses in Oncogenesis." *Cancer Research* 32(3):583–89.
Jackson, Robert et al. 2016. "Functional Variants of Human Papillomavirus Type 16 Demonstrate Host Genome Integration and Transcriptional Alterations Corresponding to Their Unique Cancer Epidemiology." *BMC Genomics* 17(1):851.
Jelen, M. M. et al. 2014. "Global Genomic Diversity of Human Papillomavirus 6 Based on 724 Isolates and 190 Complete Genome Sequences." *Journal of Virology* 88(13):7307–16.
Kabekkodu, Shama Prasada et al. 2015. "Prevalence of Human Papillomavirus Types and Phylogenetic Analysis of HPV-16 L1 Variants from Southern India." *Asian Pacific Journal of Cancer Prevention* 16(5):2073–80.
Kammer, C., U. Warthorst, N. Torrez-Martinez, C. M. Wheeler, and H. Pfister. 2000. "Sequence Analysis of the Long Control Region of Human Papillomavirus Type 16 Variants and Functional Consequences for P97 Promoter Activity." *J Gen Virol* 81(Pt 8):1975–81.
Kanodia, Shreya, Laura M. Fahey, and W. Martin Kast. 2007. "Mechanisms Used by Human Papillomaviruses to Escape the Host Immune Response." *Current Cancer Drug Targets* 7(1):79–89.
Khouadri, S. et al. 2006. "Human Papillomavirus Type 33 Polymorphisms and High-Grade Squamous Intraepithelial Lesions of the Uterine Cervix." *J Infect Dis* 194(7):886–94.
Kliewer, Erich V et al. 2009. "Twenty-Year Trends in the Incidence and Prevalence of Diagnosed Anogenital Warts in Canada." *Sexually Transmitted Diseases* 36(6):380–86.
KLUG, A. and J. T. FINCH. 1965. "STRUCTURE OF VIRUSES OF THE PAPILLOMA-POLYOMA TYPE. I. HUMAN WART VIRUS." *Journal of Molecular Biology* 11:403–23.
Von Knebel Doeberitz, Magnus. 2016. "HPV 16 and Its Variants: Minor Genome Variations Make a Big Difference." *Journal of the National Cancer Institute*.
Kocjan, B. J. et al. 2009. "Prevaccination Genomic Diversity of Human Papillomavirus Genotype 6 (HPV 6)." *Virology* 391(2):274–83.
Kocjan, Boštjan J., Mateja M. Jelen, Polona J. Maver, Katja Seme, and Mario Poljak. 2011. "Pre-

Vaccination Genomic Diversity of Human Papillomavirus Genotype 6 (HPV 6): A Comparative Analysis of 21 Full-Length Genome Sequences." *Infection, Genetics and Evolution* 11(7):1805–10.
Kocjan, Boštjan J., Katja Seme, and Mario Poljak. 2008. "Detection and Differentiation of Human Papillomavirus Genotypes HPV-6 and HPV-11 by FRET-Based Real-Time PCR." *Journal of Virological Methods* 153(2):245–49.
Kokkat, Theresa J., Miral S. Patel, Diane McGarvey, Virginia A. LiVolsi, and Zubair W. Baloch. 2013. "Archived Formalin-Fixed Paraffin-Embedded (FFPE) Blocks: A Valuable Underexploited Resource for Extraction of DNA, RNA, and Protein." *Biopreservation and Biobanking* 11(2):101–6.
Komloš, Kristina Fujs et al. 2012. "Tumor-Specific and Gender-Specific Pre-Vaccination Distribution of Human Papillomavirus Types 6 and 11 in Anogenital Warts and Laryngeal Papillomas: A Study on 574 Tissue Specimens." *Journal of Medical Virology* 84(8):1233–41.
Komloš, Kristina Fujs, Pavle Košorok, Boštjan J. Kocjan, and Mario Poljak. 2013. "Genetic Diversity of HPV-6 in Concurrent Multiple Anogenital Warts." *Acta Dermatovenerologica Alpina, Pannonica et Adriatica* 22(1):31–33.
de Koning, Maurits N. C., Wim G. V Quint, and Edyta C. Pirog. 2008. "Prevalence of Mucosal and Cutaneous Human Papillomaviruses in Different Histologic Subtypes of Vulvar Carcinoma." *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 21(3):334–44.
Kosary, C. L. 1994. "FIGO Stage, Histology, Histologic Grade, Age and Race as Prognostic Factors in Determining Survival for Cancers of the Female Gynecological System: An Analysis of 1973-87 SEER Cases of Cancers of the Endometrium, Cervix, Ovary, Vulva, and Vagina." *Seminars in Surgical Oncology* 10(1 994):31–46.
Kukimoto, Iwao et al. 2013. "Genetic Variation of Human Papillomavirus Type 16 in Individual Clinical Specimens Revealed by Deep Sequencing." *PLoS ONE* 8(11).
Kukimoto, Iwao and Masamichi Muramatsu. 2015. "Genetic Variations of Human Papillomavirus Type 16: Implications for Cervical Carcinogenesis." *Japanese Journal of Infectious Diseases* 68(3):169–75.
Lace, Michael J. et al. 2009. "Upstream Regulatory Region Alterations Found in Human Papillomavirus Type 16 (HPV-16) Isolates from Cervical Carcinomas Increase Transcription, Ori Function, and HPV Immortalization Capacity in Culture." *Journal of Virology* 83(15):7457–66.
Lai, C. H. et al. 1999. "Are Adenocarcinomas and Adenosquamous Carcinomas Different from Squamous Carcinomas in Stage IB and II Cervical Cancer Patients Undergoing Primary Radical Surgery?" *International Journal of Gynecological Cancer* 9(1):28–36.
Lange, Christian E. et al. 2009. "Three Novel Canine Papillomaviruses Support Taxonomic Clade Formation." *Journal of General Virology* 90(11):2615–21.
Lange, Christian E. et al. 2011. "Novel Snake Papillomavirus Does Not Cluster with Other Non-Mammalian Papillomaviruses." *Virology Journal* 8(1):436.
Larson, Daniel a and Craig S. Derkay. 2010. "Epidemiology of Recurrent Respiratory Papillomatosis." *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 118(6-7):450–54.
Larsson, Gabriella Lillsunde et al. 2012. "Human Papillomavirus (HPV) and HPV 16-Variant Distribution in Vulvar Squamous Cell Carcinoma in Sweden." *International Journal of Gynecological Cancer* 22(8):1413–19.
Larsson, Gabriella Lillsunde, Gisela Helenius, Sören Andersson, Bengt Sorbe, and Mats G. Karlsson. 2013. "Prognostic Impact of Human Papilloma Virus (HPV) Genotyping and HPV-16 Subtyping in Vaginal Carcinoma." *Gynecologic Oncology* 129(2):406–11.

Lavezzo, Enrico et al. 2016. "Characterization of Intra-Type Variants of Oncogenic Human Papillomaviruses by next-Generation Deep Sequencing of the E6/E7 Region." *Viruses* 8(3).
Lee, K. et al. 2008. "Human Papillomavirus 16 E6, L1, L2 and E2 Gene Variants in Cervical Lesion Progression." *Virus Research* 131(1):106–10.
Lehoux, Michaël, Claudia M. D'Abramo, and Jacques Archambault. 2009. "Molecular Mechanisms of Human Papillomavirus-Induced Carcinogenesis." *Public Health Genomics* 12(5-6):268–80.
Li, Ni, Silvia Franceschi, Rebecca Howell-Jones, Peter J. F. Snijders, and Gary M. Clifford. 2011. "Human Papillomavirus Type Distribution in 30,848 Invasive Cervical Cancers Worldwide: Variation by Geographical Region, Histological Type and Year of Publication." *International Journal of Cancer* 128:927–35.
Liu, Xia et al. 2015. "Association of Human Papillomavirus Type 31 DNA Load with Risk of Cervical Intraepithelial Neoplasia Grades 2-3." *Journal of Clinical Microbiology*.
Liu, Ying, Yaqi Pan, Weijiao Gao, Yang Ke, and Zheming Lu. 2017. "Whole-Genome Analysis of Human Papillomavirus Types 16, 18, and 58 Isolated from Cervical Precancer and Cancer Samples in Chinese Women." *Scientific Reports* 7(1):263.
Lizano, Marcela. 2006. "Distribution of HPV16 and 18 Intratypic Variants in Normal Cytology, Intraepithelial Lesions, and Cervical Cancer in a Mexican Population." *Gynecologic Oncology* 102(2):230–35.
Longworth, Michelle S. and Laimonis A. Laimins. 2004. "Pathogenesis of Human Papillomaviruses in Differentiating Epithelia." *Microbiology and Molecular Biology Reviews : MMBR* 68(2):362–72.
Lopera, Esteban A. et al. 2014. "Unexpected Inverse Correlation between Native American Ancestry and Asian American Variants of HPV16 in Admixed Colombian Cervical Cancer Cases." *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 28:339–48.
López-Romero, Ricardo et al. 2013. "HPV Frequency in Penile Carcinoma of Mexican Patients: Important Contribution of HPV16 European Variant." *International Journal of Clinical and Experimental Pathology* 6(7):1409–15.
Lübbe, J. et al. 1996. "HPV-11- and HPV-16-Associated Oral Verrucous Carcinoma." *Dermatology* 192(3):217–21.
Ma, Barbara, Richard B. S. Roden, Chien Fu Hung, and T. C. Wu. 2011. "HPV Pseudovirions as DNA Delivery Vehicles." *Therapeutic Delivery*, 427–30.
Madeleine, Margaret M. et al. 2008. "Comprehensive Analysis of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 Loci and Squamous Cell Cervical Cancer Risk." *Cancer Research* 68(9):3532–39.
Madsen, B. S., H. L. Jensen, A. J. van den Brule, J. Wohlfahrt, and M. Frisch. 2008. "Risk Factors for Invasive Squamous Cell Carcinoma of the Vulva and Vagina--Population-Based Case-Control Study in Denmark." *Int J Cancer* 122(12):2827–34.
Maglennon, Gareth Adam and John Doorbar. 2012. "The Biology of Papillomavirus Latency." *The Open Virology Journal* 6:190–97.
Marangon, Amanda Vansan et al. 2013. "The Association of the Immune Response Genes to Human Papillomavirus-Related Cervical Disease in a Brazilian Population." *BioMed Research International* 2013:1–11.
de Martel, Catherine et al. 2012. "Global Burden of Cancers Attributable to Infections in 2008: A Review and Synthetic Analysis." *The Lancet Oncology* 13(6):607–15.
Martens, Jolise E., Janwillem Arends, Paul J. Q. Van Der Linden, Bert A. G. De Boer, and Theo J. M. Helmerhorst. 2004. "Cytokeratin 17 and p63 Are Markers of the HPV Target Cell, the Cervical Stem Cell." *Anticancer Research* 24(2 B):771–75.

Matos, R. P. et al. 2013. "Nucleotide and Phylogenetic Analysis of Human Papillomavirus Types 6 and 11 Isolated from Recurrent Respiratory Papillomatosis in Brazil." *Infect Genet Evol*.
Matthews, Kate et al. 2003. "Depletion of Langerhans Cells in Human Papillomavirus Type 16-Infected Skin Is Associated with E6-Mediated down Regulation of E-Cadherin." *Journal of Virology* 77(15):8378–85.
Maver, Polona J. et al. 2011. "Prevaccination Genomic Diversity of Human Papillomavirus Genotype 11: A Study on 63 Clinical Isolates and 10 Full-Length Genome Sequences." *Journal of Medical Virology* 83(3):461–70.
Maxwell Parkin, D., Freddie Bray, Jacques Ferlay, and Paola Pisani. 2001. "Estimating the World Cancer Burden: Globocan 2000." *International Journal of Cancer* 94(2):153–56.
Mccullagh, P. and Ja Nelder. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society.* 526.
McMurray, H. and D. McCance. 2004. "Degradation of p53, Not Telomerase Activation, by E6 Is Required for Bypass of Crisis and Immortalization by Human Papillomavirus Type 16 E6/E7." *Journal of Virology* 78(11):5698–5706.
Mcmurray, H. R., D. Nguyen, T. F. Westbrook, and D. J. Mcance. 2001. "Biology of Human Papillomaviruses." *International Journal of Experimental Pathology* 82(1):15–33.
Melkert, P. W. et al. 1992. "Multiple HPV 16-Related Squamous Cell Carcinomas of the Vulva, Vagina, Anus, Skin and Cervix in a 31-Year-Old Woman." *Eur J Obstet Gynecol Reprod Biol* 46(1):53–56.
Mengual-Chuliá, Beatriz, Stéphanie Bedhomme, Guillaume Lafforgue, Santiago F. Elena, and Ignacio G. Bravo. 2016. "Assessing Parallel Gene Histories in Viral Genomes." *BMC Evolutionary Biology* 16(1).
Mengual-Chuliá, Beatriz, Raquel García-Pérez, Marc Gottschling, Ingo Nindl, and Ignacio G. Bravo. 2012. "Novel Animal Papillomavirus Sequences and Accurate Phylogenetic Placement." *Molecular Phylogenetics and Evolution* 65(3):883–91.
Middleton, Kate et al. 2003. "Organization of Human Papillomavirus Productive Cycle during Neoplastic Progression Provides a Basis for Selection of Diagnostic Markers." *Journal of Virology* 77(19):10186–201.
Mirabello, Lisa et al. 2016. "HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women." *Journal of the National Cancer Institute* 108(9).
Mirabello, Lisa et al. 2017. "HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis." *Cell* 170(6):1164–74.e6.
Mirabello, Lisa et al. 2018. "The Intersection of HPV Epidemiology, Genomics and Mechanistic Studies of HPV-Mediated Carcinogenesis." *Viruses* 10(2).
Mittal, R., A. Pater, and M. M. Pater. 1993. "Multiple Human Papillomavirus Type 16 Glucocorticoid Response Elements Functional for Transformation, Transient Expression, and DNA-Protein Interactions." *Journal of Virology* 67:5656–59.
Moody, Cary A. and Laimonis A. Laimins. 2010. "Human Papillomavirus Oncoproteins: Pathways to Transformation." *Nature Reviews. Cancer* 10:550–60.
Moreno-Acosta, P. et al. 2017. "HPV-16 Variants' Impact on Uterine Cervical Cancer Response to Radiotherapy: A Descriptive Pilot Study." *Cancer/Radiothérapie* 21(2):104–8.
Morshed, Kamal, Małgorzata Polz-Dacewicz, Marcin Szymański, and Agata Smoleń. 2010. "Usefulness and Efficiency of Formalin-Fixed Paraffin-Embedded Specimens from Laryngeal Squamous Cell Carcinoma in HPV Detection by IHC and PCR/DEIA." *Folia Histochemica et Cytobiologica* 48(3):398–402.
Moscicki, Anna Barbara et al. 2012. "Updating the Natural History of Human Papillomavirus and Anogenital Cancers." *Vaccine* 30(SUPPL.5).
Mounts, P. and H. Kashima. 1984. "Association of Human Papillomavirus Subtype and Clinical

Course in Respiratory Papillomatosis." *Laryngoscope* 94:28–33.
Mounts, P., K. V Shah, and H. Kashima. 1982. "Viral Etiology of Juvenile- and Adult-Onset Squamous Papilloma of the Larynx." *Proc Natl Acad Sci U S A* 79(17):5425–29.
Munday, John S., Magda Dunowska, Simon F. Hills, and Rebecca E. Laurie. 2013. "Genomic Characterization of Felis Catus Papillomavirus-3: A Novel Papillomavirus Detected in a Feline Bowenoid in Situ Carcinoma." *Veterinary Microbiology* 165(3-4):319–25.
Munk, C., E. I. Svare, P. Poll, J. E. Bock, and S. K. Kjaer. 1997. "History of Genital Warts in 10,838 Women 20 to 29 Years of Age from the General Population. Risk Factors and Association with Papanicolaou Smear History." *Sexually Transmitted Diseases* 24(10):567–72.
Muñoz, N. et al. 1992. "The Causal Link between Human Papillomavirus and Invasive Cervical Cancer: A Population-Based Case-Control Study in Colombia and Spain." *International Journal of Cancer. Journal International Du Cancer* 52(5):743–49.
Nafz, Julia et al. 2008. "A Novel Rodent Papillomavirus Isolated from Anogenital Lesions in Its Natural Host." *Virology* 374(1):186–97.
Nasir, Lubna and M. Saveria Campo. 2008. "Bovine Papillomaviruses: Their Role in the Aetiology of Cutaneous Tumours of Bovids and Equids." *Veterinary Dermatology* 19(5):243–54.
Nebesio, C. L., G. W. Mirowski, and T. Y. Chuang. 2001. "Human Papillomavirus: Clinical Significance and Malignant Potential." *International Journal of Dermatology* 40(6):373–79.
Omland, Turid et al. 2014. "Recurrent Respiratory Papillomatosis: HPV Genotypes and Risk of High-Grade Laryngeal Neoplasia." *PLoS ONE* 9(6).
Orlando, Paul A., Joel S. Brown, Robert A. Gatenby, and Anna R. Guliano. 2013. "The Ecology of Human Papillomavirus-Induced Epithelial Lesions and the Role of Somatic Evolution in Their Progression." *Journal of Infectious Diseases* 208(3):394–402.
Orth, G. et al. 1978. "Characterization of Two Types of Human Papillomaviruses in Lesions of Epidermodysplasia Verruciformis." *P Natl Acad Sci USA* 75(3):1537–41.
Osterhaus, A. D., D. J. Ellens, and M. C. Horzinek. 1977. "Identification and Characterization of a Papillomavirus from Birds (Fringillidae)." *Intervirology* 8(6):351–59.
Ouhoummane, N. et al. 2013. "Squamous Anal Cancer: Patient Characteristics and HPV Type Distribution." *Cancer Epidemiology* 37(6):807–12.
Paavonen, J. et al. 2009. "Efficacy of Human Papillomavirus (HPV)-16/18 AS04-Adjuvanted Vaccine against Cervical Infection and Precancer Caused by Oncogenic HPV Types (PATRICIA): Final Analysis of a Double-Blind, Randomised Study in Young Women." *The Lancet* 374(9686):301–14.
Pande, Shailja et al. 2008. "Human Papillomavirus Type 16 Variant Analysis of E6, E7, and L1 Genes and Long Control Region in Biopsy Samples from Cervical Cancer Patients in North India." *Journal of Clinical Microbiology* 46(3):1060–66.
Peng, Shiwen et al. 2007. "HLA-DQB1*02-Restricted HPV-16 E7 Peptide-Specific CD4+ T-Cell Immune Responses Correlate with Regression of HPV-16-Associated High-Grade Squamous Intraepithelial Lesions." *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 13(8):2479–87.
Pérez, Sonia, Ana Cid, Amparo Iñarrea, Mónica Pato, María José Lamas, Bárbara Couso, Margarita Gil, María Jesús Álvarez, et al. 2014. "Prevalence of HPV 16 and HPV 18 Lineages in Galicia, Spain." *PLoS ONE* 9(8).
Pérez, Sonia, Ana Cid, Amparo Iñarrea, Mónica Pato, María José Lamas, Bárbara Couso, Margarita Gil, María Jesús Alvarez, et al. 2014. "Prevalence of HPV 16 and HPV 18 Lineages in Galicia, Spain." *PloS One* 9(8):e104678.
Peters, N. and B. Watermann. 1979. "Three Types of Skin Papillomas of Flatfishes and Their Causes." *Marine Ecology Progress Series* 1:269–76.

Piegorsch, Walter W. and A. John Bailer. 2005. "Generalized Linear Models." *Analyzing Environmental Data* 103–69.
Pimenoff, Ville N., Cristina Mendes de Oliveira, and Ignacio G. Bravo. 2016. "Transmission Between Archaic and Modern Human Ancestors During the Evolution of the Oncogenic Human Papillomavirus 16." *Molecular Biology and Evolution* msw214.
Prétet, Jean Luc et al. 2008. "Human Papillomavirus Genotype Distribution in Low-Grade Squamous Intraepithelial Lesions in France and Comparison with CIN2/3 and Invasive Cervical Cancer. The EDiTH III Study." *Gynecologic Oncology* 110(2):179–84.
Prosperi, Alice et al. 2016. "Identification and Characterization of Fringilla Coelebs Papillomavirus 1 (FcPV1) in Free-Living and Captive Birds in Italy." *Journal of Wildlife Diseases* 52(3):756–58.
Qmichou, Zineb et al. 2013. "Analysis of Mutations in the E6 Oncogene of Human Papillomavirus 16 in Cervical Cancer Isolates from Moroccan Women." *BMC Infectious Diseases* 13:378.
Quint, Koen D., Maurits N. C. de Koning, Leen Jan van Doorn, Wim G. V Quint, and Edyta C. Pirog. 2010. "HPV Genotyping and HPV16 Variant Analysis in Glandular and Squamous Neoplastic Lesions of the Uterine Cervix." *Gynecologic Oncology* 117(2):297–301.
Rabelo-Santos, Silvia Helena et al. 2006. "Variants of Human Papillomavirus Types 16 and 18: Histological Findings in Women Referred for Atypical Glandular Cells or Adenocarcinoma in Situ in Cervical Smear." *International Journal of Gynecological Pathology : Official Journal of the International Society of Gynecological Pathologists* 25(4):393–97.
Read, Tim R. H. et al. 2011. "The near Disappearance of Genital Warts in Young Women 4 Years after Commencing a National Human Papillomavirus (HPV) Vaccination Programme." *Sexually Transmitted Infections* 87(7):544–47.
Reich, Olaf and Sigrid Regauer. 2015. "Two Major Pathways of Recurrent High-Grade Squamous Intraepithelial Lesions of the Cervix." *International Journal of Cancer* 137(10):2520–21.
dos Reis, Ricardo et al. 2007. "Adenosquamous Carcinoma versus Adenocarcinoma in Early-Stage Cervical Cancer Patients Undergoing Radical Hysterectomy: An Outcomes Analysis." *Gynecologic Oncology* 107(3):458–63.
Roden, R. B. et al. 1997. "Characterization of a Human Papillomavirus Type 16 Variant-Dependent Neutralizing Epitope." *Journal of Virology* 71(8):6247–52.
Ronco, Guglielmo et al. 2014. "Efficacy of HPV-Based Screening for Prevention of Invasive Cervical Cancer: Follow-up of Four European Randomised Controlled Trials." *The Lancet* 383(9916):524–32.
Rous, P. and J. W. Beard. 1935. "The Progression To Carcinoma of Virus-Induced Rabbit Papillomas (Shope)." *The Journal of Experimental Medicine* 62(4):523–48.
Rowson, K. E. and B. W. Mahy. 1967. "Human Papova (Wart) Virus." *Bacteriological Reviews* 31(2):110–31.
Rübben, a et al. 1992. "Rearrangements of the Upstream Regulatory Region of Human Papillomavirus Type 6 Can Be Found in Both Buschke-Löwenstein Tumours and in Condylomata Acuminata." *The Journal of General Virology* 73 (Pt 12:3147–53.
Rubin, M. A. et al. 2001. "Detection and Typing of Human Papillomavirus DNA in Penile Carcinoma: Evidence for Multiple Independent Pathways of Penile Carcinogenesis." *The American Journal of Pathology* 159(4):1211–18.
Ryding, J., L. Dahlberg, M. Wallen-Ohman, and J. Dillner. 2007. "Deletion of a Major Neutralizing Epitope of Human Papillomavirus Type 16 Virus-like Particles." *The Journal of General Virology* 88(Pt 3):792–802.
Sanborn, Barbara M., Berel Held, and Hsu S. Kuo. 1976. "Hormonal Action in Human Cervix-II

Specific Progestogen Binding Proteins in Human Cervix." *Journal of Steroid Biochemistry* 7(9):665–72.
de Sanjose, S. et al. 2010. "Human Papillomavirus Genotype Attribution in Invasive Cervical Cancer: A Retrospective Cross-Sectional Worldwide Study." *Lancet Oncol* 11(11):1048–56.
de Sanjosé, Silvia et al. 2013. "Worldwide Human Papillomavirus Genotype Attribution in over 2000 Cases of Intraepithelial and Invasive Lesions of the Vulva." *European Journal of Cancer (Oxford, England : 1990)* 49(16):3450–61.
de Sanjosé, Silvia, Laia Bruni, and Laia Alemany. 2014. "HPV in Genital Cancers (at the Exception of Cervical Cancer) and Anal Cancers." *Presse Medicale (Paris, France : 1983)* 43(12P2):e423–28.
Santegoets, Lindy a M. et al. 2012. "Different DNA Damage and Cell Cycle Checkpoint Control in Low- and High-Risk Human Papillomavirus Infections of the Vulva." *International Journal of Cancer. Journal International Du Cancer* 130(12):2874–85.
Sasieni, Peter, Alejandra Castanon, and Jack Cuzick. 2009. "Screening and Adenocarcinoma of the Cervix." *International Journal of Cancer* 125(3):525–29.
Scase, T. et al. 2010. "Equus Caballus Papillomavirus-2 (EcPV-2): An Infectious Cause for Equine Genital Cancer." *Equine Veterinary Journal* 42(8):738–45.
Schiffman, Mark et al. 2010. "A Population-Based Prospective Study of Carcinogenic Human Papillomavirus Variant Lineages, Viral Persistence, and Cervical Neoplasia." *Cancer Research* 70(8):3159–69.
Schiffman, Mark et al. 2016. "Carcinogenic Human Papillomavirus Infection." *Nature Reviews Disease Primers* 2:16086.
Scotto, Joseph and John C. Bailar. 1969. "Rigoni-Stern and Medical Statistics. A Nineteenth-Century Approach to Cancer Research." *Journal of the History of Medicine and Allied Sciences* 24(1):65–75.
Seliger, Barbara, Ulrike Ritz, and Soldano Ferrone. 2006. "Molecular Mechanisms of HLA Class I Antigen Abnormalities Following Viral Infection and Transformation." *International Journal of Cancer* 118(1):129–38.
Shang, Qinglong et al. 2011. "Human Papillomavirus Type 16 Variant Analysis of E6, E7, and L1 [Corrected] Genes and Long Control Region in [Corrected] Cervical Carcinomas in Patients in Northeast China." *Journal of Clinical Microbiology* 49(7):2656–63.
Shope, R. E. and E. W. Hurst. 1933. "INFECTIOUS PAPILLOMATOSIS OF RABBITS : WITH A NOTE ON THE HISTOPATHOLOGY." *The Journal of Experimental Medicine* 58(5):607–24.
Sichero, L. and L. L. Villa. 2006. "Epidemiological and Functional Implications of Molecular Variants of Human Papillomavirus." *Brazilian Journal of Medical and Biological Research* 39:707–17.
Sichero, Laura et al. 2007. "High Grade Cervical Lesions Are Caused Preferentially by Non-European Variants of HPVs 16 and 18." *International Journal of Cancer* 120(8):1763–68.
Sichero, Laura, Eduardo Luis Franco, and Luisa Lina Villa. 2005. "Different P105 Promoter Activities among Natural Variants of Human Papillomavirus Type 18." *The Journal of Infectious Diseases* 191(5):739–42.
Smith, H. O., M. F. Tiffany, C. R. Qualls, and C. R. Key. 2000. "The Rising Incidence of Adenocarcinoma Relative to Squamous Cell Carcinoma of the Uterine Cervix in the United States--a 24-Year Population-Based Study." *Gynecologic Oncology* 78(2):97–105.
Smith, Jennifer S. et al. 2007. "Human Papillomavirus Type Distribution in Invasive Cervical Cancer and High-Grade Cervical Lesions: A Meta-Analysis Update." *International Journal of Cancer* 121(3):621–32.
Snijders, P. J. et al. 1992. "Human Papillomavirus (HPV) Type 16 and 33 E6/E7 Region Transcripts in Tonsillar Carcinomas Can Originate from Integrated and Episomal HPV

DNA." *J Gen Virol* 73 (Pt 8):2059–66.
Stamatakis, A. 2006. "RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models." *Bioinformatics* 22(21):2688–90.
Stanley, M. a. 2012. "Epithelial Cell Responses to Infection with Human Papillomavirus." *Clinical Microbiology Reviews* 25(2):215–22.
Stanley, M., D. R. Lowy, and I. Frazer. 2006. "Prophylactic HPV Vaccines: Underlying Mechanisms." *Vaccine* 24:106–13. Retrieved (<Go to ISI>://000240470000014).
Stanley, Margaret A., David M. Winder, Jane C. Sterling, and Peter K. C. Goon. 2012. "HPV Infection, Anal Intra-Epithelial Neoplasia (AIN) and Anal Cancer: Current Issues." *BMC Cancer* 12:398.
Stark, M., S. A. Berger, A. Stamatakis, and C. von Mering. 2010. "MLTreeMap--Accurate Maximum Likelihood Placement of Environmental DNA Sequences into Taxonomic and Functional Reference Phylogenies." *BMC Genomics* 11:461.
Sturegård, Erik et al. 2013. "Human Papillomavirus Typing in Reporting of Condyloma." *Sexually Transmitted Diseases* 40(2):123–29.
Sun, Min et al. 2012. "Whole Genome Sequencing and Evolutionary Analysis of Human Papillomavirus Type 16 in Central China." *PloS One* 7(5):e36577.
Swan, David C., Josef R. Limor, Kara L. Duncan, Mangalathu S. Rajeevan, and Elizabeth R. Unger. 2006. "Human Papillomavirus Type 16 Variant Assignment by Pyrosequencing." *Journal of Virological Methods* 136(1-2):166–70.
Syrjänen, Stina. 2010. "Current Concepts on Human Papillomavirus Infections in Children." *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 118(6-7):494–509.
Tan, Sarah E. et al. 2013. "Investigating a Cluster of Vulvar Cancers in Young Women: Distribution of Human Papillomavirus and HPV-16 Variants in Vulvar Dysplastic or Neoplastic Biopsies." *Sexual Health* 10:18–25.
Terai, Masanori, Rob DeSalle, and Robert D. Burk. 2002. "Lack of Canonical E6 and E7 Open Reading Frames in Bird Papillomaviruses: Fringilla Coelebs Papillomavirus and Psittacus Erithacus Timneh Papillomavirus." *Journal of Virology* 76(19):10020–23.
The 1000 Genomes Project Consortium. 2012. "An Integrated Map of Genetic Variation from 1,092 Human Genomes." *Nature* 491(7422):56–65.
Thomas, D. B. and R. M. Ray. 1996. "Oral Contraceptives and Invasive Adenocarcinomas and Adenosquamous Carcinomas of the Uterine Cervix. The World Health Organization Collaborative Study of Neoplasia and Steroid Contraceptives." *American Journal of Epidemiology* 144(3):281–89.
Tobler, Kurt, Claude Favrot, Gilles Nespeca, and Mathias Ackermann. 2006. "Detection of the Prototype of a Potential Novel Genus in the Family Papillomaviridae in Association with Canine Epidermodysplasia Verruciformis." *The Journal of General Virology* 87(Pt 12):3551–57.
Togtema, Melissa, Robert Jackson, Christina Richard, Sarah Niccoli, and Ingeborg Zehbe. 2015. "The Human Papillomavirus 16 European-T350G E6 Variant Can Immortalize but Not Transform Keratinocytes in the Absence of E7." *Virology* 485:274–82.
Tornesello, M. L. et al. 2011. "Human Papillomavirus (HPV) Genotypes and HPV16 Variants and Risk of Adenocarcinoma and Squamous Cell Carcinoma of the Cervix." *Gynecologic Oncology* 121:32–42.
Tornesello, Maria Lina et al. 2004. "Analysis of Human Papillomavirus Type-16 Variants in Italian Women with Cervical Intraepithelial Neoplasia and Cervical Cancer." *Journal of Medical Virology* 74:117–26.
Tornesello, Maria Lina et al. 2008. "Human Papillomavirus Genotypes and HPV16 Variants in Penile Carcinoma." *International Journal of Cancer* 122(1):132–37.

Tyring, S. K. 2000. "Human Papillomavirus Infections: Epidemiology, Pathogenesis, and Host Immune Response." *Journal of the American Academy of Dermatology* 43(1 Pt 2):S18–26.
Vartanian, Jean-Pierre, Denise Guétard, Michel Henry, and Simon Wain-Hobson. 2008. "Evidence for Editing of Human Papillomavirus DNA by APOBEC3 in Benign and Precancerous Lesions." *Science (New York, N.Y.)* 320(5873):230–33.
Veldman, T., X. Liu, H. Yuan, and R. Schlegel. 2003. "Human Papillomavirus E6 and Myc Proteins Associate in Vivo and Bind to and Cooperatively Activate the Telomerase Reverse Transcriptase Promoter." *Proc Natl Acad Sci U S A* 100(14):8211–16.
Venkatesan, Naren N., Harold S. Pine, and Michael P. Underbrink. 2012. "Recurrent Respiratory Papillomatosis." *Otolaryngologic Clinics of North America* 45(3):671–94.
Viens, Laura J. et al. 2016. "Human Papillomavirus–Associated Cancers — United States, 2008–2012." *MMWR. Morbidity and Mortality Weekly Report* 65(26):661–66.
Villa, L. L. et al. 2000. "Molecular Variants of Human Papillomavirus Types 16 and 18 Preferentially Associated with Cervical Neoplasia." *J Gen Virol* 81(Pt 12):2959–68.
de Villiers, E. M., L. Gissmann, and H. zur Hausen. 1981. "Molecular Cloning of Viral DNA from Human Genital Warts." *Journal of Virology* 40(3):932–35.
De Villiers, Ethel Michele, Claude Fauquet, Thomas R. Broker, Hans Ulrich Bernard, and Harald Zur Hausen. 2004. "Classification of Papillomaviruses." *Virology* 324(1):17–27.
Vinh-Hung, Vincent et al. 2007. "Prognostic Value of Histopathology and Trends in Cervical Cancer: A SEER Population Study." *BMC Cancer* 7:164.
Virgin, H. W., E. J. Wherry, and R. Ahmed. 2009. "Redefining Chronic Viral Infection." *Cell* 138(1):30–50.
Visioli, Carmen B., Marco Zappa, Stefano Ciatto, Anna Iossa, and Emanuele Crocetti. 2004. "Increasing Trends of Cervical Adenocarcinoma Incidence in Central Italy despite Extensive Screening Programme, 1985-2000." *Cancer Detection and Prevention* 28(6):461–64.
Vizcaino, a P. et al. 2000. "International Trends in Incidence of Cervical Cancer: II. Squamous-Cell Carcinoma." *International Journal of Cancer. Journal International Du Cancer* 86(3):429–35.
Vrtačnik Bokal, Eda, Boštjan J. Kocjan, Mario Poljak, Zeljka Bogovac, and Nina Jančar. 2010. "Genomic Variants of Human Papillomavirus Genotypes 16, 18, and 33 in Women with Cervical Cancer in Slovenia." *The Journal of Obstetrics and Gynaecology Research* 36(6):1204–13.
Waggoner, Steven E. 2003. "Cervical Cancer." Pp. 2217–25 in *Lancet*, vol. 361.
Wang, Zhe et al. 2014. "APOBEC3 Deaminases Induce Hypermutation in Human Papillomavirus 16 DNA upon Beta Interferon Stimulation." *Journal of Virology* 88(2):1308–17.
van der Weele, Pascal, Chris J. L. M. Meijer, and Audrey J. King. 2017. "Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections." *Journal of Virology* 91(19).
Welters, Marij J. P. et al. 2003. "Frequent Display of Human Papillomavirus Type 16 E6-Specific Memory T-Helper Cells in the Healthy Population as Witness of Previous Viral Encounter." *Cancer Research* 63(3):636–41.
Wiatrak, Brian J., Deborah W. Wiatrak, Thomas R. Broker, and Linda Lewis. 2004. "Recurrent Respiratory Papillomatosis: A Longitudinal Study Comparing Severity Associated with Human Papilloma Viral Types 6 and 11 and Other Risk Factors in a Large Pediatric Population." *The Laryngoscope* 114(11 Pt 2 Suppl 104):1–23.
Winer, Rachel L. et al. 2005. "Development and Duration of Human Papillomavirus Lesions, after Initial Infection." *The Journal of Infectious Diseases* 191:731–38.
Xi, Juemin et al. 2017. "Genetic Variability and Functional Implication of the Long Control

Region in HPV-16 Variants in Southwest China." *PLoS ONE* 12(8).
Xi, Long Fu et al. 1997. "Genomic Variation of Human Papillomavirus Type 16 and Risk for High Grade Cervical Intraepithelial Neoplasia." *Journal of the National Cancer Institute* 89(11):796–802.
Xi, Long Fu et al. 1998. "Risk of Anal Carcinoma in Situ in Relation to Human Papillomavirus Type 16 Variants Risk of Anal Carcinoma in Situ in Relation to Human Papillomavirus." *Journal of the National Cancer Institute* 3839–44.
Xi, Long Fu et al. 2006. "Human Papillomavirus Type 16 and 18 Variants: Race-Related Distribution and Persistence." *Journal of the National Cancer Institute* 98(15):1045–52.
Xi, Long Fu et al. 2007. "Risk for High-Grade Cervical Intraepithelial Neoplasia Associated with Variants of Human Papillomavirus Types 16 and 18." *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 16(1):4–10.
Xi, Long Fu et al. 2012. "Association of Human Papillomavirus Type 31 Variants with Risk of Cervical Intraepithelial Neoplasia Grades 2-3." *International Journal of Cancer* 131(10):2300–2307.
Xi, Long Fu et al. 2013. "Persistence of Newly Detected Human Papillomavirus Type 31 Infection, Stratified by Variant Lineage." *International Journal of Cancer* 132(3):549–55.
Xin, C. Y. et al. 2001. "Analysis of E6 Variants of Human Papillomavirus Type 33, 52 and 58 in Japanese Women with Cervical Intraepithelial Neoplasia/cervical Cancer in Relation to Their Oncogenic Potential." *Cancer Letters* 170(1):19–24.
Yamada, T. et al. 1997. "Human Papillomavirus Type 16 Sequence Variation in Cervical Cancers: A Worldwide Perspective." *Journal of Virology* 71:2463–72.
Yan, Ming, Yan Na Zhang, Jie Hua He, and H. Huang. 2008. "Prognosis Analysis of 83 Cases of Cervical Adenosquamous Carcinoma." *Ai Zheng = Aizheng = Chinese Journal of Cancer* 27(9):956–61.
Yang, Eric J. et al. 2015. "Microanatomy of the Cervical and Anorectal Squamocolumnar Junctions: A Proposed Model for Anatomical Differences in HPV-Related Cancer Risk." *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 28(7):994–1000.
Yim, Eun-Kyoung and Jong-Sup Park. 2005. "The Role of HPV E6 and E7 Oncoproteins in HPV-Associated Cervical Carcinogenesis." *Cancer Research and Treatment : Official Journal of Korean Cancer Association* 37(6):319–24.
Zehbe, I. et al. 2001. "Human Papillomavirus 16 E6 Polymorphisms in Cervical Lesions from Different European Populations and Their Correlation with Human Leukocyte Antigen Class II Haplotypes." *International Journal of Cancer. Journal International Du Cancer* 94:711–16.
Zehbe, I. et al. 2003. "Association between Human Papillomavirus 16 E6 Variants and Human Leukocyte Antigen Class I Polymorphism in Cervical Cancer of Swedish Women." *Hum Immunol* 64(5):538–42.
Zehbe, Ingeborg et al. 2009. "Human Papillomavirus 16 E6 Variants Differ in Their Dysregulation of Human Keratinocyte Differentiation and Apoptosis." *Virology* 383(1):69–77.
Zehbe, Ingeborg, Erik Wilander, Hajo Delius, and Massimo Tommasino. 1998. "Human Papillomavirus 16 E6 Variants Are More Prevalent in Invasive Cervical Carcinoma than the Prototype." *Cancer Research* 58:829–33.
Zhang, Xiao et al. 2016. "Functional Assessment and Structural Basis of Antibody Binding to Human Papillomavirus Capsid." *Reviews in Medical Virology* 115–28.
Zuna, Rosemary E. et al. 2009. "Association of HPV16 E6 Variants with Diagnostic Severity in Cervical Cytology Samples of 354 Women in a US Population." *International Journal of*

Cancer. Journal International Du Cancer 125(11):2609–13.
Zuna, Rosemary E. et al. 2011. "HPV16 Variant Lineage, Clinical Stage, and Survival in Women with Invasive Cervical Cancer." *Infectious Agents and Cancer* 6:19.