Supporting your research with our capabilities

BD Accuri™ C6 Plus Personal Flow Cytometer
BD FACSCelesta™ Cell Analyzer
BD LSRFortessa™ X-20 Cell Analyzer
BD FACSMelody™ Cell Sorter
One of the largest portfolios of reagents

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

S. Nicolás-Parraga, L. Alemany, S. de Sanjosé, F.X. Bosch, I.G. Bravo

Infectious Causes of Cancer

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, being responsible for approximately 266,000 deaths per year worldwide. 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. Persistent infections by oncogenic Human Papillomaviruses (HPVs) are the etiologic cause of virtually all cervical cancers. 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) (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 squamocolumnar 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 Fundación Deus ex Women’s Health grant to IGB, the Agència de Gestió d’Ajuts Universitaris i de Investigació, AGAUR, Generalitat de Catalunya (2014SGR1077 to XFB), and the Fondo de Investigaciones Sanitarias (F112/00142 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, Génétique, Evolution et Contrôle (MIVEGEC), UMR CNRS 5290, IRD 224, UM, 911 Avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France, E-mail: ignacio.bravo@ird.fr; Tel. +33 467 41 5123

squamous epithelium of the ectocervix and the non-ciliated simple columnar epithelium of the endocervix. Instead, most cervical ADCs originate mostly from endocervix glandular precursor lesions. Finally, cervical ADSC is a mixed histological type amalgamating malignant glandular and squamous components consisting of intermingled ADC and SCC. 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 clinicopathological factor that influences prognosis. After radical hysterectomy, both ADC and ADSC present a poorer prognosis than SCC, with nearly 10–20% difference in 5-year overall survival rates.

The most common presentation of ICC is SCC, accounting for 80–85% of all ICC cases, compared to 10–15% of ADC and 2–3% ADSC. 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. 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, 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.

Certain exceptions to this trend are remarkable, as it is the case of Ireland. 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, especially among young adult women. The forces driving this increase in ADC and ADSC detection remain nevertheless unclear.

Not all HPV's 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); 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. 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). Thus, oncogenic HPV's 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. 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 ADSCs.

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.
Infectious Causes of Cancer

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

<table>
<thead>
<tr>
<th>EUR-CSA-AS-AF SAMPLES</th>
<th>Initial</th>
<th>Amplified</th>
<th>Classified(^1)</th>
<th>Unclassified(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCC</td>
<td>118</td>
<td>111</td>
<td>109</td>
<td>2</td>
</tr>
<tr>
<td>ADSC</td>
<td>53</td>
<td>32</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>ADC</td>
<td>120</td>
<td>97</td>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>291</td>
<td>240</td>
<td>233</td>
<td>7</td>
</tr>
</tbody>
</table>

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.\(^3\) 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+Γ model,\(^3\)\(^9\) as previously described.\(^3\)\(^8\) 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.\(^4\) 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 Rates (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.\(^3\)\(^5\) 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)\(^3\)\(^5\) 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 under-represented 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)

<table>
<thead>
<tr>
<th>Df</th>
<th>Res. Dev.</th>
<th>Df</th>
<th>Res. Dev.</th>
<th>% exp. Dev.</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>29</td>
<td>382.41</td>
<td>Variant</td>
<td>4</td>
<td>298.3</td>
</tr>
<tr>
<td>Histology</td>
<td>1</td>
<td>1.192</td>
<td>24</td>
<td>82.96</td>
<td>0.31</td>
</tr>
<tr>
<td>Geography</td>
<td>2</td>
<td>2.635</td>
<td>22</td>
<td>80.32</td>
<td>0.69</td>
</tr>
<tr>
<td>Variant:histology</td>
<td>4</td>
<td>35.193</td>
<td>18</td>
<td>46.13</td>
<td>9.20</td>
</tr>
<tr>
<td>Variant:geography</td>
<td>8</td>
<td>38.598</td>
<td>10</td>
<td>6.53</td>
<td>10.1</td>
</tr>
<tr>
<td>History:geography</td>
<td>2</td>
<td>3.789</td>
<td>8</td>
<td>2.74</td>
<td>1</td>
</tr>
<tr>
<td>Variant:history:geography</td>
<td>8</td>
<td>2.742</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
</tr>
</tbody>
</table>

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]
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

![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 × Inter-quartile). Numbers in parentheses at the bottom indicate sample size for each location.](image)

### 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).

<table>
<thead>
<tr>
<th>Histology</th>
<th>Variants</th>
<th>EUR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>SCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1-3</td>
<td>32</td>
<td>97</td>
<td>36</td>
</tr>
<tr>
<td>A4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>3.6</td>
<td>7</td>
</tr>
<tr>
<td>Sub-total</td>
<td>33</td>
<td>100</td>
<td>43</td>
</tr>
<tr>
<td>ADC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1-3</td>
<td>19</td>
<td>67.9</td>
<td>11</td>
</tr>
<tr>
<td>A4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>3.6</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td>28.6</td>
<td>19</td>
</tr>
<tr>
<td>Sub-total</td>
<td>28</td>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1-3</td>
<td>51</td>
<td>83.6</td>
<td>47</td>
</tr>
<tr>
<td>A4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>9</td>
<td>14.8</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>100</td>
<td>73</td>
</tr>
</tbody>
</table>

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): H₀ 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): H₀ 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)

<table>
<thead>
<tr>
<th>Variants</th>
<th>SCC Ref (n var)</th>
<th>ADC (n’var)</th>
<th>PR</th>
<th>PR Wald-test</th>
<th>95% CI</th>
<th>Fisher-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR A1-3</td>
<td>32</td>
<td>19</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>0.004</td>
</tr>
<tr>
<td>A4/B/C/D</td>
<td>1 (-/-/1)</td>
<td>9 (-/-/1/8)</td>
<td>2.42</td>
<td>0.002</td>
<td>1.60-3.65</td>
<td></td>
</tr>
<tr>
<td>CSA A1-3</td>
<td>36</td>
<td>11</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>A4/B/C/D</td>
<td>7 (-/-/7)</td>
<td>19 (-/-/19)</td>
<td>3.12</td>
<td>&lt;0.0001</td>
<td>1.77-5.51</td>
<td></td>
</tr>
<tr>
<td>AS A1-3</td>
<td>20</td>
<td>11</td>
<td>Ref</td>
<td>Ref</td>
<td>Ref</td>
<td>0.006</td>
</tr>
<tr>
<td>A4/B/C/D</td>
<td>6 (3/-/1/2)</td>
<td>18 (8/-/-/10)</td>
<td>2.11</td>
<td>0.004</td>
<td>1.25-3.58</td>
<td></td>
</tr>
</tbody>
</table>

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...
Infectious Causes of Cancer

...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, as this histological type remains the most prevalent ICC. A number of studies had addressed other cancer histologies but had focused on data from a restricted geographic origin. Globally, our results confirm and expand previous reports. 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; 41.7% of 24 ADCs compared to 2.4% of 42 SCCs; 67.5% of 40 ADCs compared to 15.9% of 69 SCCs) and from Spain (85.7% from 7 glandular pathologies compared to 28.6% from 7 SCCs). 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; 75.4% prevalence in 69 SCCs compared to 25% prevalence in 40 ADCs), and in other geographically more extended works (60% of 98 SCC compared to 42% in ADC). 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). Regarding variation in HPV16 lineage prevalence depending on the geographical origin of the samples, our results largely confirm the best data available, 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. Some authors have identified polymorphisms in HPV16_D variants glucocorticoid response elements (GREs) that confer facilitated activation of promoter p97, leading to an enhanced E6-E7 transcription activity. 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. 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. However, other large studies did not find differences in age between glandular and squamous ICCs. Because distinct HPVs are differentially associated with either cancer presentation and because more aggressive HPVs such as HPV16, 18 or 45 cause cancers in younger ages than other HPVs, 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. 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. 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. 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. But exfoliation cytology may be less efficient at capturing the early cytopathologic signs of ADC because it tends to occur in the endocervical canal. 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 reassessed 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 HPV's 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**

Infectious Causes of Cancer

HPV16 variants in different cervical cancer histologies


