Received: 5 February 2024 | Accepted: 30 September 2024

DOI: 10.1002/jmv.70002

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Viral whole genome sequencing reveals high variations in APOBEC3 editing between HPV risk categories

Valentine Marie Ferré[1,2] | Romain Coppée[1] | Fifonsi A. Gbeasor-Komlanvi[3,4] | Sophie Vacher[5] | Antoine Bridier-Nahmias[1] | Margot Bucau[6] | Mounerou Salou[7] | Sonia Lameiras[8] | Anne Couvelard[6,9] | Anoumou Claver Dagnra[7,10] | Ivan Bieche[5,11] | Diane Descamps[1,2] | Didier K. Ekouevi[3,12] | Jade Ghosn[1,13] | Charlotte Charpentier[1,2]

[1]Université Paris Cité, Inserm IAME UMR 1137, Paris F-75018, France

[2]Service de Virologie, AP-HP, Hôpital Bichat – Claude Bernard, Paris F-75018, France

[3]Département de Santé Publique, Université de Lomé, Faculté des Sciences de la Santé, Lomé, Togo

[4]Centre Africain de Recherche en Epidémiologie et en Santé Publique (CARESP), Lomé, Togo

[5]Department of Genetics, Institut Curie, PSL Research University, Paris, France

[6]Département de Pathologie, AP-HP, Hôpital Bichat-Claude Bernard, Paris F-75018, France

[7]Université de Lomé, Centre de Biologie Moléculaire et d'Immunologie, Lomé, Togo

[8]Institut Curie, Genomics of Excellence (ICGex) Platform, PSL Research University, Paris, France

[9]Université de Paris, Centre of Research on Inflammation, Paris INSERM U1149, France

[10]Programme national de lutte contre le sida et les infections sexuellement transmissibles, Lomé, Togo

[11]INSERM U1016, Faculty of Pharmaceutical and Biological Sciences, Paris Cité University, Paris, France

[12]ISPED, Université de Bordeaux & Centre INSERM U1219 - Bordeaux Population Health, Bordeaux, France

[13]Service de Maladies Infectieuses et Tropicales, AP-HP, Hôpital Bichat-Claude Bernard, Paris F-75018, France

Correspondence
Romain Coppée and Valentine Marie Ferré, Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France.
Email: romain.coppee@inserm.fr and valentinemarie.ferre@aphp.fr

Abstract

High-risk human papillomavirus (HPV) infections are responsible for cervical cancer. However, little is known about the differences between HPV types and risk categories regarding their genetic diversity and particularly APOBEC3-induced mutations – which contribute to the innate immune response to HPV. Using a capture-based next-generation sequencing, 156 HPV whole genome sequences covering 43 HPV types were generated from paired cervical and anal swabs of 30 Togolese female sex workers (FSWs) sampled in 2017. Genetic diversity and APOBEC3-induced mutations were assessed at the viral whole genome and gene levels. Thirty-four pairwise sequence comparisons covering 24 HPV types in cervical and anal swabs revealed identical infections in the two anatomical sites. Differences in genetic diversity among HPV types was observed between patients. The E6 gene was significantly less

Valentine Marie Ferré and Romain Coppée contributed equally to the article as co-first authors

conserved in low-risk HPVs (lrHPVs) compared to high-risk HPVs (hrHPVs) ($p = 0.009$). APOBEC3-induced mutations were found to be more common in lrHPVs than in hrHPVs ($p = 0.005$), supported by our data and by using large HPV sequence collections from the GenBank database. Focusing on the most common lrHPVs 6 and 11 and hrHPVs 16 and 18, APOBEC3-induced mutations were predominantly found in the E4 and E6 genes in lrHPVs, but were almost absent in these genes in hrHPVs. The variable APOBEC3 mutational signatures could contribute to the different oncogenic potentials between HPVs. Further studies are needed to conclusively determine whether APOBEC3 editing levels are associated to the carcinogenic potential of HPVs at the type and sublineage scales.

KEYWORDS
anal, APOBEC3, cervical, genetic diversity, human papillomavirus, whole genome sequencing

1 | INTRODUCTION

Human papillomavirus (HPV) infection is the most common sexually transmitted infection. Most cervical HPV infections are asymptomatic and cleared by the immune system within 18 months or less.[1,2] However, about 5% of all cancers worldwide are caused by HPV.[3,4] Cervical cancer is associated with persistence and multiple infections with different high-risk HPV types (hrHPVs), and is the leading cause of cancer death among women in sub-Saharan Africa.[5] Most high-income countries have implemented vaccination and screening strategies for cervical cancer, but these are still poorly implemented in sub-Saharan Africa.[6,7] For other HPV-induced cancers, the anal pathogenesis of HPV infection is much less well described. For anal cancer, precancer stages have been defined based on cervical pathogenesis, but knowledge of lesion clearance or progression to cancer at each stage in the general population is limited.

At the molecular level, HPV is a circular genome of 7.9-kbp double-stranded DNA consisting of more than 200 variants classified according to their oncogenic properties.[8–10] Low-risk HPVs (lrHPVs) are responsible for benign genital warts, with HPV types 6 and 11 being involved in more than 90% of cases. On the other hand, hrHPVs are systematically found in invasive cancers, with HPV16 and 18 causing more than 70% of cervical cancers.[11,12] The carcinogenicity of hrHPV is driven by cellular alterations caused by the viral oncoproteins E6 and E7 and their overexpression through various mechanisms such as HPV and host genome methylation and viral genome integration.[12] Mutations in viral genes have been associated with cervical cancer in various populations,[13,14] highlighting the importance of studying mutation frequencies along the HPV genome. In particular, the cytidine deaminase apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3 (APOBEC3) proteins are part of the antiviral host defense. A strong enrichment of APOBEC3 signature mutations has been reported in cervical cancer exomes.[15–18] APOBEC3A has been described as a viral restriction factor[19,20]; both APOBEC3A and B appear to be upregulated by HPVs[19,21,22]; and a lower rate of APOBEC3 signature mutations has been reported in high-grade cervical lesions compared to benign infections.[23] Several questions remain regarding the impact of APOBEC3 on HPV: Is a low rate of APOBEC3 mutation frequency a driver of viral evolution? Are these mutations involved in the carcinogenic potential of HPVs or are they responsible for viral restriction? [24]

In a previous study, we assessed the prevalence of HPV types among female sex workers (FSWs) in Togo, a country with a high burden of cervical cancer.[25] Our results showed a prevalence of 34.1% and 20.7% of hrHPVs in cervical and anal smears, respectively.[26] To complement the clinical study, we performed high-throughput capture-based sequencing to obtain whole viral genome sequences from cervical and anal swabs of 30 Togolese FSWs and to explore in depth the inter- and intra-patient genetic diversity of HPVs, with a strong focus on APOBEC3-induced mutations in the viral genome.

2 | RESULTS

2.1 | Participant characteristics and description of HPV types detected

This study included samples from 30 Togolese FSWs with at least one similar HPV type in both anal and cervical anatomical sites. The median age of the participants was 21 years (IQR = 19-25) and seven (23%) were infected with human immunodeficiency virus (HIV). Regarding cervical cytology results, 22 had normal results (73%), 3 had atypical squamous cells of undetermined significance (ASC-US; 10%), and 5 had low-grade squamous intraepithelial lesions (LSIL; 17%) (Supplementary Table S1).

Whole genome sequencing was successful in at least one of the two anatomical sites in 28 participants (Figure 1A). HPV sequencing was significantly more successful in cervical than in anal swabs ($n = 28$ vs 19; $p = 0.012$, chi-squared test; Figure 1B). Forty-four

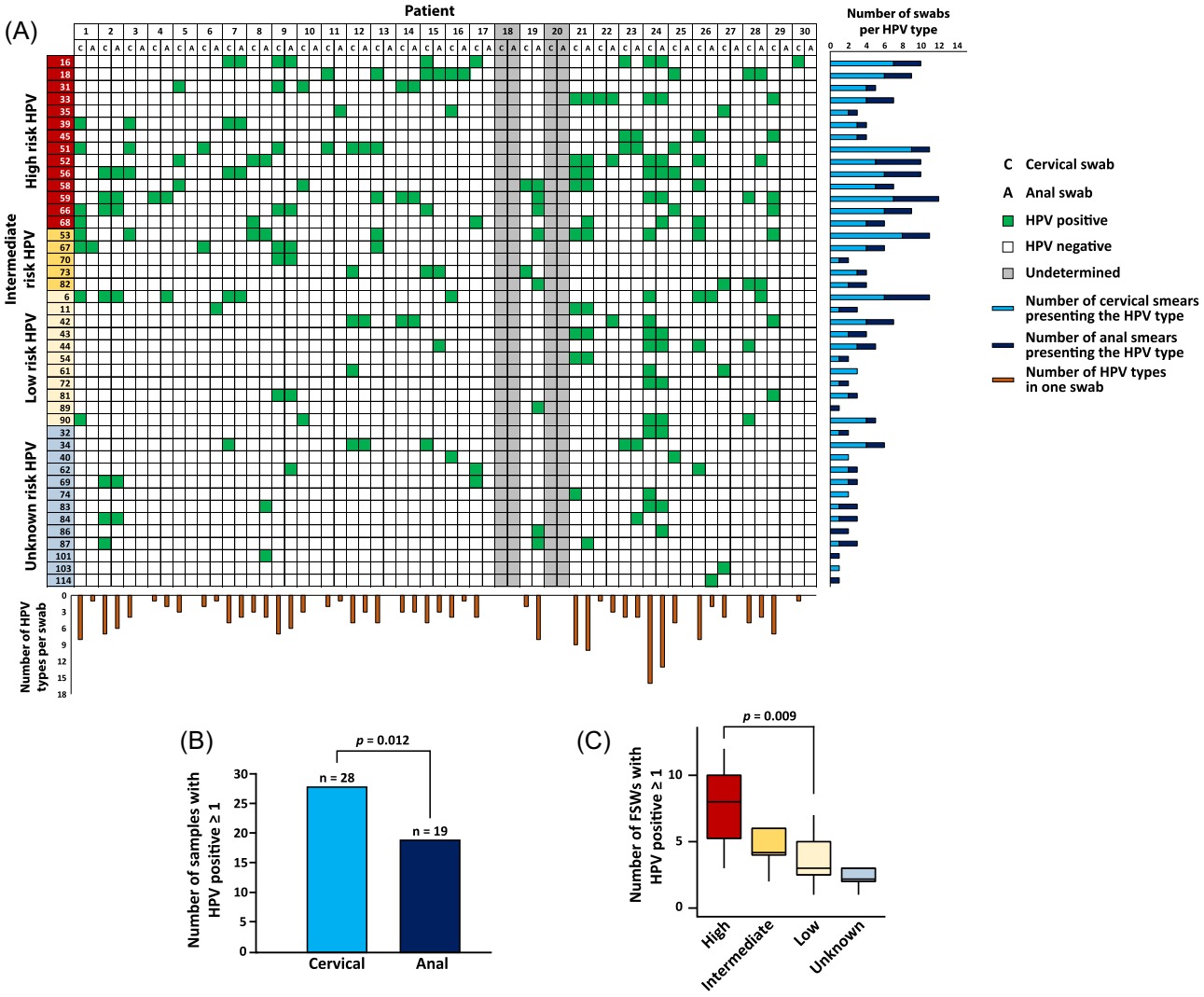

FIGURE 1 Anal and cervical HPV types detected in the study. (A) Distribution of HPV types among the 30 Togolese FSW patients included in the study. Green boxes indicate HPVs that were partially or fully sequenced by capture-based next-generation sequencing. The *right* bar plot shows the number of swabs positive for a specific HPV type. The *lower* bar plot shows the number of HPV types detected in one swab. Grey columns correspond to patients in whom no HPV was detected. (B) Number of cervical and anal swabs positive for at least one HPV type. Statistical comparison was performed using the chi-squared test. (C) Number of FSW patients positive for at least one HPV type by risk category. The box borders represent the first and third quartiles and the length of the whiskers corresponds to 1.5 times the interquartile range. Comparison between hrHPVs and lrHPVs was performed using the Mann-Whitney *U* test.

different HPV types were sequenced in cervical and anal swabs. The sequencing data details (number of reads mapped and coverage) for each HPV type in each sample are described in Supplementary Table S2. All types sequenced except one (HPV101) belonged to the genus Alphapapillomavirus and all 14 hrHPV types were represented. The capture-based next-generation sequencing method was able to detect up to 16 different HPV types in a single swab, and the majority of samples had more than one HPV at least partially sequenced (median = 3). The most commonly detected hrHPVs were HPV59 ($n = 12$), HPV51 (n = 11), HPV16, HPV52 and HPV56 (n = 10); and the most commonly detected lrHPV was HPV6 ($n = 11$) (Figure 1A). Overall, hrHPV types were significantly more common than lrHPV types regardless of anatomical site ($p = 0.009$, Mann-Whitney *U* test; Figure 1C). The comparison between traditional PCR HPV detection results and capture NGS HPV detection is shown in Supplementary Table S3.

2.2 | Variable levels of genetic diversity among HPV types

Consensus genome sequences ($n = 156$) covering 43 different HPV types (Supplementary Table S4) were generated to assess the genetic diversity of HPVs at both inter- and intra-patient levels. We confirmed that each generated sequence clustered with the expected HPV-type reference genome by constructing a maximum likelihood phylogenetic tree based on the highly conserved E6 oncoprotein and L1 major capsid protein (Supplementary Figure S1A). For HPV16 (the

most carcinogenic hrHPV), all study samples ($n = 7$) belonged to the C1 sublineage which is common in North Africa, consistent with the geographical origin of this study (Togo) (Supplementary Figure S1B).

We then compared the sequences of the same HPV type found in both anatomical sites in the same patient. Thirty-four pairwise sequence comparisons covering 24 HPV types were performed and systematically showed 100% sequence identity, suggesting that the anal and cervical sites shed the same HPV in the same patient in our data set. In contrast, using the number of nucleotides that differed from the reference genomes, we observed high genetic diversity between patients for the same HPV type, with HPV44, HPV68 and HPV82 associated with the highest standard deviations (Figure 2A). On average, we found no difference in the total number of nucleotide substitutions between hrHPV and lrHPV categories ($p = 0.88$, Mann-Whitney U test; Figure 2B).

We next examined the distribution of nucleotide substitutions among viral genes after normalization to gene size in the different HPV risk categories. Both hrHPV and lrHPV categories were associated

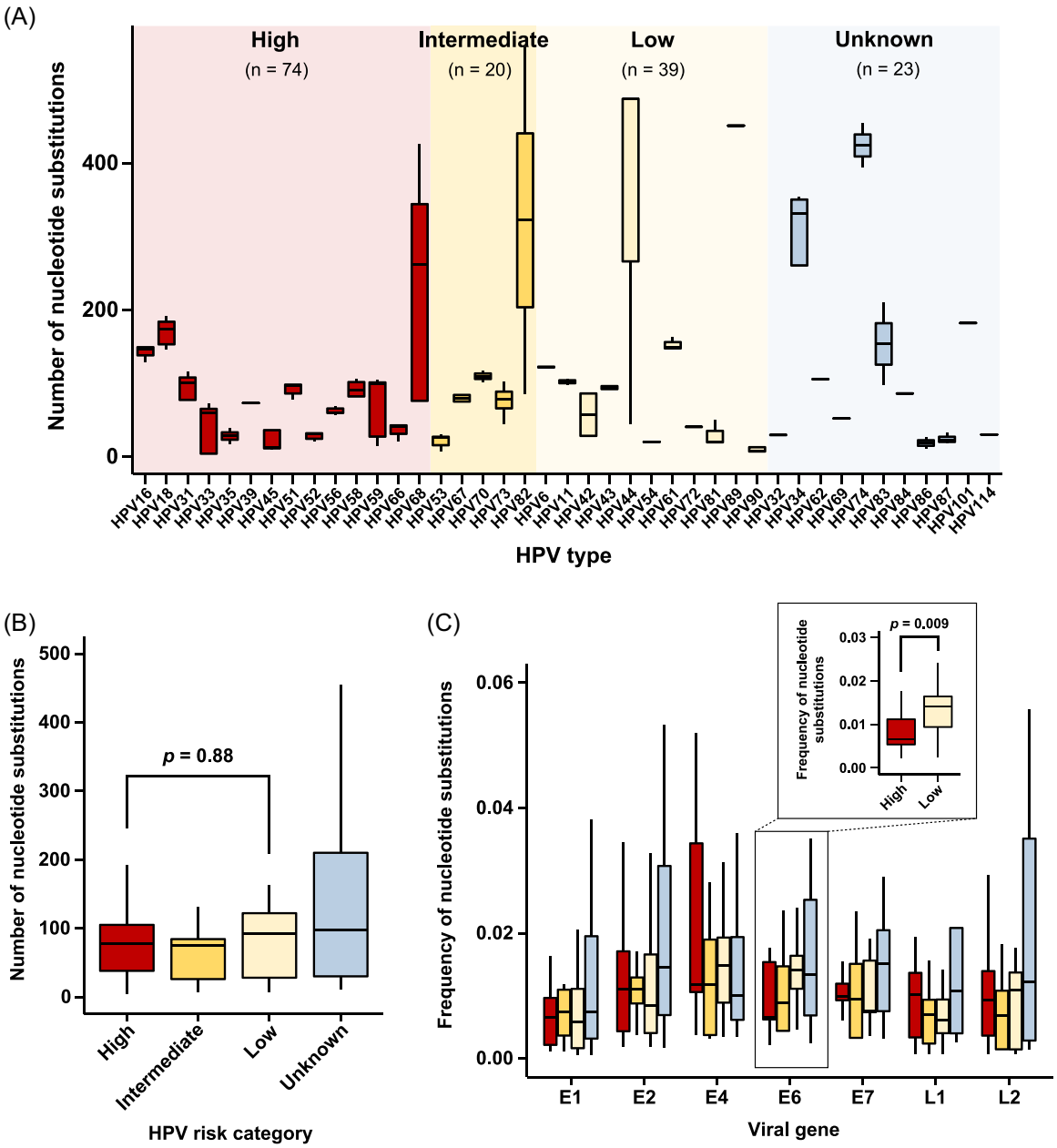

FIGURE 2 Genetic diversity of HPV types sequenced in the study. (A) Number of nucleotide substitutions in the consensus sequences generated for each HPV type. Each HPV risk category is represented by a specific color. The number of sequences included for each risk category is indicated above the plot. (B) Number of nucleotide substitutions per HPV risk category. (C) Frequency of nucleotide substitutions for each viral gene according to HPV risk category. Frequencies were normalized to gene size. The *inset* plot shows the statistical difference in mutation frequency between hrHPVs and lrHPVs for the E6 gene. The box borders represent the first and third quartiles and the length of the whiskers corresponds to 1.5 times the interquartile range. Outliers were masked.

with a significant variation in the number of nucleotide substitutions between genes ($p < 0.001$, Kruskal-Wallis H test; Figure 2C). The E4 viral gene was the least conserved regardless of risk category. For the gene encoding the oncoprotein E6, mutations were twice as frequent in lrHPVs than in hrHPVs ($p = 0.009$, Mann-Whitney U test). Furthermore, the E6 gene was the most conserved gene in hrHPVs, whereas it was the second least conserved gene after E4 in lrHPVs. Altogether, the results indicated that nucleotide substitutions are not randomly distributed along the HPV genome.

2.3 | A higher accumulation of APOBEC3-induced mutations in lrHPV than in hrHPV types

We next focused on the occurrence of APOBEC3-induced mutations in HPV genomes. The TCW motif (i.e. the target nucleotide sequence of APOBEC3 editing) was similar in number across HPV risk categories ($p = 0.77$, Kruskal-Wallis H test; Supplementary Figure S2A). In the viral genes, the TCW motif was predominantly found in L1 and L2 genes (Supplementary Figure S2B). Between risk categories, the motif was found in similar proportions in all genes except E7, which was more frequent in hrHPVs ($p = 0.01$).

To account for C > T mutations being due to APOBEC3 activity rather than random mutations, we developed a conservative strategy, which meant that the proportion of C > T mutations among the total number of TCW motifs should be two times higher than the proportion of C > T mutations among all other contexts (hereafter referred to as $Ratio^{C>T}$). Most lrHPV types had a $Ratio^{C>T} \geq 2$, with the exception of HPV54, HPV81 and HPV90 (Figure 3A). In contrast, half of the hrHPV types had a $Ratio^{C>T} < 2$, including the highly prevalent HPV18. Importantly, the $Ratio^{C>T}$ was significantly higher for lrHPV types than for hrHPV types in our data set ($p = 0.005$, Mann-Whitney U test; Figure 3B). To support this observation, we performed a similar analysis using a large collection of sequences retrieved from the GenBank database, including the most prevalent lrHPVs (HPV6 and HPV11, n = 375) and hrHPVs (HPV16 and HPV18, n = 777). The analysis was repeated in 100 replicates, where each replicate corresponded to a random draw of 200 equally distributed lrHPV 6 and 11 and hrHPV 16 and 18 sequences. We observed that HPV6 and HPV11 statistically accumulated more C > T mutations than HPV16 in 69 and 47 of the replicates, respectively. The remaining replicates showed no difference in C > T mutation accumulation between these HPV types. Compared to HPV18, both HPV6 and HPV11 had statistically more C > T mutations in all replicates. Taken together, the results support a higher accumulation of APOBEC3-induced mutations in lrHPVs than in hrHPVs.

Finally, we evaluated the distribution of these APOBEC3-induced mutations among the viral genes. We found that APOBEC3-induced mutations were heterogeneously distributed for both lrHPV and hrHPV types ($p < 0.001$; Kruskal-Wallis H test), and that most mutations were found in the E1 and L2 genes. By merging HPV6, HPV11, HPV16, and HPV18 sequences from our study and those from GenBank, we found that all four HPVs contained similar proportions of APOBEC3-induced mutations in the E1 gene and to a lesser extent in the L2 gene, as assessed by the $Ratio^{C>T}$. However, the $Ratio^{C>T}$ revealed a high number of APOBEC3-induced mutations in the E6 (72.6% of sequences) and E4 (91.9%) genes in lrHPV types 6 and 11, respectively (Figure 3C). Such mutations were almost absent in the same genes for hrHPV types 16 and 18. Rather, HPV16 had a $Ratio^{C>T} \geq 2$ mainly in the L1 gene (21.6% of sequences), whereas this ratio was almost systematically <2 in lrHPV types 6 and 11 (0.0% and 3.0% of sequences, respectively; Figure 3C). Altogether, these results revealed differences in the number and distribution of APOBEC3-induced mutations between viral genes and HPV risk categories.

3 | DISCUSSION

In this study, we compared HPV sequencing data from two anatomical sites, anus and cervix, collected simultaneously in a population of 30 FSWs from Togo, Africa, in 2017. We showed a high genetic diversity among HPV types and different levels of APOBEC3-induced mutations in the viral genome between hrHPVs and lrHPVs.

Consistent with the fact that anal smears often show less cellularity and DNA quantity after extraction than cervical smears, sequencing of HPVs from anal samples was less successful than from cervical samples. Several studies were carried out about HPV detection and typing in paired anal and cervical samples from the same patients,[26–32] but to our knowledge, none have performed whole viral genome sequencing of both samples. A comparison of sequences of similar HPV types found in paired samples from both anatomical sites revealed identical sequences. This is consistent with studies supporting autoinoculation of anal HPV infection from the cervix to the anus, with the number of different sexual partners being more related to anal infection than the practice of anal sex per se.[31,33]

By comparing the HPV sequences generated in this study, we found a similar number of nucleotide substitutions between HPV risk categories (although this varied between HPV types). The nucleotide substitutions were not randomly distributed along the HPV genome. In fact, the E4 viral gene was the least conserved, consistent with previous studies reporting that this gene is highly divergent between HPV types.[34] More interestingly, we observed different patterns of nucleotide conservation that are likely to depend on the HPV risk category. The most striking difference was the extreme conservation of the E6 gene in hrHPV types compared to lrHPV types. It can be assumed that the genetic conservation of E6 in hrHPVs is directly related to its carcinogenic potential, as previously reported for the other HPV oncoprotein E7.[35] The overexpression and activity of E6 and E7 viral proteins are implicated in the development and establishment of cervical cancer.[36] Among many oncogenic molecular mechanisms, these oncoproteins contribute to cell cycle re-entry through the interaction of E7 with the retinoblastoma protein, and also to generate genome instability through the ability of E6 to induce degradation of the tumor suppressor protein p53 via the ubiquitin pathway and the proteasome.[37] HPV oncoproteins are

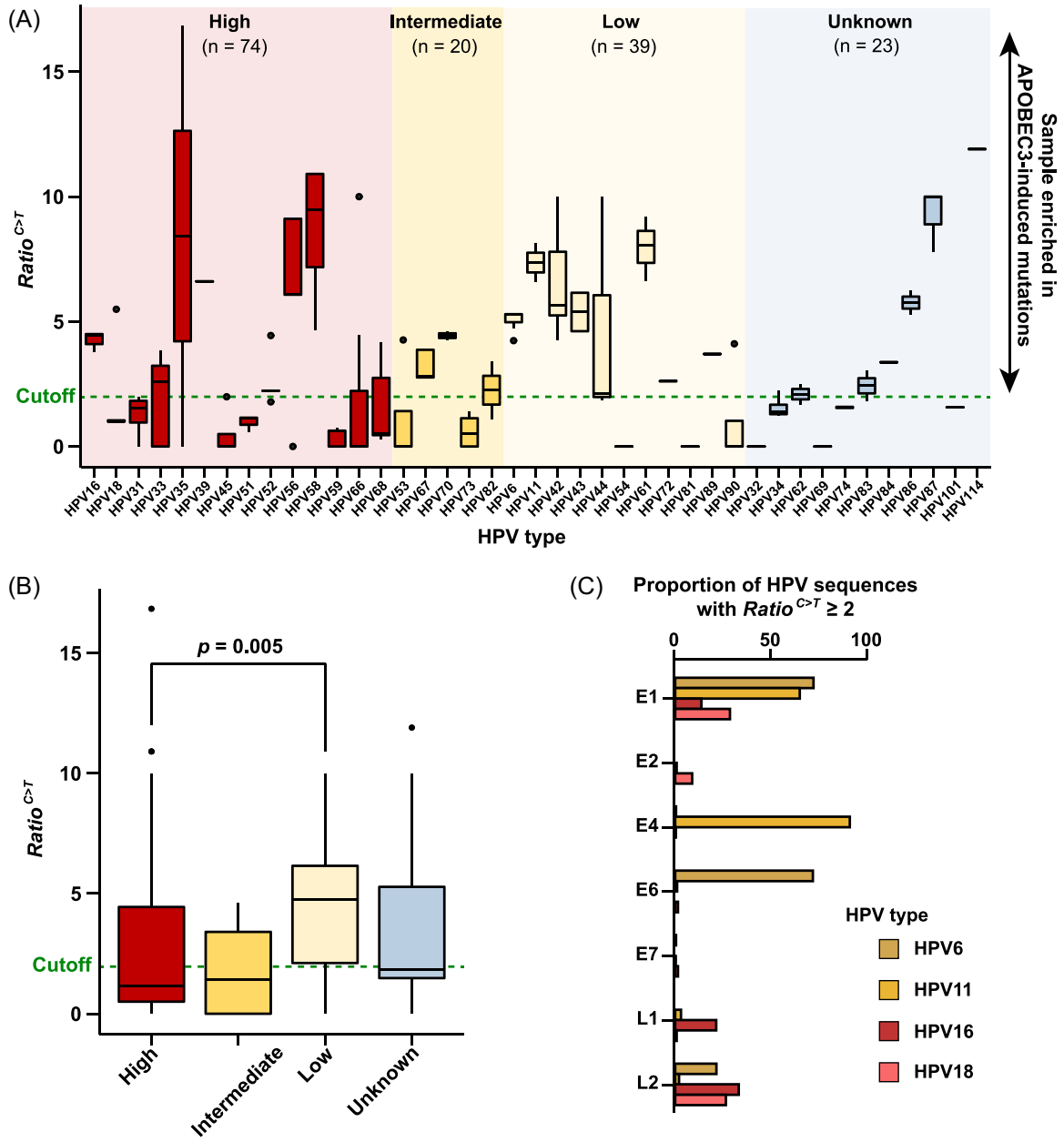

FIGURE 3 Variation in APOBEC3-induced mutations across the genome of HPV types. A. $Ratio^{C>T}$ for each HPV type. The $Ratio^{C>T}$ corresponds to the difference between the proportion of C > T mutations among the total number of TCW motifs and the proportion of C > T mutations among all other nucleotide contexts. (A) $Ratio^{C>T} \geq 2$ indicates that the HPV genome is enriched in APOBEC3-induced mutations (cutoff). Each color corresponds to an HPV risk category. The number of sequences included for each risk category is indicated above the plot. (B) $Ratio^{C>T}$ according to HPV risk category. A comparison between hrHPVs and lrHPVs was performed using the Mann-Whitney *U* test. (C) Proportion of HPV sequences with a $Ratio^{C>T} \geq 2$ for each viral gene using sequences from this study and the GenBank database for lrHPV types 6 and 11 and hrHPV types 16 and 18. Each color corresponds to an HPV type.

responsible for the development of multiple strategies to evade immune surveillance and support viral persistence.[38–40]

Innate immunity may be important in explaining HPV clearance.[41] Expression of human APOBEC3 is upregulated following HPV infection and mediates mutations in both the host and viral genomes.[42] Reported APOBEC3-induced mutation rates in HPV genomes are lower than those observed for other viruses restricted by this family of enzymes, such as HIV or hepatitis B virus (HBV).[43] This may be due to the high replication fidelity with which HPVs hijack the host DNA repair machinery. A current hypothesis is that a low rate of APOBEC3 editing of HPV genomes may be less effective in in vivo viral restriction, but may contribute to the generation of mutations that modulate the carcinogenic potential of subsequent viruses. Indeed, cross-sectional studies sequencing HPV in different grades of cervical intraepithelial neoplasia (CIN) lesions and cervical carcinoma biopsies reported a higher proportion of APOBEC3-induced mutations in

low-grade lesions than in higher-grade lesions.[23,44] Here we observed different APOBEC3 editing patterns even between HPVs of the same risk category, consistent with a previous whole HPV genome study reporting APOBEC3-induced mutations for HPV16 but none for HPV52 and HPV58.[45] Similarly, it is well established that hrHPVs do not share the same carcinogenic potential, even within the high-risk category.[46,47]

A large study was performed by Zhu et al. (2020) on HPV16 APOBEC3 editing in smears from women with different outcomes (from HPV clearance to cancer diagnosis).[23] The authors found that intracellular APOBEC3 mutation rates were higher in women whose lesions regressed or did not progress beyond CIN1 than in those who subsequently progressed from CIN2 to cancerous lesions. This is consistent with the hypothesis that APOBEC3 editing is involved in HPV viral restriction. Despite obvious differences in viral pathogenesis between lrHPV and hrHPV types, they share similar molecular pathways to achieve viral genome maintenance.[48] Therefore, the different proportion of APOBEC3-induced mutations in the viral genome between lrHPVs and hrHPVs reported in this study may be part of a distinct immune response involved in their different pathogenic potential.

In addition to the different frequencies of APOBEC3-induced mutations between hrHPV and lrHPV types, the mutations were not evenly distributed along the viral genome. For most HPV types, we found that APOBEC3-induced mutations, when present, were mostly located in the E1 and L2 genes. E1 is reported to play a role in viral replication within the host, while the minor capsid protein L2 plays an important role in the nuclear localization of viral components, DNA binding, and capsid formation. The role of APOBEC3-induced mutations in these genes remains to be determined, but they could interfere with viral replication or cell entry and/or facilitate viral immune escape. A striking difference we observed between hrHPV types 16 and 18 and lrHPV types 6 and 11 was the high frequency of APOBEC3-induced mutations in the E6 and E4 genes, found only in lrHPV types 6 and 11. As E6 expression is important for the establishment of a malignant phenotype, it can be suggested that there may be a threshold of APOBEC3-induced mutations above which these mutations play a role in viral restriction, while a lower frequency of such mutations would favor viral immune escape, like the hypothesis of sublethal APOBEC3 editing of HIV.[49]

Our study has several limitations. First, the women enrolled were rather young and did not have high-grade cervical lesions, which did not allow analysis according to cytology results. Second, the number of samples – and therefore the number of HPV sequences generated per type – was low, which may limit interpretations and prevent the extrapolation of our results to other populations, except for the results obtain on larger database as GenBank. Third, too few FSWs living with HIV were included in this study to allow analysis of the APOBEC-3 signature in relation to HIV status. And fourth, the depth of coverage for the different samples was limited, preventing the search for minor APOBEC3-induced mutations within the host. A study of APOBEC3 editing on the HPV genome in larger populations and from different geographical areas is needed to stratify the analysis for different HPV types and different grades of lesions to cancer, as has been done specifically for HPV16.[23]

In conclusion, this study reports heterogeneity in the location of nucleotide substitutions and APOBEC3-induced mutations among HPV types along the viral genome. The different proportions of APOBEC3 mutation signatures between hrHPVs and lrHPVs may contribute to their different oncogenic potentials.

4 | MATERIAL AND METHODS

4.1 | Ethical statement

This study was carried out according to the principles of the Declaration of Helsinki. Ethical approvals were obtained from the National Bioethics Committee for Health Research in Togo (No. 19/2017/CBRS) on June 22, 2017. The recruitment method and inclusion criteria are described in Ferré et al. (2019).[26]

4.2 | Study population and sample collection

Anal and cervical swabs were collected from a previously described population-based cross-sectional study conducted in 2017 among FSWs in Togo, Africa.[26] Samples with at least one common hrHPV between cervical and anal swabs, as assessed by the Anyplex™ II HPV28 Detection Assay (Seegene), were selected for this study.

4.3 | Capture preparation and HPV whole genome sequencing

A double-capture HPV method followed by next-generation sequencing, first described by Holmes et al.[50] using over 22,000 probes targeting more than 235 HPV types, was used to sequence the whole viral genome (Supplementary Method S1). DNA sequences captured by biotin-streptavidin interaction were then amplified by PCR and sequenced using an Illumina MiSeq system with 150 paired-end reads. Raw sequence data were subjected to standard Illumina quality control procedures and then aligned to HPV reference genomes obtained from the PAVE database (https://pave.niaid.nih.gov/)[51] (Supplementary Table S5) using BWA-mem (Burrows-Wheeler Aligner).[52] Aligned reads were sorted and indexed using SAMtools v.1.4.[53] HPV consensus sequences (covering at least 70% of the genome) were inferred by generating a pileup file using the *mpileup* function of SAMtools, then an R script was developed to define the per-nucleotide consensus and incorporate indels. Phylogenetic trees of the sequences, including the reference genomes, were constructed to identify types and, for HPV16, sublineages (Supplementary Method S2).

4.4 | Genetic diversity of HPVs

From the HPV consensus sequences, we first checked whether the genome sequences of cervical and anal swabs from the same patient and

for the same HPV type were identical using pairwise sequence alignments using Clustal Omega (default parameters).[54] The level of genetic diversity for the same HPV type from different patients was assessed using an in-house pipeline. Briefly, we counted the number of nucleotide substitutions for each generated sequence compared to the reference genome of the same type and then measured the standard deviation for each HPV type. The number of nucleotide substitutions between HPV risk categories (low, intermediate, high and unknown) was also assessed. The classification of HPV types is shown in Supplementary Table S6.

APOBEC3-induced mutations were identified as C > T mutations in the TCW nucleotide context (where W is either A or T).[17] We developed a pipeline that *i*) takes a pileup file as input; *ii*) compares the HPV consensus sequence to that of the reference HPV type; *iii*) counts the number of TCW motifs and C nucleotides along the reference genome; and *iv*) counts the number of C > T mutations within and outside the TCW motif along the sample genome, as described by Zhu et al.[23] We also counted the number of missing bases, particularly in the TCW context, to avoid underestimating APOBEC3-induced mutations in an incomplete HPV genome. To consider that C > T mutations were likely attributed to APOBEC3 activity rather than random mutations, we assumed that the ratio of C > T mutations to the total number of TCW motifs was two times higher than the ratio of C > T mutations among all other contexts ($Ratio^{C>T}$). Applying the strategy using either the reference sequences from the PAVE database or calculating the ancestral sequences led to similar results (Supplementary Method S3).

4.5 | Measurement of APOBEC3-induced mutations in the HPV genome from GenBank collections

The proportion of C > T mutations was assessed in large collections of HPV genomes obtained from the GenBank database. For this analysis, we focused on the most prevalent lrHPVs (HPV6 and HPV11) and hrHPVs (HPV16 and HPV18). Sequences were obtained using the blastn search (default parameters)[55] with the reference genomes as query entries. Only sequences with an unambiguous description were included. In total, 276, 99, 631, and 146 whole genome sequences were obtained for HPV types 6, 11, 16, and 18, respectively (Supplementary Table S7). For each HPV type, genome sequences were aligned using Clustal Omega (default parameters),[54] followed by manual curation. C > T mutations were then counted for each sequence compared to the reference genome sequence using an in-house R script. To avoid any bias, we performed a series of 100 replicates, where each replicate corresponded to a random draw of 200 equally distributed lrHPVs 6 and 11 and hrHPVs 16 and 18 sequences.

4.6 | Statistical analyses

The number of positive HPVs between cervical and anal swabs was evaluated using the chi-squared test with Yates' correction for continuity. Comparisons in the number of nucleotide substitutions between lrHPVs and hrHPVs were assessed with the Mann-Whitney *U* test. Variation in the number of mutations in the different HPV genes was assessed by Kruskal-Wallis one-way analysis of variance. *P*-values less than 0.05 were considered statistically significant. All statistical analyses and graphs were performed using R software version 4.1.1.

AUTHOR CONTRIBUTIONS

All authors were involved in conceptualization of the study. Investigation: FAGK, MS, ACD, and DKE recruited the participants and collected the samples and data; VMF, SV, MB, SL, AC, IB, and CC were involved in technical development and sample analyses. VMF, RC, SV, SL, IV, and CC curated the data and performed formal analysis (did the primary sequence analysis and bioinformatics). VMF, RC, and CC drafted the first version of the paper and were responsible for the decision to submit the manuscript. The corresponding author had access to all data at all times. All authors contributed to the revision and editing of the paper and approved the final version.

ACKNOWLEDGMENTS

We thank the Agence Nationale de recherche sur le SIDA et les hépatites virales (ANRS)-Maladies Infectieuses Emergentes (ANRS | MIE) for their support.

COMPETING INTERESTS

The authors declare that they have no competing interests with the current work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions. Deidentified data will be available for sharing after manuscript publication upon request to the corresponding author. Next-generation sequence files of the HPVs are accessible on the European Nucleotide Archive (ENA) under the accession numbers ERR9836278 to ERR9836337 (project: PRJEB53434). All consensus sequences generated in this study have been submitted to NCBI GenBank (accession number OP711959 to OP712113). Scripts developed for this study were deposited in the GitHub repository: https://github.com/Rcoppee/HPV_genetic_diversity.

ORCID

Valentine Marie Ferré http://orcid.org/0000-0002-2196-6141
Romain Coppée http://orcid.org/0000-0002-3024-5928
Jade Ghosn http://orcid.org/0000-0003-2914-959X

REFERENCES

1. Rodriguez AC, Schiffman M, Herrero R, et al. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. *J Natl Cancer Inst*. 2008;100:513-517.
2. Plummer M, Schiffman M, Castle PE, Maucort-Boulch D, Wheeler CM. A 2-Year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or Low-Grade squamous intraepithelial lesion. *J Infect Dis*. 2007;195:1582-1589.

3. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer*. 2017;141:664-670.
4. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941-1953.
5. Bray F, Parkin DM, Gnangnon F, et al. Cancer in sub-Saharan Africa in 2020: a review of current estimates of the national burden, data gaps, and future needs. *Lancet Oncol*. 2022;23:719-728.
6. IARC. Cervix Uteri, Globocan 2020. Available at: https://gco.iarc.fr/today/data/factsheets/cancers/23-Cervix-uteri-fact-sheet.pdf. Accessed 19 October 2022.
7. Cervical Cancer Survival Rates | Cancer 5 Year Survival Rates. Available at: https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/survival.html. Accessed 16 May 2022.
8. Muñoz N, Bosch FX, de Sanjosé S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003;348:518-527.
9. Bzhalava D, Eklund C, Dillner J. International standardization and classification of human papillomavirus types. *Virology*. 2015;476:341-344.
10. Bernard H-U, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer*. 2006;118:1071-1076.
11. Latsuzbaia A, Wienecke-Baldacchino A, Tapp J, et al. Characterization and diversity of 243 complete human papillomavirus genomes in cervical swabs using next generation sequencing. *Viruses*. 2020;12:1437.
12. Wakabayashi R, Nakahama Y, Nguyen V, Espinoza JL. The Host-Microbe interplay in human Papillomavirus-Induced carcinogenesis. *Microorganisms*. 2019;7:199.
13. Antaño-Arias R, Del Moral-Hernández O, Ortiz-Ortiz J, et al. E6/E7 variants of human papillomavirus 16 associated with cervical carcinoma in women in Southern Mexico. *Pathogens*. 2021;10:773.
14. Yao Y, Yan Z, Dai S, et al. Human papillomavirus type 16 E1 mutations associated with cervical cancer in a han Chinese population. *Int J Med Sci*. 2019;16:1042-1049.
15. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415-421.
16. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genet*. 2013;45:977-983.
17. Roberts SA, Lawrence MS, Klimczak LJ, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*. 2013;45(9):970-976. doi:10.1038/ng.2702
18. Vartanian J-P, Guétard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science*. 2008;320:230-233.
19. Warren CJ, Xu T, Guo K, et al. APOBEC3A functions as a restriction factor of human papillomavirus. *J Virol*. 2015;89:688-702.
20. Chen S, Li X, Qin J, et al. APOBEC3A possesses anticancer and antiviral effects by differential inhibition of HPV E6 and E7 expression on cervical cancer. *Int J Clin Exp Med*. 2015;8:10548-10557.
21. Warren C, Westrich J, Doorslaer K, Pyeon D. Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression. *Viruses*. 2017;9:233.
22. Vieira VC, Leonard B, White EA, et al. Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B. *mBio*. 2014;5:e02234-14.
23. Zhu B, Xiao Y, Yeager M, et al. Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance. *Nat Commun*. 2020;11:886.
24. Warren CJ, Santiago ML, Pyeon D. APOBEC3: friend or foe in human papillomavirus infection and oncogenesis? *Annual Review of Virology*. 2022;9:375-395.
25. ICO/IARC Information Centre on HPV and Cancer. Togo: human papillomavirus and related cancers, fact sheet. 2023. 2023. Accessed October 8, 2024. https://hpvcentre.net/statistics/reports/TGO_FS.pdf
26. Ferré VM, Ekouevi DK, Gbeasor-Komlanvi FA, et al. Prevalence of human papillomavirus, human immunodeficiency virus and other sexually transmitted infections among female sex workers in Togo: a national cross-sectional survey. *Clin Microbiol Infect*. 2019;25:1560.e1-1560.e7.
27. Lin C, Slama J, Gonzalez P, et al. Cervical determinants of anal HPV infection and high-grade anal lesions in women: a collaborative pooled analysis. *Lancet Infect Dis*. 2019;19:880-891.
28. Nasioutziki M, Chatzistamatiou K, Loufopoulos P-D, et al. Cervical, anal and oral HPV detection and HPV type concordance among women referred for colposcopy. *Infect Agent Cancer*. 2020;15:22.
29. Guler T, Uygur D, Uncu M, et al. Coexisting anal human papilloma virus infection in heterosexual women with cervical HPV infection. *Arch Gynecol Obstet*. 2013;288:667-672.
30. Thorsteinsson K, Storgaard M, Katzenstein TL, et al. Prevalence of cervical, oral, and anal human papillomavirus infection in women living with HIV in Denmark - the SHADE cohort study. *J Clin Virol*. 2018;105:64-71.
31. Hernandez BY, McDuffie K, Zhu X, et al. Anal human papillomavirus infection in women and its relationship with cervical infection. *Cancer Epidemiol Biomarkers Prevent*. 2005;14:2550-2556.
32. Morhason-Bello IO, Baisley K, Pavon MA, et al. Oral, genital and anal human papillomavirus infections among female sex workers in Ibadan, Nigeria. *PLoS One*. 2022;17:e0265269.
33. Kojic EM, Cu-Uvin S, Conley L, et al. Human papillomavirus infection and cytologic abnormalities of the anus and cervix among HIV-infected women in the study to understand the natural history of HIV/AIDS in the era of effective therapy (the SUN study). *Sex Transm Dis*. 2011;38:253-259.
34. Doorbar J. The E4 protein; structure, function and patterns of expression. *Virology*. 2013;445:80-98.
35. Mirabello L, Yeager M, Yu K, et al. HPV16 E7 genetic conservation is critical to carcinogenesis. *Cell*. 2017;170:1164-1174.e6.
36. Schwarz E, Freese UK, Gissmann L, et al. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature*. 1985;314:111-114.
37. Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci*. 2006;110:525-541.
38. Tummers B, Van Der Burg S. High-Risk human papillomavirus targets crossroads in immune signaling. *Viruses*. 2015;7:2485-2506.
39. Trujillo-Cirilo L, Torres-Corioriles EI, Rangel-Corona R, Corona-Ortega MT, Weiss-Steider B. Evidence that the viral oncoproteins E6 and E7 of HPV induce the expression of a functional IL-2R on cervical cancer cells. *Cytokine*. 2021;148:155592.
40. Chiang C, Pauli E-K, Biryukov J, et al. The human papillomavirus E6 oncoprotein targets USP15 and TRIM25 to suppress RIG-I-Mediated innate immune signaling. *J Virol*. 2018;92:e01737-17.
41. Amador-Molina A, Hernández-Valencia J, Lamoyi E, Contreras-Paredes A, Lizano M. Role of innate immunity against human papillomavirus (HPV) infections and effect of adjuvants in promoting specific immune response. *Viruses*. 2013;5:2624-2642.
42. Smith NJ, Fenton TR. The APOBEC3 genes and their role in cancer: insights from human papillomavirus. *J Mol Endocrinol*. 2019;62:R269-R287.
43. Wakae K, Aoyama S, Wang Z, et al. Detection of hypermutated human papillomavirus type 16 genome by next-generation sequencing. *Virology*. 2015;485:460-466.
44. Hirose Y, Onuki M, Tenjimbayashi Y, et al. Within-Host variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome. *J Virol*. 2018;92:e00017-e00018.

45. Mariaggi A-A, Péré H, Perrier M, et al. Presence of human papillomavirus (HPV) apolipoprotein B messenger RNA editing, catalytic Polypeptide-Like 3 (APOBEC)-related minority variants in HPV-16 genomes from anal and cervical samples but not in HPV-52 and HPV-58. *J Infect Dis*. 2018;218:1027-1036.
46. Clifford GM, Tully S, Franceschi S. Carcinogenicity of human papillomavirus (HPV) types in HIV-Positive women: a meta-analysis from HPV infection to cervical cancer. *Clin Infect Dis*. 2017;64:1228-1235.
47. Guan P, Howell-Jones R, Li N, et al. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *Int J Cancer*. 2012;131:2349-2359.
48. Murakami I, Egawa N, Griffin H, et al. Roles for E1-independent replication and E6-mediated p53 degradation during low-risk and high-risk human papillomavirus genome maintenance. *PLoS Pathog*. 2019;15:e1007755.
49. Venkatesan S, Rosenthal R, Kanu N, et al. Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann Oncol*. 2018;29:563-572.
50. Holmes A, Lameiras S, Jeannot E, et al. Mechanistic signatures of HPV insertions in cervical carcinomas. *NPJ Genom Med*. 2016;1:16004.
51. Van Doorslaer K, Li Z, Xirasagar S, et al. The papillomavirus episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res* 2017;45:D499-D506.
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
53. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10:giab008.
54. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ferré VM, Coppée R, Gbeasor-Komlanvi FA, et al. Viral whole genome sequencing reveals high variations in APOBEC3 editing between HPV risk categories. *J Med Virol*. 2024;96:e70002. doi:10.1002/jmv.70002