14 (2022) 200247

Contents lists available at ScienceDirect

Tumour Virus Research

journal homepage: www.journals.elsevier.com/tumour-virus-research

Differences in integration frequencies and APOBEC3 profiles of five high-risk HPV types adheres to phylogeny

Alexander Hesselberg Løvestad [a], Adina Repesa [a], Jean-Marc Costanzi [b], Sonja Lagström [b,c,d], Irene Kraus Christiansen [b,f], Trine B. Rounge [c,e,**], Ole Herman Ambur [a,*]

[a] Department of Life Sciences and Health, Faculty of Health Sciences, OsloMet - Oslo Metropolitan University, Oslo, Norway
[b] Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway
[c] Department of Research, Cancer Registry of Norway, Oslo, Norway
[d] Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[e] Centre for Bioinformatics, Department of Pharmacy, University of Oslo, Oslo, Norway
[f] Department of Clinical Molecular Biology (EpiGen), Division of Medicine, Akershus University Hospital and University of Oslo, Lørenskog, Norway

ARTICLE INFO

Keywords:
Human papillomavirus
HPV16
HPV18
HPV31
HPV33
HPV45
Minor nucleotide variation
Chromosomal integration
APOBEC3
Alpha-7
Alpha-9
Cancer-related genes

ABSTRACT

Persistent infection with Human Papillomavirus (HPV) is responsible for almost all cases of cervical cancers, and HPV16 and HPV18 associated with the majority of these. These types differ in the proportion of viral minor nucleotide variants (MNVs) caused by APOBEC3 mutagenesis as well as integration frequencies. Whether these traits extend to other types remains uncertain. This study aimed to investigate and compare genomic variability and chromosomal integration in the two phylogenetically distinct Alpha-7 and Alpha-9 clades of carcinogenic HPV types. The TaME-seq protocol was employed to sequence cervical cell samples positive for HPV31, HPV33 or HPV45 and combine these with data from a previous study on HPV16 and HPV18. APOBEC3 mutation signatures were found in Alpha-9 (HPV16/31/33) but not in Alpha-7 (HPV18/45). HPV45 had significantly more MNVs compared to the other types. Alpha-7 had higher integration frequency compared to Alpha-9. An increase in integration frequency with increased diagnostic severity was found for Alpha-7. The results highlight important differences and broaden our understanding of the molecular mechanisms behind cervical cancer induced by high-risk HPV types from the Alpha-7 and Alpha-9 clades.

1. Introduction

Human papillomaviruses (HPVs) are a group of small, double-stranded DNA viruses with a genome size of ~7.9 kb that contains eight genes. The genome consists of early region genes (E1, E2, E4-E7), late region genes (L1, L2), and two non-coding regions, the upstream regulatory region (URR) and the non-coding region (NCR) [1]. Of the early region genes, E5, E6, and E7 encode oncoproteins that promote the transformation of the host cell through induction of cell proliferation and inactivation of cell cycle regulatory and tumour-suppressor mechanisms [2,3]. To date there are over 200 characterised HPVs [4], commonly distinguished by at least 10% nucleotide differences in the L1 gene [5,6] and further divided into lineages (1 > 10% whole-genome nucleotide differences) and sublineages (0,5 > 1% difference) [7–9]. There are at least 12 HPV types that are carcinogenic (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) [10]. Persistent infection with one of these is considered a necessary cause for cervical cancer development [11]. Still, only a minority of persistent infections progress to cancer [12], indicating that additional factors are necessary for cancer progression. All the oncogenic HPV types belong to the genus *Alphapapillomavirus* (Alpha-PV) where they cluster within the species-level clades Alpha-5, Alpha-6, Alpha-7, Alpha-9 and Alpha-11 [13]. The carcinogenic HPVs of the different clades exhibit differences in carcinogenicity and tissue tropism, among other characteristics, suggesting different evolutionary histories that have given rise to their carcinogenic potential as well as differences in the molecular mechanisms behind HPV-induced cancers.

HPVs are considered slowly evolving viruses [14]. Recent studies have uncovered nucleotide variation below the consensus level in HPV genomes present within an infected person [15–20]. HPV genomic

* Corresponding authors.
** Corresponding author. Department of Research, Cancer Registry of Norway, Oslo, Norway.
E-mail address: olam@oslomet.no (O.H. Ambur).

https://doi.org/10.1016/j.tvr.2022.200247
Received 1 July 2022; Received in revised form 5 September 2022; Accepted 6 September 2022
Available online 11 September 2022

variation can have consequences for the infection outcome, and HPV16 sublineages together with the host ethnicity/genetic background has been shown to be associated with different risks of developing cervical cancer [21]. Intra-host HPV nucleotide variation is not uniformly distributed across the genome, as has been revealed in HPV16 cervical cancer cases where the E7 gene has been shown to have few non-synonymous mutations compared to the rest of the genome [16]. Additionally, persistent infections that progress to high-grade lesions or cancer are associated with less intra-host variation relative to infections that are cleared by the immune system [17,20,22].

The mutagenic processes behind HPV intra-host nucleotide variation are currently not fully understood, although it is clear that members of the gene family anti-viral host-defence enzyme apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3 (APOBEC3) contributes to the variation [17,19,20]. APOBEC3 are cytidine deaminases activated in response to viral infections and induces C > T substitutions in the trinucleotide context TCN, where N is any nucleotide (with the exception of APOBEC3G which has a preferred CCN target motif) [23]. APOBEC3-induced mutations can inhibit viral replication and generally lower viral fitness [24]. Accordingly, the APOBEC3 mutation signature is more often observed in HPV genomes from transient infections and pre-cancerous lesions compared to cancer samples and is associated with viral clearance [17,20]. Studies have also reported frequent T > C substitutions [17,19,20]; however, the mutagenic process behind this transition and its role in infection outcome is currently not understood. Another possible source for HPV intra-host nucleotide variation might be the recruitment of low-fidelity polymerases during the replication stage of the viral life cycle [25–27].

A well-studied molecular event related to HPV-induced carcinogenesis is the full or partial integration of the HPV genome into human chromosomes [28]. This is a common genomic event observed in more than 80% of HPV positive tumours and is considered a driving event in cervical carcinogenesis [29,30]. Integrations involving a deletion or disruption of E1 or E2 will lead to overexpression of oncogenes E6 and E7 [31,32], which in turn can lead to an accumulation of mutations and unregulated clonal cell division with a selective growth advantage [28]. Furthermore, integrations can also promote genomic instability in an E6/E7-independent manner by integrating within, or in close proximity to, host oncogenes or tumour-suppressor genes to functionally knock them out or affect their expression levels [33,34]. HPV integrations are associated with local altered genomic landscapes and changes in host gene expression in their vicinity, which might promote genomic instability and carcinogenesis depending on the integration site [35–39]. Integration "hot spots" have been observed repeatedly in high-grade lesions and tumours, indicating that integration in certain chromosomal loci might confer selective growth advantages and increase the risk of developing HPV-induced cancers [33,40,41].

Most studies on molecular mechanisms behind HPV-induced cancers have been conducted on HPV16 and HPV18 genotypes due to their high prevalence and carcinogenic potential [42]. By comparison, the remaining carcinogenic HPV-types are understudied. HPV16 and HPV18 have shown differences in integration frequencies and APOBEC3 interaction, suggesting that HPV-induced cancer development follow dissimilar type-dependent routes [43–45]. Within the Alpha-PVs, HPV16 sorts under Alpha-9 together with HPV31 and HPV33, while HPV18 and HPV45 sort under Alpha-7. It has been shown in previous studies that HPV45 has a high integration frequency (IF) like HPV18, reflecting similarities between these evolutionary closely related HPV types [45].

In this study, we aim to investigate genomic variability and chromosomal integration in cervical cell samples with different morphologies positive for HPV31, HPV33, and HPV45 utilizing the TaME-seq protocol [46]. Additionally, this study will include comparisons to reanalysed HPV16 and HPV18 data from a previous study [43] to gain a more comprehensive understanding of specific characteristics of the distinct clades Alpha-7 (HPV18 and HPV45) and Alpha-9 (HPV16, HPV31 and HPV33). A study going deeper into the nature of genomic events in these lesser studied carcinogenic HPV types allows for a phylogenetic approach to better understand the molecular mechanisms of host-responses to infections and those responsible for HPV-induced carcinogenesis.

2. Materials and methods

2.1. Sample selection

Cervical cell samples were collected from women attending the Norwegian cervical cancer screening program between January 2005 and April 2008. Recruitment criteria, HPV detection, and genotyping have been described previously [47,48]. In total, 156 HPV16, 75 HPV18, 117 HPV31, 104 HPV33, and 66 HPV45 samples were categorized based on the HPV type and diagnostic category. The diagnostic categories were defined as "non-progressive", cervical interepithelial neoplasia grade 2 (CIN2) and CIN3+ (Table 1). The non-progressive category consisted of samples with normal cytology (normal cytology the preceding two years and with no previous history of treatment for cervical neoplasia) or samples with atypical squamous cells of undetermined significance (ASC-US) or low-grade squamous intraepithelial lesions (LSIL) with no follow-up diagnosis within four years. The CIN3+ category consisted of samples with CIN3/adenocarcinoma in situ (AIS) and cancer.

2.2. Sample preparation and DNA extraction

Cervical cell samples had previously been collected in ThinPrep PreservCyt solution (Hologic, Marlborough, MA) and pelleted before storage −80 °C to retain DNA quality and integrity. Collected samples were stored as both cell material and extracted DNA in a research biobank at Akershus University Hospital. DNA from some samples had to be re-extracted from cell material for this study, and an easyMAG® (Biomérieux, USA) was used for the extraction and the eluate stored in a biobank at −80 °C. The DNA concentration was measured on Qubit® 3.0 Fluorometer (Life Technologies, USA) to ensure optimal DNA quantity in every sample before the PCR reaction.

2.3. Library preparation and sequencing

Library preparation was done according to the TaME-seq protocol as described previously [46]. To summarise, the samples were tagmented using Nextera DNA library prep kit (Illumina, Inc., San Diego, CA). Tagmented DNA underwent target enrichment by multiplex PCR using respective HPV31, 33 and 45 type-specific HPV primers and a combination of i7 index primers and i5 index primers [49] from the Nextera index kit (Illumina, Inc., San Diego, CA). Each sample underwent separate PCR amplifications for the forward and reverse reactions. Sequencing was performed on the Illumina HiSeq2500 platform using 125 bp paired-end reads.

2.4. Sequence alignment

Data were analysed by an in-house bioinformatics pipeline as described previously [46], with some slight changes to the reference genomes and variant calling. The pipeline can be accessed here: https://github.com/jean-marc-costanzi/TaME-seq/. Briefly, reads were mapped to the human genome (GRCh38/hg38) using HISAT2 (v2.1.0)[50]. Reference genomes for HPV16, HPV18, HPV31, HPV33, and HPV45 were obtained from the PaVE database [51] and 1 kb overhangs were added to account for the circular HPV genome. BCFtools was used to calculate mapping statistics and coverage. Samples with a mean sequencing depth of <300x were excluded from the analysis.

Table 1
Number of samples sequenced and analysed and mean mapping statistics for each diagnostic category of HPV16, HPV18, HPV31, HPV33 and HPV45 infections.

Diagnostic category	Sequenced samples	Analysed samples	Mean numbers in the analysed samples			
			Raw reads	Reads mapped to target HPV	Mean HPV genome coverage	Fraction of HPV genome covered by min. 100x
HPV16						
Non-progressive	55	21	1.3 M	1.3 M	11148	0.77
CIN2	46	25	0.6 M	0.5 M	4462	0.70
CIN3+	55	31	1.3 M	1.2 M	9483	0.74
HPV18						
Non-progressive	16	12	39 M	26 M	48129	0.86
CIN2	13	9	77 M	40 M	55097	0.86
CIN3+	46	28	24 M	13 M	29138	0.82
HPV31						
Non-progressive	18	10	10 M	6 M	26508	0.87
CIN2	22	20	14 M	5 M	23695	0.89
CIN3+	77	54	9.3 M	5.7 M	26657	0.88
HPV33						
Non-progressive	12	9	16 M	10 M	24350	0.90
CIN2	15	9	20 M	10 M	30699	0.95
CIN3+	77	70	11 M	7 M	38731	0.97
HPV45						
Non-progressive	25	21	22 M	8 M	21593	0.83
CIN2	14	12	39 M	19 M	27608	0.92
CIN3+	27	23	32 M	13 M	32066	0.85

2.5. Sequence variation analysis

Nucleotide counts mapped to the HPV reference genomes were retrieved from the HISAT sequence alignment and average nucleotide mapping quality values were retrieved from the BCFtools mpileup ouput using an in-house R (v3.5.1) script as described in [46]. Briefly, for a variant to be called it had to be present in more than two reads in a position with ≥100x depth, have a Phred quality score ≥30 and a frequency ≥1%. In addition, the variant calling of minor nucleotide variants (MNVs) was done in a reference-independent manner where the most frequent base in each position was termed the major variant followed by the second most frequent as the MNV. The MNVs had to be present in both the independently amplified F and R reactions, unless where there was discordance between the F and R reactions – then the MNV with the highest coverage was called. HPV NCR have regions of homopolymeric T tracts (HPV16:4156–4173 and 4183–4212, HPV18:4198–4234, HPV31:4072–4077 and 4145–4167, HPV33: 4149–4167 and 4186–4195, HPV45:4184–4219), which can cause polymerase or sequencing errors at high frequencies and were therefore filtered out during the variation analysis.

The ratio of non-synonymous to synonymous substitutions (dN/dS) was calculated to indicate whether genes in the different diagnostic categories were more or less prone to amino acid changes.

For mutational signature analysis, all nucleotide substitutions were classified into six base substitutions, C > A, C > G, C > T, T > A, T > C, and T > G, and further into 96 trinucleotide substitution types that include information on the bases immediately 5′ and 3’ of the substituted base. An extended mutation signature analysis was also done to investigate mutations in the APOBEC3A-favoured genomic context YTCA and APOBEC3B-favoured genomic context RTCA.

To investigate if the number of APOBEC3 target sequences differed between HPV types, FUZZNUC from the EMBOSS package (http://emboss.toulouse.inra.fr/cgi-bin/emboss/fuzznuc) was employed using reference genomes for HPV16, HPV18, HPV31, HPV33, and HPV45 obtained from the PaVE database. Both strands were investigated, and the search patterns used were TCA, YTCA, RTCA and NCN.

To calculate the proportion of TCA motifs out of all available NCN motifs, the number of TCA motifs were divided by the number of NCN motifs for each HPV type, and this proportion was treated as the expected proportion of C > T substitutions in the TCA motifs. To calculate the difference of observed vs expected proportions, the following formula was used (Observed proportions/Expected proportions)^2/Expected proportions. These values were used in a Wilcoxon rank sum test to investigate whether the difference in observed/expected proportions between the clades differed significantly. Variation and dN/dS analyses were performed using an in-house R (v3.5.1) script.

2.6. Detection of chromosomal integration sites and validation by sanger sequencing

Integration site detection was performed as described previously [46]. Briefly, a two-step analysis was employed. First, read pairs with one of the reads mapping to the human genome and the other to HPV were identified using HISAT2. Second, unmapped reads were re-mapped using LAST (v876) aligner (options -M -C2) [52] to increase detection of human-HPV read pairs. Reads sharing identical start and end coordinates were considered likely PCR duplicates and excluded from the analysis.

Validation of integration sites for HPV16 and HPV18 is previously described [43]. The Illumina reads from the respective HPV31, HPV33 and HPV45 sequencing reactions were used to make *in silico* DNA templates for design of integration-targeting primers suited for PCR and Sanger sequencing. Hybrid sequences containing human and HPV-specific sequences spanning the reported integration breakpoint, were used as templates. Primer3 [53] was used to create optimal primer pairs that included a human-specific forward primer and an HPV-specific reverse primer. Phusion™ Master Mix (Thermo Scientific, USA) was used to prepare the PCR reaction mix. The PCR conditions were as follows: initial denaturation at 98 °C for 30 s; 30 cycles at 98 °C for 10 s, at 60 °C for 30 s and 72 °C for 15 s; final extension at 72 °C for 10 min.

Samples were sequenced on the ABI® 3130xl/3100 Genetic Analyzer 16-Capillary Array (Thermo Fisher Scientific Inc., Waltham, MA) using BigDye™ Terminator v1.1 cycle sequencing kit (Thermo Fisher Scientific Inc., Waltham, MA). Sanger sequencing data was further processed in Geneious Prime (v2020.2.2) and if the sequence was homologous to the same chromosomal locus and HPV type as reported, the HPV integration was considered confirmed. Samples with inconclusive Sanger sequencing results that showed several unspecific bands on the agarose gel were re-run using a touchdown PCR with an additional 6 extra cycles. If the samples still had unspecific bands, individual bands were cut out from the agarose gel and extracted using Wizard® SV Gel and PCR Clean-Up System kit (Promega, USA) following the manufacturer’s instructions before Sanger sequencing.

2.7. Determining microhomology regions

BLASTn and/or BLAT were used to identify short homologous sequences at the integration breakpoint in the Sanger-confirmed HPV integrations. If > 3 nt overlapping sequences were present between the human and HPV genome, it was designated a microhomology sequence. The overlapping bases were identified using the Geneious Prime genome browser after the assembly of Sanger reads.

2.8. Functional annotation of genes within 10 kb of reported integration sites

All genes 10 kb upstream or downstream of the reported integration site were identified by visual inspection in Geneious Prime and their molecular function annotated using Genecards (https://www.genecards.org). Genes were classified as cancer-related genes (CRGs) if they were involved in cell cycle regulation, apoptosis, tumour suppressor mechanisms, cancer-related pathways, genes interacting with these pathways, or if a cancer-related SNP association was assigned.

2.9. Statistical methods

Non-parametric Chi-square of independence was used to determine whether there was a significantly ($p < 0.05$) higher number of breakpoints in the E1, E2, E4 and L2 genes for Alpha-7 and in E1, L2 and NCR in Alpha-9 than would be expected by chance.

To understand the relationship between the dependent variables (MNVs, samples with integrations, integration breakpoints) and the independent variables (HPV type or diagnostic category), a generalized linear model (glm) was used. The glm used a negative binomial distribution for the number of integrations and MNV model and binomial distribution for the other models. Following this, multiple comparisons of means using Tukey HSD was done using the R package multcomp [54] to test the differences between the categories. To test the differences in APOBEC3 signature mutations and clades, a Wilcoxon rank sum test were used. All statistical tests were done in R (v3.6.3). The output of the tests can be found in the supplementary material D.

Ethical approval

This study was approved by the regional committee for medical and health research ethics, Oslo, Norway (REK 2017/447).

3. Results

3.1. Characteristics and sequencing statistics

In total, 518 HPV16, HPV18, HPV31, HPV33 and HPV45 positive cervical cell samples stratified into the diagnostic categories non-progressive, CIN2 and CIN3+ were sequenced. Six samples were removed from further analysis based on a MNV pattern suggestive of co-infection of different viral sublineages (Supplementary figure A1). After removing samples that did not pass the filtering criteria (n = 164), 77 HPV16, 49 HPV18, 84 HPV31, 88 HPV33 and 56 HPV45 samples underwent comparative MNV and integration analysis. The mean sequencing depth for samples within the different diagnostic categories ranged from 4462 for HPV16 CIN2 to 55097 for HPV18 CIN2. The proportion of the genome covered with a minimum depth of 100x within the categories ranged from 0.70 for the HPV16 CIN2 category to 0.97 for the HPV33 CIN3+ categories (Table 1).

3.2. Minor nucleotide variation profiles reveal a higher number of MNVs in HPV45 positive samples

A total of 10664 MNVs were identified in the 354 analysed samples. Most of the MNVs were low-frequency variants with 1716 MNVs having a minor base frequency ≥5% and 850 with a minor base frequency ≥10%. The number of MNVs were significantly higher in HPV45 samples compared to other HPV types ($p < 0.001$), the average being 47.9, 47.92 and 34.91 in the HPV45 non-progressive, CIN2 and CIN3+ categories, respectively (Fig. 1, Table 2). Excluding HPV45, HPV16 non-progressive was the category with the highest average number of MNVs with 34.57, while HPV33 CIN2 category had the lowest number of MNVs with an average of 19. The standard deviation of the number of MNVs within categories was also found to be highest in HPV45 samples. The same were true when investigating the MNV frequencies, with HPV45 having the highest mean and SD MNV frequency in all categories (Supplementary figure A2 and Table 2). No significant differences were found in comparisons of diagnostic categories.

HPV MNVs were found within genomic elements of all the different categories. Most variation was found in HPV45 positive samples, where the genomic elements E4, URR and NCR in non-progressive samples and the E4 and E2 in CIN2 samples had the highest number of mean variants (Supplementary figure A3).

In total, 5734 nonsynonymous and 244 nonsense mutations were observed in the dataset. Most genes, across all categories, had a dN/dS ratio >1 or close to 1 (Supplementary figure A4).

3.3. APOBEC-related mutational signatures

The two most common substitutions in the dataset, C > T and T > C, were observed across all the diagnostic categories for all HPV types (Supplementary figure A5 and A6). To investigate APOBEC3-induced mutations, C > T mutations in the APOBEC3-preferred trinucleotide context TCN found within different diagnostic categories for the different HPV types were compared against each other. We observed that TCA was the trinucleotide context with the highest proportion across all diagnostic categories in Alpha-9 samples, while no such pattern was observed for Alpha-7 samples (Fig. 2). C > T mutations in the TCA context were found to differ significantly between the Alpha-7 and Alpha-9 clades ($p < 0.001$), and the results were found to be consistent when the analysis was extended to include the TCW trinucleotide context and the inclusion of C > G substitutions, but not when investigating C > G substitutions by themselves. When investigating each HPV-type separately, HPV16 showed a decrease in TCA proportion of C > T MNVs with increased lesion severity compared to HPV31 and HPV33 (Supplementary figure A6). The extended mutation signature analysis did not reveal any strong signal for either APOBEC3A (YTCA) or APOBEC3B (RTCA) preference in the dataset (Supplementary figure A7).

The number of TCA motifs found within each HPV types differed between Alpha-7 (164 HPV18, 159 HPV45) and Alpha-9 (186 HPV16, 190 HPV31, 208 HPV33) (Supplementary table B1), with most motifs being present on the minus strand (Supplementary figure A8). Values for RTCA and YTCA motifs can be seen in supplementary figure A9 and supplementary table B1. To further investigate if the number of C > T substitutions in the TCA context occurred more frequently than expected the proportion of TCA motifs out of all available NCN motifs were calculated. The proportions of TCA motifs were found to range from 0.05 (HPV45) to 0.07 (HPV33) and were interpreted as the expected proportion of C > T substitutions in that trinucleotide context, assuming that substitutions are equally likely in all NCN motifs (Supplementary table B2). The difference in observed/expected C > T substitutions in the TCA context were found to be significantly larger for Alpha-9 samples ($p < 0.001$), thus there were found to be significantly more C > T mutations in the TCA context than would be expected in Alpha-9 relative to Alpha-7 (Fig. 3).

3.4. Higher integration frequencies (IFs) in Alpha-7s compared to Alpha-9s

The number of integrations in the Alpha-7 significantly outnumber those of the Alpha-9 clade although more than twice as many Alpha-9

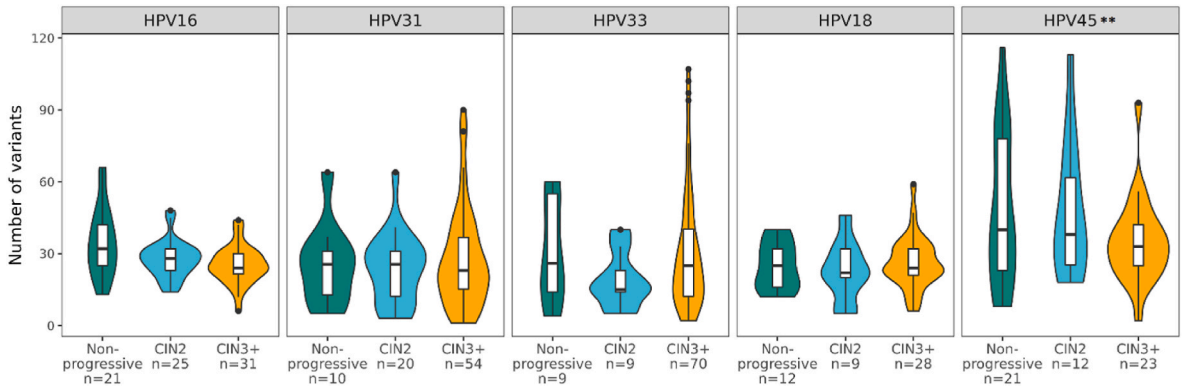

Fig. 1. Number of variants in HPV16, HPV18, HPV31, HPV33 and HPV45 positive samples. Violin plots representing the number of variants in the different diagnostic categories on the x-axis. Box-and-whisker plots are added to show the median number of MNVs (horizontal line), the box represents the 25% and 75% percentiles, and the whiskers represents the minimum and maximum number of MNVs found within one sample. The number of samples (n) is indicated below each category. Double asterix (**) represents that HPV45 has significantly more MNVs compared to all other types ($p < 0.01$).

Table 2
Number of samples analysed and different statistics of MNVs for HPV16, HPV18, HPV31, HPV33 and HPV45 positive samples stratified across the diagnostic categories.

Diagnostic category	Analysed samples	Mean number of variants	Minimum number of variants	Maximum number of variants	Standard deviation of number of variants	Mean MNV frequency	Standard deviation of MNV frequency
HPV16							
Non-progressive	21	34.57	13	66	14.25	2.74	3.65
CIN2	25	27.80	14	48	7.88	2.80	4.82
CIN3+	31	25.35	6	44	8.00	2.67	3.66
HPV18							
Non-progressive	12	25.08	12	40	9.97	3.21	4.69
CIN2	9	23.00	5	46	12.68	2.89	4.43
CIN3+	28	26.36	6	59	11.05	4.35	6.79
HPV31							
Non-progressive	10	25.70	5	64	17.12	4.53	6.72
CIN2	20	23.90	3	64	14.84	3.30	5.40
CIN3+	54	27.63	1	90	20.55	4.29	7.05
HPV33							
Non-progressive	9	30.78	4	60	21.16	2.59	2.50
CIN2	9	19.00	5	40	11.47	2.49	4.68
CIN3+	70	30.74	2	107	24.77	3.36	4.77
HPV45							
Non-progressive	21	47.90	8	116	31.28	6.79	9.07
CIN2	12	47.92	18	113	28.52	4.16	6.64
CIN3+	23	34.91	2	93	17.89	4.63	7.37

samples were sequenced (Fig. 4 and Table 3). In total, 42.8% of Alpha-7 samples had at least one integration site reported, significantly more than Alpha-9 with 6.4% ($p < 0.001$). Overall, 154 integration sites were observed in the whole dataset, of which 85% (131/154) were Alpha-7 and 15% (23/154) Alpha-9 (Table 3, Fig. 4). Alpha-7 also had the highest IF in all diagnostic categories with 21% of non-progressive, 33% of CIN2 and 61% of CIN3+ samples having integrations, thus higher IF correlated significantly with diagnostic severity ($p < 0.01$). Alpha-7 samples had significantly more integration sites in samples with integration compared to Alpha-9 ($p < 0.001$), with an average of 3.4, 3.14 and 2.74 integrations in the non-progressive, CIN2 and CIN3+ categories, respectively, compared to Alpha-9 with 1.25, 1.5 and 1.5.

Comparing the HPV-types within the two clades revealed some differences between the related types. Overall, HPV18 had significantly more samples with integrations than any other type, while HPV45 CIN3+ (the only diagnostic category with HPV45 integrations) had significantly more than all other Alpha-9 types. HPV18 also had significantly more integrations per sample than all other types, as well as the sample with the most reported integrations with 21, compared to 4, 3, 1 and 1 for HPV45/16/31/33, respectively. Within Alpha-9, HPV16 reported higher IF than both HPV31 and HPV33, as well as higher average number of integrations per sample (Supplementary table B3). A complete list of annotated integration breakpoints can be found in supplementary table C1.

3.5. Deletions and breakpoints in the HPV genome

In Alpha-7 samples with integrations, breakpoints were found in all genetic elements of the HPV genome, except NCR, while Alpha-9 integrations lacked breakpoints in E7 and E4 (Supplementary figure A10). Any breakpoint location bias was investigated using the number of reported Alpha-7 and -9 integrations divided by the average gene lengths in each clade. It was observed that Alpha-7 samples had more breakpoints in the E1, E2, E4 and L2 genes than would be expected by chance, however, this difference was not statistically significant ($p = 0.24$). Breakpoints in Alpha-9 samples were observed in E1, L2 and NCR more

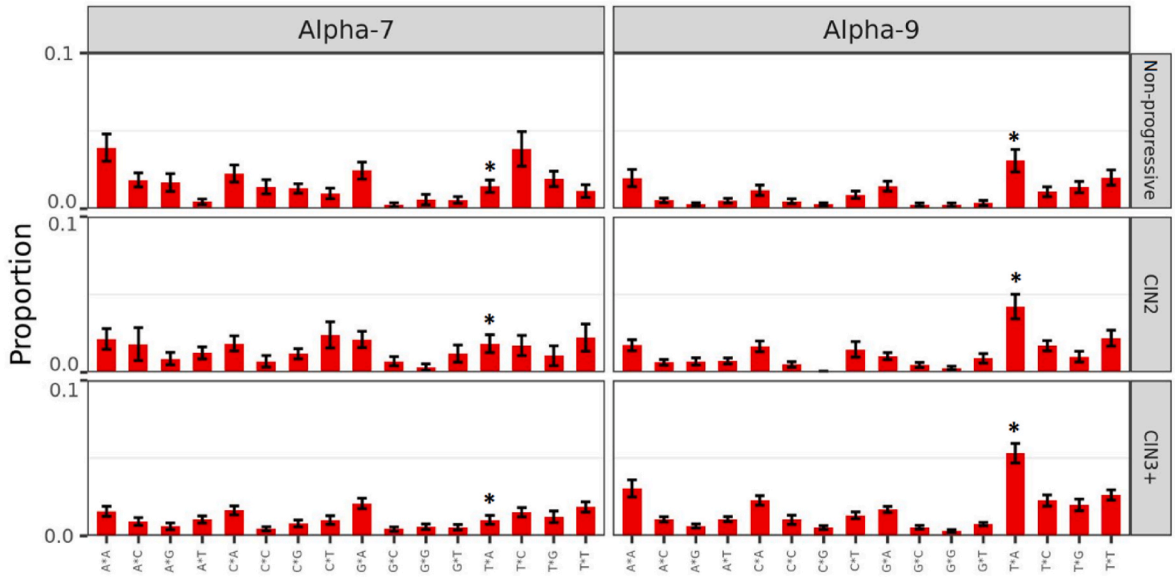

Fig. 2. C > T mutational signatures in Alpha-7 (HPV18 and HPV45) and Alpha-9 (HPV16, HPV31 and HPV33) positive samples across the different diagnostic categories. The mean proportion of C > T mutations is shown on the y-axis and the different trinucleotide contexts are shown on the x-axis. Error bars represent the standard error of the mean. Asterix (*) denotes C > T substitutions found in the TCA context and was found to be overall significant between the clades.

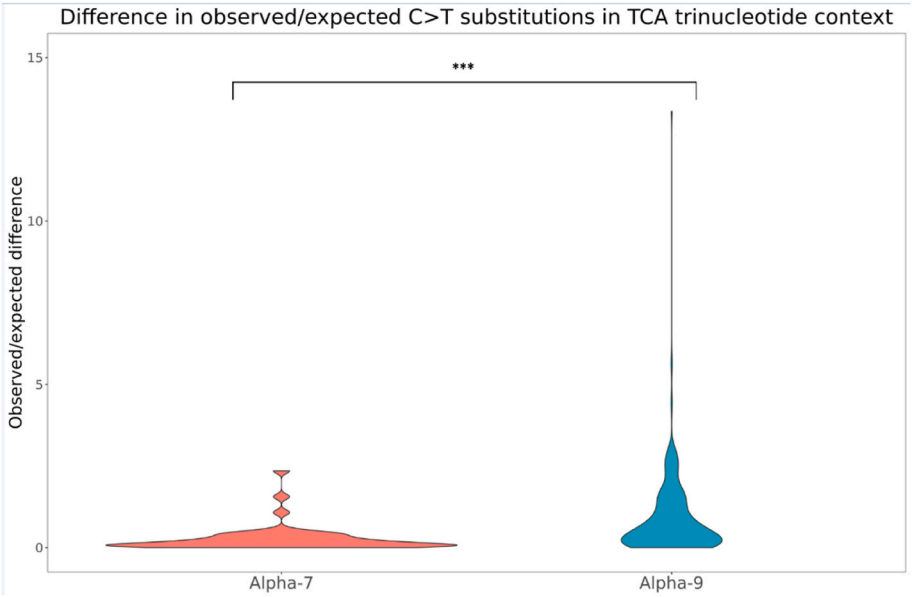

Fig. 3. Violin plots representing the difference in observed/expected number of C > T substitutions in the TCA context for individual samples in Alpha-7 and Alpha-9.

often than expected by chance, but was also not significant ($p = 0.20$). Combining and investigating the percentage of integrations with breakpoints in E1/E2 across the different diagnostic categories revealed that 38%, 36% and 51% of the integrations caused breakpoints in E1/E2 in the non-progressive, CIN2 and CIN3+ categories, respectively (Fig. 5a). A detailed figure showing sample integrations and number of breakpoints in E1/E2 can be seen in Supplementary figure A11.

To investigate large HPV genomic deletions, coverage plots for each sample were inspected for extended regions without or with exceptionally low relative sequencing coverage. In addition to the six samples with deletions reported in the previous study (one HPV16 and five HPV18), the coverage plots revealed 11 additional samples with deletions or partial deletions (One HPV31 and 10 HPV45). All HPV45 positive samples with deletions were reported as having integrations while the single HPV31 sample did not.

3.8 Presence of human cancer-related genes within ±10 kb of integration sites.

Due to the uneven number of integrations found within Alpha-7 (131 integrations) and Alpha-9 (23 integrations), all integrations in the dataset were combined when investigating integrations in human genes within ±10 kb of the integration sites. The results revealed that CRGs were present in 41% (12/29), 40% (10/25) and 59% (59/100) of non-progressive, CIN2 and CIN3+ samples, respectively (Fig. 5b, see Supplementary figure A12 and supplementary table C1 for more details). Of the integrations within ±10 kb of CRGs, 58% (7/12), 80% (8/10) and 78% (46/59) were integrated inside the ORF of the reported CRG in the non-progressive, CIN2 and CIN3+ categories (Supplementary figure A13). Three CRGs were each found twice ±10 kb of integrations sites in

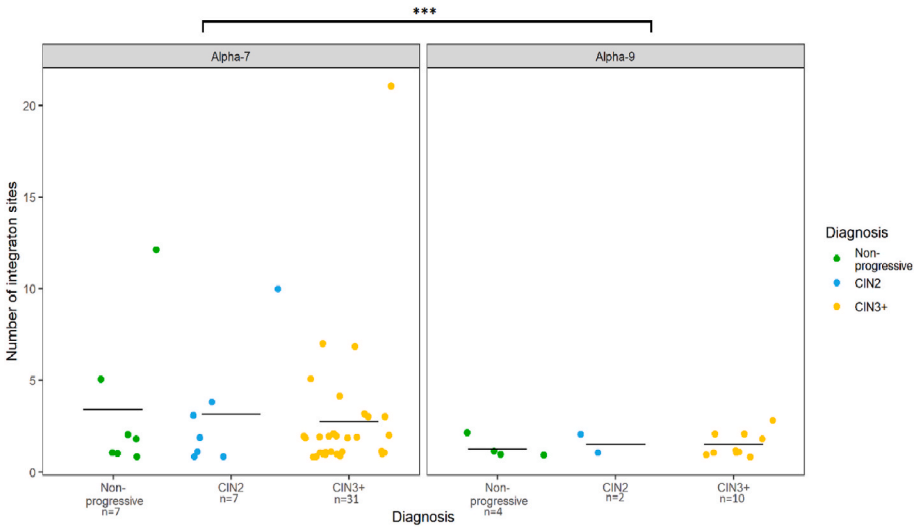

Fig. 4. Number of integrations in Alpha-7 and Alpha-9 samples with reported integration(s), stratified by diagnostic category. Horizontal line represents the mean number of integrations and n denotes number of samples in each category with reported integrations.

Table 3
Number of integrations in Alpha-7 and Alpha-9 positive samples, stratified by diagnostic category.

Diagnostic categories (samples)	Number of samples with integrations (Frequency %)	Total number of integration sites	Mean number of integrations per sample
Alpha-7			
Non-progressive (n = 33)	7 (21.2%)	24	3.43
CIN2 (n = 21)	7 (33.3%)	22	3.14
CIN3+ (n = 51)	31 (60.8%)	85	2.74
Total (n = 105)	45 (42.8%)	131	2.91
Alpha-9			
Non-progressive (n = 40)	4 (10%)	5	1.25
CIN2 (n = 54)	2 (3.7%)	3	1.5
CIN3+ (n = 155)	10 (6.4%)	15	1.5
Total (n = 249)	16 (6.4%)	23	1.44

different samples, being RCAN2 (HPV18 CIN2 and HPV18 CIN3+), MIR205 (HPV16 CIN3+ and HPV45 CIN3+) and KLHL29 (HPV18 CIN2 and HPV31 CIN3+).

The percentages of integrations in the human genome with annotated genes present within ±10 kb of the integration site were 76% (22/29), 68% (17/25) and 75% (75/100) in the non-progressive, CIN2 and CIN3+ categories, respectively (Supplementary figure A14). The percentages of integrations inside the ORF of human genes were 45%, 44% and 56% in the non-progressive, CIN2 and CIN3+ categories (Supplementary figure A15).

Alpha-7 samples had integrations in all human chromosomes except chromosome 18 and 21 (Supplementary figure A16). 28/131 Alpha-7 integrations were in chromosome 1 and 2. Alpha-9 samples had integrations in 15 different chromosomes, with chromosome 5 and 8 having most.

3.6. Validation of reported integration sites by sanger sequencing

In total, 31 reported integration sites in 21 patient samples qualified for validation by Sanger sequencing after QC filtering based on read mapping and sequencing artefacts [55]. Of the integration sites, 5 integrations were detected in HPV31 samples (84 analysed samples), 1 integration in HPV33 samples (88 analysed samples) and 25 integrations in HPV45 samples (56 analysed samples). In total, 21 of the 31 reported integrations sites were validated by Sanger sequencing which confirmed correct chromosomal coordinates and HPV type. The remaining 10 did not yield high-quality chromatograms, possibly due to low DNA concentrations, suboptimal PCR amplification, unspecific primer hybridization or genomic structural rearrangements often associated with HPV integrations [36,56]. Microhomology regions were identified in 19% (5/21) of the confirmed HPV integrations, the length ranging from 3 bp to 12 bp and are presented in supplementary table C2.

4. Discussion

This study aimed to investigate and compare type-specific intra-host variation and integration characteristics of five high-risk HPV types belonging to Alpha-7 and Alpha-9 across diagnostic categories of increasing severity. We observed differences between the diagnostic categories, as well as between Alpha-7 and Alpha-9 clades. The differences adhere to their phylogenetic assortment, where there is a statistically significant signal of APOBEC3-induced C > T mutations in Alpha-9 samples that is not found in Alpha-7. IF is also significantly higher in samples positive for Alpha-7 HPV-types relative to Alpha-9.

Minor nucleotide variants (MNVs) are variants found below the consensus level that might play a role in the development of cancer [16]. In our dataset, we observe that HPV45 samples have more MNVs compared to the other four investigated HPV types, and that Alpha-7s have more variation in E4 compared to the Alpha-9s. The biological significance of these results is currently unclear. The MNVs are called in a reference-independent manner, and we rule out that these results are artefacts that could arise from mapping to divergent reference genomes. Co-infections with two or more variants of the same HPV type could be a likely alternative source of diversity [57] and six apparent co-infections were excluded from the analysis based on indicative MNV patterns (Supplementary figure A1). When comparing the amount of MNVs, we observed no differences between the diagnostic categories, suggesting that the total quantity of MNVs in a sample is not directly associated with carcinogenic risk. Rather than quantity, the quality of MNVs in their HPV genomic context may be of significance, as has been shown for the HPV16 E7 gene in cervical cancer cases and for certain positions in HPV16 URR where MNVs have been shown associated with developing CIN3+ [16].

C > T substitutions in the trinucleotide context TCA and TCT is correlated with APOBEC3A/B activity [20]. While they are part of the innate immune system in response to viral infections, their mutation signatures are also commonly found in host genomes of HPV-positive

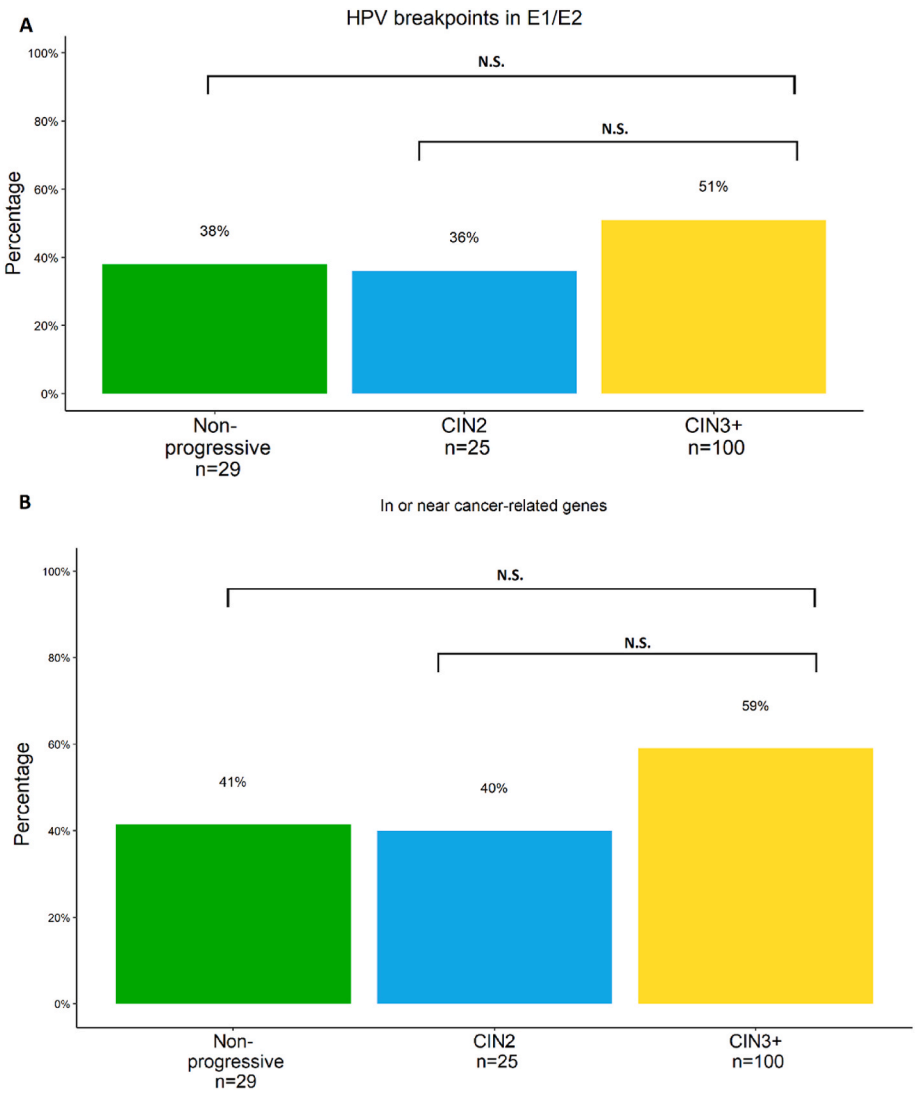

Fig. 5. A) Percentage of integrations with breakpoints in E1/E2, stratified by diagnostic category. B) Percentage of integrations with CRGs within ±10 kb of integration site or inside the ORF, stratified by diagnostic category. n is the number of integrations in each category.

cancer cells as well as in viral HPV genomes [58]. Most research investigating APOBEC3-HPV interactions have had a focus on HPV16, while less research has been done on the other high risk types [17,20,59,60]. In our study, APOBEC3A/B induced C > T mutations were found to be most common in the trinucleotide context TCA for Alpha-9s, while the Alpha-7s did not have this pattern. This is to our knowledge the first study to show that differences in APOBEC3-induced mutation profiles between any HPV types are highly significant and that the difference is in concordance with phylogeny. The number of all available TCA motifs in the HPV genomes studied here differs between the types, with Alpha-7s (average 162) having less than Alpha-9s (average 195). To investigate whether this difference in the number of available motifs affected the observed mutational C > T patterns, the proportion of TCA motifs out of all available NCN motifs were compared against the observed proportion of C > T mutations in TCA motifs found within samples. The difference between observed and expected proportions of C > T mutations in TCA motifs were found to be significantly higher for Alpha-9s than Alpha-7s. Thus, Alpha-9 samples were found to have relatively more C > T mutations in the TCA-motifs, even when correcting for the higher abundance of TCA motifs in the genomes relative to Alpha-7. This finding suggests that the Alpha-9 infections trigger a detectable APOBEC3-response not found in Alpha-7 infections; a finding that warrants further investigation considering their differences in clinical epidemiology [10], human molecular and genomic cancer characteristics [44,61], evolutionary histories [62] and impact of viral life-cycle factors and expected tropisms [63]. Our previous study looked at HPV16 and HPV18 and by including one additional Alpha-7 and two more Alpha-9 types in the comparative analysis, the phylogenetic dichotomy in mutational signature was established more broadly and emphasises the biological significance of the results. Using different whole genome sequencing protocols Hirose et al., 2018 identified the same APOBEC signature (C > T in the TCA context) across the three Alpha-9s HPV16/52/58, and the exact same signature was described in HPV16 (signature A) in Zhu et al., 2020 [17,20]. The combined presence of the APOBEC signature in the Alpha-9s now therefore encompass HPV16/31/33/52/58. Yet, contrary to current understanding it cannot be established that APOBEC3-induced mutagenesis in viral genomes is a general detectable feature in high-risk HPV infections. In HPV16 positive samples, the number of APOBEC3-related nucleotide substitutions decreases with lesion severity, which also has been shown in previous studies [17,20,43]. However, this decrease is not observed in HPV31 and HPV33 positive samples, suggesting this is not an Alpha-9 specific tendency, but rather a feature of HPV16 carcinogenesis. Our results are in corroboration with previous studies that have observed that the number of APOBEC3-related nucleotide substitutions decreases with lesion severity in HPV16 positive samples [17,20,43].

The lack of APOBEC3 activity found in Alpha-7 genomes might be driven by differences in host-response to viral infections between the different cell types and/or by genomic variation between the HPV types. Alpha-7 types have been found to be significantly more common in cases of adenocarcinoma (ADC) than squamous cell carcinoma (SCC), while Alpha-9 is the predominant type found in SCC [64,65]. These differences are likely reflecting type-specific tropisms and cells of cancer origin, where Alpha-9 is found to associate with lesions in squamous cells while Alpha-7 predominantly cause lesions in glandular cells [66]. Additionally, expression profiles of SCC tumours caused by Alpha-7 and Alpha-9 types have been shown to differ in expression levels of keratin gene family members [44]. APOBEC3 substitutions are associated with viral clearance in HPV16 infections [20], and it is interesting to note that while HPV31 and HPV33 have been shown to have a high risk to progress to CIN3, their risk to progress from CIN3 to invasive cervical cancer is relatively low compared to HPV16/18/45 [66]. We observe a decrease in APOBEC3-related nucleotide substitutions in HPV16 positive CIN3+ samples, however, this pattern is not present in HPV31/33. Thus, one can speculate that the different HPV types possess different abilities to trigger APOBEC3-activity, and that this differs both between Alpha-7 and Alpha-9 types as well as between HPV16 and HPV31/33. In a genome evolution perspective, we note that the preferred APOBEC TCN-motif is underrepresented in Alpha-HPV genomes generally and more so in the URR of HPV16 than HPV18/31 which includes the origin or replication and promoter of the E6/7 oncogenes [67]. Further comparative studies into APOBEC-HPV interaction mechanisms and evolutionary dynamics are warranted in order to better understand the trade-off between immune exposure and oncogenicity in individual HPV types.

Integration of HPV-genomes in the human genome is a suggested driver event during HPV-induced carcinogenesis and previous studies have shown that the IF differs between genotypes in cancers [30]. Our results are in line with these studies showing that Alpha-7s HPV18/45 have a higher IF than Alpha-9s HPV16/31/33, and that HPV16 have a higher IF than HPV31 and HPV33 [45]. Women with invasive cervical cancer caused by HPV16, HPV18 and HPV45 are typically younger than women with cervical cancer caused by other HPV types, suggesting that infection with these three genotypes progress to invasive cervical cancer faster than other types [64]. In this study, Apha-7-positive samples show a significant increase in IF with increase in diagnostic severity, and HPV45-positive samples only had reported integrations in the CIN3+ category. Alpha-7 positive samples with integrations also had significantly more integrations per sample compared to Alpha-9 positive samples. These findings might reflect conserved differences in the biology of Alpha-7 and Alpha-9 types that affects IFs, for example in that Alpha-7s more often than Alpha-9s can warrant integration(s) as a contributing factor to drive oncogenic transformation.

Integrations are associated with increased genomic instability, mainly through overexpression of viral oncogenes E6/E7, but also by triggering host oncogenes and disrupting of tumour suppressor genes [28,68,69]. One of the mechanisms by which integrations can cause overexpression of E6/E7 is the disruption of the E2 gene upon linearization of the circular HPV genome. When we looked at all HPV breakpoints combined, 51% of samples with integrations in the CIN3+ category had breakpoints in E1/E2, compared to 38% and 36% in the non-progressive and CIN2 categories, respectively. In addition, when investigating the coverage plots, nine of out fifteen HPV45 samples with reported integrations showed either full deletions or partial deletions in regions encompassing E1/E2.

Another mechanism by which HPV integrations can step up carcinogenic transformation is by integrating inside the ORF or in genomic proximity of host oncogenes and tumour suppressor genes, and thereby disrupting their function or altering their expression levels, respectively [30,68–70]. We found that the presence of at least one human CRG within ±10 kb of integration sites increased with almost 20% in the most severe category (CIN3+ at 59%). Chromosomal integrations have been shown to cause genomic instability in the vicinity of the integration sites by causing structural rearrangements and affecting host gene expression [37,56]. Several observed CRGs in our dataset have previously been reported in studies investigating HPV integrations and are associated with cervical cancer, including TP63, MIR205HG, MMP12 and ENO1 [30,71–73]. Additionally, RCAN2, KLHL29 and MIR205HG were observed close to integrations twice in independent samples. Decrease in RCAN2 expression has been associated with tumour proliferation in colorectal cancers, while differential methylation patterns of KLHL29 have been observed in small and large anal cancer tumours [74,75]. MIR205HG, on the other hand, has been implicated in playing a role in the development in cervical cancer by targeting and regulating genes involved in proliferation, migration and apoptosis of cervical cancer cells [71,76,77]. While the role of MIR205HG in HPV-induced carcinogenesis is established, the presence of RCAN2 and KLHL29 close to integration sites in more than one sample may also suggest their involvement in HPV-induced carcinogenesis. These findings warrant further investigation.

Taken together, the increased number of integrations with the increase in diagnostic severity observed for Alpha-7s and the general tendencies for having breakpoints in E1/E2 and integrating within ±10 kb of CRGs, supports the notion that integrations are key molecular events in driving HPV-induced carcinogenesis. The differences between the diagnostic categories regarding integration breakpoints in E1/E2 and proximity to CRGs, were not statistically significant when a glm model was applied to the data. However, the number of observations in the three categories differ substantially, and studies including more samples in the "non-progressive" category combined with follow-up data, should be conducted to ascertain their role in HPV-induced carcinogenesis. Furthermore, the difference in IFs of the Alpha-7s and -9s suggests that IF is a consistent phenomenon within phylogenetically related HPV types. What drives this difference between the clades is currently poorly understood. Different high risk HPVs produce different splice isoforms of viral oncogenes E6 and E7 and do also have different capabilities of inducing p53-degradation among other differences [62,63,78,79]. It is possible that the discrepancy observed is due to Alpha-7 oncoproteins having weaker oncogenic potential relative to Alpha-9 oncoproteins and therefore Alpha-7 infections to a larger extent require integrations to drive carcinogenesis, reflecting that nearly all Alpha-7-induced tumours have integrated viral DNA and Alpha-9 can induce cancer in episomal form [44]. HPVs are also known to induce DNA damage and uses DNA damage response pathways for the amplification of the viral genome, which could consequently lead to integration of HPV DNA by nonhomologous end joining and/or microhomology-mediated recombination [32,80–84]. To our knowledge there is nothing in the existing literature that directly compares the ability of different HPV types to induce DNA damage. Since viral proteins have been shown to induce DNA damage an alternative explanation to these clade-specific differences in IFs beyond oncogenic potential of E6/7 remains an option. However, more research into the subject is needed to better understand the molecular mechanisms which drive different manifestation of IFs.

5. Conclusions

This study reveals differences in APOBEC3-mediated mutations in concordance with evolutionary related HPV types, where Alpha-9 positive samples have a clear APOBEC3 mutation signature not observed for Alpha-7. Additionally, Alpha-7 samples are shown to have significantly more integrations and an increase in number of integrations with increased diagnostic severity. This study expands our knowledge, beyond HPV 16 and HPV18, by including three additional high risk HPV types and shows that the type-specific patterns for these molecular events extends to more closely related carcinogenic HPV types within the Alpha-7 and Alpha-9 clades. The results broaden our understanding of the molecular mechanisms behind HPV-induced cancers while also

shedding light on some of the similarities and differences between the HPV types investigated.

Authors' contributions

AHL designed and performed the experiments, analysed the results and drafted the manuscript text. AR performed experiments, analysed the results and contributed to drafting the manuscript. JMC contributed to the data analysis and performed the statistical analysis. IKC managed the sample material, contributed to the study design and result interpretation. TBR and OHA contributed to the study design, data analysis and result interpretation. All authors contributed to writing and approved the final version of the manuscript.

Funding

This work was supported by a PhD grant to AHL and a Research track grant to AR from Faculty of Health Sciences, Oslo Metropolitan University, and by a post-doctoral research grant to JMC from the South-Eastern Norway Regional Health Authority, project number 2020010. The work was also supported by innovation grants from the South-Eastern Norway Regional Health Authority 2019 and the Research Council of Norway (FORNY2020, project number 296671). The funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Data statement

The data presented in this article are not readily available because of the principles and conditions set out in the General Data Protection Regulation (GDPR), with additional national legal basis as per the Regulations on population-based health surveys and ethical approval from the Norwegian Regional Committee for Medical and Health Research Ethics (REC). Requests to access the data should be directed to the corresponding authors.

Declaration of competing interest

We have no competing interests to declare.

Data availability

The data presented in this article are not readily available because of the principles and conditions set out in the GDPR regulations. Request to access the data should be directed to the corr.authors

Acknowledgments

We thank Karin Helmersen for her help with the Sanger sequencing and DNA extraction, Milan Stosic for helpful discussions and assistance with uncooperative Genbank files, and Sinan U. Umu for bioinformatic advise. The sequencing service was provided by the Norwegian Sequencing Centre (https://www.sequencing.uio.no), a national technology platform hosted by Oslo University Hospital and the University of Oslo supported by the Research Council of Norway and the South-Eastern Regional Health Authority.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tvr.2022.200247.

References

[1] A. Harari, Z. Chen, R.D. Burk, Human papillomavirus genomics: past, present and future, Curr. Probl. Dermatol. (Basel) 45 (2014) 1–18, https://doi.org/10.1159/000355952.
[2] D. DiMaio, L.M. Petti, The E5 proteins, Virology 445 (1–2) (2013) 99–114, https://doi.org/10.1016/J.VIROL.2013.05.006.
[3] K. Hoppe-Seyler, F. Bossler, J.A. Braun, A.L. Herrmann, F. Hoppe-Seyler, The HPV E6/E7 oncogenes: key factors for viral carcinogenesis and therapeutic targets, Trends Microbiol. 26 (2) (2018), https://doi.org/10.1016/j.tim.2017.07.007.
[4] Human Reference clones – hpvcenter. https://www.hpvcenter.se/human_reference_clones/. (Accessed 5 March 2020).
[5] D. Bzhalava, C. Eklund, J. Dillner, International standardization and classification of human papillomavirus types, Virology 476 (2015) 341–344, https://doi.org/10.1016/j.virol.2014.12.028.
[6] PaVE, Papilloma virus genome database. https://pave.niaid.nih.gov/#explore/reference_genomes/human_genomes. (Accessed 6 February 2020).
[7] Z. Chen, M. Schiffman, R. Herrero, et al., Classification and evolution of human papillomavirus genome variants: alpha-5 (HPV26, 51, 69, 82), alpha-6 (HPV30, 53, 56, 66), alpha-11 (HPV34, 73), alpha-13 (HPV54) and alpha-3 (HPV61), Virology 516 (2018) 86–101, https://doi.org/10.1016/j.virol.2018.01.002.
[8] R.D. Burk, A. Harari, Z. Chen, Human papillomavirus genome variants, Virology 445 (1–2) (2013) 232–243, https://doi.org/10.1016/j.virol.2013.07.018.
[9] H.U. Bernard, R.D. Burk, Z. Chen, K. van Doorslaer, H zur Hausen, E.M. de Villiers, Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments, Virology 401 (1) (2010) 70–79, https://doi.org/10.1016/j.virol.2010.02.002.
[10] C. De Martel, M. Plummer, J. Vignat, S. Franceschi, Worldwide burden of cancer attributable to HPV by site, country and HPV type, Int Agency Res Cancer (IARC/WHO 141 (2017) 664–670, https://doi.org/10.1002/ijc.30716.
[11] F.X. Bosch, A. Lorincz, N. Muñoz, L.M. Meijer, The Causal Relation between Human Papillomavirus and Cervical Cancer, vol. 55, 2002. http://www.ipvsoc.org. (Accessed 2 February 2019).
[12] H. Zur Hausen, Papillomaviruses and cancer: from basic studies to clinical application, Nat. Rev. Cancer 2 (5) (2002) 342–350, https://doi.org/10.1038/nrc798.
[13] R.D. Burk, Z. Chen, K. Van Doorslaer, Human papillomaviruses: genetic basis of carcinogenicity, Public Health Genomics 12 (5–6) (2009) 281–290, https://doi.org/10.1159/000214919.
[14] K. Van Doorslaer, Evolution of the papillomaviridae, Virology 445 (1–2) (2013) 11–20, https://doi.org/10.1016/j.virol.2013.05.012.
[15] CM de Oliveira, I.G. Bravo, NCS.e. Souza, et al., High-level of viral genomic diversity in cervical cancers: a Brazilian study on human papillomavirus type 16, Infect. Genet. Evol. 34 (2015) 44–51, https://doi.org/10.1016/j.meegid.2015.07.002.
[16] L. Mirabello, M. Yeager, K. Yu, et al., HPV16 E7 genetic conservation is critical to carcinogenesis, Cell 170 (6) (2017) 1164–1174, https://doi.org/10.1016/j.cell.2017.08.001, e6.
[17] Y. Hirose, M. Onuki, Y. Tenjimbayashi, et al., Within-host variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome, J. Virol. 92 (12) (2018) e00017–e00018, https://doi.org/10.1128/jvi.00017-18.
[18] R.S. Dube Mandishora, K.S. Gjøtterud, S. Lagström, et al., Intra-host sequence variability in human papillomavirus, Papillomavirus Res 5 (2018) 180–191, https://doi.org/10.1016/j.pvr.2018.04.006.
[19] S. Lagström, P. van der Weele, T.B. Rounge, I.K. Christiansen, A.J. King, O. H. Ambur, HPV16 whole genome minority variants in persistent infections from young Dutch women, J. Clin. Virol. (August 2019), https://doi.org/10.1016/J.JCV.2019.08.003.
[20] B. Zhu, Y. Xiao, M. Yeager, et al., Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance, Nat. Commun. 11 (1) (2020) 886, https://doi.org/10.1038/s41467-020-14730-1.
[21] G.M. Clifford, V. Tenet, D. Georges, et al., Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: whole viral genome sequences from 7116 HPV16-positive women, Papillomavirus Res 7 (2019) 67–74, https://doi.org/10.1016/j.pvr.2019.02.001.
[22] L.S. Arroyo-Mühr, C. Lagheden, E. Hultin, et al., The HPV16 genome is stable in women who progress to in situ or invasive cervical cancer: a prospective population-based study, Cancer Res. 79 (17) (2019) 4532–4538, https://doi.org/10.1158/0008-5472.CAN-18-3933.
[23] N.A. Wallace, K. Münger, The curious case of APOBEC3 activation by cancer-associated human papillomaviruses, PLoS Pathog. 14 (1) (2018), e1006717, https://doi.org/10.1371/journal.ppat.1006717.
[24] C.J. Warren, T. Xu, K. Guo, et al., APOBEC3A functions as a restriction factor of human papillomavirus, J. Virol. 89 (1) (2015) 688–702, https://doi.org/10.1128/jvi.02383-14.
[25] J.R. Chapman, M.R.G. Taylor, S.J. Boulton, Playing the end game: DNA double-strand break repair pathway choice, Mol. Cell. 47 (4) (2012) 497–510, https://doi.org/10.1016/j.molcel.2012.07.029.
[26] S Do Kang, S. Chatterjee, S. Alam, et al., Effect of productive human papillomavirus 16 infection on global gene expression in cervical epithelium, J. Virol. 92 (20) (2018), https://doi.org/10.1128/JVI.01261-18.
[27] C.C. Spriggs, L.A. Laimins, Human Papillomavirus and the DNA Damage Response: Exploiting Host Repair Pathways for Viral Replication, 2017, https://doi.org/10.3390/v9080232.

[28] A.A. McBride, A. Warburton, The role of integration in oncogenic progression of HPV-associated cancers, PLoS Pathog. 13 (4) (2017), e1006211, https://doi.org/10.1371/journal.ppat.1006211.
[29] M. Pett, N. Coleman, Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? J. Pathol. 212 (4) (2007) 356–367, https://doi.org/10.1002/path.2192.
[30] R.D. Burk, Z. Chen, C. Saller, et al., Integrated genomic and molecular characterization of cervical cancer, Nature 543 (7645) (2017) 378–384, https://doi.org/10.1038/nature21386.
[31] S. Jeon, B.L. Allen-Hoffmann, P.F. Lambert, Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells, J. Virol. 69 (5) (1995) 2989–2997, https://doi.org/10.1128/jvi.69.5.2989-2997.1995.
[32] C. Ziegert, N. Wentzensen, S. Vinokurova, et al., A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques, Oncogene 22 (25) (2003) 3977–3984, https://doi.org/10.1038/sj.onc.1206629.
[33] C. Bodelon, M.E. Untereiner, M.J. Machiela, S. Vinokurova, N. Wentzensen, Genomic characterization of viral integration sites in HPV-related cancers, Int. J. Cancer 139 (9) (2016) 2001–2011, https://doi.org/10.1002/ijc.30243.
[34] M. Dürst, C.M. Croce, L. Gissmann, et al., Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas, Proc. Natl. Acad. Sci. U. S. A. 84 (4) (1987) 1070–1074, https://doi.org/10.1073/pnas.84.4.1070.
[35] A.I. Ojesina, L. Lichtenstein, S.S. Freeman, et al., Landscape of genomic alterations in cervical carcinomas, Nature 506 (7488) (2014) 371–375, https://doi.org/10.1038/nature12881.
[36] M. Peter, N. Stransky, J. Couturier, et al., Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma, J. Pathol. 221 (3) (2010) 320–330, https://doi.org/10.1002/path.2713.
[37] K. Akagi, J. Li, T.R. Broutian, et al., Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability, Genome Res. 24 (2) (2014) 185–199, https://doi.org/10.1101/gr.164806.113.
[38] A. Warburton, C.J. Redmond, K.E. Dooley, et al., HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression, in: M. Ott (Ed.), PLOS Genet, vol. 14, 2018, e1007179, https://doi.org/10.1371/journal.pgen.1007179, 1.
[39] G. Gao, J. Wang, J.L. Kasperbauer, et al., Whole genome sequencing reveals complexity in both HPV sequences present and HPV integrations in HPV-positive oropharyngeal squamous cell carcinomas, BMC Cancer 19 (1) (2019) 352, https://doi.org/10.1186/s12885-019-5536-1.
[40] I. Kraus, C. Driesch, S. Vinokurova, et al., The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes, Cancer Res. 68 (7) (2008) 2514–2536, https://doi.org/10.1158/0008-5472.CAN-07-2776.
[41] A. Warburton, T.E. Markowitz, J.P. Katz, J.M. Pipas, A.A. McBride, Recurrent integration of human papillomavirus genomes at transcriptional regulatory hubs, 2021 61, npj Genomic Med 6 (1) (2021) 1–15, https://doi.org/10.1038/s41525-021-00264-y.
[42] E.J. Crosbie, M.H. Einstein, S. Franceschi, H.C. Kitchener, Human papillomavirus and cervical cancer, Lancet 382 (9895) (2013) 889–899, https://doi.org/10.1016/S0140-6736(13)60022-7.
[43] S. Lagström, A.H. Løvestad, S.U. Umu, et al., HPV16 and HPV18 type-specific APOBEC3 and integration profiles in different diagnostic categories of cervical samples, Tumour Virus Res (2021) 12, https://doi.org/10.1016/J.TVR.2021.200221.
[44] R.D. Burk, Z. Chen, C. Saller, et al., Integrated genomic and molecular characterization of cervical cancer, Nature 543 (7645) (2017) 378–384, https://doi.org/10.1038/nature21386.
[45] S. Vinokurova, N. Wentzensen, I. Kraus, et al., Type-dependent integration frequency of human papillomavirus genomes in cervical lesions, Cancer Res. 68 (1) (2008) 307–313, https://doi.org/10.1158/0008-5472.CAN-07-2754.
[46] S. Lagström, S.U. Umu, M. Lepistö, et al., TaME-seq: an efficient sequencing approach for characterisation of HPV genomic variability and chromosomal integration, Sci. Rep. 9 (1) (2019) 524, https://doi.org/10.1038/s41598-018-36669-6.
[47] A. Tropé, K. Sjøborg, A. Eskild, et al., Performance of human papillomavirus DNA and mRNA testing strategies for women with and without cervical neoplasia, J. Clin. Microbiol. 47 (8) (2009) 2458–2464, https://doi.org/10.1128/JCM.01863-08.
[48] A. Tropé, K.D. Sjøborg, M. Nygård, et al., Cytology and human papillomavirus testing 6 to 12 months after ASCUS or LSIL cytology in organized screening to predict high-grade cervical neoplasia between screening rounds, J. Clin. Microbiol. 50 (6) (2012) 1927–1935, https://doi.org/10.1128/JCM.00265-12.
[49] J.J. Kozich, S.L. Westcott, N.T. Baxter, S.K. Highlander, P.D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform, Appl. Environ. Microbiol. 79 (17) (2013) 5112–5120, https://doi.org/10.1128/AEM.01043-13.
[50] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (4) (2015) 357–360, https://doi.org/10.1038/nmeth.3317.
[51] K. Van Doorslaer, Q. Tan, S. Xirasagar, et al., The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis, Nucleic Acids Res. 41 (Database issue) (2013) D571–D578, https://doi.org/10.1093/nar/gks984.
[52] S.M. Kiełbasa, R. Wan, K. Sato, P. Horton, M.C. Frith, Adaptive seeds tame genomic sequence comparison, Genome Res. 21 (3) (2011) 487–493, https://doi.org/10.1101/gr.113985.110.
[53] A. Untergasser, I. Cutcutache, T. Koressaar, et al., Primer3–new capabilities and interfaces, Nucleic Acids Res. 40 (15) (2012), https://doi.org/10.1093/NAR/GKS596.
[54] T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models, Biom. J. 50 (3) (2008) 346–363, https://doi.org/10.1002/BIMJ.200810425.
[55] A. Repesa, HPV Chromosomal Integration as a Biomarker for Cancer Progression, Master's thesis, OsloMet-Storbyuniversitetet., 2021 (May), https://hdl.handle.net/11250/2836505.
[56] M. Rusan, Y.Y. Li, P.S. Hammerman, Genomic landscape of human papillomavirus-associated cancers, Clin. Cancer Res. 21 (9) (2015) 2009–2019, https://doi.org/10.1158/1078-0432.CCR-14-1101.
[57] M. Cullen, J.F. Boland, M. Schiffman, et al., Deep sequencing of HPV16 genomes: a new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection, Papillomavirus Res (2015;1(June) 3–11, https://doi.org/10.1016/j.pvr.2015.05.004.
[58] D.L. Faden, K.A.L. Kuhs, M. Lin, et al., APOBEC mutagenesis is concordant between tumor and viral genomes in HPV-positive head and neck squamous cell carcinoma, 2021, Vol 13, Page 1666, Viruses 13 (8) (2021) 1666, https://doi.org/10.3390/V13081666.
[59] C.J. Warren, J.A. Westrich, K. Van Doorslaer, D. Pyeon, Roles of APOBEC3A and APOBEC3B in human papillomavirus infection and disease progression, Viruses 9 (8) (2017), https://doi.org/10.3390/v9080233.
[60] K. Chatfield-Reed, S. Gui, W.Q. O'Neill, T.N. Teknos, Q. Pan, HPV33+ HNSCC is associated with poor prognosis and has unique genomic and immunologic landscapes, Oral Oncol. 100 (2020), 104488, https://doi.org/10.1016/J.ORALONCOLOGY.2019.104488.
[61] A. Chakravarthy, I. Reddin, S. Henderson, et al., Integrated analysis of cervical squamous cell carcinoma cohorts from three continents reveals conserved subtypes of prognostic significance, 2020.04.02, *bioRxiv*. December (2021), 019711, https://doi.org/10.1101/2020.04.02.019711.
[62] A. Willemsen, I.G. Bravo, Origin and evolution of papillomavirus (onco)genes and genomes, Philos Trans R Soc B Biol Sci 374 (1773) (2019), https://doi.org/10.1098/rstb.2018.0303.
[63] N. Egawa, Q. Wang, H.M. Griffin, et al., HPV16 and 18 genome amplification show different E4-dependence, with 16E4 enhancing E1 nuclear accumulation and replicative efficiency via its cell cycle arrest and kinase activation functions, PLoS Pathog. 13 (3) (2017), e1006282, https://doi.org/10.1371/journal.ppat.1006282.
[64] S. de Sanjose, W.G.V. Quint, L. Alemany, et al., Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, Lancet Oncol. 11 (11) (2010) 1048–1056, https://doi.org/10.1016/S1470-2045(10)70230-8.
[65] M. Demarco, N. Hyun, O. Carter-Pokras, et al., A study of type-specific HPV natural history and implications for contemporary cervical cancer screening programs, EClinicalMedicine 22 (2020), 100293, https://doi.org/10.1016/J.ECLINM.2020.100293/ATTACHMENT/A07C2470-AA58-47A1-9593-962D28F8F692/MMC1.DOCX.
[66] P. Guan, R. Howell-Jones, N. Li, et al., Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer, Int. J. Cancer 131 (10) (2012) 2349–2359, https://doi.org/10.1002/IJC.27485.
[67] F. Poulain, N. Lejeune, K. Willemart, N.A. Gillet, Footprint of the host restriction factors APOBEC3 on the genome of human viruses, PLoS Pathog. 16 (8) (2020), e1008718, https://doi.org/10.1371/JOURNAL.PPAT.1008718.
[68] Z. Hu, D. Zhu, W. Wang, et al., Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism, Nat. Genet. 47 (2) (2015) 158–163, https://doi.org/10.1038/ng.3178.
[69] R. Zhang, C. Shen, L. Zhao, et al., Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration contributes to HPV-related cervical carcinogenesis, Int. J. Cancer 138 (5) (2016) 1163–1174, https://doi.org/10.1002/ijc.29872.
[70] M. Dürst, C.M. Croce, L. Gissmann, E. Schwarz, K. Huebner, Papillomavirus Sequences Integrate Near Cellular Oncogenes in Some Cervical Carcinomas (Viral DNA Integration/c-Myc/Genital Cancer), vol. 84, 1987.
[71] M. Dong, Z. Dong, X. Zhu, Y. Zhang, L. Song, Long non-coding RNA MIR205HG regulates KRT17 and tumor processes in cervical cancer via interaction with SRSF1, Exp. Mol. Pathol. 111 (2019), https://doi.org/10.1016/J.YEXMP.2019.104322.
[72] W. Xu, W. Yang, C. Wu, X. Ma, H. Li, J. Zheng, Enolase 1 correlated with cancer progression and immune-infiltrating in multiple cancer types: a pan-cancer analysis, Front. Oncol. 10 (2021) 3391, https://doi.org/10.3389/FONC.2020.593706/BIBTEX.
[73] C.L. Lin, T.H. Ying, S.F. Yang, et al., MTA2 silencing attenuates the metastatic potential of cervical cancer cells by inhibiting AP1-mediated MMP12 expression via the ASK1/MEK3/p38/YB1 axis, 2021 125, Cell Death Dis. 12 (5) (2021) 1–12, https://doi.org/10.1038/s41419-021-03729-1.
[74] E.M. Siegel, S. Eschrich, K. Winter, et al., Epigenomic characterization of locally advanced anal cancer: an RTOG 98-11 specimen study, Dis. Colon Rectum 57 (8) (2014) 941, https://doi.org/10.1097/DCR.0000000000000160.
[75] H. Niitsu, T. Hinoi, Y. Kawaguchi, et al., KRAS mutation leads to decreased expression of regulator of calcineurin 2, resulting in tumor proliferation in colorectal cancer, 2016 58, Oncogene 5 (8) (2016) e253, https://doi.org/10.1038/oncsis.2016.47. e253.
[76] Y. Li, H. Wang, H. Huang, Long non-coding RNA MIR205HG function as a ceRNA to accelerate tumor growth and progression via sponging miR-122–5p in cervical

cancer, Biochem. Biophys. Res. Commun. 514 (1) (2019) 78–85, https://doi.org/10.1016/J.BBRC.2019.04.102.
[77] L. Yin, Y. Zhang, L. Zheng, Analysis of differentially expressed long non-coding RNAs revealed a pro-tumor role of MIR205HG in cervical cancer, Mol. Med. Rep. 25 (2) (2022) 1–8, https://doi.org/10.3892/MMR.2021.12558/HTML.
[78] Y. Zheng, X. Li, Y. Jiao, C. Wu, High-risk human papillomavirus oncogenic E6/E7 mRNAs splicing regulation, Front. Cell. Infect. Microbiol. (2022) 790, https://doi.org/10.3389/FCIMB.2022.929666, 0.
[79] T. Mesplède, D. Gagnon, F. Bergeron-Labrecque, et al., p53 degradation activity, expression, and subcellular localization of E6 proteins from 29 human papillomavirus genotypes, J. Virol. 86 (1) (2012) 94–107, https://doi.org/10.1128/jvi.00751-11.
[80] J.E. Leeman, Y. Li, A. Bell, et al., Human papillomavirus 16 promotes microhomology-mediated end-joining, Proc. Natl. Acad. Sci. U. S. A. 116 (43) (2019) 21573–21579, https://doi.org/10.1073/PNAS.1906120116/-/DCSUPPLEMENTAL.
[81] N. Sakakibara, R. Mitra, A.A. McBride, The papillomavirus E1 helicase activates a cellular DNA damage response in viral replication foci, J. Virol. 85 (17) (2011) 8981–8995, https://doi.org/10.1128/JVI.00541-11/ASSET/A7D40B6F-A2F2-49CA-A9B3-AAE6F2F9516B/ASSETS/GRAPHIC/ZJV9990949670009.JPEG.
[82] V.M. Williams, M. Filippova, V. Filippov, K.J. Payne, P. Duerksen-Hughes, Human papillomavirus type 16 E6* induces oxidative stress and DNA damage, J. Virol. 88 (12) (2014) 6751–6761, https://doi.org/10.1128/JVI.03355-13.
[83] R. Senapati, N.N. Senapati, B. Dwibedi, Molecular mechanisms of HPV mediated neoplastic progression, Infect. Agents Cancer 11 (1) (2016), https://doi.org/10.1186/S13027-016-0107-4.
[84] S. Duensing, K. Münger, The human papillomavirus type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability, Cancer Res. 62 (23) (2002) 7075–7082.