



## HPV16 and HPV18 type-specific APOBEC3 and integration profiles in different diagnostic categories of cervical samples

Sonja Lagström<sup>a,b,c</sup>, Alexander Hesselberg Løvestad<sup>d</sup>, Sinan Uğur Umu<sup>b</sup>, Ole Herman Ambur<sup>d</sup>, Mari Nygård<sup>b</sup>, Trine B. Rounge<sup>b,e,\*\*,1</sup>, Irene Kraus Christiansen<sup>a,f,\*,1</sup>

<sup>a</sup> Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway

<sup>b</sup> Department of Research, Cancer Registry of Norway, Oslo, Norway

<sup>c</sup> Institute of Clinical Medicine, University of Oslo, Oslo, Norway

<sup>d</sup> Faculty of Health Sciences, OsloMet, Oslo Metropolitan University, Oslo, Norway

<sup>e</sup> Department of Informatics, University of Oslo, Oslo, Norway

<sup>f</sup> Department of Clinical Molecular Biology (EpiGen), Division of Medicine, Akershus University Hospital and University of Oslo, Lørenskog, Norway

### ARTICLE INFO

#### Keywords:

Human papillomavirus  
Minor nucleotide variation  
APOBEC3  
Chromosomal integration  
Viral genomic deletion

### ABSTRACT

Human papillomavirus (HPV) 16 and 18 are the most predominant types in cervical cancer. Only a small fraction of HPV infections progress to cancer, indicating that additional factors and genomic events contribute to the carcinogenesis, such as minor nucleotide variation caused by APOBEC3 and chromosomal integration.

We analysed intra-host minor nucleotide variants (MNVs) and integration in HPV16 and HPV18 positive cervical samples with different morphology. Samples were sequenced using an HPV whole genome sequencing protocol TaME-seq. A total of 80 HPV16 and 51 HPV18 positive samples passed the sequencing depth criteria of 300× reads, showing the following distribution: non-progressive disease (HPV16 n = 21, HPV18 n = 12); cervical intraepithelial neoplasia (CIN) grade 2 (HPV16 n = 27, HPV18 n = 9); CIN3/adenocarcinoma *in situ* (AIS) (HPV16 n = 27, HPV18 n = 30); cervical cancer (HPV16 n = 5).

Similar numbers of MNVs in HPV16 and HPV18 samples were observed for most viral genes, with the exception of HPV18 E4 with higher numbers across clinical categories. APOBEC3 signatures were observed in HPV16 lesions, while similar mutation patterns were not detected for HPV18. The proportion of samples with integration was 13% for HPV16 and 59% for HPV18 positive samples, with a noticeable portion located within or close to cancer-related genes.

### 1. Introduction

A persistent infection with one of the carcinogenic HPV genotypes is accepted as a necessary cause of cervical cancer development [1]. Of the 12 carcinogenic types [2], HPV16 and HPV18 are associated with about 70% of all cervical cancers [3]. HPV16 is predominantly associated with squamous cell carcinomas (SCC), while HPV18 is more often detected in adenocarcinomas [3], suggesting that these HPV types differ in their target cell specificity [4]. Nevertheless, only a small fraction of HPV infections will persist and progress to cancer [5], indicating that additional factors and genomic events are necessary for the HPV-induced carcinogenic process.

The 7.9 kb double stranded HPV DNA genome consists of early region (E1, E2, E4-7) genes, late region (L1, L2) genes, an upstream regulatory region (URR) and a short non-coding region (NCR) between the genes E5 and L2 [6,7]. To date, more than 200 HPV genotypes have been identified, based on at least 10% difference within the conserved L1 gene sequence [8]. HPV types harbouring minor genetic variation are grouped into lineages (1–10% whole genome nucleotide difference) and sublineages (0.5–1.0% difference) [9]. HPV evolve slowly partly since the HPV genome replication is dependent on host cell high-fidelity polymerases [10]. However, recent studies have revealed variability below the level of HPV sublineages. These are non-lineage genetic variants, which may at low frequencies indicate intra-host viral diversification and evolution [11–13].

\* Corresponding author. Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway.

\*\* Corresponding author. Department of Research, Cancer Registry of Norway, Oslo, Norway.

E-mail addresses: [trine.rounge@krefregisteret.no](mailto:trine.rounge@krefregisteret.no) (T.B. Rounge), [irene.kraus.christiansen@ahus.no](mailto:irene.kraus.christiansen@ahus.no) (I.K. Christiansen).

<sup>1</sup> Equal contribution.

### Abbreviations

AID	activation-induced cytidine deaminase
AIS	adenocarcinoma <i>in situ</i>
ASC-US	atypical squamous cells of undetermined significance
CIN	cervical intraepithelial neoplasia
dN/dS	ratio of non-synonymous to synonymous substitutions
HPV	human papillomavirus
LSIL	low-grade squamous intraepithelial lesion
MNV	minor nucleotide variant
NCR	non-coding region
ncRNA	non-coding RNA
NGS	next-generation sequencing
SCC	squamous cell carcinoma
URR	upstream regulatory region
UTR	untranslated region

The generation of viral genetic variants is caused by various stochastic or targeted mutagenic processes [14]. One of the targeted mechanisms suggested to cause MNVs and impact HPV mutational drift involves the anti-viral host-defence enzyme apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3 (APOBEC3) proteins [15]. APOBEC3 proteins are cytidine deaminases causing deoxycytidine (C) to deoxythymidine (T) mutations during viral replication. The mutations can lead to defects in viral genome replication necessary for the viral life cycle [16]. APOBEC3 mutational signatures have been found in the human genome in cervical cancers [17], as well as in HPV genomes in cervical pre-cancerous and cancer samples [11,18,19], and has recently been associated with viral clearance [20]. APOBEC3A may function as a HPV restriction factor [15] and APOBEC3B has been shown to be upregulated by HPV [21]. The two enzymes APOBEC3A and APOBEC3B display preference for the motifs YTCA (Y = pyrimidine) and RTCA (R = purine), respectively [22]. Findings of hypovariability of the E7 gene suggest negative selection opposite of APOBEC3-related editing and an essential gene conservation for progression to cancer [23,24].

HPV integration into the host genome is regarded as a driving event in cervical carcinogenesis and is observed in >80% of HPV-induced cancers [25]. Integrations causing disruption or complete deletion of the E1 or E2 gene result in constitutive expression of the viral E6 and E7 oncogenes [26], leading to inactivation of cell cycle checkpoints and genomic instability [27]. Integration may also lead to disruption of host genes, such as tumour-suppressor genes or negative regulators of oncogenes, modified expression of adjacent genes, as well as other genomic alterations, which may promote HPV-induced carcinogenesis [28–30]. In high-grade lesions and cancers, integrations in certain chromosomal loci, including loci 3q28, 8q24.21 and 13q22.1, have been reported more often than in other loci [31], suggesting selective growth advantages for cells with site-specific integrations in e.g. important regulatory genes. Increasing integration frequencies have been reported upon comparison of cervical precancerous and cancer lesions [32,33].

Recently, we developed a novel next-generation sequencing (NGS) strategy TaME-seq for simultaneous analysis of HPV genomic variability and chromosomal integration [34]. Employing the TaME-seq method, we have explored HPV16 and HPV18 intra-host genomic variability and integration in HPV positive cervical samples with different morphologies. Differences in HPV variability between the diagnostic categories may shed light on intra-host viral genome dynamics and evolution processes in cervical carcinogenesis. In addition, integration analysis will contribute to a better understanding of this event during HPV-induced carcinogenesis.

## 2. Material and methods

### 2.1. Sample selection

Cervical cell samples have previously been collected from women attending the cervical cancer screening program in Norway between January 2005 and April 2008. Samples were collected in ThinPrep PreservCyt solution (Hologic, Marlborough, MA) and pelleted before storage at  $-80^{\circ}\text{C}$ . The samples were stored in a research biobank at Akershus University Hospital, consisting of both the cell material and extracted DNA. Recruitment criteria and HPV detection and genotyping have been described previously [35,36]. Cytology samples were previously analysed for HPV using the AmpliCor HPV DNA test (Roche Diagnostics, Switzerland) followed by genotyping by Linear Array (Roche Diagnostics, Switzerland) and PreTect HPV-Proofer (PreTect AS, Norway).

In this study, primarily DNA was used for downstream analyses; for some samples, DNA extraction had to be performed from the cell material. DNA extraction was performed using the automated NucliSENS easyMag platform (BioMerieux Inc., France) with off-board lysis. All samples in the biobank that were positive for HPV16 and/or HPV18, alone or together with other HPV types, by one or both of the genotyping methods were included in the study, with the exception of HPV16 CIN3 samples for which a random selection of 50 samples were included. In total, 157 HPV16 positive samples and 75 HPV18 positive samples were subjected to sequencing (Table 1). All samples were allocated to mutually exclusive categories based on the HPV type and the diagnostic categories of non-progressive disease, histologically confirmed cervical intraepithelial neoplasia (CIN) grade 2 (CIN2), CIN3/adenocarcinoma *in situ* (AIS) and cancer. The non-progressive disease category included samples from women with normal cytology also having normal cytology the preceding two years and with no previous history of treatment for cervical neoplasia (HPV16  $n = 24$ , HPV18  $n = 3$ ), and samples from women with atypical squamous cells of undetermined significance (ASC-US) or low-grade squamous intraepithelial lesions (LSIL) with no follow-up diagnosis within four years subsequent to the diagnosis (HPV16  $n = 31$ , HPV18  $n = 13$ ). For the CIN2, CIN3/AIS and cancer categories, sequencing was performed on cell samples taken at the time of conisation; cytological examination of these samples was not performed. The cancer category included SCC ( $n = 4$ ) and adenocarcinoma ( $n = 1$ ) samples.

### 2.2. Library preparation and sequencing

Library preparation was performed using the TaME-seq method as described previously [34]. In brief, samples were subjected to tagmentation using Nextera DNA library prep kit (Illumina, Inc., San Diego, CA), following target enrichment performed by multiplex PCR using HPV primers and a combination of i7 index primers [37] and i5 index primers from the Nextera index kit (Illumina, Inc., San Diego, CA). Sequencing was performed on the HiSeq2500 platform with 125 bp paired-end reads.

### 2.3. Sequence alignment

Data was analysed by an in-house bioinformatics pipeline as described previously [34]. Reads were mapped to human genome (GRCh38/hg38) using HISAT2 (v2.1.0) [38]. HPV16 and HPV18 reference genomes were obtained from the PaVE database (<https://pave.niaid.nih.gov>). Mapping statistics and sequencing coverage were calculated using the Pysam package [39] with an in-house Python (v3.5.4) script. Downstream analysis was performed using an in-house R (v3.5.1) script. Samples with a mean sequencing depth of  $<300\times$  were excluded from the further analysis.

**Table 1**  
Number of samples and mean mappings statistics in each HPV16 and HPV18 diagnostic category.

Diagnostic category	Sequenced samples	Analysed samples	Mean age	Mean numbers in the analysed samples				
				Raw reads	Reads mapped to target HPV	Mean coverage	Fraction of genome covered by min. 100×	
<b>HPV16</b>								
Normal <sup>a</sup>	24	2 <sup>c</sup>	21	49 (32–68)	1.4 M	1.1 M	13516	0.78
ASC-US/LSIL <sup>b</sup>	31	19 <sup>c</sup>		33 (19–54)				
CIN2 <sup>c</sup>	47		27	31 (17–61)	0.6 M	0.4 M	4711	0.69
CIN3/AIS <sup>c</sup>	50		27	34 (22–54)	1.0 M	0.8 M	9616	0.76
Cancer <sup>c,d</sup>	5		5	30 (25–39)	2.4 M	1.7 M	20850	0.67
Total	157		80					
<b>HPV18</b>								
Normal <sup>a</sup>	3	1 <sup>e</sup>	12	49 (47–52)	38.8 M	23.4 M	292143	0.86
ASC-US/LSIL <sup>b</sup>	13	11 <sup>e</sup>		33 (20–49)				
CIN2 <sup>c</sup>	13		9	34 (20–44)	77.1 M	36.5 M	431649	0.86
CIN3/AIS <sup>c</sup>	46		30	34 (24–54)	25.5 M	12.2 M	147747	0.82
Cancer	0	–	–					
Total	75		51					

<sup>a</sup> By cytology.

<sup>b</sup> By cytology; no cell abnormalities within 4-year follow-up.

<sup>c</sup> Cytology taken at the time of consiation, with the histological diagnosis presented.

<sup>d</sup> Includes cases of SCC (n = 4) and adenocarcinoma (n = 1).

<sup>e</sup> Non-progressive category, samples combined for analysis.

#### 2.4. Sequence variation analysis

Mapped nucleotide counts over the HPV genomes and average mapping quality values for each nucleotide were retrieved from the HISAT sequence alignment. Variant calling was performed using an in-house R (v3.5.1) script. Nucleotides seen  $\leq 2$  times in each position and nucleotides with mean Phred quality score of  $< 20$  were filtered out. Since the analysis focused on the intra-host MNVs, the variant calling was performed independent of the reference genome; the most frequent base in each position was called as the major nucleotide and the second most abundant base as the MNV. Both F and R nucleotide counts from the same sample, obtained independently from separate amplification reactions, were combined and variant allele frequencies were calculated for each genomic position. If MNVs called from the two separate reactions were discordant, the highest covered MNV was used. Genomic positions covered with  $< 100\times$  were filtered out. MNVs were called if the MNV frequency was  $> 1\%$ . HPV16 and HPV18 have homopolymeric T tracts in NCR (HPV16:4156–4173, HPV16:4183–4212, HPV18:4198–4234); these regions may be prone to polymerase or sequencing errors and were filtered out.

The ratio of non-synonymous to synonymous substitutions (dN/dS) was calculated to indicate potential positive (new MNVs favoured) or negative (new MNVs eliminated) selection affecting protein-coding genes. 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, including information on the bases immediately 5' and 3' of the mutated base. To differentiate APOBEC3A and APOBEC3B activity, an extended mutational signature analysis was conducted on mutations in the genomic context YTCA and RTCA, respectively. Analysis was performed using an in-house R (v3.5.1) script.

#### 2.5. Detection of chromosomal integration

Integration site detection was performed as described previously [34]. In brief, a two-step analysis strategy was employed to identify read pairs spanning integration sites. First, read pairs with one read mapped to HPV and the other to the human chromosome were identified using HISAT2. Second, unmapped reads were re-mapped using the LAST (v876) aligner (options -M -C2) [40] to increase detections of the above mentioned read pairs. Reads sharing the same start and end coordinates

were considered as potential PCR duplicates and were excluded. Selected integration sites were confirmed by PCR amplification and Sanger sequencing 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). Samples with a mean depth of  $> 1000\times$  and  $< 85\%$  of the genome covered by minimum 100× were manually inspected using IGV (v2.3.90) to detect HPV genomic deletions.

#### 2.6. Functional annotation of genes within or close to integration sites

Nearest gene, with a transcription start site within 100 kb from the integration site, was identified using Ensembl. Gene2function (<http://www.gene2function.org>) and Genecards (<https://www.genecards.org>) were used to annotate the molecular function and disease phenotype of each gene. SNP associations in the GWAS Catalog [41] were retrieved from Genecards. Genes involved in cell cycle regulation, cell proliferation, apoptosis, tumour suppressor mechanisms, cancer-related pathways, or genes interacting with these pathways, or genes with direct cancer-related SNP associations, were termed as cancer-related genes. The integration sites were manually inspected using Geneious Prime (v.2019.0.4) to investigate whether the integration site was located in exons, introns or UTRs. Information regarding regulatory elements, including promoters, promoter flanking regions, enhancers and CTCF-binding sites, was retrieved from Ensembl regulatory build [42]. Integration sites in retained introns, ncRNA and anti-sense RNA were reported if they had a transcript support level of 1 or 2.

#### 2.7. Statistical analysis

Statistical analyses were done in R (v3.5.1). The Kruskal-Wallis test was used to examine differences in numbers and frequencies of MNVs and integrations between the groups. A p-value of  $< 0.05$  was considered statistically significant.

#### 2.8. Ethical considerations

This study was approved by the Regional Committee for Medical and Health Research Ethics, Oslo, Norway (REK 2017/447). Written informed consent has been obtained from all study participants.

### 3. Results

#### 3.1. Characteristics and sequencing statistics

This study included 232 HPV16 and HPV18 positive cervical cell samples which were categorised according to cytology or histology diagnosis. A total of 80 HPV16 positive samples and 51 HPV18 positive samples, allocated to diagnostic categories of non-progressive disease, CIN2, CIN3/AIS and cancer, passed the strict sequencing depth criteria necessary for further analyses of minor nucleotide variation and integration. In total, 1.05 billion read pairs were analysed. The mean sequencing coverage per sample in the different categories ranged from 4711 (CIN2) to 20850 (cancer) for HPV16 positive samples and from 147747 (CIN3/AIS) to 431649 (CIN2) for HPV18 positive samples. On average, the samples had 77.7% of the genome covered with a minimum depth of  $100\times$  (Table 1).

#### 3.2. Minor nucleotide variation profiles similar for HPV16 and HPV18

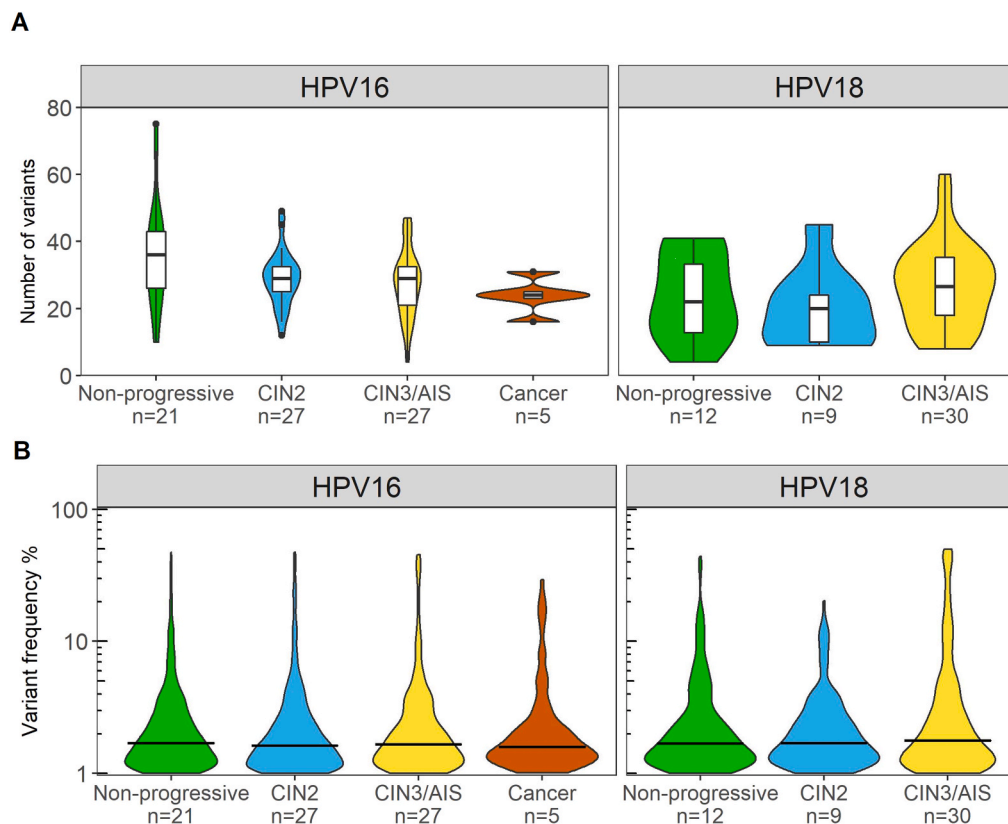
Overall, the number of MNVs was similar in HPV16 and HPV18 positive samples, and between the diagnostic categories. In total, 3669 MNVs were found in all 131 samples. In HPV16 positive samples, the mean number of MNVs found in the non-progressive category was 36 per sample, 29 in the CIN2 category, 27 in the CIN3/AIS category, and 24 in the cancer category. Corresponding numbers for HPV18 positive samples were 24, 20, and 27 for the non-progressive, CIN2 and CIN3/AIS categories, respectively (Fig. 1A). HPV16 positive samples had mean MNV frequencies of 2.8% for non-progressive, 2.9% for CIN2, 3.3% for CIN3/AIS and 3.0% for cancer categories. For HPV18 positive samples, the mean MNV frequencies were 3.1% for non-progressive, 2.6% for CIN2 and 5.0% for CIN3/AIS categories (Fig. 1B). Statistical analysis was performed; the mean numbers and MNV frequencies were not statistically different between the HPV types or the diagnostic groups within an HPV type.

#### 3.3. Different level of variation in HPV16 and HPV18 genes

HPV MNVs occurred throughout all HPV genes (Fig. 2A). A higher degree of variation was observed in the HPV18 E4 gene throughout the different diagnostic categories. The dN/dS patterns for HPV16 showed mostly nonsynonymous variants ( $dN/dS > 1$ ), while a considerable part of HPV18 genes had equal amounts of nonsynonymous and synonymous variants ( $dN/dS \approx 1$ ) (Fig. 2B). Strikingly, several HPV16 genes showed signs of positive selection, i.e. a preference for non-synonymous mutations (dN) over synonymous mutations (dS). HPV16 E6 had the most pronounced dN/dS ratio of 6. In contrast, the E7 gene in the same samples had a dN/dS ratio of 0.4, indicating neutral or negative selection. Over all, diagnostic categories and in both HPV types, the E2 gene displayed the highest dN/dS ratio, which for HPV18 were consistently  $>2$ . For the other HPV18 genes, the dN/dS ratio was close to 1 across diagnostic categories.

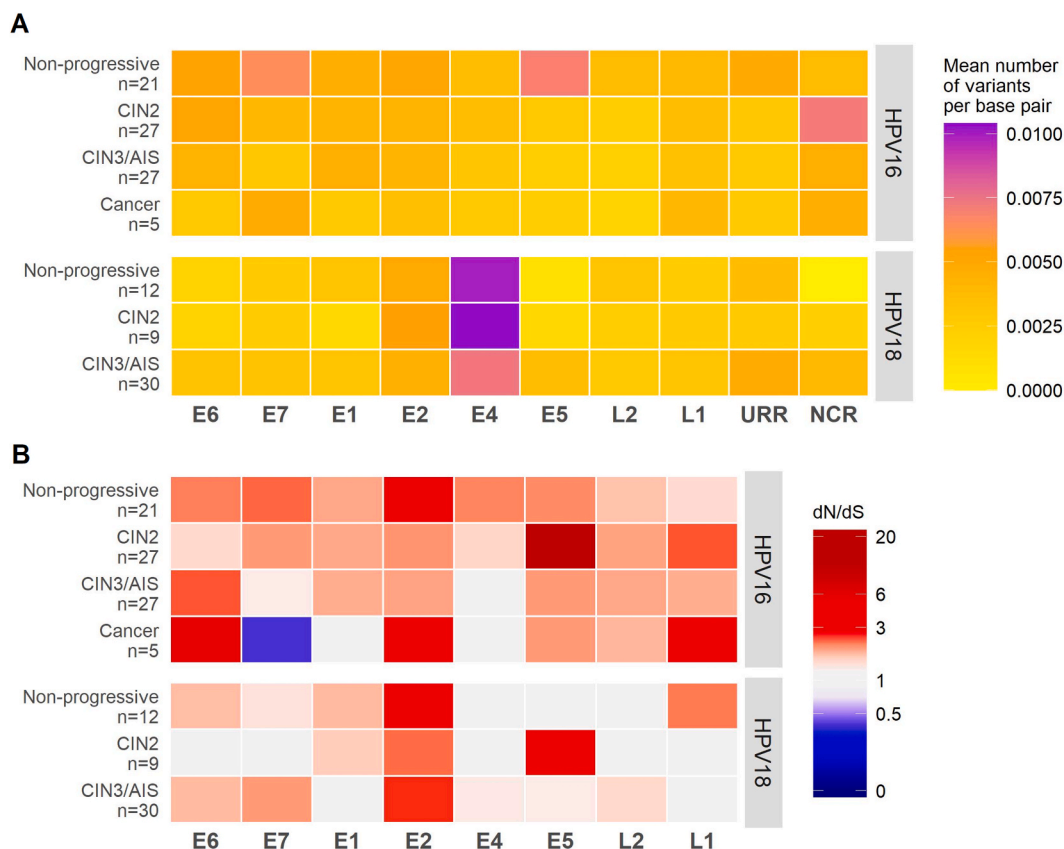
#### 3.4. APOBEC3-related mutational signatures identified in non-progressive and CIN2 samples

Among nucleotide substitutions, predominantly C > T and T > C substitutions were observed across all diagnostic categories (Supplementary Figure S1). The APOBEC3-related C > T substitutions were compared between the different categories and HPV types (Fig. 3). C > T substitutions in the trinucleotide context TCW (W is A or T), a preferred target sequence for the APOBEC3 proteins [43] and a more stringent motif than TCN (N is any nucleotide [44]), was the most prevalent mutational signature type in HPV16 non-progressive samples and to a slightly less extent in HPV16 CIN2 samples. HPV16 CIN3/AIS and cancer samples did not show any preferred signature patterns. Interestingly, HPV18 samples showed different C > T trinucleotide substitution patterns compared to HPV16 samples. In all HPV18 diagnostic categories, C > T substitutions in the trinucleotide context ACA was predominantly observed, while C > T substitutions in the trinucleotide context GCA was



**Fig. 1.** Number of variants and variant frequencies in HPV16 and HPV18 positive samples. A) Number of variants presented as violin plots across the different diagnostic categories shown on x-axis. Violin plot shows the probability density of the data, using kernel density estimation. Box-and-whisker plots are added to show the median number (horizontal line), 25% and 75% percentiles (box), minimum and maximum values (whiskers). Black dots represent outliers. B) Variant frequencies (%) of detected minor variants shown as violin plots across the different diagnostic categories shown on x-axis. The horizontal bar indicates the median variant frequency.





**Fig. 2.** Number of variants, nonsynonymous and synonymous variations in the different HPV genes. A) Heat map with yellow-orange-purple gradient colour-coding representing mean number of variants per sample in HPV16 and HPV18 genomic regions. Number of variants is normalised by the gene length and stratified by the diagnostic category. B) Heat map with blue-white-red gradient colour-coding representing the ratio of non-synonymous to synonymous substitutions (dN/dS) in HPV16 and HPV18 genomic regions across the different diagnostic categories.

the second most prevalent in non-progressive and CIN2 samples. For the extended signature mutational analysis, there were only 15 instances of mutations in the YTCA context in 8 samples while mutations in the RTCA context were not found in any samples in the dataset.

### 3.5. Higher HPV integration frequencies in HPV18 than in HPV16 positive samples

The proportion of samples with integration was 13% (10/80) for HPV16 and 59% (30/51) for HPV18 positive samples (Table 2). The integration frequency was higher in all HPV18 positive diagnostic categories compared to the HPV16 categories. Of the HPV16 positive samples, HPV integration was detected in 4%, 7% and 60% in CIN2, CIN3/AIS and cancer samples, respectively. Corresponding numbers in HPV18 samples were 78% and 53% for CIN2 and CIN3/AIS categories, respectively. The total number of integration sites found in each diagnostic category was in general higher for HPV18 positive samples, ranging from 22 (CIN2) to 60 (CIN3/AIS), while for HPV16 samples, a total of 17 integration sites were identified (Table 2).

In Fig. 4A, the difference between HPV16 and HPV18 positive samples in terms of number of integration sites is illustrated, stratified by diagnostic category. Combined for all diagnostic groups, HPV18 samples had significantly more integration sites than HPV16 samples ( $p$ -value < 0.001). The mean numbers of integration sites per HPV18 positive sample were 3.4, 3.1 and 3.8 for the non-progressive, CIN2 and CIN3/AIS categories, respectively. The mean numbers of integration sites per HPV16 positive sample with observed integration, were 1.3, 2, 1.5 and 2.3 for the non-progressive, CIN2, CIN3/AIS and cancer categories, respectively (Fig. 4A). In total, six HPV16 positive samples and 18 HPV18 positive samples had more than one integration site observed

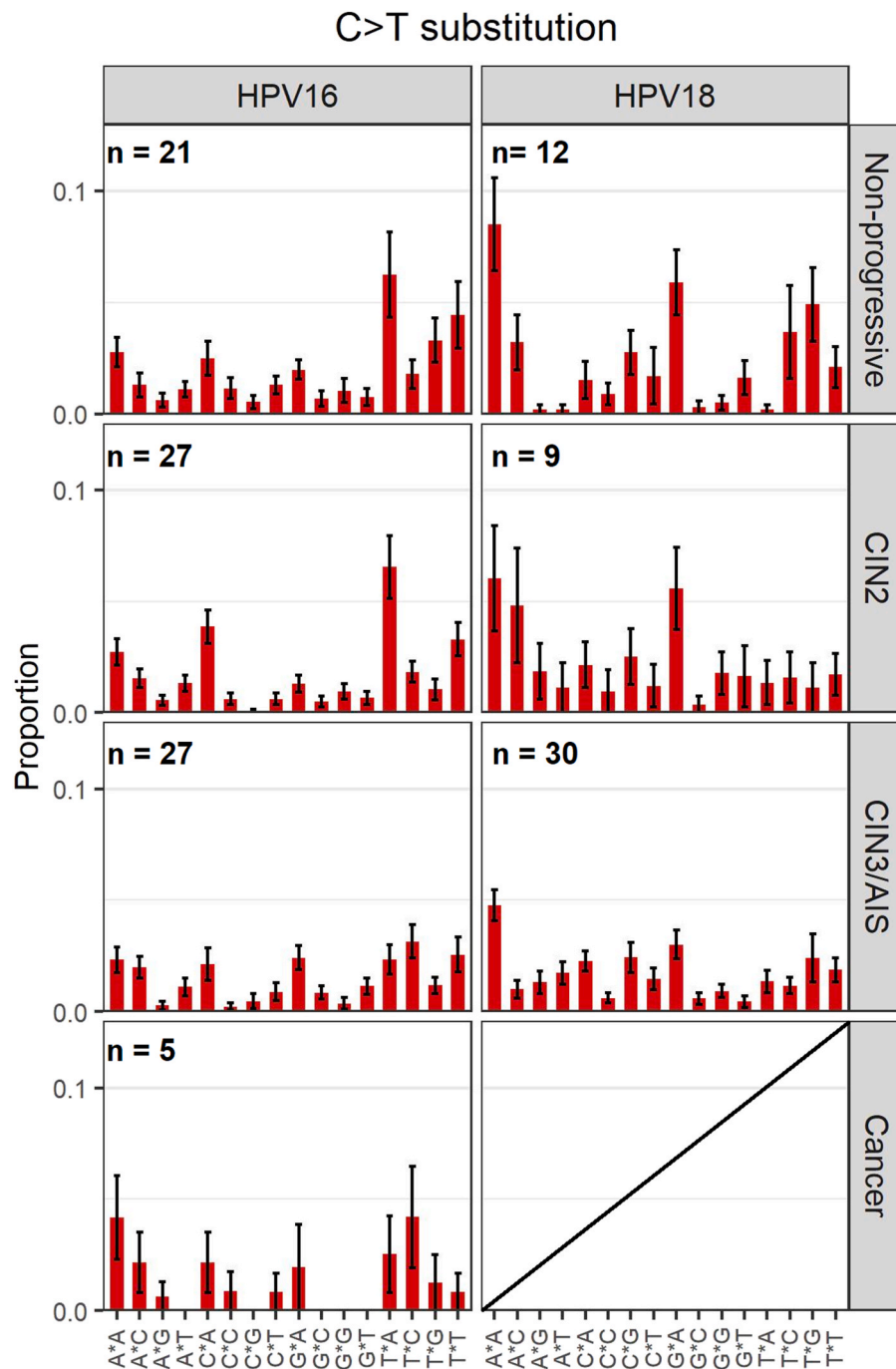
(Supplementary Table S1).

The validation rates of integration sites using Sanger sequencing (good quality chromatograms produced) was 44% (7/16 samples) (Supplementary Table S1, Supplementary Table S2). A PCR product or a smear was identified on agarose gel but no clean chromatogram was seen in additional 44% (7/16) of the reactions (Supplementary Figure S2). Two integration sites, one in HPV16 and one in HPV18 positive sample, both in the non-progressive category, could not be confirmed (Supplementary Table S1).

### 3.6. Break points and deletions in the HPV genome

For HPV16, integration-associated break points in the viral genome were detected in all genes except E4 and E7. Notably, NCR between the E5 and L2 genes, harboured two break points in one cancer sample (Fig. 4B, Supplementary Table S1). In the HPV18 positive samples, break points were located in all HPV genomic regions except NCR. Expected number of break points in each gene relative to gene lengths was estimated with regard to randomness by dividing the total number of break points within a HPV type by the length of the gene. Based on this, breaks were more frequently observed in E1 and NCR in HPV16 samples and in E2, E4 and L2 in HPV18 samples, while L1 and URR were less prone to break (Fig. 4B). For HPV16 and HPV18 combined, break points were located in E1 or E2 in 38%, 38%, 48%, and 57% of all the breaks in non-progressive, CIN2, CIN3/AIS, and cancer categories, respectively (Supplementary Figure S3). All cancer samples had at least one break point in E1 or E2 (Supplementary Table S1).

HPV genomic regions covered with very few or no sequencing reads were considered as deletions according to previous validations [34]. Such deletions were observed in six samples; one HPV16 positive sample



**Fig. 3.** C > T mutational signatures in HPV16 and HPV18 positive samples. The mean proportion of 16 trinucleotide substitution types is shown below the plots across the different diagnostic categories. Error bars represent the standard error of the mean.

(cancer) and five HPV18 positive samples (Supplementary Figure S4). For these samples, human sequences were detected flanking the deleted regions, indicating chromosomal integration. In all six samples, the genomic deletion encompassed the region between E1/E2 and L2. The deletions were either partial, suggesting the presence of both episomal and integrated HPV DNA, or complete with no reads detected for the deleted region.

### 3.7. Integration sites in the human genome

In HPV16 positive samples, integration sites (n = 17) were distributed on 10 chromosomes; for the cancer samples, all integration sites (n = 7) were located on chromosomes 1, 8 or 10 (Fig. 4C). Interestingly, the

integration sites on chromosome 8 were located in the *PVT1* oncogene, in the chromosomal locus 8q24.21 (Supplementary Table S1), previously being defined as an HPV integration hotspot [31]. For the HPV18 positive samples, integration sites (n = 106) were found in all chromosomes except chromosomes 18 and 21 (Fig. 4C). Most HPV18 integration sites were observed on chromosomes 2 and 4. In HPV18 samples, 36% (4/11) of the integration sites on chromosome 4 were located in the previously defined hotspot locus 4q13.3 [31], all from samples diagnosed with CIN2 or CIN3/AIS.

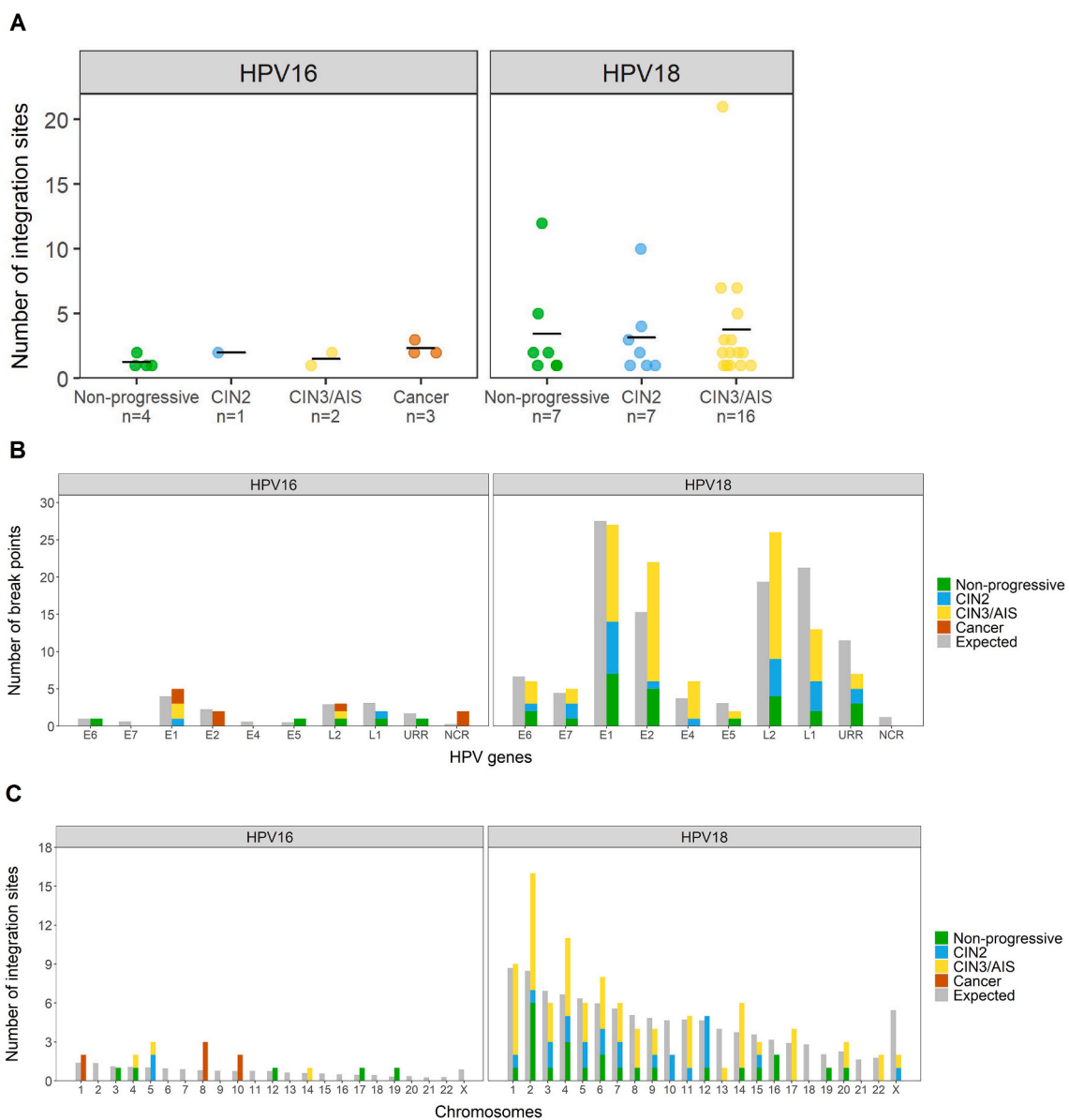
Due to a low frequency of integration events in HPV16 positive samples, HPV16 and HPV18 samples were combined for reporting HPV integrations affecting different human genetic elements. The frequency of integration sites located in human genes ranged from 50 to 71%, with

**Table 2**  
Number of HPV16 and HPV18 positive samples with integration, stratified by the diagnostic categories.

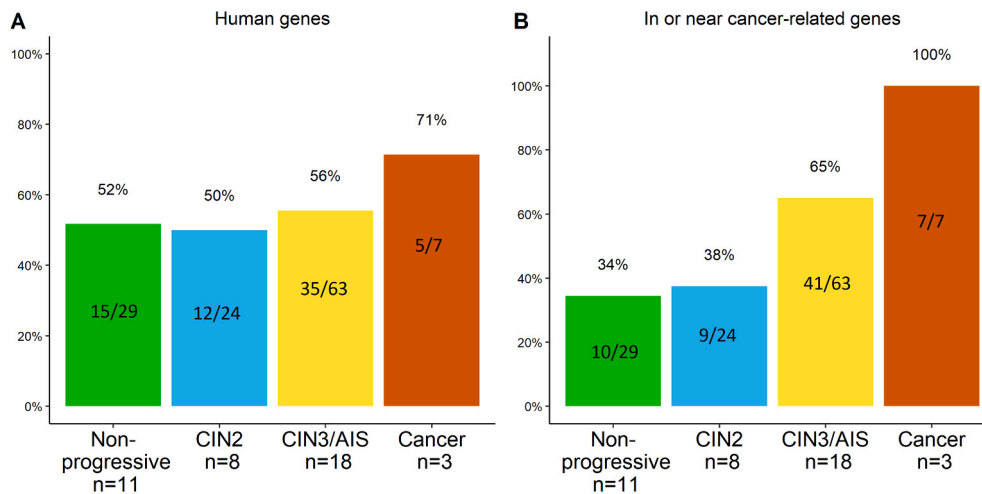
Diagnostic category	Number of samples with integration (Frequency %)	Total number of integration sites
<b>HPV16</b>		
Non-progressive (n = 21)	4 (19%)	5
CIN2 (n = 27)	1 (4%)	2
CIN3/AIS (n = 27)	2 (7%)	3
Cancer (n = 5)	3 (60%)	7
Total (n = 80)	10 (13%)	17
<b>HPV18</b>		
Non-progressive (n = 12)	7 (58%)	24
CIN2 (n = 9)	7 (78%)	22
CIN3/AIS (n = 30)	16 (53%)	60
Total (n = 51)	30 (59%)	106

the highest frequency observed in cancer samples (Fig. 5A). Integration sites were detected in or close to cancer-related genes (Supplementary Table S3) in 100% (7/7) of cancer samples (n = 3), in 65% (41/63) of CIN3/AIS samples (n = 18), in 38% (9/24) of CIN2 samples (n = 8), and in 34% (10/29) in non-progressive samples (n = 11) (Fig. 5B). In individual samples, the highest numbers of integration sites located in or near cancer-related genes was 13/21 in CIN3/AIS, 3/10 in CIN2, and 5/12 in non-progressive samples, all being HPV18 positive (Supplementary Figure S5).

Integration located in exons, introns, regulatory regions, retained introns, non-coding RNA (ncRNA), antisense RNA and untranslated regions (UTRs) varied between the diagnostic groups (Supplementary Table S1). Integration frequency in exons and regulatory regions decreased with lesion severity, while the integration frequency in introns, retained introns and ncRNA increased with lesion severity. Antisense and UTR showed only few integrations in certain diagnostic groups (Supplementary Figure S6).



**Fig. 4.** Chromosomal integration sites and HPV break points in HPV16 and HPV18 positive samples. A) Number of integration sites in samples with observed integration. Each spot in the plot indicates one sample. Total number of samples with integration is specified for each diagnostic category on x-axis. Vertical lines indicate the mean number of integration sites. B) Break points in HPV genes. C) Integration sites in human chromosomes compared to expected number of break points assuming random viral genome integration.



**Fig. 5.** The frequency of integration sites combined for HPV16 and HPV18 A) in human genes, and B) in or near human cancer-related genes. Number of integration sites is indicated inside the bars and total number of samples with integration (n) for each diagnostic category is specified on x-axis.

#### 4. Discussion

This study compares HPV16- and HPV18-associated genomic events, i.e. MNVs and integrations, in normal/ASC-US/LSIL samples from women with no clinical progression; CIN2, CIN3/AIS and cervical cancer samples. We find that these genomic events are strikingly different between HPV16 and HPV18 positive samples. In line with other studies [11,20], we show decreases in APOBEC3-related nucleotide substitutions in HPV16 positive samples of increasing severity. As previously reported [25,45], HPV18 samples show higher integration frequencies compared to HPV16, while we also found an increase in integration frequencies in or in close proximity to cancer-related genes with increasing lesion severity.

In this study, the number and frequency of intra-host MNVs was similar between HPV genotypes and morphological categories. Recent HPV deep sequencing studies, exploring HPV genomic variation with various PCR-based NGS approaches and different variant calling thresholds, show slightly divergent numbers of MNVs [11,20,34]. We found a total of 3669 MNVs in the 131 samples, being in line with studies reporting a high number of HPV variation at the population level [24,46,47], within infected hosts [11,12,34]. A recent study on HPV16 genome stability analysed possible HPV16 sublineage co-infections and observed 20–38 variants in each sample [48], corresponding to the mean numbers of MNVs in this study. The variation was reported not to be due to co-infections, but interpretation of the nucleotide variation source was not further elaborated [48]. The prevalence of sub-lineage co-infections is expected to be low [49].

When investigating the number of MNVs for each region or gene in the HPV genomes, normalised by the gene length, HPV18 E4 showed a higher degree of variation relative to other genomic regions. This is an interesting observation which should be further examined. For HPV16, a higher degree of variation in the NCR was initially observed in the categories CIN2, CIN3/AIS and cancer. However, when filtering out the homopolymeric T tracts in the NCR, the differences between categories subsided. This filtering was done since the T tracts are inherently unstable making it challenging to assign mutations to methodological factors or true biology. Similar variation was not seen for HPV18 positive samples with less homopolymeric tracts. Recent studies document high degrees of variation in HPV16 NCR, but without any biological interpretation [11,23]. The NCR in HPV16 has been characterised to portray a weak promoter activity specific to L2 mRNA expression [50]. Repeat sequences of varying length in NCR have been reported [51] and the NCR has been shown to harbour miRNA binding sites [52]. The loss of miRNA binding sites due to nucleotide variation in NCR was suggested to serve

as a novel mechanism to sustain L2 expression, and thereby justify the potential role of L2 in HPV-induced carcinogenesis [52]. However, an opposite finding has also been reported, showing more variation in NCR in clearing than in persistent HPV16 infections [46].

Ratio of nonsynonymous to synonymous variants (dN/dS) is used as indicator of positive or negative selection occurring over generations within hosts [14]. This ratio may indicate non-random occurrence and persistence of minor nucleotide variability in genes. In this study, the observed nucleotide variations in the HPV16 and HPV18 genes were biased toward nonsynonymous substitutions, being in line with previous results showing a high ratio of non-synonymous nucleotide variation [11]. Only HPV16 E7 had a dN/dS ratio of <1, indicating negative selection and conservation of function. Interestingly, two recent studies reported similar results on strict conservation of the HPV16 E7 gene at the population level [23,24]. A potential source of synonymous and non-synonymous substitutions may be APOBEC3 activity creating C > T substitutions [16]. APOBEC3-related mutations have previously been reported in cervical cancer lesions [11,19,20]. Our finding of APOBEC3-related signatures in the HPV16 positive non-progressive samples indicates that this mechanism is active also in an early stage of infection. The relative amount of variants related to APOBEC3 may at a more severe stage of disease disappear, due to an increase in non-APOBEC3 mutations caused by e.g. hampered DNA repair mechanisms in an increasingly cancerous environment [53]. This study was the first to characterise mutational patterns in HPV18 samples, showing mutation patterns in the trinucleotide context RCA (R is A or G), a target motif for the activation-induced cytidine deaminase (AID) that is a member of the APOBEC protein family [54].

HPV-induced carcinogenesis is a multi-step process that may be facilitated through the disruption of host genes and genomic instability caused by viral integration [28–30]. A high number of integrations in a sample may in itself be a sign of genomic instability, which may further accelerate such events. In our dataset, multiple integration sites were observed in 24 samples, with the maximum of 21 integration sites in one HPV18 sample in the CIN3/AIS category, possibly promoting a higher degree of chromosomal instability. Our results showed a higher number of integration events in HPV18 positive samples compared to HPV16 positive samples, being consistent with previous observations [25,45]. Genomic instability as a consequence of multiple integrations, is further strengthened by finding integrations in the E1 and E2 genes, which might result in overexpression of the viral oncogenes E6 and E7. Previous studies using NGS methodology for HPV integration analysis report disruptions mainly in E1 and E2 genes in samples that have progressed to cancer [55,56]. In addition, we found HPV genomic



deletions in one HPV16 positive cancer sample and in five HPV18 positive samples of all categories. In all of these, the genomic deletion always led to partial or complete loss of E1, E2 and L2. Similar results showing HPV genomic deletions have been reported in cervical carcinomas [57] and HPV positive oropharyngeal squamous cell carcinomas [58]. Interestingly, we also observed integration with break points in NCR in one cancer sample. To our knowledge, this is the first study to report break points in NCR.

Due to the low frequency of integration events in HPV16 positive samples, HPV16 and HPV18 integrations were combined for the analysis of integrations in or close to cancer-related genes. We observed an overall increase in the proportion of integration sites within or close to cancer-related genes with increasing lesion severity. All integrations in the cancer samples occurred within or near the cancer-related genes *PVT1*, *WAC* and *miR-205*. The *PVT1* oncogene, a long non-coding RNA gene, has been associated with multiple cancers including cervical cancer [59]. The *PVT1* gene is located in the chromosomal locus 8q24.21, which is one of the regions previously reported to contain integration sites in cervical carcinomas more often than other loci [31]. Transcription of *PVT1* is regulated by the key tumour suppressor protein p53 and *PVT1* is implicated in regulating the *MYC* oncogene [60]. The *WAC* protein regulates the cell-cycle checkpoint activation in response to DNA damage and is a positive regulator of mTOR, which functions as a key player in the regulation of cell growth and metabolism [61]. The miRNA miR-205 has been implicated in many cancers and targets genes involved in DNA repair, cell cycle control and cancer-related pathways [62]. In the CIN2 and CIN3/AIS categories, 38% and 65% of the integration sites were observed in or close to cancer-related genes, respectively. Interestingly, integration sites in or close to cancer-related genes were also observed in the non-progressive disease category. Whether this might represent one of several components for risk stratification remains to be determined. Our results, together with a recent study [63], have shown that viral integrations may also occur in other genetic elements that are involved in regulation of gene expression, such as ncRNA and UTRs.

NGS protocols with comprehensive analyses of whole HPV genomes, their variability and integrations, enable greater understanding of the role of genomic events during cancer development. By comparatively analysing genomic events, we get a broader picture of the dynamic changes in the HPV genome during malignant cell transformation. HPV16 and HPV18 are to a certain degree associated with different types of invasive cervical cancers [3,4] and may utilise different molecular mechanisms to induce carcinogenesis. Firstly, HPV18 is suggested to cause more genomic instability [4,45] and HPV18 lesions are more aggressively progressing from CIN3 to cancer than HPV16 positive lesions [4]. Furthermore, previously reported results show different DNA methylation patterns [64] and mechanistic signatures of integrations [57] for HPV16 and HPV18, which strengthens the hypothesis of different underlying mechanisms for HPV16- and HPV18-induced cervical carcinogenesis.

Despite the large sample number in total, the sample size in certain diagnostic categories was low, limiting us from performing statistical analyses and drawing conclusions from the given part of the dataset. Some samples, mainly in the non-progressive category, had low sequencing coverage for the HPV genome. This is most likely explained by low viral load, which was not measured in the samples. Low viral load has previously been observed to affect the sequencing yield [13]. Two integration sites in non-progressive samples were not confirmed by Sanger sequencing. This may be explained by sub-optimal PCR primers, PCR conditions, low viral load or may reflect repeated integrations or other genomic structures affecting the PCR reaction. Still, since the NGS data showed clear results, both integration sites were included in the analysis.

## 5. Conclusions

To summarise, we have in this study analysed intra-host HPV minor nucleotide variation, chromosomal integration and genomic deletions in cervical cell samples with different morphology by utilising the TaME-seq protocol [34]. The results show a high number of low-frequency variation, distinct variation patterns and integration frequencies, providing initial insight into dissimilar genomic alterations between HPV16 and HPV18, possibly reflecting differences in the mechanisms of cell transformation induced by the two genotypes. In addition, the study adds to the growing evidence of within-host HPV genomic variability. Cancer registry data with information on future cervical disease or longitudinal studies including patient outcome, preferably with a larger sample size for all diagnostic categories, are needed for further interpretation of different HPV whole genome MNV signatures and to validate the role and importance of viral integrations.

## CRedit authorship contribution statement

**Sonja Lagström:** Writing – original draft, Formal analysis, designed and performed the experiments, analysed the results and drafted the manuscript text. . All authors contributed to writing and approved the final version of the manuscript. **Alexander Hesselberg Løvestad:** Writing – original draft, Formal analysis, analysed the results and contributed to drafting the manuscript. . All authors contributed to writing and approved the final version of the manuscript. **Sinan Uğur Umu:** Writing – original draft, Data curation, Formal analysis, contributed to the data analysis. . All authors contributed to writing and approved the final version of the manuscript. **Ole Herman Ambur:** Writing – original draft, contributed to the study design and result interpretation. . All authors contributed to writing and approved the final version of the manuscript. **Mari Nygård:** Writing – original draft, contributed to the clinical interpretation of the results. . All authors contributed to writing and approved the final version of the manuscript. **Trine B. Rounge:** Writing – original draft, Data curation, Formal analysis, contributed to the study design, data analysis and result interpretation. . All authors contributed to writing and approved the final version of the manuscript. **Irene Kraus Christiansen:** Writing – original draft, managed the sample material, contributed to the study design and result interpretation. All authors contributed to writing and approved the final version of the manuscript.

## Acknowledgements

We thank Mona Hansen for DNA extraction, Hanne Kristiansen for sequencing library preparation, Karin Helmersen for Sanger sequencing, and Marcin W. Wojewodziec for his help with gene annotation of the chromosomal integration sites.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tvr.2021.200221>.

## Funding

This work was funded by a grant from South-Eastern Norway Regional Health Authority (project number 2016020).

## 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.

### Authors' contributions

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

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### References

- [1] F.X. Bosch, et al., The causal relation between human papillomavirus and cervical cancer, *J. Clin. Pathol.* 55 (2002) 244–265.
- [2] IARC working group on the evaluation of carcinogenic risks to humans, biological agents. Volume 100 B. A review of human carcinogens, IARC Monogr. Eval. Carcinog. Risks Hum. 100 (2012) 1–441.
- [3] S. de Sanjose, et al., Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, *Lancet Oncol.* 11 (2010) 1048–1056, [https://doi.org/10.1016/s1470-2045\(10\)70230-8](https://doi.org/10.1016/s1470-2045(10)70230-8).
- [4] W.A. Tjalma, et al., Differences in human papillomavirus type distribution in high-grade cervical intraepithelial neoplasia and invasive cervical cancer in Europe, *Int. J. Canc.* 132 (2013) 854–867, <https://doi.org/10.1002/ijc.27713>.
- [5] H. zur Hausen, Papillomaviruses and cancer: from basic studies to clinical application, *Nat. Rev. Canc.* 2 (2002) 342–350, <https://doi.org/10.1038/nrc798>.
- [6] H.U. Bernard, Taxonomy and phylogeny of papillomaviruses: an overview and recent developments, *Infect. Genet. Evol.* 18 (2013) 357–361, <https://doi.org/10.1016/j.meegid.2013.03.011>.
- [7] B. Smith, et al., Sequence imputation of HPV16 genomes for genetic association studies, *PLoS One* 6 (2011), e21375, <https://doi.org/10.1371/journal.pone.0021375>.
- [8] 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>.
- [9] R.D. Burk, A. Harari, Z. Chen, Human papillomavirus genome variants, *Virology* 445 (2013) 232–243, <https://doi.org/10.1016/j.virol.2013.07.018>.
- [10] K. Van Doorslaer, Evolution of the papillomaviridae, *Virology* 445 (2013) 11–20, <https://doi.org/10.1016/j.virol.2013.05.012>.
- [11] Y. Hirose, et al., Within-host variations of human papillomavirus reveal APOBEC-signature mutagenesis in the viral genome, *J. Virol.* (2018), <https://doi.org/10.1128/jvi.00017-18>.
- [12] R.S. Dube Mandishora, et al., Intra-host sequence variability in human papillomavirus, *Papillomavirus Res* (2018), <https://doi.org/10.1016/j.pvr.2018.04.006>.
- [13] S. Lagström, et al., HPV16 whole genome minority variants in persistent infections from young Dutch women, *J. Clin. Virol.* 119 (2019) 24–30, <https://doi.org/10.1016/j.jcv.2019.08.003>.
- [14] E. Domingo, J. Sheldon, C. Perales, Viral quasispecies evolution, *Microbiol. Mol. Biol. Rev.* 76 (2012) 159–216, <https://doi.org/10.1128/MMBR.05023-11>.
- [15] C.J. Warren, et al., APOBEC3A functions as a restriction factor of human papillomavirus, *J. Virol.* 89 (2015) 688–702, <https://doi.org/10.1128/JVI.02383-14>.
- [16] R.S. Harris, J.P. Dudley, APOBECs and virus restriction, *Virology* 479–480 (2015) 131–145, <https://doi.org/10.1016/j.virol.2015.03.012>.
- [17] L.B. Alexandrov, et al., Signatures of mutational processes in human cancer, *Nature* 500 (2013) 415–421, <https://doi.org/10.1038/nature12477>.
- [18] J.P. Vartanian, et al., Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions, *Science* 320 (2008) 230–233, <https://doi.org/10.1126/science.1153201>.
- [19] A.A. Mariaggi, 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.* 218 (2018) 1027–1036, <https://doi.org/10.1093/infdis/jiy287>.
- [20] B. Zhu, et al., Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance, *Nat. Commun.* 11 (2020) 886, <https://doi.org/10.1038/s41467-020-14730-1>.
- [21] V.C. Vieira, et al., Human papillomavirus E6 triggers upregulation of the antiviral and cancer genomic DNA deaminase APOBEC3B, *mBio* 5 (2014), <https://doi.org/10.1128/mBio.02234-14>.
- [22] K. Chan, et al., An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers, *Nat. Genet.* 47 (2015) 1067–1072, <https://doi.org/10.1038/ng.3378>.
- [23] L.S. Arroyo-Muhr, et al., Human papillomavirus type 16 genomic variation in women with subsequent in situ or invasive cervical cancer: prospective population-based study, *Br. J. Canc.* 119 (2018) 1163–1168, <https://doi.org/10.1038/s41416-018-0311-7>.
- [24] L. Mirabello, et al., HPV16 E7 genetic conservation is critical to carcinogenesis, *Cell* 170 (2017) 1164–1174, <https://doi.org/10.1016/j.cell.2017.08.001>, e6.
- [25] Cancer Genome Atlas Research Network, Integrated genomic and molecular characterization of cervical cancer, *Nature* 543 (2017) 378–384, <https://doi.org/10.1038/nature21386>.
- [26] J. Doorbar, et al., Human papillomavirus molecular biology and disease association, *Rev. Med. Virol.* 25 (Suppl 1) (2015) 2–23, <https://doi.org/10.1002/rmv.1822>.
- [27] A.A. McBride, A. Warburton, The role of integration in oncogenic progression of HPV-associated cancers, *PLoS Pathog.* 13 (2017), e1006211, <https://doi.org/10.1371/journal.ppat.1006211>.
- [28] K. Akagi, et al., Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability, *Genome Res.* 24 (2014) 185–199, <https://doi.org/10.1101/gr.164806.113>.
- [29] C. Bodelon, et al., Genomic characterization of viral integration sites in HPV-related cancers, *Int. J. Canc.* 139 (2016) 2001–2011, <https://doi.org/10.1002/ijc.30243>.
- [30] M. Peter, et al., Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma, *J. Pathol.* 221 (2010) 320–330, <https://doi.org/10.1002/path.2713>.
- [31] I. Kraus, et al., The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes, *Canc. Res.* 68 (2008) 2514–2522, <https://doi.org/10.1158/0008-5472.CAN-07-2776>.
- [32] Y. Liu, et al., Genome-wide profiling of the human papillomavirus DNA integration in cervical intraepithelial neoplasia and normal cervical epithelium by HPV capture technology, *Sci. Rep.* 6 (2016) 35427, <https://doi.org/10.1038/srep35427>.
- [33] J. Huang, et al., Comprehensive genomic variation profiling of cervical intraepithelial neoplasia and cervical cancer identifies potential targets for cervical cancer early warning, *J. Med. Genet.* 56 (2019) 186–194, <https://doi.org/10.1136/jmedgenet-2018-105745>.
- [34] S. Lagström, et al., TaME-seq: an efficient sequencing approach for characterisation of HPV genomic variability and chromosomal integration, *Sci. Rep.* 9 (2019) 524, <https://doi.org/10.1038/s41598-018-36669-6>.
- [35] A. Trope, et al., Performance of human papillomavirus DNA and mRNA testing strategies for women with and without cervical neoplasia, *J. Clin. Microbiol.* 47 (2009) 2458–2464, <https://doi.org/10.1128/JCM.01863-08>.
- [36] A. Trope, 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 (2012) 1927–1935, <https://doi.org/10.1128/JCM.00265-12>.
- [37] J.J. Kozich, et al., 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 (2013) 5112–5120, <https://doi.org/10.1128/AEM.01043-13>.
- [38] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods* 12 (2015) 357–360, <https://doi.org/10.1038/nmeth.3317>.
- [39] H. Li, et al., The sequence alignment/map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079, <https://doi.org/10.1093/bioinformatics/btp352>.
- [40] S.M. Kielbasa, et al., Adaptive seeds tame genomic sequence comparison, *Genome Res.* 21 (2011) 487–493, <https://doi.org/10.1101/gr.113985.110>.
- [41] D. Welter, et al., The NHGRI GWAS Catalog, a curated resource of SNP-trait associations, *Nucleic Acids Res.* 42 (2014) D1001–D1006, <https://doi.org/10.1093/nar/gkt1229>.
- [42] D.R. Zerbino, et al., The ensembl regulatory build, *Genome Biol.* 16 (2015) 56, <https://doi.org/10.1186/s13059-015-0621-5>.
- [43] S.A. Roberts, et al., An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers, *Nat. Genet.* 45 (2013) 970–976, <https://doi.org/10.1038/ng.2702>.
- [44] M.B. Burns, et al., APOBEC3B is an enzymatic source of mutation in breast cancer, *Nature* 494 (2013) 366–370, <https://doi.org/10.1038/nature11881>.
- [45] S. Vinokurova, et al., Type-dependent integration frequency of human papillomavirus genomes in cervical lesions, *Canc. Res.* 68 (2008) 307–313, <https://doi.org/10.1158/0008-5472.CAN-07-2754>.
- [46] P. van der Wee, C. Meijer, A.J. King, Whole-genome sequencing and variant analysis of human papillomavirus 16 infections, *J. Virol.* 91 (2017), <https://doi.org/10.1128/jvi.00844-17>.
- [47] P. van der Wee, C. Meijer, A.J. King, High whole-genome sequence diversity of human papillomavirus type 18 isolates, *Viruses* 10 (2018), <https://doi.org/10.3390/v10020068>.
- [48] L.S. Arroyo-Muhr, et al., The HPV16 genome is stable in women who progress to in situ or invasive cervical cancer: a prospective population-based study, *Canc. Res.* 79 (2019) 4532–4538, <https://doi.org/10.1158/0008-5472.CAN-18-3933>.

- [49] D.T. Geraets, et al., Long-term follow-up of HPV16-positive women: persistence of the same genetic variant and low prevalence of variant co-infections, *PLoS One* 8 (2013), e80382, <https://doi.org/10.1371/journal.pone.0080382>.
- [50] H. Maki, K. Fujikawa-Adachi, O. Yoshie, Evidence for a promoter-like activity in the short non-coding region of human papillomaviruses, *J. Gen. Virol.* 77 (Pt 3) (1996) 453–458, <https://doi.org/10.1099/0022-1317-77-3-453>.
- [51] B. Bhattacharjee, et al., Characterization of sequence variations within HPV16 isolates among Indian women: prediction of causal role of rare non-synonymous variations within intact isolates in cervical cancer pathogenesis, *Virology* 377 (2008) 143–150, <https://doi.org/10.1016/j.virol.2008.04.007>.
- [52] P. Mandal, et al., Differential expression of HPV16 L2 gene in cervical cancers harboring episomal HPV16 genomes: influence of synonymous and non-coding region variations, *PLoS One* 8 (2013), e65647, <https://doi.org/10.1371/journal.pone.0065647>.
- [53] K. McFadden, M.A. Luftig, Interplay between DNA tumor viruses and the host DNA damage response, *Curr. Top. Microbiol. Immunol.* 371 (2013) 229–257, [https://doi.org/10.1007/978-3-642-37765-5\\_9](https://doi.org/10.1007/978-3-642-37765-5_9).
- [54] P. Pham, et al., Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation, *Nature* 424 (2003) 103–107, <https://doi.org/10.1038/nature01760>.
- [55] Z. Hu, 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 (2015) 158–163, <https://doi.org/10.1038/ng.3178>.
- [56] B. Xu, et al., Multiplex identification of human papillomavirus 16 DNA integration sites in cervical carcinomas, *PLoS One* 8 (2013), e66693, <https://doi.org/10.1371/journal.pone.0066693>.
- [57] A. Holmes, et al., Mechanistic signatures of HPV insertions in cervical carcinomas, *npj Genomic Medicine* 1 (2016), <https://doi.org/10.1038/npjgenmed.2016.4>.
- [58] G. Gao, et al., Whole genome sequencing reveals complexity in both HPV sequences present and HPV integrations in HPV-positive oropharyngeal squamous cell carcinomas, *BMC Canc.* 19 (2019) 352, <https://doi.org/10.1186/s12885-019-5536-1>.
- [59] M. Iden, et al., The lncRNA PVT1 contributes to the cervical cancer phenotype and associates with poor patient prognosis, *PLoS One* 11 (2016), e0156274, <https://doi.org/10.1371/journal.pone.0156274>.
- [60] S.W. Cho, et al., Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element, *Cell* 173 (2018) 1398–1412, <https://doi.org/10.1016/j.cell.2018.03.068>, e22.
- [61] G. David-Morrison, et al., WAC regulates mTOR activity by acting as an adaptor for the TTT and pontin/reptin complexes, *Dev. Cell* 36 (2016) 139–151, <https://doi.org/10.1016/j.devcel.2015.12.019>.
- [62] A.Y. Qin, et al., MiR-205 in cancer: an angel or a devil? *Eur. J. Cell Biol.* 92 (2013) 54–60, <https://doi.org/10.1016/j.ejcb.2012.11.002>.
- [63] D. Tang, et al., VISDB: a manually curated database of viral integration sites in the human genome, *Nucleic Acids Res.* 48 (2020) D633–D641, <https://doi.org/10.1093/nar/gkz867>.
- [64] S.M. Amaro-Filho, et al., HPV DNA methylation at the early promoter and E1/E2 integrity: a comparison between HPV16, HPV18 and HPV45 in cervical cancer, *Papillomavirus Res* 5 (2018) 172–179, <https://doi.org/10.1016/j.pvr.2018.04.002>.