Experimental and Molecular Pathology 137 (2024) 104906

Contents lists available at ScienceDirect

Experimental and Molecular Pathology

journal homepage: www.elsevier.com/locate/yexmp

Accurate detection of copy number aberrations in FFPE samples using the mFAST-SeqS approach

Aude Jary [a,b], Yongsoo Kim [a,b], Kirsten Rozemeijer [a,b], Paul P. Eijk [a,b], Ramon P. van der Zee [a,b,c], Maaike C.G. Bleeker [a,b], Saskia M. Wilting [d], Renske D.M. Steenbergen [a,b,*]

[a] Department of Pathology, Amsterdam UMC, location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[b] Cancer Center Amsterdam, Biomarkers and Imaging, Amsterdam, the Netherlands
[c] Department of Internal Medicine, division of Infectious Diseases, Amsterdam UMC, location Universiteit van Amsterdam, Amsterdam, the Netherlands
[d] Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands

ARTICLE INFO

Editor: Dr. Marco Giudici

Keywords:
mFAST-SeqS
LINE-1
Copy number aberrations
FFPE
HPV-related diseases
Shallow sequencing
Anal lesions
Vulva lesions

ABSTRACT

Background: Shallow whole genome sequencing (Shallow-seq) is used to determine the copy number aberrations (CNA) in tissue samples and circulating tumor DNA. However, costs of NGS and challenges of small biopsies ask for an alternative to the untargeted NGS approaches. The mFAST-SeqS approach, relying on LINE-1 repeat amplification, showed a good correlation with Shallow-seq to detect CNA in blood samples. In the present study, we evaluated whether mFAST-SeqS is suitable to assess CNA in small formalin-fixed paraffin-embedded (FFPE) tissue specimens, using vulva and anal HPV-related lesions.

Methods: Seventy-two FFPE samples, including 36 control samples (19 vulva;17 anal) for threshold setting and 36 samples (24 vulva; 12 anal) for clinical evaluation, were analyzed by mFAST-SeqS. CNA in vulva and anal lesions were determined by calculating genome-wide and chromosome arm-specific z-scores in comparison with the respective control samples. Sixteen samples were also analyzed with the conventional Shallow-seq approach.

Results: Genome-wide z-scores increased with the severity of disease, with highest values being found in cancers. In vulva samples median and inter quartile ranges [IQR] were 1[0–2] in normal tissues (n = 4), 3[1–7] in premalignant lesions (n = 9) and 21[13–48] in cancers (n = 10). In anal samples, median [IQR] were 0[0–1] in normal tissues (n = 4), 14[6–38] in premalignant lesions (n = 4) and 18[9–31] in cancers (n = 4). At threshold 4, all controls were CNA negative, while 8/13 premalignant lesions and 12/14 cancers were CNA positive. CNA captured by mFAST-SeqS were mostly also found by Shallow-seq.

Conclusion: mFAST-SeqS is easy to perform, requires less DNA and less sequencing reads reducing costs, thereby providing a good alternative for Shallow-seq to determine CNA in small FFPE samples.

1. Background

High-risk HPV (hrHPV) infection is the first step in the development of HPV-induced malignancies, which develop via precursor stages known as intraepithelial neoplasia (i.e.: cervical intraepithelial neoplasia (CIN); anal intraepithelial neoplasia (AIN); vulva intraepithelial neoplasia (VIN)) to cancer. These premalignant lesions are graded 1–3 based on their severity. During the course of HPV-induced carcinogenesis epigenetic and genetic changes accumulate, which are driven by the viral oncoproteins E6 and E7 (Duensing and Münger, 2004; Wilting and Steenbergen, 2016). For instance, hrHPV E6 and E7 are closely associated with and boost the activity of the DNA methyltransferases (DNMTs), promoting DNA methylation of tumor suppressor genes (Au Yeung et al., 2010; Burgers et al., 2007). Additionally, these oncoproteins cause centrosome abnormalities, chromosomal segregation errors, and DNA damage, which result in chromosomal instability (Moody and Laimins, 2010). HPV-induced anogenital cancers and premalignant lesions, are characterized by multiple copy number aberrations (CNA), with a gain of 3q being reported most frequently (Cacheux et al., 2018; Gagne et al., 2005; Heselmeyer et al., 1997; Jee et al., 2001; Swarts et al., 2018; Thomas et al., 2014; Wilting et al., 2009; Wilting and Steenbergen, 2016).

Our earliest studies on CNA in HPV-induced cancers used comparative genomic hybridization (Wilting et al., 2009). The advent of Next

* Corresponding author at: Amsterdam UMC, Vrije Universiteit Amsterdam, Pathology, De Boelelaan 1117, 1081 HV Amsterdam, the Netherlands.
E-mail address: r.steenbergen@amsterdamumc.nl (R.D.M. Steenbergen).

https://doi.org/10.1016/j.yexmp.2024.104906
Received 7 June 2023; Received in revised form 17 May 2024; Accepted 22 May 2024
Available online 30 May 2024

Generation Sequencing (NGS) has improved the sensitivity and the resolution for the detection of CNA in human tumors (Morozova and Marra, 2008). This progress also allows the monitoring of tumor genomes by minimally invasive means, such as by the analysis of cell-free circulating tumor DNA (ctDNA) in blood (Chen and Zhao, 2019). Most untargeted approaches (meaning whole genome sequencing) provide the genome wide pattern of CNA, but require high amounts of sequencing reads and are expensive. Moreover, in comparison to targeted approaches, the sensitivity of genome-wide approaches to detect chromosomal alterations or mutations is lower, requiring a CNA or mutant allele rate of at least 5–10% (Belic et al., 2015). As of to date, several amplicon-based targeted approaches have been developed to identify CNA (Douville et al., 2020; Grasso et al., 2015). The Fast Aneuploidy Screening Test-sequencing System (FAST-SeqS), a targeted approach to determine chromosomal aberrations by amplifying the repeat regions Long Interspersed Nucleotide Element-1 (LINE-1), was first described by Kinde et al. in 2012 to detect Trisomy 18 and 21 in fetal-circulating DNA in the blood of maternal samples (Kinde et al., 2012). A slightly modified version of FAST-seqS (mFAST-SeqS) was then applied in the context of cancer to detect CNA in ctDNA in fresh blood samples and FFPE tumor tissues (Angus et al., 2021; Belic et al., 2016; Douville et al., 2018; Mendelaar et al., 2022). Interestingly, Belic et al. reported a very good correlation of the results between the targeted mFAST-SeqS and untargeted plasma-Seq approaches to detect CNA in blood samples (Belic et al., 2016).

As CNA are seen in virtually all HPV-induced cancers (Thomas et al., 2014), as well as in more advanced stages of premalignant lesions (Swarts et al., 2018; Thomas et al., 2014), the detection of CNA or specific gains or losses of chromosomal arms (for instance gain of 3q) might be of interest for cancer risk stratification (Tian et al., 2019). However, current challenges in terms of limited DNA availability due to the small biopsies and costs of NGS, ask for an alternative to the untargeted NGS approaches (Scheinin et al., 2014). Towards this goal, we evaluated the potential of using targeted mFAST-SeqS method to determine CNA in small formalin-fixed paraffin-embedded (FFPE) tissue biopsies representing various stages of anal and vulva disease. For validation mFAST-SeqS results were compared to results obtained by the untargeted Shallow-seq method (Scheinin et al., 2014).

2. Materials and methods

2.1. Control samples

To determine the best threshold for calling of CNA over the background when using mFAST-SeqS, we used a set of 36 normal FFPE tissues, including 19 vulva and 17 male anal samples.

2.2. Clinical samples

CNA analysis using the mFAST-SeqS approach was clinically evaluated on 36 FFPE samples including 8 normal tissues (4 vulva, 4 anal), 13 premalignant lesions (4 VIN3, 5 VIN3 adjacent to cancer, 4 AIN3) and 15 cancers (11 vulva squamous cell carcinoma (VSCC) and 4 anal squamous cell carcinomas (ASCC). VIN adjacent to cancer (VINadjVSCC) is considered to be a surrogate of the most advanced premalignant stage of VIN (Swarts et al., 2018). To compare this targeted approach with the current gold standard of CNA determination, the untargeted Shallow Whole Genome Sequencing (Shallow-seq) was also performed on a subset of 16 samples. A description of the different samples tested is provided in Supplementary Table S1.

All FFPE samples were retrospectively identified and retrieved from the pathology archives of the Amsterdam University Medical Centers (Amsterdam UMC), as previously reported (Swarts et al., 2018; van der Zee et al., 2021a) (see Supplementary Table S1). Ethical approval for anal samples was granted under reference number 05/031 (normal control samples) and waived for use of archived ASCC specimens (reference no. 17/151) and AIN biopsies (reference no. 18/341). Ethical approval for vulva samples was granted under reference number 2017.626 (normal control samples) and waived for use of archived VIN and VSCC tissues (reference number no. 2017.561).

We adhered to the Declaration of Helsinki and Code of Conduct for Responsible Use of Left-over Material of the Dutch Federation of Biomedical Scientific Societies. Ethical approval was granted or waived as reported above.

2.3. mFAST-SeqS analysis

The mFAST-SeqS technique targets the long interspersed nucleotide element-1 (LINE-1) retrotransposons in the human genome to identify CNA in samples, as previously reported (Belic et al., 2016). Sequencing and analysis were performed by either the Department of Medical Oncology, Erasmus MC Cancer Institute, on the MiSeq system (Illumina San Diego, CA, USA) or by the Department of Pathology, Cancer Center Amsterdam on the iSeq system (Illumina San Diego, CA, USA). To control for potential sequencing bias 2 samples were tested by both sequencing methods.

Briefly, 10 ng of DNA was amplified for a first PCR targeting LINE-1. To increase complexity of the resulting sequencing libraries a random spacer was introduced to the primers as described by Fadrosh et al. (Fadrosh et al., 2014). Then, a second PCR was performed to add specific barcodes identifying each sample before normalization, pooling and sequencing on the Illumina sequencer (150 base-pairs (bp) single reads). After trimming of the reads using Trimmomatic, reads were mapped to the human genome hg19 using the Burrow's Wheeler Alignmer (BWA aligner) (Li and Durbin, 2009). After normalization of read counts on total reads per sample, calculation of the chromosome arm-specific z-score and the genome-wide z-score (squared sum of all chromosome arm-specific z-score as an overall measure of aneuploidy in a sample) was performed by comparison of anal samples with normal anal controls and vulva samples with normal vulva controls, as previously reported (Belic et al., 2016). The full protocol is provided in the Supplementary Table S2 and S3.

2.4. Design of new reverse primers with different Index/barcodes PCR

The available method contained 24 reverse primers, allowing for a total of 24 samples to be tested simultaneously (Belic et al., 2015). To increase the sample size per NGS run, a set of 22 new reverse primers for the Index PCR (including a specific barcoding sequence to distinguish each sample) were designed. First, we selected a set of 402 barcodes of 6 bp (available at Download barcode sets at http://hannonlab.cshl.edu/nxCode/nxCode/Ready_made_sets.html) containing a GC-content between 25% and 75% and without homopolymers of length 4 or more. Then, the identification of the optimal index sequences, in addition to the 24 reverse index PCR already published (Belic et al., 2015), was performed using the BARCOSEL online tool (http://ekhidna2.biocenter.helsinki.fi/barcosel). This tool was used to find the optimal set by checking the pairwise sequence distances and the nucleotide balance depending on Illumina sequencer and chemistry used. Finally, the verification of the full barcode set (46 indexes in total) was performed using CheckMyIndex online tool (https://checkmyindex.pasteur.fr/) which provides the best association of Index according to the following parameters: total number of samples to sequence, multiplexing rate, constraint on the indexes, minimum number of red/green lights required at each position, compatibility with the Illumina chemistry to be used. The Agilent Bioanalyzer using a 7500 DNA kit (Agilent, Santa Clara, USA) was used to validate the 22 new primers and to visualize the LINE-1 library (sequences available in Supplementary Table S2).

2.5. Shallow-seq analysis

CNA by Shallow-seq was performed and analyzed as previously

reported (Swarts et al., 2018). Briefly, 250 ng of input DNA was fragmented and then sequenced using the TruSeq nano-kit (Illumina San Diego, CA, USA) on the Illumina HiSeq 2000, 50 bp single-read, to reach a yield of about 8 million reads per sample. Sequencing results were analyzed as described previously (Scheinin et al., 2014). Briefly, the Bioconductor script QDNAseq was used to map reads to the human genome (GRCh37/hg19) using BWA (Li and Durbin, 2009). Reads with a mapping quality <37 and PCR duplicates were filtered out. Reads counts were quantified in nonoverlapping 30 kb windows, followed by a simultaneous loess correction for sequence mappability and GC content. Segmentation ($\alpha = 0.00000000001$, number of SD's between means = 2) and subsequent calling of gained, amplified and lost regions was done using CGHcall. Segments with a probability score of ≥ 0.5 were considered gained, amplified or lost (van de Wiel et al., 2007).

2.6. Statistical analysis

Continuous variables were expressed in median and interquartile range [IQR] and discrete variables as number and percentage (n, %). The Shapiro test was performed to assess if a normal distribution of the genome-wide and chromosome arm-specific z-scores were present. If not, non-parametric Kruskal-Wallis and Mann-Whitney *U* tests were performed to compare genome-wide and chromosome arm-specific z-scores. Results of the 2 samples tested with both iSeq and MiSeq sequencing were compared with the Wilcoxon Signed Rank Test, the agreement was assessed with a Bland-Altman plot and the correlation with a Pearson correlation test. Results were considered to be significant if $p < 0.05$. We compared the results obtained with the 2 approaches and considered that results were concordant if the z-score was above 4 (or below 4) with the mFAST-seqS approach and a gain (or a loss) of the same arm-chromosome was also seen with the Shallow-seq. Statistical analysis and figures were generated with RStudio 2022.02.0 software and packages *gt*, *gtsummary*, *rstatix* and *ggplot2*.

3. Results

3.1. Technical aspect of mFAST-SeqS developments

We first increased the maximum number of 24 samples per sequencing run based on the available 24 reverse index primers, by designing 22 new reverse index primers. Their capacity to amplify the LINE-1 region was assessed on one FFPE vulva sample, and all gave expected profiles (See Supplementary Table S4 and Fig. S1).

Next, we evaluated whether sequencing devices may influence the mFAST-SeqS results, by sequencing of two samples (VU13 and VU15) using both MiSeq and iSeq machines. Analysis and controls were kept the same. For both samples, the genome-wide z-score was almost identical between machines: for VU13, the genome-wide z-score was 4.09 and 4.46, on the MiSeq and iSeq respectively, and for VU15, 21.22 and 21.15, respectively. By comparing the chromosome arm-specific z-scores, variances were not significantly different from 0 for both samples (Wilcoxon Signed Rank Test, VU13 $p = 0.44$ and VU15 $p = 0.24$) (Figure1, a and b). Finally, results between the two techniques were significantly positively correlated for VU13 ($r = 0.80$, $p < 0.001$) and for VU15 ($r = 0.97$, $p < 0.001$) (Fig. 1, c and d).

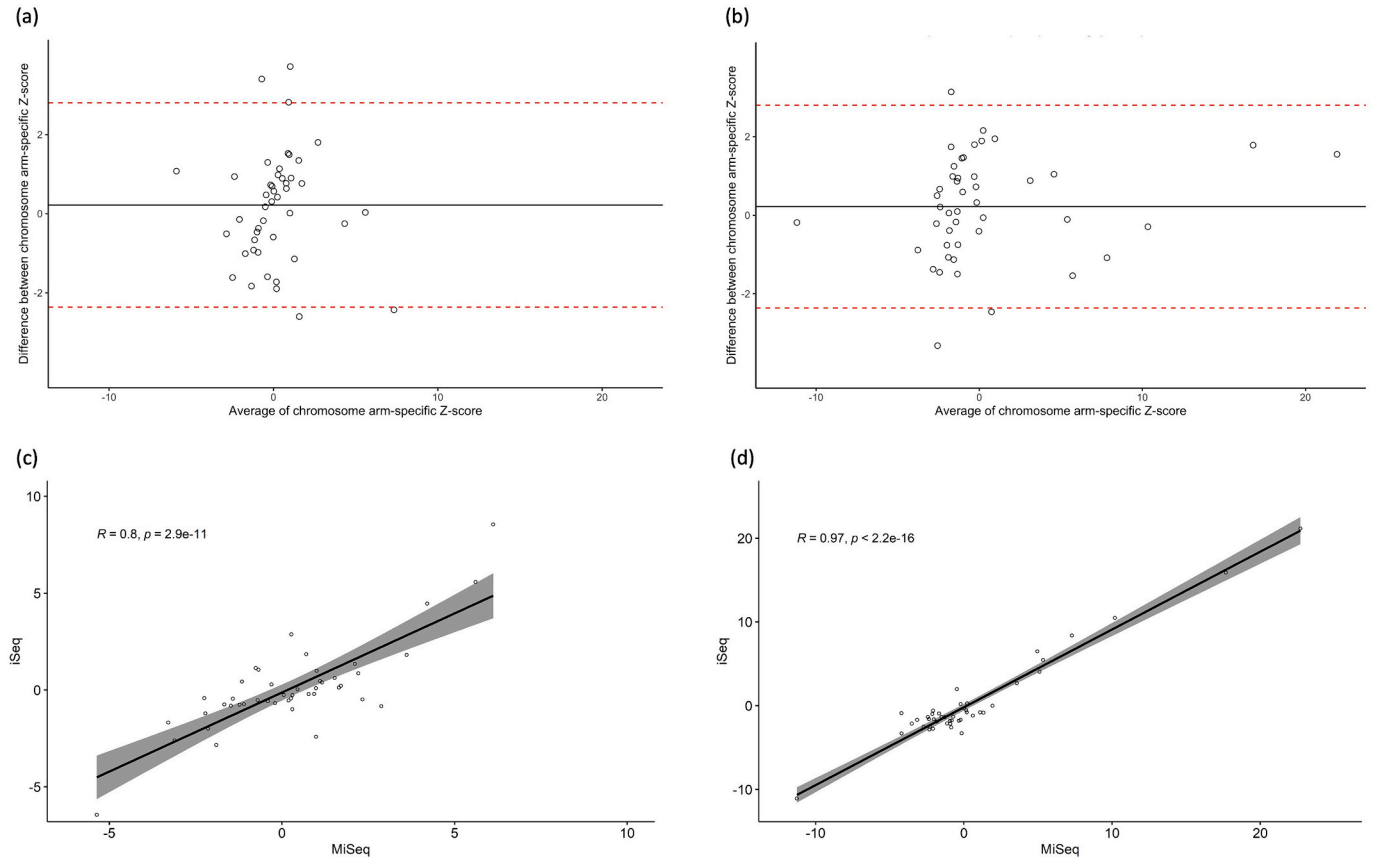

Fig. 1. Comparing chromosome arm-specific z-scores between the 2 sequencing Illumina devices. Samples VU13 (a, c) and VU15 (b, d) were sequenced with both the iSeq and MiSeq Illumina machine and chromosome arm-specific z-scores were compared by plotting a Bland-Altman (a and b).
Red dotted line (a and b): 95% confidence interval of the z-score difference. Graphs c and d: Pearson correlation test. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Sequencing of control samples and threshold setting

To determine the threshold to call a CNA in FFPE tissue, 36 normal FFPE tissues (19 vulva and 17 male anal samples) were sequenced and analyzed with the mFAST-SeqS pipeline. A median of 109,245 [95,315–125,306] mapped reads were generated and after normalization, chromosome arm-specific z-scores for each sample were calculated by comparison with the other controls. A total of 1440 chromosome arm-specific z-scores were determined after exclusion of the chromosome arms of males (Yp and Yq), as well as 13p, 14p, 15p, 21p and 22p which have insufficient LINE1 elements to be analyzed with this approach. For each sample, we calculated a genome-wide z-score as a general measure for aneuploidy and the median was −0.32 [−0.53–0.26].

By setting the z-score threshold at 5 (established on plasma samples by Belic et al.), 4 and 3, 0.07% (1/1440), 0.14% (2/1440) and 1.25% (18/1440) of chromosome arm-specific z-scores were detected as abnormal, corresponding to 2.78% (1/36), 5.56% (2/36) and 19.44% (7/36) of the controls (see Supplementary Table S5 and Table S6). For further analyses, we chose to continue with threshold 4 as it yields approximately 95% specificity.

3.3. mFAST-SeqS analysis to determine CNA in normal, premalignant and cancer tissues

For further analysis, eight additional controls i.e. 4 normal anal epithelia and 4 normal vulva epithelia were analyzed and compared with premalignant lesions and cancers, i.e. 4 VIN3, 5 VINadjVSCC, 11 VSCC, 4 AIN3 and 4 ASCC. The mFAST-SeqS approach was effective for 35 of 36 samples tested with a median [IQR] of reads mapped of 110,504 [99,013–127,645]; sample VU14 did not yield the required amount of reads for reliable analysis ($n = 16,907$).

In vulva FFPE samples, genome-wide z-scores increased with the severity of the disease ($p = 0.009$) with highest values in cancers, i.e. median and IQR of 1 [0–2] in normal tissues, 3 [1–7] in VIN3, 17 [16–58] in VINadjVSCC, and 21 [13–48] in VSCC (Table 1) (Fig. 2, a). In anal FFPE samples, genome-wide z-scores also tended to increase with the severity of the disease ($p = 0.058$). Highest values were found in cancers, with median and IQR of 0 [0–1] in normal tissues, 14 [6–38] in AIN3, and 18 [9–31] in ASCC (Fig. 2, b).

At threshold 4, CNA were not detected in any controls (8/8), while 8/13 premalignant lesions and 12/14 cancers were CNA positive. In vulva FFPE tissues, all controls were CNA negative, while 5/9 VIN3 and 9/10 VSCC were CNA positive. In anal FFPE tissues, all controls were CNA negative as well, while 3/4 AIN3 and 3/4 ASCC were CNA positive. Positivity rates for different thresholds are presented in Table 2.

3.4. Comparison of mFAST-SeqS and Shallow-seq approaches

To validate the mFAST-SeqS method applied to FFPE biopsies, sixteen premalignant and cancer samples were additionally analyzed by conventional Shallow sequencing.

First, we compared chromosome arm-specific z-score from mFAST-SeqS analysis with CNA (gain and loss >10 million base-paired) from Shallow-seq analysis (Supplementary Table S7). At a threshold of 4, only 4.5% (25) of chromosome arm aberrations detected with mFAST-SeqS were not seen by Shallow-seq. At the same threshold, six out of 14 vulva samples showed a good agreement between the 2 methods without any additional aberrations detected with the mFAST-SeqS (VU13, VU15, VU25, VU26, VU33 and VU35) (Fig. 3 and Supplementary Fig. S2). Six others samples also showed a good agreement with only 1, 2 or 3 additional chromosome arm aberrations being found by mFAST-SeqS (1: VU16 and VU27, 2: VU30 and VU34, 3: VU28 and VU29) (Supplementary Fig. S2).

Extra aberrations per sample detected with the mFAST-SeqS increased with the number of CNA identified with Shallow-seq (Supplementary Table S7).

Conversely, in all the samples, at least 1 CNA of >10 million base pairs detected by Shallow-seq was missed by the mFAST-SeqS approach (gain or loss). These mismatches took place when (i) gains and losses co-occurred at the same chromosome arm (e.g. 12p gain and loss in VU15) and (ii) the amount of reads gained or lost was low (reflecting the level of the CNA in the sample) (e.g., gain 12p in VU25).

Focusing on the main CNA associated with anogenital premalignant lesions and cancers, 8 out of 13 vulva cases (3 VINadjVSCC and 5 VSCC) harbored a gain of 3q, of which 75% (6/8) were also detected with the mFAST-SeqS approach. In anal samples, a gain of 3q was found in the 2 samples by both sequencing approaches.

Table 1
Overall measure of aneuploidy according to the genome-wide z-score determined with the mFAST-SeqS approach in normal epithelial samples, premalignant lesions and cancers of the vulva and the anus.

	Sample ID	Sample type	Mapped Reads	Genome-wide z-score* (FFPE controls)
	VU10	Normal	92,288	−0.43
	VU11	Normal	208,790	−0.14
	VU9	Normal	102,528	0.08
	VU12	Normal	116,863	3.68
	VU19	VIN3	235,376	−1.53
	VU17	VIN3	120,511	1.54
	VU18	VIN3	99,915	3.56
	VU20	VIN3	126,373	16.20
	VU26	VINadjVSCC	202,536	2.26
	VU28	VINadjVSCC	137,287	15.85
	VU27	VINadjVSCC	179,903	16.90
Vulva samples	VU29	VINadjVSCC	156,284	58.00
	VU25	VINadjVSCC	260,504	71.51
	VU34	VSCC	89,104	3.82
	VU13	VSCC	105,641	4.09
	VU33	VSCC	107,878	12.31
	VU30	VSCC	79,580	13.57
	VU32	VSCC	80,513	20.87
	VU15	VSCC	101,070	21.22
	VU35	VSCC	121,619	37.44
	VU31	VSCC	80,617	51.32
	VU16	VSCC	111,554	72.72
	VU36	VSCC	104,710	97.22
	VU14	VSCC	16,907	uninterpretable
	VU22	Normal	100,579	−0.11
	VU23	Normal	125,058	0.31
	VU21	Normal	120,864	1.74
	VU24	Normal	105,487	2.94
	VU8	AIN3	115,956	0.69
Anal samples	VU7	AIN3	109,455	7.77
	VU6	AIN3	131,464	20.72
	VU2	AIN3	113,544	88.31
	VU3	ASCC	226,739	1.13
	VU5	ASCC	93,126	11.33
	VU4	ASCC	94,476	24.35
	VU1	ASCC	96,310	50.76

VIN3: vulva intraepithelial neoplasia; VIN adjVSCC: VIN adjacent to VSCC; VSCC: vulva squamous cell cellular; AIN3: anal intraepithelial neoplasia; ASCC: anal squamous cell carcinoma.

* Genome-wide z-score was calculated for each sample by comparison with ratio means and standard deviation means from a set of 19 vulva controls and 17 anal controls, respectively.

4. Discussion

Chromosomal aberrations occur during carcinogenesis and have previously been detected in HPV-induced premalignant vulva lesions with the untargeted Shallow-seq reference approach (Swarts et al., 2018). Frequently, biopsies available for CNA detection are small, with low DNA concentration, which particularly in FFPE samples is also frequently fragmented. As such Shallow-seq can be challenging. In this paper, we evaluated the mFAST-SeqS target approach to identify CNA on

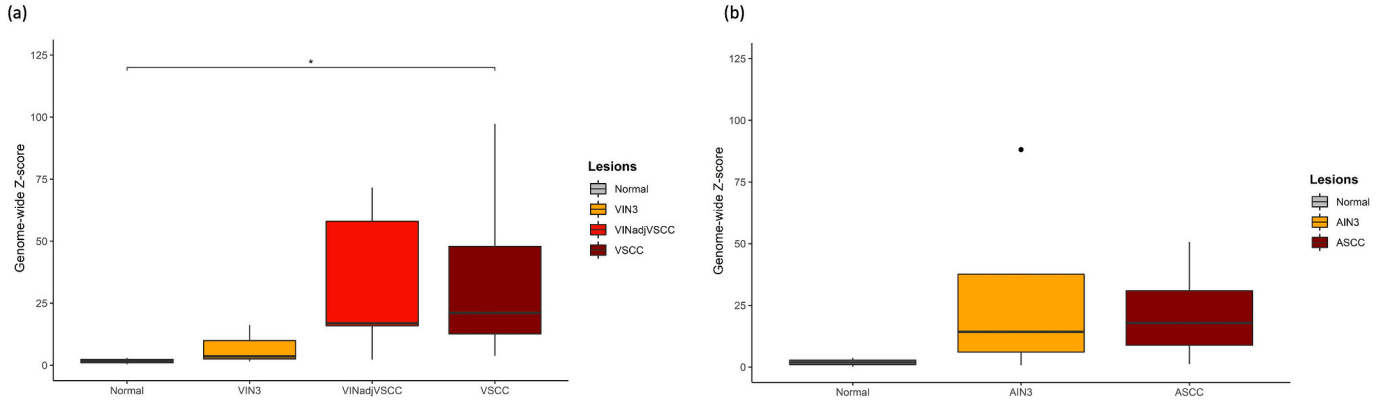

Fig. 2. Genome-wide z-score determined by mFAST-SeqS. Genome-wide z-scores increased with the severity of the diseases in both vulva (a) and anal (b) lesions. The threshold as ameasure of aneuploidy was set at 4. The boxplot was constructed with the ggplot2 package of R software: bold line is the median, quartile 1 and quartile 3 the lower and upper sides of the square and whiskers "min" and "max" corresponding to 1.5*IQR. Outliers are represented by a black circle. Statistical test: Mann-Whitney *U* test. $P < 0.05$ was considered to be significant.

Table 2
Positivity rates of the genome-wide z-score determined by the mFAST-SeqS approach at different thresholds.

	Threshold at 3		Threshold at 4		Threshold at 5	
	+	−	+	−	+	−
Control[a]	1	7	0	8	0	8
Cases[b]	22	5	20	7	19	8

[a] control: 4 normal anal tissue and 4 normal vulva tissue.
[b] cases: 4 VIN3, 6 VIN adjacent to cancer, 10 VSCC, 4 AIN3 and 4 ASCC.

FFPE samples. We effectively designed 22 new index primers next to the existing 24, allowing for the analysis of up to 46 samples simultaneously. Also, a good agreement was found between the MiSeq and the iSeq sequencing machines of Illumina, which underlines the robustness of the method. Using the mFAST-SeqS method, we could successfully determine CNA in 35 out of 36 FFPE specimens. At a genome-wide z score threshold of 4, none of the normal tissues and 12 out of 14 cancers were found to be CNA positive, in line with literature based on Shallow-seq results (Swarts et al., 2018). The validity of using mFAST-SeqS as an alternative to Shallow-seq for CNA analysis in FPPE materials, is further supported by the fact that a good agreement was found between Shallow-seq and mFAST-SeqS to detect CNA.

By mFAST-SeqS of both vulva and anal samples, an increase of the genome-wide z-score, corresponding to an overall measure of aneuploidy, from normal to premalignant lesions and cancers was found. This is in line with previous publications on CNA in HPV-induced lesions (Heselmeyer et al., 1997; Jee et al., 2001; Swarts et al., 2018; Thomas et al., 2014; Wilting et al., 2009). Similarly, also epigenetic modifications, such as increasing DNA methylation levels of certain host genes, were found to be associated with an increase of lesion severity in cervical, anal and vulva HPV-induced lesions (Thuijs et al., 2021; van der Zee et al., 2021b; Verlaat et al., 2018).

With a threshold of aneuploidy set at 4, we found that all the controls were CNA negative while 8/13 premalignant lesions and 12/14 cancers were CNA positive. A subset of VIN3 and AIN3 lesions did not show any CNA. This apparent discrepancy between molecular and histological patterns in HPV-induced premalignant lesions has previously been reported in multiple studies, including studies using the Shallow-seq approach (Steenbergen et al., 2014; Swarts et al., 2018; Thomas et al., 2014). Interestingly in studies on CIN3 lesions and VIN3 lesions, an increased number of CNA was associated with an advanced stage of premalignant disease and a higher risk of progression to cancer (Bierkens et al., 2012; Swarts et al., 2018). For instance, Swarts et al. reported that the number of CNA was higher in VIN of women that later on

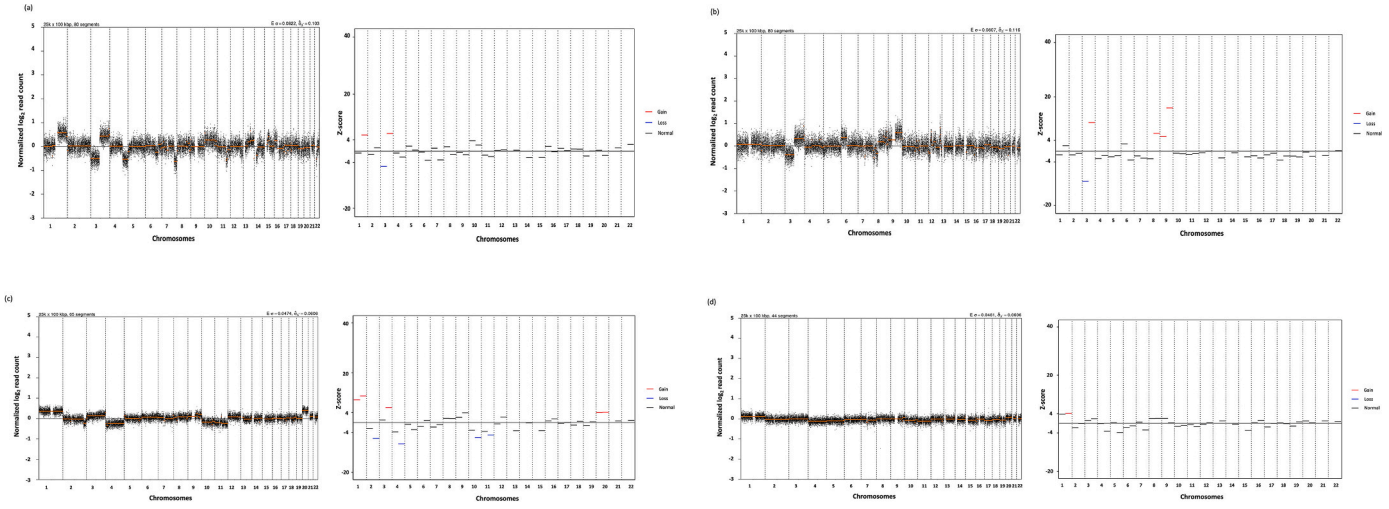

Fig. 3. Chromosomal aberrations detected by Shallow sequencing (left) and the mFAST-SeqS (right) for samples VU13 (a), VU15 (b), VU25 (c) and VU26 (d). X-axis corresponds to the 22 human chromosomes, left y-axis to the normalized $\log_2$ reads counts mapped on the reference human genome, and right y-axis corresponds to the z-score calculated for each arm-chromosome by comparison to a group control.

developed VSCC than in women with VIN which did not progress to cancer (Swarts et al., 2018). This indicates that a molecular approach might be more relevant to identify more advanced high-grade lesions with a higher cancer risk than histology (Steenbergen et al., 2014). It should be noted that as for most other molecular tests, also for mFAST-SeqS it is of utmost importance to use tissue matched controls for threshold setting. For this reason, we used anal controls for analysis of anal lesions and vulva controls for evaluation of vulva lesions.

Analysis of 16 samples (14 VIN3$^+$ and 2 AIN3$^+$) by both mFAST-SeqS and Shallow-seq showed a good concordance of gains and losses of chromosomal regions between the 2 methods. The mFAST-SeqS approach missed some chromosomal aberrations when gains and losses co-occurred at the same arm-chromosome or when gains and losses were present with a small deviation from the neutral state, reflecting the small concentration of the CNA in the sample. This bias was already reported by Belic et al. with this technique (Belic et al., 2016).

For some samples, we also identified gains or losses of chromosome arms which were not found with the Shallow-seq approach. This may be explained by the characteristics of the Line 1 elements used for mFAST-SeqS. The Line 1 elements are transposable elements dispersed throughout the human genome, with allele frequencies and level of heterozygosity being variable in different human populations (Sheen et al., 2000). Additionally, the Line 1 elements are subject to retro-transposition events leading to aberrant Line 1 integration which can induce complex translocations and large-scale duplications found in cancers (Rodriguez-Martin et al., 2020).

Interestingly, with a threshold of genome-wide z-score set at 4, all the cases of premalignant and cancer lesions had CNA detected by the analysis of the chromosome arm-specific z-score. Among them, 14 out of 16 cases showed a genome-wide z-score above 4 while the 2 remaining cases (1 VINadjVSCC and 1 VSCC) harbored at least one chromosome arm-specific z-score above 4.

The primary chromosomal abnormality associated with HPV-induced malignancies of the ano-genital area is a gain of 3q (Swarts et al., 2018; Thomas et al., 2014). In high-grade AIN, a gain of 3q is also reported as the most common abnormalities (Gagne et al., 2005). In our work, in 8 vulva cases a gain on 3q was found by Shallow-seq, 6 of which were also detected with the mFAST-SeqS approach. For the 2 samples in which mFAST-SeqS did not detect a 3q gain, a gain and loss occurred on the same chromosome arm (VU30) or only a small fragment of 3q was lost (VU32). In anal samples, a gain of 3q as found in 2 samples was detected by both methods. Since gain of 3q in particular is associated with advanced high-grade disease (Bierkens et al., 2012; Swarts et al., 2018), employing the mFAST-SeqS technology to detect 3q gain may be an alternative to detect advanced stages of HPV-induced premalignant lesions, and is potentially helpful for cancer risk assessment. With evolving technologies also combined CNA and methylation sequencing methods requiring low input DNA, such as EM-sequencing, may turn out valuable for small FFPE biopsies (Sun et al., 2023).

Our study has some limitations: (i) the number of lesional samples was relatively low and did not allow an evaluation of the diagnostic value of the mFAST-SeqS approach. However, we mainly aimed to evaluate the mFAST-SeqS approach on small biopsies and to validate it by comparison with the gold standard method. Additional studies with more samples are required to evaluate the clinical accuracy of the mFAST-SeqS approach as a prognostic or a cancer risk stratification marker; (ii) we only included HPV-positive samples, representing the far majority of anal lesions and VIN3 lesions, while the majority of VSCC are HPV-negative.

In conclusion, we reported the ability of mFAST-seqS, a promising tool, requiring less DNA, less sequencing reads and costs than conventional Shallow-seq, to directly analyze chromosomal aberrations in small FFPE tissue samples.

Ethics statement

This study followed the ethical guidelines of the Institutional Review Board of the Amsterdam UMC and we adhered to the Code of Conduct for Responsible Use of Left-over Material of the Dutch Federation of Biomedical Scientific Societies. Ethical approval was granted or waived for the use of archived biopsies (see materials and methods and supplementary materials).

Authors contributions

Study design: AJ, RS, SW.
Data collection: RVDZ, MCGB.
Data management: AJ, YK.
Laboratory experiments: AJ, PPE.
Statistical analysis: AJ, KR, SW.
Data interpretation: AJ, KR, SW.
Writing first draft of manuscript: AJ.
All authors were involved in writing the manuscript and gave final approval of the submitted and published version of the manuscript.

CRediT authorship contribution statement

Aude Jary: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Yongsoo Kim:** Formal analysis. **Kirsten Rozemeijer:** Methodology, Formal analysis. **Paul P. Eijk:** Formal analysis. **Ramon P. van der Zee:** Data curation. **Maaike C.G. Bleeker:** Data curation. **Saskia M. Wilting:** Methodology, Formal analysis, Conceptualization. **Renske D.M. Steenbergen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

(1) RDMS is a minority shareholder of Self-screen B.V., a spin-off company of VUmc; Self-screen B.V. develops, manufactures and licenses high-risk HPV and methylation marker assays for cervical cancer screening and holds patents on these tests. SMW has received research funding from Pfizer.

Data availability

Data will be made available on request.

Acknowledgements

This work has been supported by the Fondation ARC pour la recherche sur le cancer and the Dutch Cancer Society (grants 10382 and 10781).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yexmp.2024.104906.

References

Angus, L., Deger, T., Jager, A., Martens, J.W.M., de Weerd, V., van Heuvel, I., van den Bent, M.J., Sillevis Smitt, P.A.E., Kros, J.M., Bindels, E.M.J., Heitzer, E., Sleijfer, S., Jongen, J.L.M., Wilting, S.M., 2021. Detection of aneuploidy in cerebrospinal fluid from patients with breast cancer can improve diagnosis of leptomeningeal metastases. Clin. Cancer Res. 27, 2798–2806. https://doi.org/10.1158/1078-0432.CCR-20-3954.

Au Yeung, C.L., Tsang, W.P., Tsang, T.Y., Co, N.N., Yau, P.L., Kwok, T.T., 2010. HPV-16 E6 upregulation of DNMT1 through repression of tumor suppressor p53. Oncol. Rep. 24, 1599–1604. https://doi.org/10.3892/or_00001023.

Belic, J., Koch, M., Ulz, P., Auer, M., Gerhalter, T., Mohan, S., Fischereder, K., Petru, E., Bauernhofer, T., Geigl, J.B., Speicher, M.R., Heitzer, E., 2015. Rapid identification of

plasma DNA samples with increased ctDNA levels by a modified FAST-SeqS approach. Clin. Chem. 61, 838–849. https://doi.org/10.1373/clinchem.2014.234286.

Belic, J., Koch, M., Ulz, P., Auer, M., Gerhalter, T., Mohan, S., Fischereder, K., Petru, E., Bauernhofer, T., Geigl, J.B., Speicher, M.R., Heitzer, E., 2016. mFast-SeqS as a monitoring and pre-screening tool for tumor-specific aneuploidy in plasma DNA. Adv. Exp. Med. Biol. 924, 147–155. https://doi.org/10.1007/978-3-319-42044-8_28.

Bierkens, M., Wilting, S.M., van Wieringen, W.N., van Kemenade, F.J., Bleeker, M.C.G., Jordanova, E.S., Bekker-Lettink, M., van de Wiel, M.A., Ylstra, B., Meijer, C.J.L.M., Snijders, P.J.F., Steenbergen, R.D.M., 2012. Chromosomal profiles of high-grade cervical intraepithelial neoplasia relate to duration of preceding high-risk human papillomavirus infection. Int. J. Cancer 131, E579–E585. https://doi.org/10.1002/ijc.26496.

Burgers, W.A., Blanchon, L., Pradhan, S., de Launoit, Y., Kouzarides, T., Fuks, F., 2007. Viral oncoproteins target the DNA methyltransferases. Oncogene 26, 1650–1655. https://doi.org/10.1038/sj.onc.1209950.

Cacheux, W., Dangles-Marie, V., Rouleau, E., Lazartigues, J., Girard, E., Briaux, A., Mariani, P., Richon, S., Vacher, S., Buecher, B., Richard-Molard, M., Jeannot, E., Servant, N., Farkhondeh, F., Mariani, O., Rio-Frio, T., Roman-Roman, S., Mitry, E., Bieche, I., Lièvre, A., 2018. Exome sequencing reveals aberrant signalling pathways as hallmark of treatment-naive anal squamous cell carcinoma. Oncotarget 9, 464–476. https://doi.org/10.18632/oncotarget.23066.

Chen, M., Zhao, H., 2019. Next-generation sequencing in liquid biopsy: cancer screening and early detection. Hum. Genom. 13, 34. https://doi.org/10.1186/s40246-019-0220-8.

Douville, C., Springer, S., Kinde, I., Cohen, J.D., Hruban, R.H., Lennon, A.M., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Karchin, R., 2018. Detection of aneuploidy in patients with cancer through amplification of long interspersed nucleotide elements (LINEs). Proc. Natl. Acad. Sci. USA 115, 1871–1876. https://doi.org/10.1073/pnas.1717846115.

Douville, C., Cohen, J.D., Ptak, J., Popoli, M., Schaefer, J., Silliman, N., Dobbyn, L., Schoen, R.E., Tie, J., Gibbs, P., Goggins, M., Wolfgang, C.L., Wang, T.-L., Shih, I.-M., Karchin, R., Lennon, A.M., Hruban, R.H., Tomasetti, C., Bettegowda, C., Kinzler, K.W., Papadopoulos, N., Vogelstein, B., 2020. Assessing aneuploidy with repetitive element sequencing. Proc. Natl. Acad. Sci. USA 117, 4858–4863. https://doi.org/10.1073/pnas.1910041117.

Duensing, S., Münger, K., 2004. Mechanisms of genomic instability in human cancer: insights from studies with human papillomavirus oncoproteins. Int. J. Cancer 109, 157–162. https://doi.org/10.1002/ijc.11691.

Fadrosh, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M., Ravel, J., 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2, 6. https://doi.org/10.1186/2049-2618-2-6.

Gagne, S.E., Jensen, R., Polvi, A., Da Costa, M., Ginzinger, D., Efird, J.T., Holly, E.A., Darragh, T., Palefsky, J.M., 2005. High-resolution analysis of genomic alterations and human papillomavirus integration in anal intraepithelial neoplasia. J. Acquir. Immune Defic. Syndr. 40, 182–189. https://doi.org/10.1097/01.qai.0000179460.61987.33.

Grasso, C., Butler, T., Rhodes, K., Quist, M., Neff, T.L., Moore, S., Tomlins, S.A., Reinig, E., Beadling, C., Andersen, M., Corless, C.L., 2015. Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. J. Mol. Diagn. 17, 53–63. https://doi.org/10.1016/j.jmoldx.2014.09.008.

Heselmeyer, K., du Manoir, S., Blegen, H., Friberg, B., Svensson, C., Schröck, E., Veldman, T., Shah, K., Auer, G., Ried, T., 1997. A recurrent pattern of chromosomal aberrations and immunophenotypic appearance defines anal squamous cell carcinomas. Br. J. Cancer 76, 1271–1278. https://doi.org/10.1038/bjc.1997.547.

Jee, K.J., Kim, Y.T., Kim, K.R., Kim, H.S., Yan, A., Knuutila, S., 2001. Loss in 3p and 4p and gain of 3q are concomitant aberrations in squamous cell carcinoma of the vulva. Mod. Pathol. 14, 377–381. https://doi.org/10.1038/modpathol.3880321.

Kinde, I., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., 2012. FAST-SeqS: a simple and efficient method for the detection of aneuploidy by massively parallel sequencing. PLoS One 7, e41162. https://doi.org/10.1371/journal.pone.0041162.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Mendelaar, P.A.J., Robbrecht, D.G.J., Rijnders, M., de Wit, R., de Weerd, V., Deger, T., Westgeest, H.M., Aarts, M.J.B., Voortman, J., Martens, J.W.M., van der Veldt, A.A.M., Nakauma-González, J.A., Wilting, S.M., Lolkema, M., 2022. Genome-wide aneuploidy detected by mFast-SeqS in circulating cell-free DNA is associated with poor response to pembrolizumab in patients with advanced urothelial cancer. Mol. Oncol. 16, 2086–2097. https://doi.org/10.1002/1878-0261.13196.

Moody, C.A., Laimins, L.A., 2010. Human papillomavirus oncoproteins: pathways to transformation. Nat. Rev. Cancer 10, 550–560. https://doi.org/10.1038/nrc2886.

Morozova, O., Marra, M.A., 2008. From cytogenetics to next-generation sequencing technologies: advances in the detection of genome rearrangements in tumors. Biochem. Cell Biol. 86, 81–91. https://doi.org/10.1139/O08-003.

Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D., Detering, H., Li, Y., Rodriguez-Castro, J., Dueso-Barroso, A., Bruzos, A.L., Dentro, S.C., Blanco, M.G., Contino, G., Ardeljan, D., Tojo, M., Roberts, N.D., Zumalave, S., Edwards, P.A., Weischenfeldt, J., Puiggròs, M., Chong, Z., Chen, K., Lee, E.A., Wala, J.A., Raine, K.M., Butler, A., Waszak, S.M., Navarro, F.C.P., Schumacher, S.E., Monlong, J., Maura, F., Bolli, N., Bourque, G., Gerstein, M., Park, P.J., Wedge, D.C., Beroukhim, R., Torrents, D., Korbel, J.O., Martincorena, I., Fitzgerald, R.C., Van Loo, P., Kazazian, H.H., Burns, K.H., Campbell, P.J., Tubio, J.M.C., 2020. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat. Genet. 52, 306–319. https://doi.org/10.1038/s41588-019-0562-0.

Scheinin, I., Sie, D., Bengtsson, H., van de Wiel, M.A., Olshen, A.B., van Thuijl, H.F., van Essen, H.F., Eijk, P.P., Rustenburg, F., Meijer, G.A., Reijneveld, J.C., Wesseling, P., Pinkel, D., Albertson, D.G., Ylstra, B., 2014. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res. 24, 2022–2032. https://doi.org/10.1101/gr.175141.114.

Sheen, F., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A., Swergold, G.D., 2000. Reading between the LINEs: human genomic variation induced by LINE-1 Retrotransposition. Genome Res. 10, 1496–1508.

Steenbergen, R.D.M., Snijders, P.J.F., Heideman, D.A.M., Meijer, C.J.L.M., 2014. Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. Nat. Rev. Cancer 14, 395–405. https://doi.org/10.1038/nrc3728.

Sun, Z., Behati, S., Wang, P., Bhagwate, A., McDonough, S., Wang, V., Taylor, W., Cunningham, J., Kisiel, J., 2023. Performance comparisons of methylation and structural variants from low-input whole-genome methylation sequencing. Epigenomics 15, 11–19. https://doi.org/10.2217/epi-2022-0453.

Swarts, D.R.A., Voorham, Q.J.M., van Splunter, A.P., Wilting, S.M., Sie, D., Pronk, D., van Beurden, M., Heideman, D.A.M., Snijders, P.J.F., Meijer, C.J.L.M., Steenbergen, R.D.M., Bleeker, M.C.G., 2018. Molecular heterogeneity in human papillomavirus-dependent and -independent vulvar carcinogenesis. Cancer Med. 7, 4542–4553. https://doi.org/10.1002/cam4.1633.

Thomas, L.K., Bermejo, J.L., Vinokurova, S., Jensen, K., Bierkens, M., Steenbergen, R., Bergmann, M., von Knebel Doeberitz, M., Reuschenbach, M., 2014. Chromosomal gains and losses in human papillomavirus-associated neoplasia of the lower genital tract - a systematic review and meta-analysis. Eur. J. Cancer 50, 85–98. https://doi.org/10.1016/j.ejca.2013.08.022.

Thuijs, N.B., Berkhof, J., Özer, M., Duin, S., van Splunter, A.P., Snoek, B.C., Heideman, D.A.M., van Beurden, M., Steenbergen, R.D.M., Bleeker, M.C.G., 2021. DNA methylation markers for cancer risk prediction of vulvar intraepithelial neoplasia. Int. J. Cancer. https://doi.org/10.1002/ijc.33459.

Tian, R., Cui, Z., He, D., Tian, X., Gao, Q., Ma, X., Yang, J.-R., Wu, J., Das, B.C., Severinov, K., Hitzeroth, I.I., Debata, P.R., Xu, W., Zhong, H., Fan, W., Chen, Y., Jin, Z., Cao, C., Yu, M., Xie, W., Huang, Z., Bao, Y., Xie, H., Yao, S., Hu, Z., 2019. Risk stratification of cervical lesions using capture sequencing and machine learning method based on HPV and human integrated genomic profiles. Carcinogenesis 40, 1220–1228. https://doi.org/10.1093/carcin/bgz094.

van de Wiel, M.A., Kim, K.I., Vosse, S.J., van Wieringen, W.N., Wilting, S.M., Ylstra, B., 2007. CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics 23, 892–894. https://doi.org/10.1093/bioinformatics/btm030.

van der Zee, R.P., Richel, O., van Noesel, C.J.M., Ciocănea-Teodorescu, I., van Splunter, A.P., Ter Braak, T.J., Nathan, M., Cuming, T., Sheaff, M., Kreuter, A., Meijer, C.J.L.M., Quint, W.G.V., de Vries, H.J.C., Prins, J.M., Steenbergen, R.D.M., 2021a. Cancer risk stratification of anal intraepithelial neoplasia in human immunodeficiency virus-positive men by validated methylation markers associated with progression to cancer. Clin. Infect. Dis. 72, 2154–2163. https://doi.org/10.1093/cid/ciaa397.

van der Zee, R.P., van Noesel, C.J.M., Martin, I., Ter Braak, T.J., Heideman, D.A.M., de Vries, H.J.C., Prins, J.M., Steenbergen, R.D.M., 2021b. DNA methylation markers have universal prognostic value for anal cancer risk in HIV-negative and HIV-positive individuals. Mol. Oncol. https://doi.org/10.1002/1878-0261.12926.

Verlaat, W., Van Leeuwen, R.W., Novianti, P.W., Schuuring, E., Meijer, C.J.L.M., Van Der Zee, A.G.J., Snijders, P.J.F., Heideman, D.A.M., Steenbergen, R.D.M., Wisman, G.B.A., 2018. Host-cell DNA methylation patterns during high-risk HPV-induced carcinogenesis reveal a heterogeneous nature of cervical pre-cancer. Epigenetics 13, 769–778. https://doi.org/10.1080/15592294.2018.1507197.

Wilting, S.M., Steenbergen, R.D.M., 2016. Molecular events leading to HPV-induced high grade neoplasia. Papillomavirus Res. 2, 85–88. https://doi.org/10.1016/j.pvr.2016.04.003.

Wilting, S.M., Steenbergen, R.D.M., Tijssen, M., van Wieringen, W.N., Helmerhorst, T.J.M., van Kemenade, F.J., Bleeker, M.C.G., van de Wiel, M.A., Carvalho, B., Meijer, G.A., Ylstra, B., Meijer, C.J.L.M., Snijders, P.J.F., 2009. Chromosomal signatures of a subset of high-grade premalignant cervical lesions closely resemble invasive carcinomas. Cancer Res. 69, 647–655. https://doi.org/10.1158/0008-5472.CAN-08-2478.