USCAP

ARTICLE

RNA-based high-risk HPV genotyping and identification of high-risk HPV transcriptional activity in cervical tissues

Corina N. A. M. van den Heuvel[1] · Diede L. Loopik [ID][2] · Renée M. F. Ebisch[2] · Duaa Elmelik[1] ·
Karolina M. Andralojc[1,3] · Martijn Huynen[4] · Johan Bulten[5] · Ruud L. M. Bekkers[2] · Leon F. A. G. Massuger[2] ·
Willem J. G. Melchers [ID][3] · Albert G. Siebers[5] · William P. J. Leenders[1]

Abstract
Nearly all cervical cancers are initiated by a persistent infection with one of the high-risk human papillomaviruses (high-risk HPV). High-risk HPV DNA testing is highly sensitive but cannot distinguish between active, productive infections and dormant infections or merely deposited virus. A solution for this shortcoming may be the detection of transcriptional activity of viral oncogenes instead of mere presence of high-risk HPVs. In this study, fresh-frozen cervical tissues ($n = 22$) were subjected to high-risk HPV DNA detection using the line probe assay and to targeted RNA next-generation sequencing using single-molecule molecular inversion probes. Targeted RNA sequencing was applied for (1) RNA-based genotyping of high-risk HPV, giving information on specific HPV-subtype (2) discrimination of E2, E6, and E7 transcripts and (3) discovery of possible non-HPV cancer biomarkers. Data were analyzed using computational biology. Targeted RNA sequencing enabled reliable genotyping of high-risk HPV subtypes and allowed quantitative detection of E2, E6, and E7 viral gene expression, thereby discriminating cervical lesions from normal cervical tissues. Moreover, targeted RNA sequencing identified possible cervical cancer biomarkers other than high-risk HPV. Interestingly, targeted RNA sequencing also provided high-quality transcription profiles from cervical scrape samples, even after 1 week of dry storage or storage in Preservcyt fixative. This proof of concept study shows that targeted RNA sequencing can be used for high-risk HPV genotyping and simultaneous detection of high-risk HPV gene activity. Future studies are warranted to investigate the potential of targeted RNA sequencing for risk assessment for the development of cervical lesions, based on molecular analysis of cervical scrapes.

✉ William P. J. Leenders
   William.Leenders@radboudumc.nl

[1] Department of Biochemistry, Radboud Institute for Molecular Life Sciences, 6525 GA Nijmegen, The Netherlands

[2] Department of Obstetrics and Gynecology, Radboud university medical center, 6500 HB Nijmegen, The Netherlands

[3] Department of Medical Microbiology, Radboud university medical center, 6500 HB Nijmegen, The Netherlands

[4] Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, 6525 GA Nijmegen, The Netherlands

[5] Department of Pathology, Radboud university medical center, 6500 HB Nijmegen, The Netherlands

Introduction

Cervical cancer is the fourth most frequent cancer in women worldwide, with approximately half a million new cases each year of whom ~50% are lethal [1]. Nearly all cervical cancers are associated with infection with one of 14 high-risk human papillomaviruses (high-risk HPV), subtypes HPV16 and HPV18 being responsible for ~70% of cases [1, 2]. Persistent infection with high-risk HPV is considered a necessary cause for the development of squamous cell carcinoma of the cervix, while cervical adenocarcinomas are often, but not always, HPV-related [2–6].

The life-time risk for women to acquire an HPV infection is ~80%, but by far the majority of infections are cleared within 2 years. In ~10% of women the infection persists [1, 7, 8]. These cases may be associated with the development of low-grade lesions (cervical intraepithelial neoplasia), most of which however regress within a year. Only a minority of all persistent high-risk HPV infections

becomes a transforming infection that may progressively develop into high-grade cervical intraepithelial neoplasia and cancer, a process that may take years [9]. The risk of developing high-grade cervical intraepithelial neoplasia is higher after infection with HPV16/18 than with one of the other high-risk HPV subtypes [10]. Since high-risk HPV DNA testing cannot distinguish between transient and clinically relevant, transforming infections, methods to enable risk prediction in high-risk HPV DNA-positive tissues are needed.

In its episomal form HPV infections are characterized by expression of the early gene E2, a repressor of the oncogenes E6 and E7. Integration of high-risk HPV in the host genome is thought to be an important step in the oncogenesis of cervical cancer [11, 12], as it leads to reduced E2 expression and activation of the E6 and E7 oncogenes [13, 14]. E6 and E7 gene products are primarily responsible for oncogenic transformation by interference with p53 and pRb proteins, disrupting DNA repair regulation and cell cycle checkpoints [12, 15–22]. Overexpression of E6 and E7 oncogenes can therefore be used to identify transforming infections, whereas levels of E2 expression may provide additional information on viral integration state [6]. To obtain a comprehensive understanding of cervical carcinogenesis in clinical samples, methods for simultaneous detection of E2, E6, and E7 transcripts of all high-risk HPV subtypes are therefore needed and multiple tests for HPV E6/E7 mRNA detection are now available [23, 24]. The slow development from initial high-risk HPV infection to cervical cancer makes it likely that gradual accumulation of tumorigenic events in cervical intraepithelial neoplasia is required for progression to cancer [25]. Mapping of these events will further contribute to our knowledge of cancer biology in the cervix.

The aims of this study were therefore to (1) investigate the value of targeted RNA next-generation sequencing for high-risk HPV genotyping and oncogene activity monitoring in cervical tissues, (2) identify possible biomarkers other than high-risk HPV that may distinguish normal cervical tissue from cervical cancer, and (3) determine the quality of targeted RNA sequencing data originating from cervical scrape samples, possibly allowing better non-invasive risk assessment for the development of cervical cancer.

Materials and methods

Patient material

In this prospective single-center study frozen cervical tissue samples were included from 22 non-pregnant Dutch women, presenting at the department of Obstetrics and Gynecology of the Radboudumc for surgery of the uterus and/or cervix for different clinical indications. Punch biopsies from cervical tissue were taken during surgery or after extirpation. Tissues were snap frozen in liquid nitrogen and stored at −80 °C until RNA isolation. H&E stainings were performed on 4 µm cryosections and reviewed by a pathologist (JB, Table 1). Based on histopathology, which was concordant with clinical patient diagnosis, tissues were classified as normal cervix tissue ($n = 4$, N1-4) or cervical cancer ($n = 18$, composed of 15 squamous cell cancers [SCC1–15] and 3 adenocarcinomas [AdCa1–3]). Evalyn Brush (Rovers Medical Devices B.V., Oss, Netherlands) cervicovaginal samples were collected in triplicate from two healthy women and either stored dry or in Preservcyt solution (Hologic, Marlborough, MA, USA) for different time spans (dry: 1 h, 1 week; Preservcyt: 1 h, 24 h, 1 week, 2 weeks). Written informed consent was obtained from all women in this study. The protocol was approved by the regional institutional review board (No. 2014–1295). All methods were performed in accordance with the guidelines for use of human tissue of the Radboudumc and the tissue samples were anonymized to the researchers.

Table 1 Tissue samples with diagnosis and cancer stage

Sample	Tissue diagnosis	Cancer stage
N1	Normal cervix	–
N2	Normal cervix	–
N3	Normal cervix	–
N4	Normal cervix	–
SCC1	SCC	IIB
SCC2	SCC	IVB
SCC3	SCC	IIB
SCC4	SCC	IIB
SCC5	SCC	IIB
SCC6	SCC	IIB
SCC7	SCC	IIB2
SCC8	SCC	IVA
SCC9	SCC	IIB
SCC10	SCC	IIIB
SCC11	SCC	IVA
SCC12	SCC	IIB
SCC13	SCC	IB1
SCC14	SCC	IIB
SCC15	SCC	IVA
AdCa1	AdCa	IIB
AdCa2	AdCa	IB2
AdCa3	AdCa	IB2

Diagnoses were based on histopathological review of the analyzed biopsies and were concordant with clinical patient diagnosis

AdCa adenocarcinoma, SCC squamous cell carcinoma

HPV genotyping

DNA was isolated from all tissues using MagNA Pure (Roche, Bazel, Switzerland). The purified DNA was eluted in 50 µl TE-buffer. For detection and genotyping of HPV, broad spectrum HPV amplification was performed using the short-PCR-fragment line probe assay assay (SPF$_{10}$-LiPA$_{25}$; Labo Bio-medical Products B.V., Rijswijk, The Netherlands). This assay amplifies a small fragment of 65 bp from the L1 open reading frame and allows detection of a broad range of HPV genotypes with high analytical sensitivity [26].

Targeted RNA next-generation sequencing

Targeted RNA sequencing was performed essentially as described before [27], without prior knowledge on HPV status. RNA was isolated from 10 µm cryosections using TRIzol reagent (ThermoFisher Scientific, Waltham, MA, USA) and reverse transcribed with Superscript II (ThermoFisher Scientific) using random hexamer primers, according to the manufacturer's instructions. A previously described cohort of single-molecule molecular inversion probes [27] was expanded with among others probes recognizing E2, E6, and E7 transcripts from high-risk HPV types 16, 18, 31, 33, 45, and 52. Single-molecule molecular inversion probes to detect E2, E6, and E7 transcripts from HPV39 and HPV56 were initially not included in the panel, but were added at a later stage to confirm line probe assay results. Probe design was based on the MIPgen algorithm as described by Boyle et al. [28], and included a random octanucleotide unique molecule identifier (UMI).

All assay preparations were performed at 4 °C. Single-molecule molecular inversion probe pool preparation, capture and exonuclease treatment were performed using the Microlab STAR automated liquid handling workstation (Hamilton, Reno, Nevada, USA). Single-molecule molecular inversion probes were pooled at 100 µM/probe (246 transcripts, 1301 probes) and phosphorylated using T4 Polynucleotide Kinase (New England Biolabs, Ipswich, MA, USA) in T4 DNA ligase buffer (NEB) for 45 min at 37 °C, followed by 20 min inactivation at 65 °C. The capture reaction was performed using 15–50 ng of cDNA and a 1 pM pool of phosporylated probes. Capture and enzymatic circularization by primer extension and ligation were performed in a reaction mixture containing Ampligase buffer (Epicentre, Madison, WI, USA), dNTPs, Hemo KlenTaq enzyme (New England Biolabs, NEB, Ipswich, MA, USA) and DNA ligase (Ampligase, Epicentre) by incubation for 10 min at 95 °C followed by 18 h at 60 °C. Non-circularized single-molecule molecular inversion probes and remaining RNA and cDNA were removed by exonuclease treatment as described [27]. The pool of circularized probes was subjected to PCR amplification using the Microlab STARlet automated liquid handling workstation (Hamilton), with 2× iProof High-Fidelity DNA Polymerase master Mix (Bio-Rad, Hercules, CA) and a primer set including a unique barcoded reverse primer for each sample. A total of 2 µl of each sample was pooled using the Microlab STAR and 266 bp amplicons were purified using AMPureXP beads (Beckman Colter Genomics, High Wycombe, UK). After quality control on a TapeStation 2200 (Agilent Technologies, Santa Clara, CA, USA) and quantification using Qubit (Life Technologies, ThermoFisher Scientific, Waltham, MA USA), a 4 nM library was sequenced on the Illumina Nextseq platform (Illumina, San Diego, CA) at the Radboudumc sequencing facility. Reads were mapped against reference transcripts (UCSC human genome assembly hg19 and variant-specific FASTA sequences) using the SeqNext module of JSI SequencePilot version 4.2.2 build 502 (JSI Medical Systems, Ettenheim, Germany). All identical PCR products were reduced to one consensus read (unique read) using the UMI. Unique read counts for each single-molecule molecular inversion probe in a sample were divided by the total of unique read counts in that sample and multiplied by 10^6 (Fragments per Million, FPM). Individual transcript levels were expressed as mean FPM of all probes targeting that transcript. Less than 10 unique read counts for a high-risk HPV transcript were considered insignificant. Variant calling was performed within SeqNext, excluding all variants with a coverage of <10% variant reads.

Statistical analysis

Statistical analyses were performed in R (version 3.4.3). Mean FPM values were log$_2$ transformed (after addition of 0.01 to prevent log(0) transformation errors) and clustered in an unsupervised manner using the Manhattan distance and Ward.D2 method, and translated into a heatmap using ClustVis. A Wilcoxon Mann–Whitney U test was performed to study differential gene expression between clusters. Multiple testing corrections were done using Benjamini–Hochberg (FDR < 0.01).

Results

High-risk HPV DNA vs. RNA detection

Line probe assay testing detected high-risk HPV DNA in two out of four normal cervix tissues, all 15 squamous cell carcinomas, and two of the three adenocarcinoma tissues (Fig. 1a). Profiling the same set of tissues for HPV RNAs in the targeted RNA sequencing assay, blinded to the line probe assay results, showed activity of viral oncogenes in all HPV DNA-positive squamous cell carcinoma and
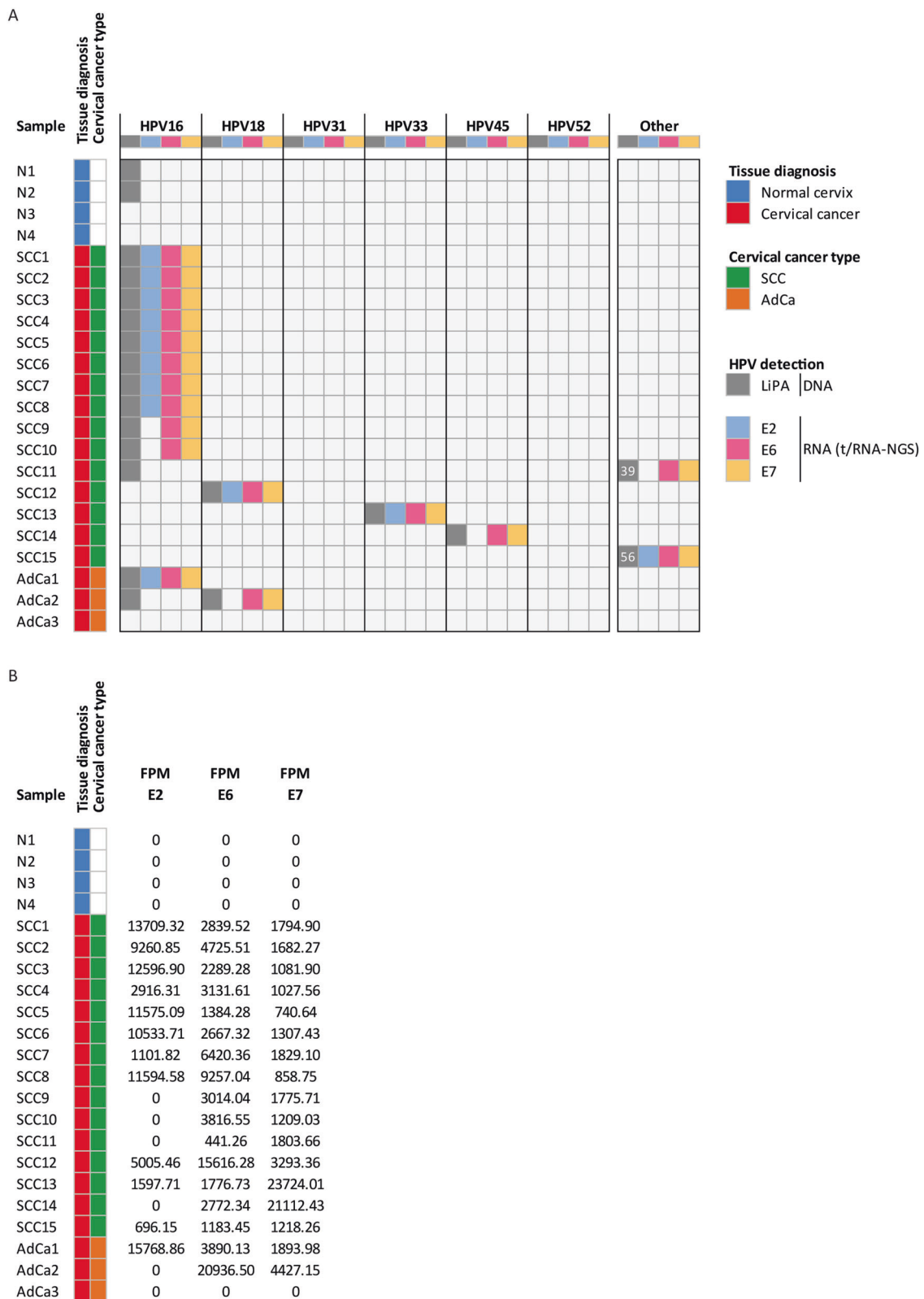
A

Sample	Tissue diagnosis	Cervical cancer type	HPV16	HPV18	HPV31	HPV33	HPV45	HPV52	Other

Tissue diagnosis
- Normal cervix
- Cervical cancer

Cervical cancer type
- SCC
- AdCa

HPV detection
- LiPA | DNA
- E2
- E6 | RNA (t/RNA-NGS)
- E7

B

Sample	Tissue diagnosis	Cervical cancer type	FPM E2	FPM E6	FPM E7
N1			0	0	0
N2			0	0	0
N3			0	0	0
N4			0	0	0
SCC1			13709.32	2839.52	1794.90
SCC2			9260.85	4725.51	1682.27
SCC3			12596.90	2289.28	1081.90
SCC4			2916.31	3131.61	1027.56
SCC5			11575.09	1384.28	740.64
SCC6			10533.71	2667.32	1307.43
SCC7			1101.82	6420.36	1829.10
SCC8			11594.58	9257.04	858.75
SCC9			0	3014.04	1775.71
SCC10			0	3816.55	1209.03
SCC11			0	441.26	1803.66
SCC12			5005.46	15616.28	3293.36
SCC13			1597.71	1776.73	23724.01
SCC14			0	2772.34	21112.43
SCC15			696.15	1183.45	1218.26
AdCa1			15768.86	3890.13	1893.98
AdCa2			0	20936.50	4427.15
AdCa3			0	0	0

Fig. 1 a DNA versus RNA detection of high-risk HPV. HPV DNA presence (line probe assay; LiPA) versus presence of high-risk HPV E2, E6, and E7 gene transcripts (targeted RNA sequencing; t/RNA-NGS) in cervical tissue samples. Targeted RNA sequencing-based high-risk HPV genotyping was concordant with line probe assay-based genotyping. All cancers in which high-risk HPV was detected by line probe assay expressed E6 and E7 oncogenes, whereas all normal tissues did not express HPV genes, irrespective of line probe assay outcome. E2 expression was not detected in four squamous cell carcinomas (SCC) and one adenocarcinoma (AdCa). **b** Mean FPM values of E2, E6, and E7 transcripts of the high-risk HPV RNA-based genotypes as shown in **a**

adenocarcinoma samples. All genotypes identified by RNA sequencing were concordant with the outcome of the line probe assay assay (Fig. 1a). In two squamous cell carcinomas the line probe assay detected HPV39 and HPV56, two subtypes for which no single-molecule molecular inversion probes were included in the original targeted RNA sequencing panel. Transcriptional activity of these subtypes was confirmed in an additional analysis (Fig. 1a). In AdCa2, HPV16 and HPV18 were detected by line probe assay, but only HPV18 E6 and E7 oncogene activity was detected by targeted RNA sequencing. In SCC11 HPV16 and HPV39 DNA was detected, but only HPV39 E6 and E7 RNA was found.

Samples that tested high-risk HPV-positive by line probe assay analysis but were transcriptionally inactive by targeted RNA sequencing were normal cervical tissues ($n = 2$). One adenocarcinoma tested negative in both DNA and RNA assays, in agreement with previous studies showing that cervical adenocarcinomas are not always caused by high-risk HPV infection [2–5].

High-risk HPV E2 versus E6/E7 expression

We next studied whether we can distinguish E2 and E6/E7 viral gene expression, which may be used to discriminate transforming infections from clinically irrelevant infections. In all 17 high-risk HPV RNA-positive tissues oncogenes E6 and E7 were expressed (Fig. 1a), with expression levels ranging from ~440 to 21,000 FPM (Fig. 1b). E6 expression exceeded E7 expression in 13/17 tissues (~75%). E2 expression was undetectable in 5 of the 17 samples (~30%), of which four were squamous cell carcinomas and one was adenocarcinoma. Coexpression of E2 (~740–23,700 FPM, Fig. 1b), E6 and E7 were observed in the other 12 tissues (~70%).

Non-HPV biomarkers of cervical cancer

To discover cervical cancer biomarkers other than high-risk HPV, we performed agglomerative unsupervised cluster analysis of all gene expression data excluding HPV (Fig. 2). Normal cervical tissues and cancer tissues clustered together in separate groups. Within the cancer cluster, adenocarcinomas clustered together, including the HPV-negative AdCa3. SCC11, the only HPV39-positive squamous cell carcinoma in the cohort, did not cluster with the other cervical cancers. We found no association of sample clusters with high-risk HPV E2 versus E6/E7 expression levels.

Statistical analysis revealed that 52 genes were differentially expressed between normal cervix and cervical cancer tissues with $p < 0.05$ (Table 2, see Supplementary Table S1 for all transcripts in the analysis). Interestingly, many tyrosine kinases that are considered targets for

precision medicines (FGFRs, PDGFRs, KDR, KIT, AXL, ERBB2, NTRK2) were expressed at lower levels in cancers than in normal cervix tissue, with the exception of the proto-oncogene MET. We previously observed a similar trend in renal cancer [29]. In general, genes involved in glycolysis were expressed at higher levels in cancer tissues (HK2 [hexokinase 2)], PGK1 [phosphoglycerate kinase 1], ENO1 [enolase 1], SLC16A3 [monocarboxylate transporter 4], CA9 [carbonic anhydrase 9]). These features are in accordance with the general concept that cancers experience hypoxia and shift energy production to glycolysis [30]. Expression of the tumor suppressor gene PTEN was lower in the cancer cluster, as was the expression of androgen receptor (AR). EpCAM (epithelial cell adhesion molecule) and the PSMA-coding gene FOLH1 were expressed at relatively high levels in the cancer group.

In addition to gene expression, targeted RNA sequencing also allows for variant detection. The mismatch repair endonuclease PMS2 mutation c.187G>A (p.Val63Met, classified as confirmed somatic in five neoplasms in the liver in the COSMIC database (https://cancer.sanger.ac.uk/cosmic/analyses) and has a likely pathogenic CADD score of 18.5) was detected in 15/18 cancer tissues and in 1/4 normal cervix tissues (not shown).

Targeted RNA sequencing on cervical scrape samples

In the context of nation-wide cervical cancer screening programs, cervical scrape samples are collected for cytology-based screening, which is increasingly preceded by molecular screening for the presence of high-risk HPV DNA. We therefore tested whether it is feasible to implement targeted RNA sequencing analysis on cervical scrape samples, collected under conditions that are relevant for population-based screening programs (with delays between scraping and analysis). Storage of scrape samples at room temperature for up to 1 week, either dry or in Preservcyt fixative, still allowed the generation of good-quality targeted RNA sequencing profiles (Fig. 3). After 2 weeks of Preservcyt fixation, the quality of targeted RNA sequencing data (expressed as total unique reads obtained in a sample) decreased.

Discussion

Infection of cervical tissues with high-risk HPV is not necessarily accompanied by oncogenic activity of high-risk HPV. This is a relevant observation because molecular testing for the presence of high-risk HPV DNA is increasingly implemented in population-based screening programs. High-risk HPV DNA screening is very sensitive (90–95% vs. 30–87% for cytology), objective and allows
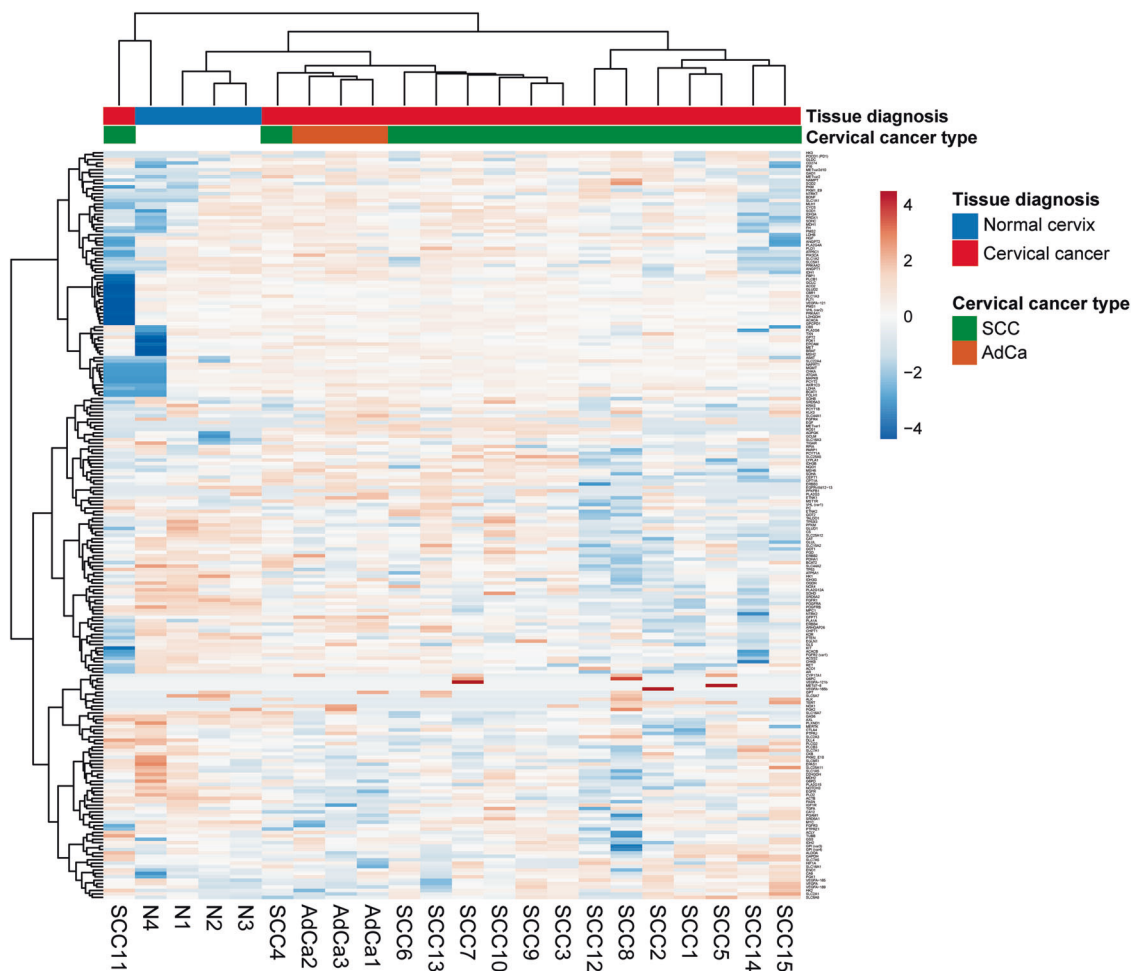
Fig. 2 Heatmap of targeted RNA sequencing outcome of gynecological tissues. Samples include normal cervical tissue ($n = 4$) and cervical cancer ($n = 18$). Histopathology diagnosis is presented in the upper bars. The heatmap of 219 genes was generated by unsupervised hierarchical clustering using the Manhattan distance and Ward.D2 clustering method. Note that normal cervical tissues and cancer tissues clustered together, with the exception of one squamous cell carcinoma clustering with the normal tissues. AdCa adenocarcinoma, SCC squamous cell carcinoma

high-throughput testing [31–33] but does not distinguish between deposited and dormant virus, transient infections and clinically relevant active and producing infections. Even with cytomorphology screening as triage test, this results in many unnecessary colposcopy referrals, overtreatment and follow-up. This problem may be solved by measuring transcripts from high-risk HPV E6 and E7 oncogenes in infected tissues [34–38]. A number of studies have now shown the increased specificity of HPV-RNA testing compared with HPV-DNA testing [23, 24, 39].

The aim of this proof of concept study was to investigate the value of targeted RNA sequencing technology to detect high-risk HPV oncogene activity in well-characterized gynecological tissues. We previously showed that targeted RNA sequencing yields reliable quantitative gene expression data and mutation data [27, 29, 40, 41].

During integration into the host genome the reading frame of the early gene E2 is often disrupted [42], alleviating repression of the E6/E7 oncogenes. Concomitant measurement of E2, E6, and E7 transcripts may therefore provide information on high-risk HPV integration status. Till now, comprehensive analysis of high-risk HPV transcripts required that for each tissue sample at least 44 RT-qPCR experiments should be performed (E2, E6, and E7 RT-qPCR for each of the 14 high-risk HPV subtypes, and a positive and negative control RT-qPCR) making such analyses labor-intensive. With targeted RNA sequencing, transcripts from the viral genes E2, E6, and E7 from all 14 high-risk HPV genotypes can be quantitatively detected in one go, whereas the sequence information concomitantly provides direct and reliable high-risk HPV subtype identification.

We observed E6/E7 gene expression in the absence of E2 in 30% of cases, suggesting HPV integration in the host genome. In the remaining cancers, E2 and E6/E7 genes were coexpressed. There are several possible explanations for this finding. E6/E7 expression levels may have become

Table 2 Gene expression in cervical cancer versus normal cervix ($p < 0.05$)

Gene	Mean FPM normal cervix	Mean FPM Cervical cancer	P-value	FDR	FC cancer/normal
CAT	796.67	348.41	0.0003	0.0000	−2.29
FGFR1	551.37	52.56	0.0003	0.0001	−10.49
PDGFRB	2316.79	432.58	0.0003	0.0001	−5.36
PDGFRA	3115.21	532.85	0.0005	0.0002	−5.85
ACTB	17688.89	6758.95	0.0011	0.0002	−2.62
MET	390.61	1979.50	0.0011	0.0003	5.07
NOTCH3	789.16	243.74	0.0011	0.0003	−3.24
GAS6	998.86	232.10	0.0019	0.0004	−4.30
GLUD2	271.61	101.39	0.0019	0.0004	−2.68
MYC	2543.53	826.28	0.0019	0.0004	−3.08
SLC1A2	87.97	12.64	0.0024	0.0005	−6.95
AR	877.29	47.97	0.0025	0.0005	−18.29
ACACB	145.22	28.81	0.0025	0.0006	−5.04
FGFR3	215.65	58.69	0.0025	0.0006	−3.67
HGF	161.87	29.68	0.0025	0.0007	−5.45
SRD5A2	94.98	2.76	0.0026	0.0007	−34.32
HK1	1826.48	946.42	0.0033	0.0007	−1.93
KDR	617.99	304.08	0.0033	0.0008	−2.03
NTRK2	1553.38	402.43	0.0033	0.0009	−3.86
PLD2	965.71	351.71	0.0033	0.0009	−2.75
PLXND1	650.33	207.36	0.0033	0.0010	−3.14
SLC16A3	321.93	1195.80	0.0049	0.0010	3.71
ACO2	408.74	170.07	0.0074	0.0011	−2.40
KIT	498.57	140.41	0.0074	0.0011	−3.55
MPC1	1455.09	734.18	0.0074	0.0011	−1.98
PLA2G12A	283.94	150.84	0.0074	0.0012	−1.88
PTEN	3456.25	1768.13	0.0074	0.0012	−1.95
CKB	1907.10	689.03	0.0104	0.0013	−2.77
ERBB2	1044.52	607.66	0.0104	0.0013	−1.72
GLUL	2661.85	1241.21	0.0104	0.0014	−2.14
PDHA1	511.81	338.04	0.0104	0.0014	−1.51
ATP5A1	2252.63	1617.85	0.0145	0.0014	−1.39
CA9	9.52	355.97	0.0145	0.0015	37.34
ETNK2	186.15	91.95	0.0145	0.0015	−2.02
GLUD1	1739.34	1042.18	0.0145	0.0016	−1.67
ACO1	352.97	230.81	0.0194	0.0016	−1.53
ENO1	8976.05	15745.72	0.0194	0.0017	1.75
FASN	412.67	192.94	0.0194	0.0017	−2.14
OGDH	1003.01	587.77	0.0194	0.0018	−1.71
PGK1	1959.84	3548.33	0.0194	0.0018	1.81
SLC16A7	381.38	176.34	0.0194	0.0018	−2.16
FOLH1	29.92	194.47	0.0215	0.0019	6.50
RET	17.38	8.54	0.0237	0.0019	−2.03
D2HGDH	290.76	117.72	0.0257	0.0020	−2.47
HK2	250.37	405.29	0.0257	0.0020	1.62
ACSS2	351.86	224.88	0.0331	0.0021	−1.56
AXL	2041.85	919.21	0.0331	0.0021	−2.22
FGFR2 (var1)	103.83	30.01	0.0370	0.0021	−3.46
EPCAM	201.30	1714.79	0.0424	0.0022	8.52
IDH3G	546.90	413.66	0.0424	0.0022	−1.32
TP53	125.99	81.80	0.0424	0.0023	−1.54
SLC5A7	20.82	1.09	0.0499	0.0023	−18.86

Gene expression differences as tested with a Wilcoxon Mann–Whitney U test and Benjamini–Hochberg correction for multiple testing ($p < 0.05$, FDR < 0.01). Note that for significance, the p-value must not exceed the FDR. Fold changes were calculated as (cluster B/A) or $-1/$(cluster B/A) to prevent fold changes >0 and <1. Here, $p < 0.05$ was considered relevant
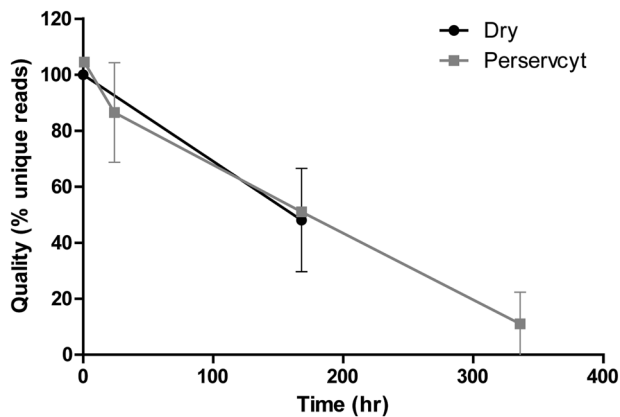
FC fold change

Fig. 3 Quality of targeted RNA sequencing data originating from cervical scrape samples. Data quality is expressed as percentage unique reads relative to 1 h sample preservation (100%). Up to 1 week of sample preservation either dry or in Preservcyt solution, targeted RNA sequencing generates good-quality transcription profiles from cervical scrapes ($n = 2$). Total number of unique reads varied per sample (depending on amount of cells in the scrapes and transcriptional activity therein) and started at excellent values of 38,000 and 225,000 for both samples, respectively, at $T = 0$

independent of E2 repression, e.g., via methylation of the E2 binding sites [12, 14, 15, 42, 43], or analyzed tissues may have been composed of mixed episomal and integrated viruses (either in the same cells or in different cell clones). Our data also demonstrate that multiple high-risk HPV types can be present in cancer tissues whereas only one of these subtypes is responsible for E6/7 oncogene expression. From this study we cannot conclude whether these cases concerned coinfected cells or multiple cell clones monoinfected by different high-risk HPV types.

Targeted RNA sequencing analysis has multiple other advantages compared with standard techniques. The test uses Illumina barcoding technology and therefore can be performed in parallel for as many samples as there are barcoded primers (currently 386, but expandable), making the test high-throughput and highly cost-effective. A further important advantage is that the current test-format measures expression levels of a number of potentially cancer-associated genes and differentiation genes that are associated with normal physiology. This is highly relevant since host genes play an essential role not only in the development of cancer but also in prevention thereof [44]. Analyzing expression profiles of gynecological tissues may therefore also allow for the identification of non-HPV-related cancers [2–5]. This notion is strengthened by our finding that AdCa3 in our cohort is high-risk HPV-negative, but still groups with the two high-risk HPV-positive adenocarcinomas in unsupervised clustering analysis. Furthermore, supervised analysis of targeted RNA sequencing data from normal tissues and cancers identified a number of genes that are differentially expressed in cancers and normal

tissues. Upregulated genes in cancers involved genes involved in metabolism and confirmed increased hypoxia in cancerous tissues. Other genes that were higher expressed in cancers were *EpCAM* and *FOLH1*, confirming previous reports [45, 46] and the proto-oncogene *MET*. Relatively low expression levels of *AR* (androgen receptor) in cancer were also in accordance with previous reports [47, 48] and illustrates that in cancers functions may be lost that are present in tissues with normal physiology. Further clinical research and computational biology on targeted RNA sequencing profiles from large numbers of samples in validation sets should confirm whether gene expression profiles can be further refined and translated into transcript classifiers that define a risk for cancer development.

Of high interest, targeted RNA sequencing also allows to investigate potential associations of gene mutations with cervical cancer. Such profiles may be of additional importance for risk assessment of cervical biopsies. Our finding that mutations in the DNA repair protein PMS2 are associated with the cancers in our cohort suggests that PMS2 mutations may play a role in HPV-induced carcinogenesis, inducing genetic instability as a first step in the carcinogenic process. PMS2 mutations are common in Lynch syndrome and a recent publication showed a slightly increased risk of developing endometrial cancer in women with PMS2 mutations [49]. Our results suggest that this PMS2 variant could possibly be a marker for cervical cancer. However, larger studies are needed to further investigate this.

Whereas this proof of concept study was performed on surgically obtained tissues, we show here that targeted RNA sequencing also works for cervical scrapes, even scrapes that have been stored dry or in Preservcyt fixative at room temperature for up to a week. This is an important finding as it illustrates that targeted RNA sequencing may allow risk assessment of women for developing gynecological cancers based on noninvasive analysis of scrapes. Including probes to detect somatic mutations such as recently identified [25] may further increase the value of such risk assessment. These studies are now ongoing.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer. Accessed 3 Jan 2019.

2. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999;189:12–9.

3. Andersson S, Rylander E, Larsson B, Strand A, Silfversvard C, Wilander E. The role of human papillomavirus in cervical adenocarcinoma carcinogenesis. Eur J Cancer. 2001;37:246–50.

4. Burd EM. Human papillomavirus and cervical cancer. Clin Microbiol Rev. 2003;16:1–17.

5. Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: variation by geographical region, histological type and year of publication. Int J Cancer. 2011; 128:927–35.

6. Nguyen HP, Ramirez-Fort MK, Rady PL. The biology of human papillomaviruses. Curr Probl Dermatol. 2014;45:19–32.

7. Ho GY, Bierman R, Beardsley L, Chang CJ, Burk RD. Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med. 1998;338:423–8.

8. Rodriguez AC, Schiffman M, Herrero R, Wacholder S, Hildesheim A, Castle PE, et al. Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections. J Natl Cancer Inst. 2008;100:513–7.

9. Vink MA, Bogaards JA, van Kemenade FJ, de Melker HE, Meijer CJ, Berkhof J. Clinical progression of high-grade cervical intraepithelial neoplasia: estimating the time to preclinical cervical cancer from doubly censored national registry data. Am J Epidemiol. 2013;178:1161–9.

10. Bulk S, Bulkmans NW, Berkhof J, Rozendaal L, Boeke AJ, Verheijen RH, et al. Risk of high-grade cervical intra-epithelial neoplasia based on cytology and high-risk HPV testing at baseline and at 6-months. Int J Cancer. 2007;121:361–7.

11. Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H. et al. Integrated genomic and molecular characterization of cervical cancer. Nature. 2017; 543:378–84.

12. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. PLoS Pathog. 2017;13: e1006211.

13. Stoler MH, Rhodes CR, Whitbeck A, Wolinsky SM, Chow LT, Broker TR. Human papillomavirus type 16 and 18 gene expression in cervical neoplasias. Hum Pathol. 1992;23:117–28.

14. Tsakogiannis D, Gortsilas P, Kyriakopoulou Z, Ruether IG, Dimitriou TG, Orfanoudakis G, et al. Sites of disruption within E1 and E2 genes of HPV16 and association with cervical dysplasia. J Med Virol. 2015;87:1973–80.

15. Oyervides-Munoz MA, Perez-Maya AA, Rodriguez-Gutierrez HF, Gomez-Macias GS, Fajardo-Ramirez OR, Trevino V, et al. Understanding the HPV integration and its progression to cervical cancer. Infect Genet Evol. 2018;61:134–44.

16. Cripe TP, Haugen TH, Turk JP, Tabatabai F, Schmid PG 3rd, Durst M, et al. Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis. EMBO J. 1987;6:3745–53.

17. Takebe N, Tsunokawa Y, Nozawa S, Terada M, Sugimura T. Conservation of E6 and E7 regions of human papillomavirus types 16 and 18 present in cervical cancers. Biochem Biophys Res Commun. 1987;143:837–44.

18. Vousden KH, Doniger J, DiPaolo JA, Lowy DR. The E7 open reading frame of human papillomavirus type 16 encodes a transforming gene. Oncogene Res. 1988;3:167–75.

19. Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I. Human papillomavirus molecular biology and disease association. Rev Med Virol. 2015;25:2–23.

20. Park TW, Fujiwara H, Wright TC. Molecular biology of cervical cancer and its precursors. Cancer. 1995;76:1902–13.

21. Steenbergen RD, Snijders PJ, Heideman DA, Meijer CJ. Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. Nat Rev Cancer. 2014;14:395–405.

22. Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. Clin Sci. 2006;110:525–41.

23. Benevolo M, Vocaturo A, Caraceni D, French D, Rosini S, Zappacosta R, et al. Sensitivity, specificity, and clinical value of human papillomavirus (HPV) E6/E7 mRNA assay as a triage test for cervical cytology and HPV DNA test. J Clin Microbiol. 2011;49:2643–50.

24. Wang HY, Kim H, Park S, Park KH, Lee H. Analytical performance evaluation of the HPV OncoCheck assay for detection of high-risk HPV infection in liquid-based cervical samples. Exp Mol Pathol. 2019;106:149–56.

25. Wu Y, Zhao J, Dong S, Wang Y, Li A, Jiang Y, et al. Whole-exome and RNA sequencing reveals novel insights into the pathogenesis of HPV associated cervical cancer. Cancer Biomark. 2019;25:341–50.

26. Melchers WJ, Bakkers JM, Wang J, de Wilde PC, Boonstra H, Quint WG, et al. Short fragment polymerase chain reaction reverse hybridization line probe assay to detect and genotype a broad spectrum of human papillomavirus types. Clinical evaluation follow. Am J Pathol. 1999;155:1473–8.

27. de Bitter T, van de Water C, van den Heuvel C, Zeelen C, Eijkelenboom A, Tops B, et al. Profiling of the metabolic transcriptome via single molecule molecular inversion probes. Sci Rep. 2017;7:11402.

28. Boyle EA, O'Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics. 2014;30: 2670–2.

29. van den Heuvel C, van Ewijk A, Zeelen C, de Bitter T, Huynen M, Mulders P, et al. Molecular Profiling of Druggable Targets in Clear Cell Renal Cell Carcinoma Through Targeted RNA Sequencing. Front. Oncol. 2019;9:117.

30. Rashmi R, Huang X, Floberg JM, Elhammali AE, McCormick ML, Patti GJ, et al. Radioresistant cervical cancers are sensitive to inhibition of glycolysis and redox metabolism. Cancer Res. 2018;78:1392–403.

31. Arbyn M, Ronco G, Anttila A, Meijer CJ, Poljak M, Ogilvie G, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine. 2012;30:F88–99.

32. Ebisch RM, Siebers AG, Bosgraaf RP, Massuger LF, Bekkers RL, Melchers WJ. Triage of high-risk HPV positive women in cervical cancer screening. Expert Rev Anticancer Ther. 2016;16:1073–85.

33. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med. 2000;132:810–9.

34. Arbyn M, Snijders PJ, Meijer CJ, Berkhof J, Cuschieri K, Kocjan BJ, et al. Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening? Clin Microbiol Infect. 2015;21:817–26.

35. Castle PE, Cuzick J, Stoler MH, Wright TC, Jr., Reid JL, et al. Detection of human papillomavirus 16, 18, and 45 in women with ASC-US cytology and the risk of cervical precancer: results from the CLEAR HPV study. Am J Clin Pathol. 2015;143:160–7.

36. Haedicke J, Iftner T. A review of the clinical performance of the Aptima HPV assay. J Clin Virol. 2016;76(Suppl 1):S40–S8.

37. Luttmer R, Berkhof J, Dijkstra MG, van Kemenade FJ, Snijders PJ, Heideman DA, et al. Comparing triage algorithms using HPV DNA genotyping, HPV E7 mRNA detection and cytology in high-risk HPV DNA-positive women. J Clin Virol. 2015;67:59–66.

38. Origoni M, Cristoforoni P, Carminati G, Stefani C, Costa S, Sandri MT, et al. E6/E7 mRNA testing for human papilloma virus-induced high-grade cervical intraepithelial disease (CIN2/CIN3): a promising perspective. Ecancermedicalscience. 2015;9:533.

39. Giorgi Rossi P, Benevolo M, Vocaturo A, Caraceni D, Ciccocioppo L, Frega A, et al. Prognostic value of HPV E6/E7 mRNA assay in women with negative colposcopy or CIN1 histology result: a follow-up study. PLoS ONE. 2013;8:e57600.

40. Lenting K, Khurshed M, Peeters TH, van den Heuvel C, van Lith SAM, de Bitter T, et al. Isocitrate dehydrogenase 1-mutated human gliomas depend on lactate and glutamate to alleviate metabolic stress. FASEB J. 2019;33:557–71.

41. van den Heuvel C, Das AI, de Bitter T, Simmer F, Wurdinger T, Molina-Vila MA, et al. Quantification and localization of oncogenic receptor tyrosine kinase variant transcripts using molecular inversion probes. Sci Rep. 2018;8:7072.

42. Thierry F, Yaniv M. The BPV1-E2 trans-acting protein can be either an activator or a repressor of the HPV18 regulatory region. EMBO J. 1987;6:3391–7.

43. Leung TW, Liu SS, Leung RC, Chu MM, Cheung AN, Ngan HY. HPV 16 E2 binding sites 1 and 2 become more methylated than E2 binding site 4 during cervical carcinogenesis. J Med Virol. 2015;87:1022–33.

44. Peng G, Dan W, Jun W, Junjun Y, Tong R, Baoli Z, et al. Transcriptome profiling of the cancer and adjacent nontumor tissues from cervical squamous cell carcinoma patients by RNA sequencing. Tumour Biol. 2015;36:3309–17.

45. Chantima W, Thepthai C, Cheunsuchon P, Dharakul T. EpCAM expression in squamous cell carcinoma of the uterine cervix detected by monoclonal antibody to the membrane-proximal part of EpCAM. BMC Cancer. 2017;17:811.

46. Wernicke AG, Kim S, Liu H, Bander NH, Pirog EC. Prostate-specific membrane antigen (PSMA) expression in the neo-vasculature of gynecologic malignancies: implications for PSMA-targeted therapy. Appl Immunohistochem Mol Morphol. 2017;25:271–6.

47. Noel JC, Bucella D, Fayt I, Simonart T, Buxant F, Anaf V, et al. Androgen receptor expression in cervical intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix. Int J Gynecol Pathol. 2008;27:437–41.

48. The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000169083-AR/pathology. Accessed 9 Jan 2019.

49. Ten Broeke SW, van der Klift HM, Tops CMJ, Aretz S, Bernstein I, Buchanan DD, et al. Cancer risks for PMS2-Associated Lynch Syndrome. J Clin Oncol. 2018;36:2961–8.