IJC
International Journal of Cancer

Genome-wide microRNA analysis of HPV-positive self-samples yields novel triage markers for early detection of cervical cancer

Barbara C. Snoek[1], Wina Verlaat[1,†], Iris Babion[1,†], Putri W. Novianti[1,2], Mark A. van de Wiel[2,3], Saskia M. Wilting[1,4], Nienke E. van Trommel[5], Maaike C.G. Bleeker[1], Leon F.A.G. Massuger[6], Willem J.G. Melchers[7], Daoud Sie[1], Daniëlle A.M. Heideman[1], Peter J.F. Snijders[1,‡], Chris J.L.M. Meijer[1] and Renske D.M. Steenbergen[1]

[1]Amsterdam UMC, Vrije Universiteit Amsterdam, Pathology, Cancer Center Amsterdam, Amsterdam, Netherlands
[2]Amsterdam UMC, Vrije Universiteit Amsterdam, Epidemiology and Biostatistics, Amsterdam, Netherlands
[3]Amsterdam UMC, Vrije Universiteit Amsterdam, Mathematics, Amsterdam, Netherlands
[4]Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, Netherlands
[5]Department of Gynecology, Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, Netherlands
[6]Department of Obstetrics and Gynaecology, Radboud University Medical Center, Nijmegen, Netherlands
[7]Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, Netherlands

Offering self-sampling for HPV testing improves the effectiveness of current cervical screening programs by increasing population coverage. Molecular markers directly applicable on self-samples are needed to stratify HPV-positive women at risk of cervical cancer (so-called triage) and to avoid over-referral and overtreatment. Deregulated microRNAs (miRNAs) have been implicated in the development of cervical cancer, and represent potential triage markers. However, it is unknown whether deregulated miRNA expression is reflected in self-samples. Our study is the first to establish genome-wide miRNA profiles in HPV-positive self-samples to identify miRNAs that can predict the presence of CIN3 and cervical cancer in self-samples. Small RNA sequencing (sRNA-Seq) was conducted to determine genome-wide miRNA expression profiles in 74 HPV-positive self-samples of women with and without cervical precancer (CIN3). The optimal miRNA marker panel for CIN3 detection was determined by GRridge, a penalized method on logistic regression. Six miRNAs were validated by qPCR in 191 independent HPV-positive self-samples. Classification of sRNA-Seq data yielded a 9-miRNA marker panel with a combined area under the curve (AUC) of 0.89 for CIN3 detection. Validation by qPCR resulted in a combined AUC of 0.78 for CIN3+ detection. Our study shows that deregulated miRNA expression associated with CIN3 and cervical cancer development can be detected by sRNA-Seq in HPV-positive self-samples. Validation by qPCR indicates that miRNA expression analysis offers a promising novel molecular triage strategy for CIN3 and cervical cancer detection applicable to self-samples.

Key words: microRNA profiling, self-sampling, human papillomavirus, cervical intraepithelial neoplasia

Abbreviations: miRNAs: microRNAs; sRNA-Seq: Small RNA sequencing; AUC: area under the curve; CIN3+: cervical cancer; hrHPV: high-risk HPV; SCC: squamous cell carcinoma; SR50: single read 50; TMM: trimmed mean of M values; GEO: Gene Expression Omnibus; GRridge: group regularized (logistic-) ridge regression; qPCR: quantitative PCR; LOOCV: leave-one-out cross-validation; ROC: receiver operating characteristics; CI: confidence intervals; RT: reverse transcription

Additional Supporting Information may be found in the online version of this article.

[†]Both authors contributed equally

[‡]In memory of

Grant sponsor: European Research Council; **Grant numbers:** ERC-AdG-2012_322986_MASSCARE; **Grant sponsor:** European Research Council; **Grant numbers:** ERC-PoC_2015_713303_MIMICS; **Grant sponsor:** Dutch Cancer Society; **Grant numbers:** KWF-10395; **Grant sponsor:** Dutch Cancer Society; **Grant numbers:** KWF-5708

DOI: 10.1002/ijc.31855

History: Received 17 May 2018; Accepted 30 Jul 2018; Online 7 Sep 2018

Correspondence to: Renske D.M. Steenbergen, PhD, Amsterdam UMC, Vrije Universiteit Amsterdam, Pathology, Cancer Center Amsterdam, PO Box 7057, 1007 MB Amsterdam; Netherlands; E-mail: r.steenbergen@vumc.nl; Tel.: 00-31-204442331; Fax: 00-31-204442964

What's new?

MicroRNAs (miRNAs) are suspected of playing a role in cervical cancer development. They are also potential markers for the identification of human papillomavirus (HPV)-infected women who are at risk of cervical cancer. Here, using small RNA sequencing of HPV-positive self-samples from women with and without cervical precancer (CIN3), the authors identify a miRNA signature consisting of multiple miRNAs that is strongly predictive of CIN3. Validation of this signature by qPCR revealed a good clinical performance for CIN3+ detection. The findings suggest that miRNA analysis is an effective means of CIN3+ prediction in HPV-positive self-samples obtained for cervical cancer screening.

Introduction

In many countries, the participation of women in population-based cervical screening is suboptimal which hampers the effectiveness of the screening program.[1] Women who do not participate (30% of invited women in The Netherlands), so-called nonattendees, are at increased risk of developing cervical intraepithelial neoplasia grade 3 or worse (CIN3+).[2] Offering self-sampling for high-risk HPV (hrHPV) testing to nonattendees has been shown to increase the efficacy of the screening program[3] and has demonstrated to be equally effective in detecting CIN3+ when compared to HPV testing on physician-taken specimens.[4,5] Therefore, self-sampling for HPV testing is now offered to nonattendees in The Netherlands. To discriminate between women with a transient (clinically irrelevant) HPV infection and women with clinically meaningful cervical disease, additional risk assessment (so-called triage) is necessary. For cervical scrapes, testing of HPV-positive women by (repeat) cytology is a generally acknowledged triage strategy.[6,7] Since cytological examination of self-samples is not reliable,[8] women testing positive for hrHPV on their self-sample need to have a physician-taken smear for triage testing by cytology. This prerequisite for cervical cytology for previous nonattendees, has shown to be associated with loss to follow-up.[9] Therefore, the identification and subsequent use of nonmorphological, molecular triage markers that can be directly applied on self-samples becomes very important.

Molecular triage methods that have been described so far for application on self-samples include HPV16/18 genotyping[10,11] and DNA methylation marker analysis.[9,10,12,13] MicroRNAs (miRNAs) have emerged as key regulators of gene expression and their involvement in carcinogenesis has now been widely accepted.[14] They are increasingly recognized as promising biomarkers for cancer diagnostics and may provide alternative molecular triage markers. Interestingly, Tian *et al.* reported that miRNA detection in HPV-positive cervical scrapes has a superior performance over cytology for the detection of high-grade CIN lesions in a referral population,[15] which indicates the potential of miRNA detection for triage testing of HPV-positive women in cervical screening programs.

A number of studies have demonstrated altered miRNA expression patterns in cervical cancer cell lines and cervical cancer tissues.[16] However, little is known about genome-wide miRNA expression patterns in CIN lesions[17,18] and no data is presently available on miRNA expression profiles in HPV-positive self-samples. Importantly, previous research has shown that the clinical performance of molecular markers may be dependent on the sample type used,[12] given that the cellular composition may differ between tissues, physician-taken cervical scrapes and self-samples and the fact that nondisease-related cells such as vaginal cells are largely overrepresented in self-samples. Accordingly, the optimum miRNAs listed so far may not be directly extrapolated to self-samples.

Our study set out to directly determine the genome-wide miRNA profiles in self-samples aiming to establish a miRNA signature for CIN3 detection to serve as a novel triage method for HPV-positive self-samples. To this end, we performed small RNA next-generation sequencing (sRNA-Seq) on HPV-positive self-samples from women with CIN3 lesions and women without cervical disease in follow-up (≤CIN1), obtained from a prospective screening trial among nonattendees.[3] Penalized logistic regression analysis resulted in the identification of a panel of 9 miRNAs that together could detect CIN3 with high discriminatory power. The performance of this miRNA signature was subsequently validated by qPCR in an independent sample series of HPV-positive self-samples.

Materials and Methods

Clinical cervical specimens and RNA extraction

For the discovery set, we used 74 hrHPV-positive cervicovaginal self-samples that were collected from nonattendees participating in the PROHTECT 1 trial (NTR792)[3] with the first generation Delphi-screener (Delphi-Bioscience, The Netherlands) between December 2006 and December 2007. Detailed characteristics of study design and clinical specimens, inclusion criteria and follow-up procedures have been described previously.[3] miRNA sequencing data from a pilot experiment of 12 self-samples for power calculations revealed a ratio of 1 (hrHPV-positive controls) to 1 (CIN3) for proper marker discovery. Therefore, the discovery set consisted of 36 control women with either histologically confirmed ≤CIN1 or that displayed hrHPV clearance combined with normal cytology in follow-up (hereafter referred to as controls; median age of 37; range 31–56), and 38 women (cases) who were histologically diagnosed with a CIN3 lesion (median age of 39; range 31–62). CIN2 lesions were not included given their ambiguous classification; they often represent a misclassified CIN1 or CIN3.[19] Controls and cases were matched according to age and HPV type. Total RNA was isolated using Direct-Zol™

RNA Miniprep (Zymo Research, Freiburg, Germany) after the manufacturer's instructions.

For the validation set, we used 155 hrHPV-positive lavage self-samples that were collected from nonattendees participating in the methylation triage arm of the PROHTECT 3A trial (NTR2606)[9] with the second generation Delphi-screener (Delphi-Bioscience, The Netherlands) between November 2010 and December 2011. Detailed characteristics of study design and clinical specimens, inclusion criteria and follow-up procedures have been described previously.[9] These 155 hrHPV-positive self-samples consisted of 101 control women with either histologically confirmed ≤CIN1 or that displayed hrHPV clearance combined with normal cytology in follow-up (hereafter referred to as controls; median age of 42; range 33–63), 49 women who were histologically diagnosed with a CIN3 lesion (median age of 40; range 33–58) and five women who were histologically diagnosed with squamous cell carcinoma (SCC) (median age of 47; range 38–58). Furthermore, 36 hrHPV-positive brush-based self-samples (median age of 46; range 27–83) were included that were obtained from women who visited the Antoni van Leeuwenhoek Hospital/Netherlands Cancer Institute or Academic Medical Center Amsterdam for treatment of SCC between 2015 and 2017 (X15MET and SOLUTION study). Total RNA was isolated using Trizol reagent (Thermo Fisher Scientific, Bleiswijk, The Netherlands) according to the manufacturer's instructions.

Samples from both PROHTECT trials were tested for HPV by GP5+/6+ PCR using the Diassay EIA HPV GP HR kit (Diassay, Voorburg, The Netherlands) as described previously.[9,20] Samples from the XMET and SOLUTION studies were tested for HPV using the HPV-Risk Assay (Self-screen BV, Amsterdam, The Netherlands).[21]

Ethical approval was obtained from the National Health Council (PROHTECT trials) and Institutional Review Board of the Antoni van Leeuwenhoek Hospital/Netherlands Cancer Institute (METC15.1468) and VU University Medical Center (METC2016.213). All participants gave informed consent.

All samples were used in an anonymous fashion in accordance with the "Code for Proper Secondary Use of Human Tissues in the Netherlands" as formulated by the Dutch Federation of Medical Scientific Organizations (http://www.fmwv.nl and www.federa.org).[22]

Library construction and deep sequencing

Preparation of cDNA libraries was performed with equal amount of input RNA per sample (1,000 ng of total RNA) using the TruSeq Small RNA Sample Preparation Kit according to the manufacturer's instruction (Illumina, San Diego, CA). In brief, cDNA libraries were measured on an Agilent 2100 Bioanalyzer (Agilent Technologies) and up to five samples were pooled in equimolar concentrations for the sequencing run. Sequencing was performed single read 50 (SR50) cycles on a HiSeq 2500 (Illumina) for 74 samples (36 controls, 38 CIN3).

Data processing miRNA sequencing

In order to obtain miRNA expression profiles, all raw small RNA sequencing data was analyzed using the miRDeep2 algorithm.[23,24] In brief, read counts were filtered for low-quality sequences, adapters and sequences shorter than 18 nucleotides followed by alignment to the human genome (hg19). Input from known miRNAs was obtained from miRBase version 20.[25] To avoid multiple counting of miRNA sequences that can originate from several genomic loci, we only counted the expression of miRNA paralogs once. As part of our quality control, we assessed the distribution of miRNA read counts and excluded those samples that depicted an atypical distribution (Supporting Information Fig. S1). Based on this criterion, 10 samples with insufficient quality were excluded. Another eight samples were already excluded on forehand because of low sequencing depth.

After quality control, miRNA read counts were normalized according to the trimmed mean of M values (TMM) normalization strategy.[26] Subsequently, normalized miRNA read counts were square root transformed to the quasi Gaussian scale followed by miRNA data standardization. Finally, from 2,577 known miRNAs according to miRBase 20, only those miRNAs that exhibited nonzero read counts in at least three samples were included. This resulted in a total set of 772 miRNAs with an average of 416 miRNAs per sample (controls: 403, CIN3: 434). Raw sequencing reads and quantified read counts are available from the NCBI Gene Expression Omnibus (GEO) through series accession number GSE104758.

Statistical analysis miRNA sequencing and microarray data

Adaptive group regularized (logistic-) ridge regression (GRridge)[27] was applied to the preprocessed miRNA sequencing data (expression values). Auxiliary information, namely abundance and conservation status,[28] was incorporated to build the omics-based prediction model. *Post hoc* feature selection was subsequently applied to the GRridge model, which resulted in the identification of nine miRNAs to be further validated by quantitative PCR (qPCR). More details of the GRridge predictive modeling as well as the results (including data preprocessing and predictive modeling) have been described elsewhere.[29] The performance of the GRridge prediction model on the miRNA expression data was assessed by leave-one-out cross-validation (LOOCV), then visualized by the receiver operating characteristics (ROC) curve and quantified by area under the curve (AUC).

In-house available genome-wide microarray miRNA profiles of cervical tissues,[30] including normal cervical tissues, CIN2/3 lesions (hereafter referred to as high-grade CIN) and SCC were utilized to evaluate the expression patterns of the selected miRNAs. Due to limitations in accurate histologic grading of high-grade CIN on frozen tissue specimens, CIN2 and CIN3 were grouped together in this sample series.[30] Since the miRNA microarray did not include probes for miR-184, the expression of miR-184 in cervical tissues could not be evaluated. For the selected miRNAs, we employed the

Kruskall-Wallis omnibus test and the post-hoc nonparametric test to compare miRNA expression between biological groups. Obtained P values were adjusted for multiple comparisons. Statistical analyses were performed in R open source software by implementing the GRridge[27] package.

Quantitative RT-PCR (qPCR)

Expression of hsa-let-7b-5p, hsa-miR-9-5p, hsa-miR-15b-5p, hsa-miR-20a-5p, hsa-miR-31-5p, hsa-miR-93-5p, hsa-miR-183-5p, hsa-miR-184, and hsa-miR-222–3p was determined using TaqMan microRNA assays (002619, 000583, 000390, 000580, 002279, 001090, 002269, 000485, 002276; Thermo Fisher Scientific). Since we most recently described hsa-miR-423-3p and hsa-miR-30b-5p as suitable normalization strategy for hrHPV-positive self-samples,[31] these miRNAs were included as reference genes (002626, 000602; Thermo Fisher Scientific).

The manufacturer's protocol was adapted in order to multiplex reverse transcription of all 9 miRNA markers and 2 reference genes (11 miRNAs in total) and validated in comparison to singleplex reverse transcription reactions. In short, the specific reverse transcription primers were combined in a primer pool and cDNA was synthesized in 16.5 μL reactions containing 9 μL primer pool, 0.45 μL dNTPs (100 mM), 2.25 μL 10× RT buffer, 0.3 μL RNase inhibitor (20 U/μL), and 4.5 μL MultiScribe Reverse Trancriptase (TaqMan microRNA Reverse Transcription kit, Thermo Fisher Scientific). We used 50 ng total RNA as reverse transcription template.

Quantitative PCR reactions were performed in 10 μL, consisting of 5 μL TaqMan® Universal Master Mix II no UNG, 0.5 μL miRNA specific TaqMan assays (Thermo Fisher Scientific), 3.5 μL H_2O and 1 μL cDNA. The ABI 7500 Fast Real-Time PCR System (Thermo Fisher Scientific) was used for quantification. Cycle conditions used for cDNA synthesis and PCR were according to the manufacturer's protocols.

Samples that exhibited Cq values >32 for the geomean of the reference genes were considered unsuitable for miRNA qPCR analysis. Based on this criterion, one CIN3 self-sample was excluded. For the remaining self-samples, miRNA expression was normalized to the geometric mean of the reference genes applying the $2^{-\Delta Ct}$ method.[32]

Detailed characteristics of physician-taken cervical scrapes (Supporting Information Fig. S3), including miRNA expression analysis by qPCR have been described previously.[33]

Statistical analysis miRNA qPCR data

Since our panel of miRNA markers should be able to detect cervical precancerous lesions with increased risk to progress to cervical cancer as well as cervical cancer with high discriminatory power, we combined CIN3 and cervical cancer samples (CIN3+) for the analysis of the validation set. Logistic regression analysis was performed to assess the prediction ability of the six miRNAs on the normalized qPCR data for detection of CIN3+. First, the Cq ratio of each miRNA was transformed to the square root scale. Next, univariable logistic regression model was fitted to observe the performance of each miRNA. Multivariable logistic regression with backward elimination was subsequently applied to define the best combination of miRNAs to differentiate self-samples from women with CIN3 + from hrHPV-positive controls. Supporting Information Table S1 shows the contribution (standardized coefficient) of each miRNA in the most optimal prediction model. The performance of the most optimal prediction model on the transformed miRNA expression data was visualized by ROC curve and evaluated by AUC, sensitivity and specificity. 95% confidence intervals (CI) were calculated for the AUC obtained for the most optimal prediction model.[34] Additionally, we employed the Kruskall-Wallis omnibus test and the post-hoc nonparametric test to compare miRNA expression between biological groups. Obtained *P* values were adjusted for multiple comparisons. All statistical analyses were performed in R open source software using the pROC[35] package.

Results

Discovery of a miRNA signature in HPV-positive self-samples

In order to identify a CIN3-specific miRNA signature in self-samples, we performed genome-wide small RNA sequencing on 74 HPV-positive self-samples (discovery set). This yielded an average of 41 million raw reads per sample (controls: 48 million, CIN3: 37 million). After preprocessing (for details please refer to Materials and Methods), we obtained an average of 25 million reads per sample (controls: 29 million, CIN3: 23 million) of which approximately 24% was mapped to the human genome (controls: 20%, CIN3: 24%). Most non-human reads mapped to *Lactobacillus*. A fraction of the mapped reads represented miRNAs, with an average of 4% per sample (controls: 5%, CIN3: 3%). Based on quality control assessment (for details please refer to Materials and Methods), 18 samples were excluded resulting in a total of 56 samples (32 controls, 24 CIN3) for further analysis.

To identify a miRNA signature with high discriminatory power for CIN3, we performed adaptive group regularized ridge regression, GRridge,[27] which enables objective use of codata. This resulted in the identification of a panel of 9 miRNAs (let-7b, miR-9, miR-15b, miR-20a, miR-31, miR-93, miR-183, miR-184, and miR-222) as strong predictor of CIN3 with an AUC of 0.89 (Fig. 1*a*). The expression of all miRNAs was found to be increased in CIN3 compared to HPV-positive controls (Fig. 1*b*).

In order to ensure that the identified miRNA markers in self-samples are related to cancer development, we next analyzed in-house available genome-wide miRNA profiles of a set of 37 cervical tissues (10 controls, 18 high-grade CIN, 9 SCC).[30] In concordance with the expression patterns observed in self-samples, we observed increased expression for miR-9, miR-15b, miR-20a, miR-93 and miR-183 in cervical tissues from women with high-grade CIN compared to controls (Fig. 2). For let-7b, miR-31 and miR-222, however, we found similar or decreased expression levels in high-grade CIN cervical tissues compared to

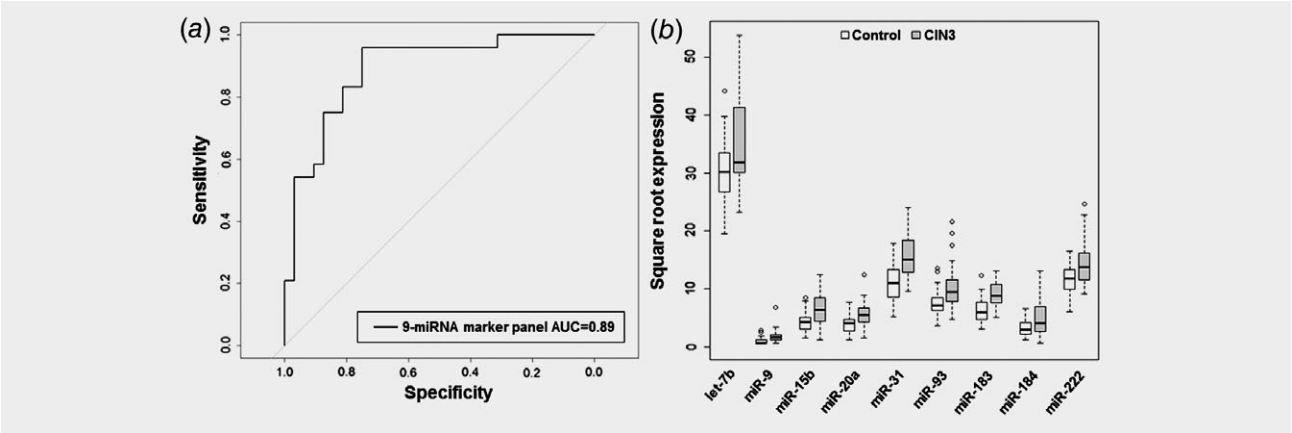

Figure 1. Identification of a CIN3-specific miRNA signature in HPV-positive self-samples (*a*) Receiver operating characteristics (ROC) curve and area under the curve (AUC) of a 9-miRNA marker panel for CIN3 detection in HPV-positive self-samples (discovery set; 32 controls and 24 CIN3). (*b*) Square root expression levels of the 9 miRNAs in HPV-positive self-samples (discovery set; 32 controls (open box) and 24 CIN3 (light gray box)).

controls, although these differences were not significant (Fig. 2). Importantly, all miRNAs except let-7b showed increased expression in SCC compared to controls.

Validation of the miRNA signature by qPCR in HPV-positive self-samples

To validate the clinical performance of the 9 miRNAs identified in the discovery phase, we performed miRNA qPCR in an independent set of 191 HPV-positive self-samples (validation set). Among the 9 miRNAs, 3 miRNAs (miR-9, miR-183 and miR-184) exhibited very low expression level, detected at an average Cq of >32. Since the quantification of low expressed transcripts can reduce reproducibility and substantially increase technical PCR noise,[31] we excluded these 3 miRNAs from subsequent analysis.

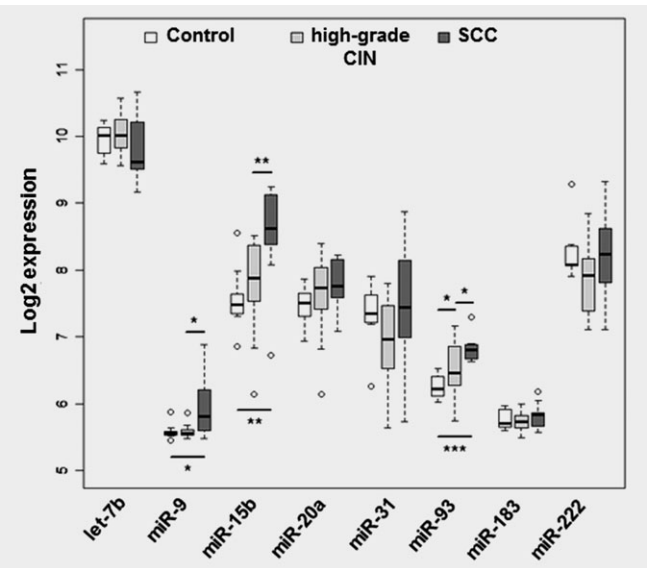

Figure 2. Expression levels of the miRNA signature in cervical tissues. Log2 expression levels obtained from miRNA microarray data of 8 out of 9 miRNAs in cervical tissues (10 controls (open box), 18 high-grade CIN (light gray box), 9 SCC (dark gray box)). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In concordance with the results presented for the discovery set, we found a significant difference for miR-15b and miR-93 between self-samples from women with CIN3 and HPV-positive controls (Fig. 3*b*). In contrast, for let-7b we observed significantly lower expression levels in CIN3 self-samples compared to the HPV-positive controls, and even lower levels in SCC (Fig. 3*b*). Interestingly, however, the results for let-7b are in line with the data from cervical tissues for which we also observed decreased expression of let-7b in SCC (Fig. 2). For miR-31 no significant difference between the biological groups was observed (Fig. 3*b*). Furthermore, all miRNAs except miR-31 and let-7b showed significantly increased expression levels in self-samples from women with SCC compared to HPV-positive controls (Fig. 3*b*).

Next, we performed univariable logistic regression to assess the performance of each miRNA and multivariable logistic regression followed by backward elimination to identify the best combination of miRNA markers. Since we set out to find a panel of miRNA markers that can be applied in cervical screening that should be able to detect both CIN3 and cervical cancer with high discriminatory power, we combined CIN3 and cervical cancer (CIN3+) for the analysis of the validation set. The best clinical performance was achieved by multivariable logistic regression with a panel of 5 miRNAs (let-7b, miR-15b, miR-20a, miR-93, and miR-222) for which we obtained an AUC of 0.78 (95% confidence interval (CI): 0.7173–0.8479) for CIN3+ detection (Fig. 3*a*). Expectedly we observed a better clinical performance for the 5-miRNA marker panel compared to the individual clinical performance of the miRNAs (Supporting Information Fig. S2). We included thresholds corresponding to 65–75% specificity to determine the sensitivity of the 5-miRNA marker panel for CIN3 and SCC detection (Table 1). At the threshold corresponding to 65% specificity in HPV-positive controls, the

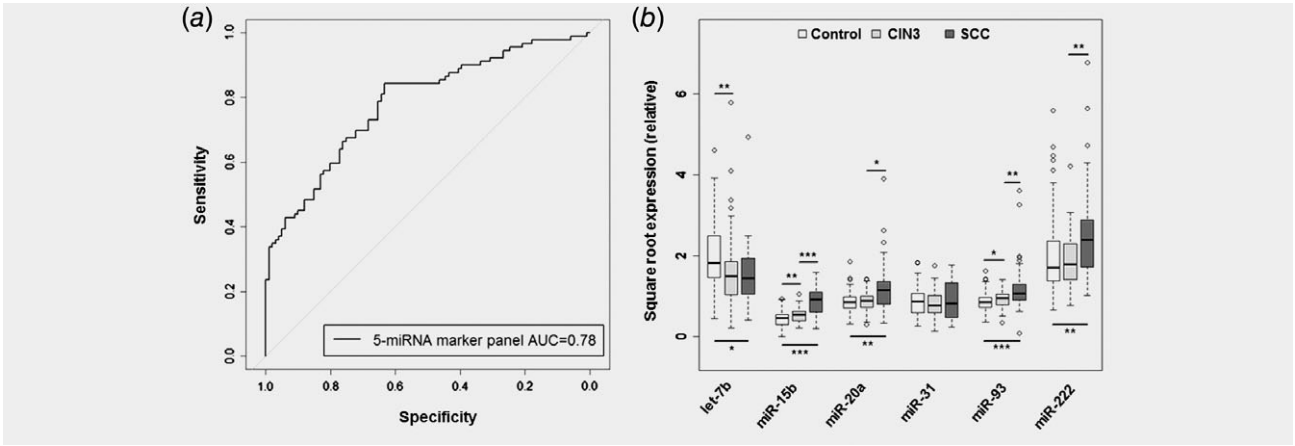

Figure 3. Clinical performance and expression levels of the miRNA markers in HPV-positive self-samples. (*a*) ROC curve and AUC of a 5-miRNA marker panel (let-7b, miR-15b, miR-20a, miR-93, and miR-222 in HPV-positive self-samples for CIN3+ detection (validation set; 101 controls, 48 CIN3, 41 SCC). (*b*) Square root expression levels (relative to miR-423 and miR-30b) of the 6 miRNAs in HPV-positive self-samples (validation set; 101 controls (open box), 48 CIN3 (light gray box), 41 SCC (dark gray box)). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5-miRNA marker panel revealed a sensitivity of 67% (32 out of 48) for CIN3 detection. Importantly, 38 out of 41 (93%) of the cervical cancers were detected (Table 1).

Discussion

This is the first study to determine genome-wide miRNA profiles in HPV-positive self-samples to allow for direct and objective triage in cervical screening. Since HPV-positive self-samples are relatively impure due to the large amount of non-cervical (i.e., vaginal cells, lymphocytes) and normal cervical epithelial cells, we applied our recently proposed GRridge model[27] to identify a miRNA signature with high discriminatory power for CIN3 detection. This method enables objective use of auxiliary data such as miRNA abundance and conservation and was previously shown to outperform other prediction methods.[29] We identified a miRNA signature consisting of 9 miRNAs (let-7b, miR-9, miR-15b, miR-20a, miR-31, miR-93, miR-183, miR-184, miR-222) with high discriminatory power for CIN3 detection in HPV-positive self-samples (AUC = 0.89).

Importantly, our study shows the feasibility of conducting sRNA sequencing for miRNA quantification on self-samples. It should be noted, however, that prior to the analysis we had to exclude a substantial amount of samples of insufficient quality (18 out of 74). This quality filtering was important to improve the miRNA selection and any potential bias introduced is countered in the validation phase where independent HPV-positive self-samples were used without any filtering. Moreover, we found that differential expression of the identified miRNAs could be confirmed in independent cervical tissue specimens analyzed by a different method.[30] In addition, recent data show increased expression of miR-15b in HPV-positive physician-taken cervical scrapes from women with CIN3 and SCC.[33] Similarly, preliminary miRNA expression data show consistent results for miR-20a, miR-93 and miR-222 in HPV-positive cervical scrapes (Supporting Information Fig. S3). Taken together, this shows that the detection of the aforementioned miRNAs in self-samples reflects cervical disease and supports the validity of our approach.

Consistent with our findings, Li and colleagues found increased expression of miR-15b, miR-20a, miR-31, miR-93, and miR-222 in cervical precancerous tissues.[36] These miRNAs have also been shown to be upregulated in cervical cancer tissues compared to controls.[16,17,37] Moreover, upregulation of miR-15b was found to be directly linked to a chromosomal gain of 3q which is often observed in cervical cancers.[30] Additionally, miR-15b expression was recently shown to gradually increase from CIN1 to CIN2-3 and cervical cancer tissues and to be associated with poor prognosis.[37]

Validation by qPCR in an independent set of HPV-positive self-samples resulted in a good clinical performance for a 5-miRNA signature consisting of let-7b, miR-15b, miR-20a, miR-93, and miR-222 (AUC = 0.78) with a sensitivity of 67% and a specificity of 65% for CIN3 detection. Importantly, 93% (38 out of 41) of the cervical cancers were detected. Since two of the three miRNA negative cervical cancer self-samples were just below the threshold this may indicate that these miRNA negative self-samples do not contain sufficient numbers of representative cells or well-sampled material and warrant additional testing. Compared to previously described molecular triage markers in HPV-positive self-samples,[9,12,38–40] we obtained a nearly comparable clinical performance for CIN3+ detection as observed for DNA methylation panels.[9,38] Unlike DNA methylation analysis,

Table 1. Clinical performance of the 5-miRNA marker panel in HPV-positive self-samples

Specificity (%)	Sensitivity CIN3 (%)	Sensitivity SCC (%)
75	50	88
70	54	88
65	67	93

miRNA quantification by qPCR does not require bisulfite conversion prior to PCR amplification and low amounts of starting material are sufficient for accurate and reproducible analysis. Unfortunately, most miRNA quantification methods only offer assays for each miRNA individually. For future (semi)high-throughput full molecular screening a combined measurement of several miRNAs into one single test would be desirable. Although we included a multiplex reverse transcription (RT) step that allows for simultaneous transcription of all our miRNAs including reference genes, further technical advancements are warranted, thereby saving self-sampled material, time and costs.

One limitation of our study that should be considered in interpreting our findings is that we were unable to validate all 9 miRNAs by qPCR due to low expression levels of 3 miRNAs (miR-9, miR-183, miR-184). This could have affected the performance of the miRNA signature in the validation set. Also, we applied a different method for validation compared to the discovery (qPCR *versus* sRNA-Seq). Although several studies have shown close correlations between qPCR and RNA-Seq data,[41] differences in miRNA quantification between the methods have been reported.[42,43] A possible explanation for discrepancies between the methods could be the existence of miRNA variants. With the rapid development of sRNA-Seq came the realization that numerous miRNAs exist as multiple length variants, which are termed isomiRs.[44] The majority of length variants are located at the 3′end of the miRNA sequence[44] which could potentially hamper their identification by qPCR.[45] This might explain our discordant results for let-7b between the discovery and validation set. We found significantly increased expression of let-7b in CIN3 compared to controls in the discovery set, while its expression was found downregulated in CIN3 and SCC in the validation set. Interestingly, reduced expression of let-7b in SCC was also observed in cervical tissue specimens, for which we have miRNA microarray data.[30] Since miRNA probes used in microarrays are designed to detect the canonical miRNA sequence as listed in miRBase,[46] and do not bind to all isomiRs as detected by sRNA-Seq, these microarray results are likely to better reflect the results obtained with qPCR. The mapping strategy that we applied for the sRNA-Seq data did not allow for the analysis of isomiRs. Therefore, extensive analysis of our sRNA-Seq data is needed in order to elucidate the presence and marker value of isomiRs in self-samples.

Despite the description of various detection methods for efficient and accurate quantification of miRNAs,[47] only some of these approaches have been tested for their ability to distinguish between highly similar miRNA sequences. Although these methods possess high sensitivity for specific sequences, they are not adequately specific and closely related isomiRs are also detected.[48] Therefore, further advancements in isomiR detection methods are needed in order to accurately quantify individual isomiRs.

Furthermore, the use of HPV-positive self-sample series from a population of nonattending women can be considered as a limitation. Therefore, future confirmation in a regular population-based sample series is warranted. Also, further research on the miRNA signature in HPV-positive clinical samples, as well as a comparison to established triage methods including cytology and HPV16/18 genotyping is needed to confirm the value of the miRNA signature for triage.

In conclusion, by genome-wide miRNA profiling on HPV-positive self-samples, we identified a CIN3-specific miRNA signature with good performance. Subsequent validation by qPCR on an independent set of HPV-positive self-samples showed a good clinical performance of this miRNA signature for CIN3+ detection. Our findings indicate that miRNA analysis on HPV-positive self-samples can greatly improve the management of HPV-positive women and facilitates the implementation of self-sampling in cervical cancer screening programs.

Acknowledgments

The authors thank Paul Eijk for excellent technical assistance.

D. Heideman, P. Snijders, C. Meijer and R. Steenbergen have a minority stake in Self-screen B.V., a spin-off company of Amsterdam UMC (formerly known as VU University Medical Center Amsterdam), which owns patents related to this work. D. Heideman has been on the speaker's bureau of Qiagen, and serves occasionally on the scientific advisory board of Pfizer and Bristol-Meyer Squibb. P. Snijders has been on the speaker's bureau of Roche, Abbott, Gen-Probe, Qiagen, and Seegene, and was a consultant for Crucell Holland B.V. C. Meijer has participated in the sponsored speaker's bureau of SPMSD/Merck, and served occasionally on the scientific advisory board of Qiagen and SPMSD/Merck. He was coinvestigator on a SPMSD sponsored HPV vaccination trial of which his institute received research funding, and was a minority shareholder of Diassay B.V till March 2016. He owns a very small number of Qiagen shares and is part time director of Self-Screen. All other authors declare that they have no competing interests.

References

1. Arbyn M, Rebolj M, De Kok IMCM, et al. The challenges of organising cervical screening programmes in the 15 old member states of the European Union. *Eur J Cancer* 2009;45:2671–8. https://doi.org/10.1016/j.ejca.2009.07.016.
2. Rozemeijer K, de Kok IMCM, Naber SK, et al. Offering self-sampling to non-attendees of organized primary HPV screening: when do harms outweigh the benefits? *Cancer Epidemiol Biomarkers Prev* 2015;24:773–82. https://doi.org/10.1158/1055-9965.EPI-14-0998.
3. Gök M, Heideman DAM, van Kemenade FJ, et al. HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. *BMJ* 2010;340:c1040. http://www.ncbi.nlm.nih.gov/pubmed/20223872. Accessed October 13, 2016.
4. Arbyn M, Verdoodt F, Snijders PJF, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. *Lancet Oncol* 2014;15:172–83. https://doi.org/10.1016/S1470-2045(13)70570-9.
5. Arbyn M, Castle PE. Offering self-sampling kits for HPV testing to reach women who do not attend in the regular cervical cancer screening program. *Cancer Epidemiol Biomarkers Prev* 2015;24:769–72. https://doi.org/10.1158/1055-9965.EPI-14-1417.
6. Rijkaart DC, Berkhof J, van Kemenade FJ, et al. Evaluation of 14 triage strategies for HPV DNA-positive women in population-based cervical screening. *Int J Cancer* 2012;130:602–10. https://doi.org/10.1002/ijc.26056.

7. Dijkstra MG, van Niekerk D, Rijkaart DC, et al. Primary hrHPV DNA testing in cervical cancer screening: how to manage screen-positive women? A POBASCAM trial substudy. *Cancer Epidemiol Biomarkers Prev* 2014;23:55–63. https://doi.org/10.1158/1055-9965.EPI-13-0173.
8. Yoshida T, Sano T, Takada N, et al. Comparison of self-collected and clinician-collected materials for cervical cytology and human papillomavirus genotyping: analysis by linear array assay. *Acta Cytol* 2011;55:106–12. https://doi.org/10.1159/000320924.
9. Verhoef VMJ, Bosgraaf RP, Van Kemenade FJ, et al. Triage by methylation-marker testing versus cytology in women who test HPV-positive on self-collected cervicovaginal specimens (PROHTECT-3): a randomised controlled non-inferiority trial. *Lancet Oncol* 2014;15:315–22. https://doi.org/10.1016/S1470-2045(14)70019-1.
10. Verhoef VMJ, Heideman DAM, van Kemenade FJ, et al. Methylation marker analysis and HPV16/18 genotyping in high-risk HPV positive self-sampled specimens to identify women with high grade CIN or cervical cancer. *Gynecol Oncol* 2014;135:58–63. https://doi.org/10.1016/j.ygyno.2014.08.003.
11. Ebisch RMF, de Kuyper-de Ridder GM, Bosgraaf RP, et al. The clinical value of HPV genotyping in triage of women with high-risk-HPV-positive self-samples. *Int J Cancer* 2016;139:691–9. https://doi.org/10.1002/ijc.30090.
12. Hesselink AT, Heideman DAM, Steenbergen RDM, et al. Methylation marker analysis of self-sampled cervico-vaginal lavage specimens to triage high-risk HPV-positive women for colposcopy. *Int J Cancer* 2014;135:880–6. https://doi.org/10.1002/ijc.28723.
13. Verlaat W, Snoek BC, Heideman DAM, et al. Identification and validation of a 3-gene methylation classifier for HPV-based cervical screening on self-samples. *Clin Cancer Res*. April 2018;24:3456–3464. doi:https://doi.org/10.1158/1078-0432.CCR-17-3615.
14. MacFarlane L-A, Murphy PR. MicroRNA: biogenesis, function and role in cancer. *Curr Genomics* 2010;11:537–61. https://doi.org/10.2174/138920210793175895.
15. Tian Q, Li Y, Wang F, et al. MicroRNA detection in cervical exfoliated cells as a triage for human papillomavirus-positive women. *J Natl Cancer Inst* 2014;106:9. https://doi.org/10.1093/jnci/dju241.
16. Satapathy S, Batra J, Jeet V, et al. MicroRNAs in HPV associated cancers: small players with big consequences. *Expert Rev Mol Diagn* 2017;17:711–22. https://doi.org/10.1080/14737159.2017.1339603.
17. Wilting SM, Verlaat W, Jaspers A, et al. Methylation-mediated transcriptional repression of microRNAs during cervical carcinogenesis. *Epigenetics* 2013;8:220–8. https://doi.org/10.4161/epi.23605.
18. Mo W, Tong C, Zhang Y, et al. microRNAs' differential regulations mediate the progress of human papillomavirus (HPV)-induced cervical intraepithelial neoplasia (CIN). *BMC Syst Biol* 2015;9:4. https://doi.org/10.1186/s12918-015-0145-3.
19. Herrington CS. The terminology of pre-invasive cervical lesions in the UKcervical screening programme. *Cytopathology* 2015;26:346–50. https://doi.org/10.1111/cyt.12307.
20. Bosgraaf RP, Verhoef VMJ, Massuger LFAG, et al. Comparative performance of novel self-sampling methods in detecting high-risk human papillomavirus in 30,130 women not attending cervical screening. *Int J Cancer*. 2014;136:646–55. doi:https://doi.org/10.1002/ijc.29026.
21. Hesselink AT, Berkhof J, van der Salm ML, et al. Clinical validation of the HPV-risk assay, a novel real-time PCR assay for detection of high-risk human papillomavirus DNA by targeting the E7 region. *J Clin Microbiol* 2014;52:890–6. https://doi.org/10.1128/JCM.03195-13.
22. Human Tissue and Medical Research: Code of conduct for responsible use (2011).
23. Friedländer MR, Chen W, Adamidi C, et al. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* 2008;26:407–15. https://doi.org/10.1038/nbt1394.
24. Friedländer MR, Mackowiak SD, Li N, et al. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* 2012;40:37–52. https://doi.org/10.1093/nar/gkr688.
25. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011;39(Database issue):D152–7. https://doi.org/10.1093/nar/gkq1027.
26. Robinson MD, Oshlack A, Wang E, et al. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;11:R25. https://doi.org/10.1186/gb-2010-11-3-r25.
27. van de Wiel MA, Lien TG, Verlaat W, et al. Better prediction by use of co-data: adaptive group-regularized ridge regression. *Stat Med* 2016;35:368–81. https://doi.org/10.1002/sim.6732.
28. Agarwal V, Bell GW, Nam J-W, et al. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015;4:101–12. https://doi.org/10.7554/eLife.05005.
29. Novianti PW, Snoek BC, Wilting SM, et al. Better diagnostic signatures from RNAseq data through use of auxiliary co-data. *Bioinformatics* 2017;33:btw837. https://doi.org/10.1093/bioinformatics/btw837.
30. Wilting SM, Snijders PJF, Verlaat W, et al. Altered microRNA expression associated with chromosomal changes contributes to cervical carcinogenesis. *Oncogene* 2013;32:106–16. https://doi.org/10.1038/onc.2012.20.
31. Babion I, Snoek BC, van de Wiel MA, et al. A strategy to find suitable reference genes for MiRNA quantitative PCR analysis and its application to cervical specimens. *J Mol Diagn* 2017;19:625–737. https://doi.org/10.1016/j.jmoldx.2017.04.010.
32. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. *Methods* 2001;25:402–8. https://doi.org/10.1006/meth.2001.1262.
33. Babion I, Snoek BC, Novianti PW, et al. Triage of high-risk HPV-positive women in population-based screening by miRNA expression analysis in cervical scrapes; a feasibility study. *Clin Epigenetics* 2018;10:76. https://doi.org/10.1186/s13148-018-0509-9.
34. Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845. http://www.ncbi.nlm.nih.gov/pubmed/3203132. Accessed March 7, 2018.
35. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77. https://doi.org/10.1186/1471-2105-12-77.
36. Li Y, Wang F, Xu J, et al. Progressive miRNA expression profiles in cervical carcinogenesis and identification of HPV-related target genes for miR-29. *J Pathol* 2011;224:484–95. https://doi.org/10.1002/path.2873.
37. Wen F, Xu J-Z, Wang X-R. Increased expression of miR-15b is associated with clinicopathological features and poor prognosis in cervical carcinoma. *Arch Gynecol Obstet* 2017;295:743–9. https://doi.org/10.1007/s00404-016-4286-4.
38. De Strooper LMA, Verhoef VMJ, Berkhof J, et al. Validation of the FAM19A4/mir124-2 DNA methylation test for both lavage- and brush-based self-samples to detect cervical (pre)cancer in HPV-positive women. *Gynecol Oncol* 2016;141:341–7. https://doi.org/10.1016/j.ygyno.2016.02.012.
39. Luttmer R, De Strooper LMA, Steenbergen RDM, et al. Management of high-risk HPV-positive women for detection of cervical (pre)cancer. *Expert Rev Mol Diagn* 2016;16:961–74. https://doi.org/10.1080/14737159.2016.1217157.
40. Steenbergen RDM, Snijders PJF, Heideman DAM, et al. Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. *Nat Rev Cancer* 2014;14:395–405. https://doi.org/10.1038/nrc3728.
41. Git A, Dvinge H, Salmon-Divon M, et al. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* 2010;16:991–1006. https://doi.org/10.1261/rna.1947110.
42. Koppers-Lalic D, Hackenberg M, de Menezes R, et al. Non-Invasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. *Oncotarget* 2014;7:22566–78. https://doi.org/10.18632/oncotarget.8124.
43. Sapre N, Hong MKH, Macintyre G, et al. Curated MicroRNAs in urine and blood fail to validate as predictive biomarkers for high-risk prostate cancer. *PLoS One* 2014;9:e91729. https://doi.org/10.1371/journal.pone.0091729.
44. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends Genet* 2012;28:544–9. https://doi.org/10.1016/j.tig.2012.07.005.
45. Schamberger A, Orbán TI. 3′ IsomiR species and DNA contamination influence reliable quantification of microRNAs by stem-loop quantitative PCR. *PLoS One* 2014;9:e106315. https://doi.org/10.1371/journal.pone.0106315.
46. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014;42(D1):D68–73. https://doi.org/10.1093/nar/gkt1181.
47. Mestdagh P, Hartmann N, Baeriswyl L, et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat Methods* 2014;11:809–15. https://doi.org/10.1038/nmeth.3014.
48. Magee R, Telonis AG, Cherlin T, et al. Assessment of isomiR discrimination using commercial qPCR methods. *Non-coding RNA* 2017;3:18. https://doi.org/10.3390/ncrna3020018.