CHAPTER 4

Genome-wide DNA methylation profiling reveals methylation markers associated with 3q gain for detection of cervical precancer and cancer

Wina Verlaat
Peter J.F. Snijders
Putri W. Novianti
Saskia M. Wilting
Lise M.A. De Strooper
Geert Trooskens
Johan Vandersmissen
Wim Van Criekinge
G. Bea A. Wisman
Chris J.L.M. Meijer
Daniëlle A.M. Heideman
Renske D.M. Steenbergen

Clinical Cancer Research 2017; 23(14): 3813-3822
**ABSTRACT**

**Background**

Epigenetic host cell changes involved in cervical cancer development following a persistent high-risk human papillomavirus (hrHPV) infection, provide promising markers for the management of hrHPV-positive women. In particular, markers based on DNA methylation of tumour suppressor gene promoters are valuable. These markers ideally identify hrHPV-positive women with precancer (CIN2/3) in need of treatment. Here, we set out to identify biologically relevant methylation markers by genome-wide methylation analysis of both hrHPV-transformed cell lines and cervical tissue specimens.

**Methods and Results**

Genome-wide discovery by next generation sequencing (NGS) of methyl-binding domain-enriched DNA (MBD-Seq) yielded 20 candidate methylation target genes. Further verification and validation by multiplex-targeted bisulphite NGS and (quantitative) methylation-specific PCR (MSP) resulted in 3 genes (GHSR, SST and ZIC1) that showed a significant increase in methylation with severity of disease in both tissue specimens and cervical scrapes (\(P < 0.005\)). The area under the ROC curve for CIN3 or worse varied between 0.86 and 0.89. Within the group of CIN2/3, methylation levels of all 3 genes increased with duration of lesion existence (\(P < 0.0005\)), characterised by duration of preceding hrHPV infection, and were significantly higher in the presence of a 3q gain (\(P < 0.05\)) in the corresponding tissue biopsy.

**Conclusion**

By unbiased genome-wide DNA methylation profiling and comprehensive stepwise verification and validation studies using in vitro and patient-derived samples, we identified 3 promising methylation markers (GHSR, SST and ZIC1) associated with a 3q gain for the detection of cervical (pre)cancer.

---

**TRANSLATIONAL RELEVANCE**

Implementation of primary high-risk HPV (hrHPV) DNA testing in cervical screening programmes in the coming years will result in better protection against cervical (pre) cancer. However, its specificity for clinically meaningful disease is lower compared with current cytology-based screening. Furthermore, current methods are unable to determine which precursor lesions will progress or regress, leading to over-referral and overtreatment. This study aimed to identify biologically relevant DNA methylation changes that enable identification of hrHPV-positive women with clinically relevant cervical lesions who are in need of referral. Following an unbiased genome-wide approach, we identified 3 novel biologically relevant methylation markers (GHSR, SST and ZIC1, all located on chromosome 3q) that provide most promising molecular triage markers for detection of cervical (pre)cancer in cervical scrapes. These methylation markers mainly detect advanced precursor lesions with a high likelihood of short-term progression risk to cancer, leading to the detection of clinically relevant precursor lesions. These findings could greatly advance the management of hrHPV-positive women.
INTRODUCTION

Cervical cancer is caused by a persistent high-risk human papillomavirus (hrHPV) infection, with HPV16 and HPV18 being the most predominant types. Cervical squamous cell carcinoma (SCC) develops through well-defined precursor lesions [cervical intraepithelial neoplasia, classically graded 1 to 3 (CIN1-3)], of which CIN3 is considered the most advanced precursor lesion of cervical cancer. Common practice is the ablation of CIN2 and CIN3 tissue to prevent cervical cancer development. However, many CIN2 and some CIN3 lesions will regress spontaneously if left untreated. Current CIN classification is not able to identify which lesions will progress or regress, leading to considerable over-referral and overtreatment with associated cervical morbidity. The progression of CIN2/3 to SCC results from the accumulation of genetic and epigenetic alterations in the host cell over a time period of 15 to 30 years. Given their variable duration of existence, CIN2/3 represents a heterogeneous disease, with various risk of progression to cancer. Recent molecular studies on CIN2/3 indicate that a subset of these lesions display a cancer-like profile including hypermethylation of tumour suppressor genes and specific chromosomal alterations. These so-called advanced CIN2/3 lesions, characterised by a longer duration (> 5 years) of preceding hrHPV infection (PHI), are most likely to have a high short-term progression risk to cancer and are in need of treatment. DNA methylation-based assays that can be used as a biomarker for the detection of advanced CIN2/3 lesions would greatly advance the management of hrHPV-positive women by reducing overdiagnosis and overtreatment. At present, a number of methylated host cell genes have been described by us and others as promising markers for the management of hrHPV-positive women, such as CADM1, MAL, miR124-2, FAM19A4 and EPB41L3. Here, we describe an unbiased genome-wide-based approach, comprised of next generation sequencing (NGS) of methyl-binding domain-enriched DNA (MBD-Seq), to identify biologically relevant methylation-based markers within patient-derived and in vitro samples representing the multistep process of hrHPV-induced carcinogenesis. Our longitudinal in vitro model of primary keratinocytes immortalised by hrHPV types closely mimics in vivo carcinogenic stages, both with respect to molecular as well as morphologic changes. This model system allows us to study the sequential order of accumulating epigenetic host cell changes associated with hrHPV-mediated carcinogenesis. Combined with the DNA methylation profiles of various stages of cervical (pre)cancer in vivo, this offers a rewarding strategy to identify biologically relevant methylation markers. In addition, stepwise verification and validation studies were performed using multiplex-targeted bisulphite NGS, methylation-specific PCR (MSP) and multiplex quantitative MSP (qMSP) on large clinical series of both cervical tissues and scrapes, comprising all stages of cervical carcinogenesis. To further understand the relation of DNA methylation levels in cervical scrapes to the severity and duration of the underlying cervical lesion, samples from women with CIN2/3 with a known duration of PHI, as a proxy of lesion duration, were included. Furthermore, to link methylation findings to other critical host cell alterations such as DNA copy number aberrations, we analysed previously obtained chromosomal profiles of these samples. Altogether, our comprehensive approach yielded 3 novel promising methylation markers associated with 3q gain for the detection of cervical (pre)cancer.
MATERIALS AND METHODS

Cell lines and clinical specimens

Discovery
For genome-wide DNA methylation profiling by MBD-Seq, various cell lines and frozen cervical tissue specimens were used. Cell lines consisted of (i) primary human foreskin keratinocytes (HFK) from two donors, (ii) HFK from one donor transduced with HPV16E6E7 at passage 7 and passage 30, representing the pre-immortal and immortal stage, respectively, (iii) consecutive passages of HPV16- (FK16A and FK16B) and HPV18- (FK18A and FK18B) immortalised keratinocytes, including early (passages 11-30), middle (passages 30-70) and late (> passage 70) immortal passages, and (iv) cervical cancer cell line SiHa. The tissue specimens comprised CIN1 (n = 3), CIN2/3 (n = 12) and SCC (n = 10). Given limitations in accurate histologic grading of high-grade CIN on frozen tissue sections, CIN2 and CIN3 were grouped together in this frozen tissue sample series.

Verification I
Multiplex-targeted bisulphite NGS was performed on a subset of hrHPV-transformed cell lines and frozen tissue specimens used for discovery with sufficient material left, that is, HFKs from two donors, early and late immortal passages of FK16B and FK18B, SiHa, CIN2/3 lesions (n = 8) and SCCs (n = 7). In addition, 2 pooled control specimens comprising pooled DNAs of 3 normal cervical formalin-fixed paraffin-embedded (FFPE) tissue specimens or 6 scrapes from women without cervical disease, respectively, were used.

Verification II
For verification by MSP, HFKs from four different donors, late immortal passages of FK16A, FK16B and FK18A, cervical cancer cell lines SiHa, HeLa and CaSki, and FFPE tissue specimens including normal cervical tissue (n = 11), CIN3 (n = 11) and SCC (n = 10) were used.

Validation
For validation by multiplex qMSP, HFKs from five different donors, middle and late immortal passages of FK16A, FK16B, FK18A and FK18B, and cervical cancer cell lines SiHa, HeLa and CaSki were used. In addition, an independent series of FFPE tissue specimens comprising normal cervical tissue (n = 41), CIN1 (n = 32), CIN2 (n = 23), CIN3 (n = 33) and SCC (n = 30), as well as a series of scrapes comprising specimens from hrHPV-positive women with normal cytology and/or without evidence of CIN2+ (n = 97) or from women who were histologically diagnosed with CIN2 (n = 21), CIN3 (n = 75), or SCC (n = 27), were used. Another 48 cervical scrapes from women with a CIN2 or CIN3 lesion in the second screening round (interval 5 years) of the POBASCAM trial were included. Details of this additional series have been described previously. In brief, the 5-year history of hrHPV infection of these women was known and used as surrogate for duration of lesion existence. Women with a PHI < 5 years were considered to have early CIN2/3 lesions (n = 19) and women with a PHI ≥ 5 years were considered to have advanced CIN2/3 lesions (n = 29). DNA copy number data were obtained during a previous study in which tissue specimens corresponding to this series of cervical scrapes were microdissected to enrich for the dysplastic areas and isolated DNA was hybridised to 105K arrays (Agilent Technologies). DNA copy number data are available from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/projects/geo/), through series accession number GSE30155.

DNA isolation, bisulphite treatment and hrHPV testing
Isolation of DNA from cell lines was performed by proteinase K digestion followed by standard phenol-chloroform extraction as described previously. Isolation of DNA from tissue specimens was performed by High Pure PCR Template Preparation Kit (Roche Diagnostics) using manufacturer’s protocol. DNA from cervical scrapes was isolated using the NucleoMag 96 Tissue kit (Macherey-Nagel) and a Microlab Star robotic system (Hamilton) according to the manufacturer’s protocol. For multiplex-targeted bisulphite NGS, MSP and qMSP analyses, DNA was bisulphite-converted using the EZ DNA Methylation kit (Zymo Research). hrHPV DNA detection was performed by GP5+/6+ PCR enzyme immunoassay or HPV-Risk assay (Self-screen B.V.)

MBD-Seq
Genome-wide DNA methylation profiling was performed by MBD-Seq, using the MethylCap kit v2 (Diagenode) with 500 ng input DNA combined with Illumina Genome Analyzer IIx.
paired-end sequencing as described before\textsuperscript{24,25}. Next, Bowtie 1.0.0. was used to map the obtained paired-end reads from fastq-files to the human reference genome GRCh37\textsuperscript{26}. These mapped sequencing results were visualised in the H2G2 genome browser (http://h2g2.ugent.be/biobix.html). From this browser, we extracted the read counts of all predefined methylated regions located in promoters (-1,000 bp to +500 bp relative to the transcription start site). The read counts of all regions in cervical cancer (SiHa, SCC) were compared with the read counts in controls (HFK, CIN1). These regions were ranked on the basis of the difference in read counts to select the hypermethylated regions. Next, we analysed hypermethylation of these regions in the precursors (HFKs transduced with HPV16E6E7, FK16A, FK16B, FK18A, FK18B and CIN2/3). From this list, we selected the top 20 most hypermethylated promoters.

**Multiplex-targeted bisulphite NGS**

For multiplex-targeted bisulphite NGS, a total of 62 primer pairs were designed to amplify the 20 selected methylated regions by PCR with on average 3 amplicons per target. Each forward and reverse primer contained a universal extension. The amplicons were pooled together per sample and purified from primer dimers using the Gel Extraction kit (Qiagen). A second PCR was performed on the pool of amplicons using primers annealing to the universal extension to add the flow cell binding sites p5 and p7 and a unique barcode to every sample. The concentration of every sample library was measured using the Bioanalyzer 2100 (Agilent Technologies). Up to 12 uniquely barcoded sample libraries were equimurally pooled prior to sequencing. This final pool contained 10% spike-in of PhiX Control v3 (Illumina) to increase the diversity and was loaded on a MiSeq Personal Sequencer (Illumina) using 5 pmol/L. A MiSeq Reagent Kit v2 (300 cycles; Illumina) was used in a paired-end 150 run according to manufacturer’s protocol. All bisulphite sequencing data described in this study were processed as follows: quality was evaluated using FastQC Version 0.11.2., the reads were aligned on the bisulphite-treated version of the human genome and methylation calls were performed in Bismark version 0.12.9\textsuperscript{27}. Reads were combined to represent a DNA methylation profile for each sample.

**MSP and multiplex qMSP**

MSP was performed as previously described using 50 ng of bisulphite-converted DNA as input material\textsuperscript{28}. MSP primers were designed to specifically amplify methylated bisulphite-treated DNA. In vitro methylated DNA (IVD) was used as positive control and unmodified SiHa DNA and H\textsubscript{2}O served as negative controls. The housekeeping gene \textit{β-actin} (\textit{ACTB}) verified sufficient DNA quality and successful bisulphite conversion of all samples.

Multiplex qMSP for 3 targets and \textit{ACTB} was designed as described previously\textsuperscript{29} and performed using 50 ng of bisulphite-converted DNA and 200-300 nmol/L of each primer and fluorescent dye-labelled probe, on the ABI 7500 Fast Real-Time PCR System (Applied Biosystems). The methylation values of the targets were normalised to reference gene \textit{ACTB} using the comparative \textit{C}_\text{t} method (2^{-ΔC_t} x 100) to obtain \textit{C}_\text{t} ratios. All samples in qMSP analysis had an \textit{ACTB} \textit{C}_\text{t} value < 32, indicating suitability for DNA methylation analysis.

**Statistical analyses**

For each methylated gene, Wilcoxon rank test was applied to test the difference in methylation levels between disease categories with a significance level of 0.05 (two-sided). For logistic regression analysis in cervical scrapes, samples from women with CIN3 or SCC (CIN3+) were categorised as cases and those from hrHPV-positive women with normal cytology and/or without evidence of CIN2+ were categorised as controls. To assess the relationship between the outcome of the samples (case or control) and the methylation markers, univariable logistic regression analysis was performed on the square root-transformed \textit{C}_\text{t} ratios of \textit{GHSR}, \textit{SST} and \textit{ZIC1}. The performance of the logistic regression model was visualised by receiver operating characteristics (ROC) curve, and was evaluated by sensitivity, specificity and AUC. Predicted probabilities, representing the risk for an underlying CIN3+, were calculated using the univariable logistic regression models of \textit{GHSR}, \textit{SST} and \textit{ZIC1}. Samples with predicted probabilities above the optimum threshold (\textit{GHSR}: 0.396; \textit{SST}: 0.358; \textit{ZIC1}: 0.354) were scored methylation-positive for the respective marker.
RESULTS

An outline of the study is given in Fig. 1 and Supplementary Table S1.

Genome-wide DNA methylation profiling by MBD-Seq

For discovery of DNA methylation profiles associated with hrHPV-induced carcinogenesis, genome-wide MBD-Seq was performed on primary keratinocytes, hrHPV-transformed cell lines and cervical tissue specimens (Fig. 1; Discovery). Differentially methylated promoter regions were identified as those with a difference in cervical cancer samples (SiHa, SCC) over normal/low-grade samples (HFK, CIN1), and precursors (hrHPV-transformed cell lines and CIN2/3 lesions) over normal/low-grade samples. Differentially methylated promoters were ranked to select the top 20 candidate hypermethylation target genes (Table 1).

Verification analyses and selection of candidate methylation markers

Verification of the top 20 candidates by multiplex-targeted bisulphite NGS was performed on a subset of samples used in the MBD-Seq screen (Fig. 1; Verification I). Increased methylation of 9 genes, \textit{GALR1, GHSR, IRX1, NEUROG3, POU4F3, PROX1, SIM2, SST} and \textit{ZIC1}, was verified in hrHPV-immortalised keratinocytes, cervical cancer cells, CIN2/3 lesions and SCCs compared with controls (Table 1). The remaining 11 genes were excluded from further studies due to the detection of methylation in controls (\textit{ANKRD2, CLDN18, GRK7, HIPK4, SLC45A2, TBX19} and \textit{TFDP3}) or technical issues (\textit{MYCL2, PDE4DIP, RXFP3} and \textit{SOX14}; Table 1).

Subsequent MSP analysis confirmed methylation of all 9 genes in late immortal passages of hrHPV-immortalised keratinocytes and cervical cancer cell lines, while HFKs were mainly methylation-negative (Fig. 1; Verification II; Table 1; Supplementary Fig. S1A). \textit{SIM2} was clearly methylation-positive in 1 of 4 HFKs. \textit{GALR1, IRX1} and \textit{NEUROG3} were ambiguously methylation-positive in some late immortal passages of hrHPV-immortalised keratinocytes. The remaining five genes (\textit{GHSR, POU4F3, PROX1, SST} and \textit{ZIC1}) were further evaluated by MSP in a series of cervical tissue specimens (Fig. 1; Verification II) confirming increased methylation frequency with increasing disease severity (Table 1; Supplementary Figs. S1B and S2). Two genes, \textit{POU4F3} and \textit{PROX1}, were methylation-negative in 1 of 10 SCCs leaving 3 genes (\textit{GHSR, SST} and \textit{ZIC1}) for a comprehensive validation analysis.

Validation of candidate markers in cell lines and cervical tissue specimens by qMSP analysis

On the basis of the verification findings, \textit{GHSR, SST} and \textit{ZIC1}, were further analysed by multiplex qMSP on cell lines and independent large series of cervical tissue specimens (Fig. 1; Validation). No \textit{SST} and \textit{ZIC1} methylation was detected in any of the HFKs, whereas \textit{GHSR} showed very low levels of methylation in 3 of 5 HFKs. Methylation of all 3 genes significantly increased with passaging of hrHPV-immortalised keratinocytes and was highest in cervical cancer cell lines (\(P < 0.05\); Fig. 2A-C). For all 3 genes, a gradual increase in methylation was observed with increasing severity of cervical disease. Methylation levels were significantly increased in CIN2 compared with normal cervical tissues (\(P < 0.005\)), in CIN3 compared with CIN2 (\(P < 0.05\)), as well as in SCCs compared with CIN3 (\(P < 0.0005\); Fig. 2D-F).
Table 1. The selected 20 novel candidate methylation markers and the summarised results of the verification analyses.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Verification by targeted bisulphite NGS</th>
<th>Verification by MSP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cell lines</td>
<td>Tissues</td>
</tr>
<tr>
<td>ANKRD2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CLDN18</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GALR1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GHSR</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>GRK7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HIPK4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>JRX1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MYCL2</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>NEUROG3</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PDE4DIP</td>
<td>+</td>
<td>na</td>
</tr>
<tr>
<td>POU4F3</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PROX1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RXFP3</td>
<td>na</td>
<td>+</td>
</tr>
<tr>
<td>SIM2</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SLCO4A2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SOX14</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>SST</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TBX19</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TFDP3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ZIC1</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

+: verified differential methylation; -: not verified differential methylation (i.e. detection of methylation in controls). Highlighted in bold: 3 methylation markers selected for multiplex quantitative MSP. NGS: next generation sequencing; MSP: methylation-specific PCR; na: not assessable.

Marker potential of GHSR, SST and ZIC1 in cervical scrapes

Next, validation analysis included multiplex qMSP on a total of 220 hrHPV-positive cervical scrapes to evaluate the potential of GHSR, SST and ZIC1 methylation analysis to detect cervical (pre)cancer (Fig. 1; Validation). For all 3 genes, fold changes in methylation levels compared with the reference group (hrHPV-positive controls) increased with severity of underlying disease (Table 2A). Largest fold changes were seen for ZIC1, showing a 10.1-fold increase in scrapes from women with CIN2, elevating to a 58.7-fold change in scrapes from women with CIN3, and a 921.1 fold-change in scrapes from women with SCC. Differences were significant for all 3 genes in hrHPV-positive scrapes from women with CIN3+ compared with hrHPV-positive controls (P < 0.0005) and hrHPV-positive scrapes from women with CIN2 (P < 0.005; Supplementary Fig. S3). Logistic regression analysis showed a similar clinical performance for CIN3+ detection in hrHPV-positive cervical scrapes for GHSR, SST and ZIC1, with an AUC of 0.87, 0.86 and 0.89, respectively (Fig. 3A; Table 3). ZIC1 presented the highest AUC, which corresponds at a certain threshold to a sensitivity of 86.3% and a specificity of 80.4% (Table 3). All cervical cancers were detected in scrapes by GHSR, whereas SST and ZIC1 detected identical 26 of 27 (96.3%) cervical cancers.

Figure 2. Validation of GHSR, SST and ZIC1 by qMSP in an in vitro model and in cervical tissue specimens. Methylation levels of (A) GHSR, (B) SST and (C) ZIC1 in an in vitro model represented by the log2-transformed Ct ratios (y-axis) in the different groups of cell lines (x-axis). Methylation levels of (D) GHSR, (E) SST and (F) ZIC1 in cervical tissue specimens represented by the log2-transformed Ct ratios (y-axis) in different histology groups (x-axis). HFK: human foreskin keratinocytes; middle immortal: FK16A/B and FK18A/B p30-p70; late immortal: FK16A/B and FK18A/B > p70; CCL: cancer cell lines; CIN: cervical intraepithelial neoplasia; SCC: squamous cell carcinoma. * P < 0.05; ** P < 0.005; *** P < 0.0005; NS: not significant.
Methylation levels of GHSR, SST and ZIC1 in cervical scrapes from women with CIN2/3 lesions with different duration of existence and known DNA copy number profile

To evaluate marker potential in relation to advanced CIN2/3, cervical scrapes from women with an underlying CIN2/3 lesion with a known duration of PHI, reflecting durations of lesion existence (19 early CIN2/3 lesions: < 5-year PHI; 29 advanced CIN2/3 lesions: ≥ 5-year PHI), were analysed by multiplex qMSP (Fig. 1; Validation). Methylation levels of GHSR, SST and ZIC1 increased with duration of PHI. Methylation levels of GHSR, SST and ZIC1 were significantly increased in cervical scrapes from women with advanced CIN2/3 compared with scrapes from women with early CIN2/3 (P < 0.0005; Fig. 3B-D), showing a 13.3-, 9.7- and 47.1-fold increase, respectively, over the reference (Table 2B), with methylation levels in scrapes from women with early CIN2/3 being comparable with the reference group. Using the above described logistic regression model, most cervical scrapes from women with advanced CIN2/3 scored methylation-positive by GHSR (20/29; 69%), SST (17/29; 58.6%) and ZIC1 (20/29; 69%), whereas all (SST and ZIC1) or majority (GHSR; 18/19; 94.7%) of scrapes from women with early CIN2/3 were methylation-negative. A combination of GHSR and ZIC1, showing methylation positivity for at least one of the markers, increased the detection of advanced CIN2/3 (22/29; 75.9%).

As GHSR, SST and ZIC1 are all located on chromosome 3q, a region that is gained in 55% of cervical SCC cases, we additionally evaluated a potential association of methylation levels with gain of the gene locus. For this purpose, copy number data obtained during previous studies on the underlying advanced and early CIN2/3 lesions were used. Methylation levels of all 3 genes in the cervical scrapes were significantly higher in the presence of a 3q gain in the corresponding CIN2/3 lesions compared with the absence of a 3q gain (P < 0.05; Fig. 3E-G), which was also reflected by their fold changes over the reference group (i.e. 50.9-153.0 in presence of 3q gain versus 2.3-29.6 in absence of 3q gain; Table 2B). For comparison, previously obtained methylation levels of FAM19A4 (chr. 3p14.1), CADM1 (chr. 11q23.2), MAL (chr. 2q11.1) and miR124-2 (chr. 8q12.3) were not significantly different between scrapes from women with advanced CIN2/3 with and without a 3q gain (Supplementary Fig. S4).

As GHSR, SST and ZIC1 are all located on chromosome 3q, a region that is gained in 55% of cervical SCC cases, we additionally evaluated a potential association of methylation levels with gain of the gene locus.
DISCUSSION

By unbiased genome-wide DNA methylation profiling and subsequent comprehensive stepwise verification and validation studies using in vitro and patient-derived samples representing the multistep process of hrHPV-induced carcinogenesis, we identified 3 new promising methylation markers (GHSR, SST and ZIC1) associated with 3q gain for the detection of cervical (pre)cancer. Methylation levels of GHSR, SST and ZIC1 showed a significant increase with severity of disease, reaching high levels in advanced CIN2/3 and SCC, both when analysing tissue specimens as well as cervical scrapes. Within the group of women with CIN2/3 lesions, methylation levels of all 3 genes in cervical scrapes increased with duration of lesion existence (P < 0.0005) and were significantly higher in the presence of a 3q gain (P < 0.05) in the corresponding tissue biopsy. Logistic regression analysis of GHSR, SST and ZIC1 in hrHPV-positive cervical scrapes for CIN3+ detection resulted in an AUC of 0.87, 0.86 and 0.89, respectively, showing a similar and good clinical performance of all 3 genes, supporting the potential of these methylation markers for the management of hrHPV-positive women. In contrast to cytology-based triage tests, such as p16/Ki-67 dual-stained cytology, methylation marker analysis has an improved reliability and it is applicable on both cervical scrapes and self-collected cervical-vaginal specimens. Methylation marker analysis may also be combined with HPV16/18 genotyping and has been found complementary in detection of CIN2/3 and cancer. DNA methylation is especially sensitive for detection of cancer and advanced CIN2/3, whereas most early CIN2/3 are detected by HPV16/18 genotyping. In addition, these 3 markers could potentially also be useful for early detection of other hrHPV-associated cancer types, such as a subset of head and neck cancers and other anogenital cancers.

The advantages of our study are the evaluation of large, independent series of different in vitro and patient-derived sample types and the use of a systematic discovery-verification-validation approach. MBD-Seq is an effective discovery methodology for biologically relevant methylation markers as it enriches for methylation targets characterised by methylation of multiple consecutive CpGs, which is a characteristic of gene silencing when affecting promoter regions, and an advantage over commonly used techniques detecting only methylation of single CpGs. Furthermore, using MBD-Seq, we could also detect methylated HPV DNA by analysis of the non-aligned reads. In the current study, both in vitro and patient-obtained samples comprising all stages of cervical carcinogenesis were used, which allowed us to evaluate the succession of methylation events and the selection of methylation targets which are biologically relevant and associated with hrHPV-mediated carcinogenesis. Furthermore, by using both sample types during stepwise verification and
Validation, we could exclude candidates showing methylation very early or relatively late during carcinogenesis, and accordingly would result in unfavourable clinical specificity or sensitivity, respectively.

High-grade CIN lesions (CIN2/3) represent a heterogeneous stage of disease, only a subset of which may progress to cancer over a period of 15 to 30 years\textsuperscript{5-7}. Previous data have shown that the heterogeneity in duration of lesion existence is reflected by their molecular profile, with significant increased copy number alterations and methylation levels in advanced CIN2/3 lesions compared with early CIN2/3 lesions\textsuperscript{5-8}. In line with previous findings\textsuperscript{5-8}, current data show that the methylation levels of \textit{GHSR}, \textit{SST} and \textit{ZIC1} are significantly increased in cervical scrapes from women with advanced CIN2/3 lesions compared with scrapes from women with early CIN2/3 lesions, which is also reflected in their frequency of methylation positivity. Of note and in addition to previous findings, the current study shows that methylation of all 3 genes, which are located at chromosome 3q, that is, \textit{GHSR} (3q26), \textit{SST} (3q28) and \textit{ZIC1} (3q24), is related to the presence of a 3q gain. Gain at 3q, which is the most frequently gained region in cervical cancer\textsuperscript{6,30}, as well as amplification of the human telomerase gene (\textit{TERC}; 3q26), have been suggested as a marker for progression to cancer\textsuperscript{31-33}. The observation that methylation levels of genes at other chromosomal locations were not significantly increased in cervical scrapes from women with advanced CIN2/3 with a 3q gain, suggests that DNA methylation of \textit{GHSR}, \textit{SST} and \textit{ZIC1} may be biologically important to compensate for increased gene dosage during cervical carcinogenesis.

At present, there are only a few studies which investigated a relation between DNA methylation and DNA copy number profiles in cervical cancer, and those were mostly focused on allelic losses, such as for chromosomes 11q and 3p\textsuperscript{35}. In lung cancer, a correlation between DNA hypermethylation and gene amplification has been described for \textit{SNOX2}, which is also located on chromosome 3q\textsuperscript{36}. Another study described a significant association between DNA methylation of the \textit{HOXA} locus and a gain of chromosome 7 in glioblastoma\textsuperscript{37}. MBD-Seq on cervical specimens by Boers and colleagues, reported 15 candidate methylation targets\textsuperscript{38}, part of which were also located on known copy number gains, that is, \textit{ST6GALNAC5} (1p), \textit{LHX8} (1p), \textit{CDH6} (5p) and \textit{IRX1} (5p). Yet, it is unclear whether DNA hypermethylation is a response reaction to copy number alterations or a cause of copy number alterations due to destabilisation of the genome. Further studies are warranted to investigate the coevolution of DNA methylation and copy number alterations, in particular copy number gains, using integrated copy number and methylation analysis. In the current study, we were unfortunately unable to acquire information on copy numbers from our sequencing data due to varying read-counts between samples using MBD-Seq analysis.
ACKNOWLEDGEMENTS

The authors thank Marjolein van Zummeren, Hanne Sørgård, Annina van Splunter, Daoud Sie, Paul Eijk, Dirk van Essen, Larissa Vork and Tine Beyaert for excellent technical assistance.

REFERENCES

Chapter 4

Methylation Markers at 3q for Cervical (Pre)cancer Detection


**SUPPLEMENTARY FIGURES AND TABLE**

**Supplementary Figure S1.** Overview of MSP results in Verification II. MSP data of methylation markers selected for verification are shown in (A) cell lines and (B) cervical tissue specimens. Samples were scored as methylated (+), weakly methylated (+/-) or unmethylated (-). ND: not determined; HFK: human foreskin keratinocytes; FK16A/B: hrHPV16-immortalised keratinocytes; FK18A: hrHPV18-immortalised keratinocytes; late immortal: > p70; CIN3: cervical intraepithelial neoplasia grade 3; SCC: squamous cell carcinoma.

**Supplementary Figure S2.** Representative MSP results in cervical tissue specimens. Data are shown for (A) GHSR and (B) reference ACTB. IVd: in *vitro* methylated DNA; CIN3: cervical intraepithelial neoplasia grade 3; SCC: squamous cell carcinoma.
Chapter 4  

Supplementary Figure S3. Validation of GHSR, SST and ZIC1 by qMSP in hrHPV positive cervical scrapes. Shown are methylation levels (i.e. log2-transformed C\textsubscript{t} ratios) of (A) GHSR, (B) SST and (C) ZIC1 in hrHPV-positive cervical scrapes measured by qMSP according to underlying disease category (x-axis). hrHPV+ ctrl: hrHPV-positive controls; CIN: cervical intraepithelial neoplasia; SCC: squamous cell carcinoma. * P < 0.05; ** P < 0.005; *** P < 0.0005; NS: not significant.

Supplementary Figure S4. No association of FAM19A4, CADM1, MAL and mir124-2 methylation levels with 3q gain. Shown are (A) FAM19A4, (B) CADM1, (C) MAL and (D) mir124-2 methylation levels (i.e. log2-transformed C\textsubscript{t} ratios) in hrHPV-positive cervical scrapes in relation to the presence of a chromosomal gain of 3q in the underlying CIN2/3 lesion. +3q gain: with 3q gain; -3q gain: without 3q gain; CIN: cervical intraepithelial neoplasia; Early CIN2/3: PHI < 5 years; Advanced CIN2/3: PHI ≥ 5 years. NS: not significant.
## Supplementary Table S1. Number of samples used per study item.

<table>
<thead>
<tr>
<th></th>
<th>Discovery</th>
<th>Verification I</th>
<th>Verification II</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell lines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFK</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>FK16A,B/FK18A,B</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Cancer cell lines</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>7</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td><strong>Tissues</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>ND</td>
<td>2</td>
<td>11</td>
<td>41</td>
</tr>
<tr>
<td>CIN1</td>
<td>3</td>
<td>ND</td>
<td>ND</td>
<td>32</td>
</tr>
<tr>
<td>CIN2</td>
<td>12</td>
<td>8</td>
<td>ND</td>
<td>23</td>
</tr>
<tr>
<td>CIN3</td>
<td>10</td>
<td>7</td>
<td>11</td>
<td>33</td>
</tr>
<tr>
<td>SCC</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>17</td>
<td>32</td>
<td>159</td>
</tr>
<tr>
<td><strong>Cervical scrapes (1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hrHPV-positive controls</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>97</td>
</tr>
<tr>
<td>CIN2</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>21</td>
</tr>
<tr>
<td>CIN3</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>75</td>
</tr>
<tr>
<td>SCC</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>220</td>
</tr>
<tr>
<td><strong>Cervical scrapes (2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early CIN2/3</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>19</td>
</tr>
<tr>
<td>Advanced CIN2/3</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>29</td>
</tr>
<tr>
<td>without 3q gain</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>16</td>
</tr>
<tr>
<td>with 3q gain</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>11</td>
</tr>
<tr>
<td>unknown CNA profile</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>48</td>
</tr>
</tbody>
</table>

HFK: human foreskin keratinocytes; FK16A/B: hrHPV16-immortalised keratinocytes; FK18A/B: hrHPV18-immortalised keratinocytes; CIN: cervical intraepithelial neoplasia; SCC: squamous cell carcinoma; Early CIN2/3: PHI < 5 years; Advanced CIN2/3: PHI ≥ 5 years. ND: not determined.