New insights in methodology of screening for cervical cancer

Wang, Rong

Chapter 5

Genome-wide methylome analysis to discover (novel) hypermethylation biomarkers for both adeno- and squamous cell cervical carcinoma

Wang R.[1,3], Schuuring E.[2], Boers A.[1], van Leeuwen R.W.[1],
van der Zee A.G.J.[1], Wisman G.B.A. [1]

[1] Department of Gynecologic Oncology, University of Groningen, University Medical Center Groningen, Cancer Research Center Groningen, Groningen, the Netherlands
[2] Department of Pathology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
[3] Department of Laboratory Medicine, Tianjin Medical University, Tianjin, China

Submitted for publication

Abstract

Aims: Cervical adenocarcinomas (ADC) are mainly diagnosed at an advanced stage of disease. In the last decade, the incidence of ADC increased in most developed countries and represents about 20% of cervical cancers. One explanation for the increase of ADC is the less effective cytomorphological detection of ADC and its precursors in population-based screening programs. Analysis of DNA methylation markers might improve the detection of ADC in earlier stages. However, no specific methylation markers have been described for the detection of ADC. The aim of this study was to discover novel methylation markers for cervical cancer detecting both ADC and SCC.

Methods: To generate a global methylation-profile of DNA from 20 normal cervices, 6 ADC and 6 SCC, methylated DNA fragments were captured using the Methyl Binding Domain (MBD) of human MeCP2 followed by next-generation-sequencing (MethylCap-seq). Differential methylated markers were selected for verification by bisulfite pyrosequencing or methylation specific PCR (MSP) on the same samples used for MethylCap-seq and validated on an independent series of FFPE specimens from normal cervices and cervical cancers. Further clinical validation was performed by quantitative methylation specific PCR (QMSP) on cervical scrapings from an independent cohort of 89 women with a normal cervix and 125 cervical cancer patients.

Results: Validation of the highest ranking 15 differentially methylated candidate markers resulted in 5 markers exhibiting different methylation between normal and cancer tissues ($p<0.05$). Using QMSP analysis on cervical scrapings, the sensitivity of these 5 markers varied from 80.5% to 91.9% to detect both ADC and SCC with almost all normal scrapings negative (specificity: 94% -98.9%).

Conclusion: Using MethylCap-seq analysis, we identified 5 new methylation markers with a high sensitivity for both ADC as well as SCC in cervical scrapings.

Introduction:

Cervical cancer is one of the common female cancers in the world. Each year, more than 500,000 new cases and around 275,000 deaths occur globally[1]. Cervical squamous cell carcinoma (SCC) and cervical adenocarcinoma (ADC) are two main histological subtypes of invasive cervical cancer, which account for 75–90% and 10–25%, respectively[2-4]. Compared to SCC, ADC is more common in European countries with an incidence ranging from 5.5% to 30.0%[5]. Currently, the incidence of SCC is declining in most developed countries. In contrast, there is a rise in the absolute and relative incidence of ADC[6]. In Europe, ADC is increasing rapidly, especially in younger women. In the Netherlands, the absolute incidence rate of ADC increased with 15.8% in women aged 15-29 years and 2.5% in women aged 30-44 years[7]. Moreover, compared to SCC, ADC is mainly diagnosed in more advanced stages, appears to be less sensitive to radiation therapy and chemotherapy and is associated with a worse prognosis than SCC[8-11].

Both the upward trend and postponed detection of ADC are probably due to the present population-based screening programs, which are more effective in detecting the precursors of SCC than those of ADC. Because of its localization higher up in the cervix it is more difficult to obtain representative cytology samples and to observe ADC or its precursors by colposcopy[12]. High risk human papillomavirus (hrHPV) associated with cervical carcinogenesis is widely known and its detection is more sensitive for the detection of cervical adenocarcinoma in situ (AdCIS) and ADC than cytology. However, in population-based screening programs hrHPV-positive women with normal cytology will require additional biomarkers, again more specific for SCC and its precursors than for ADC to enable the gynecologist to decide on whether performing endocervical curettage or not[13]. Therefore, novel biomarkers for cervical cancer are required that ideally will identify and discriminate between ADC and SCC as well as their precursors with high sensitivity.

Aberrant gene expression caused by epigenetic mechanisms are prominent features of many types of cancer[14], and promoter DNA methylation of tumor suppressor genes (TSG) has been reported to be an early event in carcinogenesis[15]. DNA methylation markers might be exploited in cancer diagnosis as variations in DNA methylation are

observed more frequently than other genetic variations[16]. Although we[17,18] and others[19] have reported many methylation markers associated with cervical cancer, many of these markers are more frequently methylated in SCC compared to ADC[17]. As to cervical cancer diagnostics, an important advantage of DNA methylation markers is that they can be detected in the same scrapings as used for HPV analyses[20,21]. However, so far only a limited number of methylated genes have been identified that are specifically associated with ADC. Most of these markers have lower sensitivity for ADC and SCC both or either one [22,23]. Recently, 4 genes (PAX1, PTPRR, SOX1, and ZNF582) were reported that are frequently methylated in ADC as well as SCC[24]. However, data on screening using these markers is missing for larger cohorts.

In the past ten years, advances in whole genome methylation profiling technologies have revolutionized the field of cancer research. In order to identify cervical cancer specific methylation markers, the pharmacological unmasking expression microarray approach[25] and chromatin immunoprecipitation (ChIP) combined with methylation-specific oligonucleotide microarray have been performed[26,27]. Nevertheless, microarray-based screening has drawbacks such as their design and production, while also the inaccurate hybridization signals and antibody-based MeDIP are rather variable, which leaves space for further improvement. Reduction in costs have spurred the adoption of next generation sequencing (NGS) platforms with higher sensitivity and accuracy compared to traditional microarray profiling[28]. Recently, affinity-based methylation capture assay coupled using methyl binding domain (MBD) complexes with NGS (MethylCap-seq) has been reported to be an effective technique to comprehensively analyze the methylome in lung cancer, ovarian cancer, and head and neck cancer[29-31], and panels of hypermethylated loci have shown to represent possible methylation markers for early detection. These technologies have facilitated the discovery of potential biomarkers for disease development and progression as well as our understanding of the complex, underlying molecular mechanisms that lead to cancer.

Until now, no DNA methylome analysis has been performed using patient material including cervical ADC. In this study, MethylCap-seq was applied to perform a

genome-wide DNA methylation screening of cervical cancer, including both ADC and SCC, and normal cervix tissues. With this approach, we sought to identify genome-wide aberrant methylation patterns of cervical cancer-specific markers with high sensitivity to detect both ADC and SCC in cervical scrapings.

Materials and Methods

General Strategy

In order to identify and validate cervical cancer-specific methylation markers both for ADC and SCC, the following strategy was applied (Fig1). **Step1,** DNA from frozen tissue of cervical cancer (ADC=6, SCC=6) and normal cervices (n=20) was analyzed using MethylCap-seq. Subsequently, the differentially methylated regions (DMRs) were identified between normal cervices and cancers by statistical analysis. **Step2,** among the methylation candidates, the top 15 were selected for verification by methylation specific PCR (MSP) or pyrosequencing on the same frozen tissue as used for step1. **Step3,** using the selected methylation candidates (in step2), MSP was performed to validate on DNA of FFPE tissue from an independent series (normal n=17, cancer n=13 composed of ADC n=6 and SCC n=7). **Step4,** methylation candidates showing a significant difference methylation frequency in normal and cancer were selected for further clinical validation on cervical scrapings from a large series of cervical cancer patients (n=125 comprising n=57 ADC and n=68 SCC) and healthy age-matched controls (n=89) by real time quantitative methylation specific PCR (QMSP).

Patients:

Patients with cervical cancer referred to the outpatient clinic of the University Medical Center Groningen (UMCG) are asked to participate in our on-going 'Methylation study' that has been approved by the Institutional Review Board (IRB) of UMCG, the Netherlands. All patients from whom material was obtained gave written informed consent. Frozen tissue, paraffin embedded tissue and scrapings for this study were prospectively collected and stored in our tissue bank.

Normal tissue samples and normal scrapings are collected from patients with non-malignant disease. All cervical tissue that was used for the normal control group was

judged as normal by histopathological examination. Patients referred with cervical

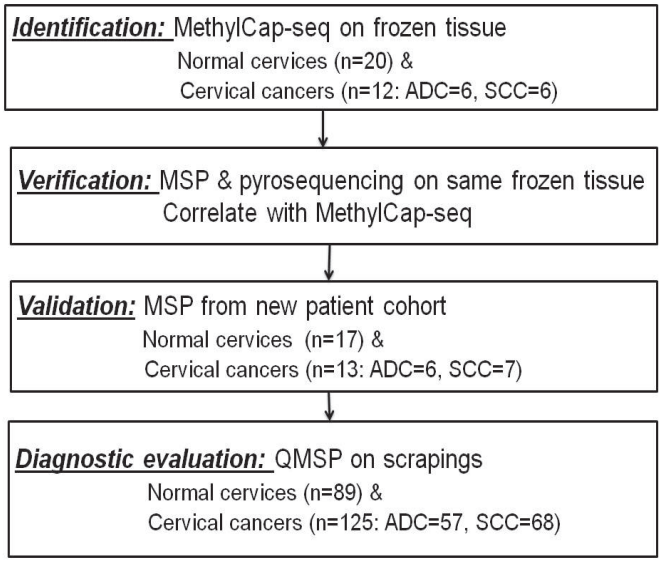

Fig1. Flow scheme for the identification of new cervical cancer markers

cancer are staged according to the FIGO criteria with pelvic examination and biopsies under general anaesthesia. Cervical scrapings from both groups (cervical cancer staging and benign gynecologic surgery) were collected before surgery under general anaesthesia. The tissue samples were scored by an experienced gynecologic pathologist and the histological classification was used as the reference standard. All clinicopathological data were retrieved from patient files and stored in our large anonymous password-protected institutional Gynecologic Oncology database.

For MethylCap-seq and pyrosequencing, frozen tissue specimens were collected from 20 normal cervices (IQR 33-45, median age: 43 years) and 12 cancers composed of 6 SCC and 6 ADC (IQR 27- 69, median age: 44 years). Stage of cervical cancer patients was for ADC: 3 FIGO stage IB1, 1 FIGO stage IB2 and 2 FIGO stage IIA. For SCC: 3 FIGO stage IB1, 1 FIGO stage IB2 and 2 FIGO stage IIA. For MSP analysis, formalin fixed paraffin embedded (FFPE) tissue was collected from 17 normal cervices (IQR 40-44, median age: 43 years), 13 cervical cancers including

6 ADC and 7 SCC (IQR 42-54, median age: 49 years). Stage of cervical cancer patients was for ADC: 2 FIGO stage IB1, 1 FIGO stage IB2, 2 FIGO stage IIA and 1 FIGO stage IIIA, for SCC: 2 FIGO stage IB1, 2 FIGO stage IB2, 1 FIGO stage IIA, 1 FIGO stage IIB and 1 FIGO stage IIIB.

For QMSP, scrapings were collected from 89 normal cervices (IQR 43-53, median age: 47 years), and from 125 cervical cancer patients (IQR 23 to 84, median age: 50 years) compromising 68 SCC and 57 ADC. Stage of cervical cancer patients was for ADC: 2 FIGO stage IA1, 1 FIGO stage IA2, 25 FIGO stage IB1, 8FIGO stage IB2, 8 FIGO stage IIA, 5 FIGO stage IIB, 1 FIGO stage IIIA, 6 FIGO stage IIIB and 1 FIGO stage IV, for SCC: 1 FIGO stage IA1, 19 FIGO stage IB1, 14 FIGO stage IB2, 13 FIGO stage IIA, 14 FIGO stage IIB, 1 FIGO stage IIIA, 5 FIGO stage IIIB and 1 FIGO stage IV.

Sample collection procedure and DNA isolation:

From the frozen tissue and FFPE samples, 10 µm tissue sections were cut and macrodissection was performed to enrich for epithelial cells. Before and after cutting hematoxylin and eosin slides were made to check presence of epithelial cells. Cervical scrapings were collected in 5 ml ice-cold phosphate buffered saline (PBS: 6.4 mM Na_2HPO_4; 1.5 mM KH_2PO_4; 0.14 M NaCl; 2.7 mM KCl) and kept on ice until further processing. Of these 5 ml cell suspension, 1 ml was used for cytomorphological assessment. The remaining part (4 ml) was centrifuged and the cell pellet was suspended in 1 ml TRAP wash buffer and divided in 4 fractions. Two fractions were stored as dry pellet at -80°C for DNA isolation.

Tissue slices from FFPE were deparaffinized using 100% xylene followed by 100% ethanol[17]. Genomic DNA from fresh-frozen macro-dissected samples and cervical scrapings was isolated by standard overnight 1% SDS and Proteinase K treatment, salt-chloroform extraction and isopropanol precipitation as described previously[18]. DNA pellets were washed with 70% ethanol and dissolved in 150 µl TE^{-4} (10 mM Tris/HCL; 0.1 mM EDTA, pH=8.0). Genomic DNA was amplified in a multiplex PCR according to the BIOMED-2 protocol, to check the DNA's structural integrity [32]. For the MethyCap-seq samples, DNA quantity was measured using Quant-i T™

PicoGreen ® dsDNA Assay Kit according to manufacturer's protocol (Invitrogen, Carlsbad, CA, USA). For cervical scrapings DNA concentrations and 260/280 ratios were measured using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA). A 260/280 ratio of >1.8 was required for all samples.

MethylCap-seq:
Genomic DNA samples (500 ng) were randomly sheared to a size range of 300-1000 bps using a Bioruptor™ UCD-200 (Diagenode, Liège, Belgium) and fragments of ~300 bp were isolated. Methylated DNA fragments were captured with methyl-binding domains using the MethylCap kit according to manufacturer's instructions (Diagenode, Liège, Belgium). The kit consists of the MBD of human MeCP2, as a C-terminal fusion with Glutathione-Stransferase (GST) containing an N-terminal His6-tag. Leukocyte DNA of 4 healthy controls was included in 2 sets of 2 samples.

Captured DNA was paired-end-sequenced on the Illumina Genome Analyzer II platform according to protocol (Illumina, San Diego, CA, USA). Results were mapped on the nucleotide sequence using Bowtie software[33], visualized using BioBix' H2G2 browser (http://h2g2.ugent.be/) and processed using the human reference genome (NCBI build 37). The paired-end fragments were unique and located within 400 bp of each other[34].

For statistical analysis, only reads of the promoter (-2000 bp to + 500 bp of transcription start site) were retrieved as these are mainly related with transcriptional silencing of genes. In order to identify differences between normal cervices and cervical cancer tissues, we dichotomised the read data into methylation positive or negative. Samples from normal cervices were considered methylation negative if a sample showed either 0 or 1 read. Cancer samples were considered methylation positive if a sample showed ≥3 reads. Subsequently, Fisher exact test was performed to determine the significant DMRs between ADC and normal or SCC and normal. To downsize the number of DMRs and to pinpoint candidate methylation markers in cervical cancer the following criteria were applied (see fig 2): 1) the methylation frequency is significantly different between normal and cancer. 2) Unmethylated (0 or

1 read) in at least 75% (15/20) of the normal cervix group. 3) Methylated (≥3 reads) in at least 50% (3/6) of ADC and SCC, respectively. 4) Low/negative reads in the leukocytes to prevent false positive results. The region was excluded if both leukocyte samples showed >1 read or if 1 leukocyte sample showed >2 reads. 5) DNA region length should ≥30bp. 6) Comparable regions in both identified histological subtype candidates group.

Bisulfite treatment of DNA:

Bisulfite treatment on denatured genomic DNA was performed as previously reported [35]. One microgram of genomic DNA per sample was modified with sodium bisulfite using the EZ DNA methylation kit according to manufacturer's protocol (Zymo Research, Corp, Irvine, CA). Leukocyte DNA from healthy women and whole genome amplified DNA (WGA) were used as negative controls, in vitro methylated (by SssI enzyme) leukocyte DNA was used as positive control.

Pyrosequencing:

Bisulfite treated DNA was amplified using PyroMark PCR kit (Qiagen, Hilden, Germany). PCR reaction and cycling conditions were according to the kit manual. Subsequently, sample preparation and pyrosequencing was performed by PyroMark Q24 instrument (Qiagen, Hilden, Germany) using the Pyro Gold Q24 Reagents (Qiagen, Hilden, Germany). Data was analyzed and quantified with the PyroMark Q24 software version 2.0.6 (Qiagen, Hilden, Germany). Non-template control (water), positive and negative controls were used in each reaction. PCR and sequence primers were available in *Supple. TableS1*.

MSP (Methylation specific PCR)

For MSP, each reaction was performed in 30 µl total reaction volume, containing: 600 nM of each MSP primer, 1.5 µl of bisulfite treated DNA (approximately 15 ng), standard PCR components (Applied Biosystems, Carlsbad, CA, USA) and 0.5 U AmpliTaq Gold DNA polymerase (Applied Biosystems, Carlsbad, CA, USA). Condition of the MSP was: 10 min hot-start at 95°C; 95°C for 60 sec, 60°C for 60 sec, 72°C 60 sec for a total of 40 cycles, with a final elongation step of 7 min at 72°C. PCR products were separated on a 2.5% agarose gel, stained with ethidium bromide

and visualized by UV illumination. Non-template control (water), positive/negative control and normal control were used in each reaction.

QMSP: Real time quantitative methylation specific PCR

QMSP was performed as we described previously with an internal (FAM-ZEN/IBFQ)-labeled hybridization probe for quantitative analyses[18]. Primer and probe sequences are available in *Supple. TableS2*. β-actin was used as a methylation independent internal reference gene. QMSP reactions were performed in 10 µl final volume in 384 well plates, containing: 300 nM of forward and reverse primers, 250 nM of hybridization probe, 5 µl of 2* QuantiTech Probe PCR Master Mix (Qiagen, Hilden, Germany) and 2.5 µl bisulfite modified DNA (approximately 25 ng). Each sample was analyzed in triplicate by ABI PRISM ® 7900HT Sequence Detection System (Applied Biosystems, Carlsbad, CA, USA). Negative and positive controls were included in each QMSP. Standard curve analysis was performed on each plate and by each primers-probe set on serial dilutions of in vitro methylated leukocyte DNA. A DNA sample was considered methylated if at least 2 out of the 3 wells were methylation positive with a Ct-value below 50 and DNA input of at least 225 pg β-actin. The relative level of methylation of the region of interest was determined by the following calculation: (average DNA quantity of methylated gene of interest / average DNA quantity for internal reference gene β-actin) x 10000[35].

Statistical analysis:

Statistical analysis was performed using SPSS Statistics 20.0 (SPSS 20, Chicago, IL, USA). Chi-square test and Fisher's exact test for small numbers were used to analyze the different methylation frequency between normal and cancer. The correlation between the average methylation level of each frozen tissue sample and MelthyCap-seq reads were analyzed using Spearman's rank test. The Mann-Whitney U test was used to determine differences in methylation ratio between 2 groups. P-value<0.05 was considered to be statistically significant. The sensitivity, specificity, receiver operating characteristic curves (ROC) and area under ROC curve (AUC) were calculated for the clinical validation[36]. The optimal threshold was calculated based on the largest Youden's index[37,38].

Results:

Identification of methylated candidates by MethylCap-seq:

To identify DMRs in ADC and SCC compared to normal cervices, genome-wide MethylCap-seq was performed. After we applied our criteria based-on methylation frequency, 6,231 DMRs showed differential methylation in ADC compared to normal cervices and 10,724 DMRs were identified in SCC compared to normal cervices (Fig 2). In ADC as well as in SCC also hypomethylation was more frequently observed compared to normal cervices (Fig 3).

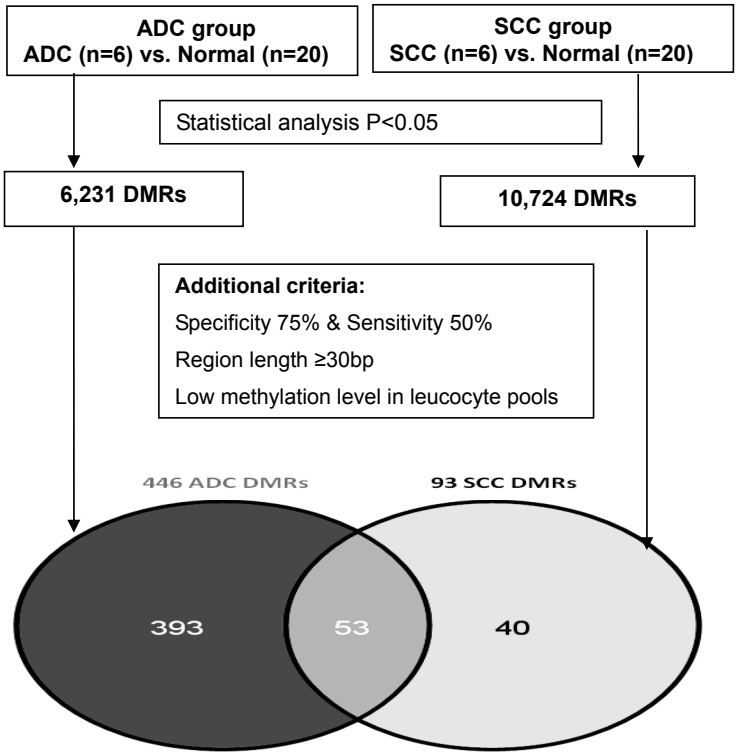

Fig2. Identification of methylated candidates by MethylCap-seq

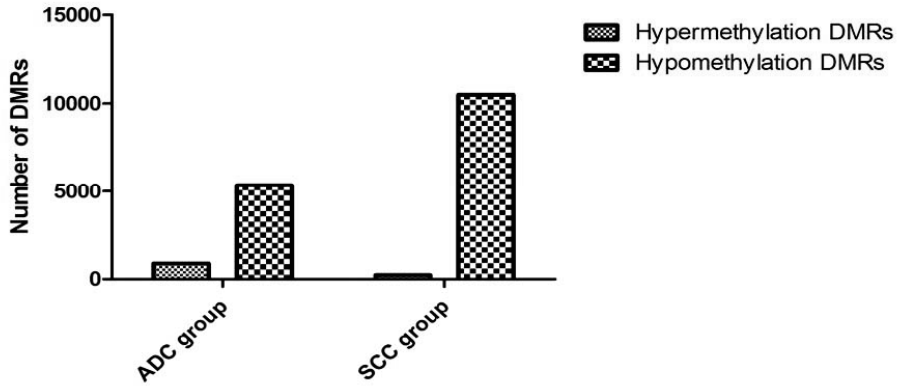

Fig3. Distribution of DMRs in both ADC and SCC groups

Additionally, we strengthened our criteria by focusing on the hypermethylated DMRs, as these are more easily to translate to MSP assays, which can be implemented as clinical diagnostic tests. Overall 446 DMRs, comprising 357 genes, were identified in ADC and 93 DMRs, comprising 89 genes, in SCC. Gene ontology (GO) functional analysis for these DMRs was performed to determine if similar pathways are affected in both histological types of cancer. There were in total 328 and 49 GO terms enriched in ADC and SCC, respectively. Most GO terms enriched in SCC were also enriched in ADC, as 37/38 of the biological processes, 4/5 of the cellular components and 5/6 of the molecular functions, respectively, were also shown in ADC. This underlines that similar pathways are disrupted in the carcinogenesis of both histological types of cervical carcinomas.

Figure 4 shows the P-values of the top5 GOs enriched in ADC together with the associated P-values in SCC. The most significant common pathways were sequence-specific DNA binding (GO:0043565), transcription factor activity (GO:0003700), transcription regulator activity (GO:0030528), plasma membrane (GO:0005886), neuron differentiation (GO:0030182), which indicates that some hyperDMRs are associated with transcription regulation.

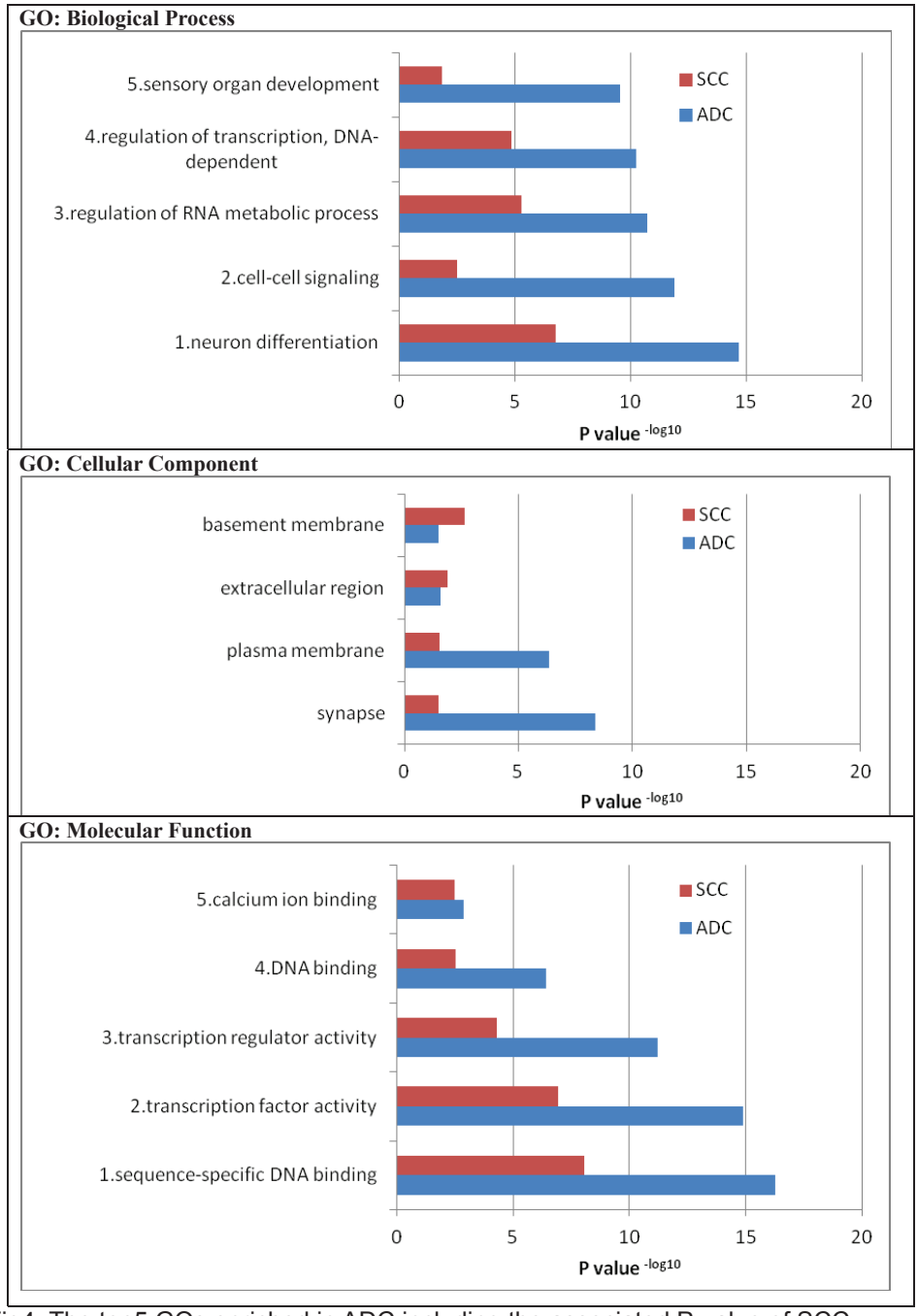

Fig4. The top5 GOs enriched in ADC including the associated P-value of SCC.

Verification and validation of the top15 candidates:

In order to identify methylation markers common in both histological groups, the DMRs identified from the two cancer groups were cross-compared, which generated 53 DMRs comprising 50 genes (Fig 2). Of these 53 candidates, the top 15 candidate markers (*Supple Table S3*) were selected to verify the MethylCap-seq data by either MSP or pyrosequencing analysis. Using the same DNA as was used for MethylCap-seq, it was shown that for 10 genes (SOX1, GFRA1, SLC6A5, TBX5, OLIG2, AC004963.1, TBX20, AC096537.2(219), CR753863.1, SOX14) a significant correlation remained between the PCR band intensity determined by MSP or the percentage of pyrosequencing and the number of reads from the MethylCap-seq. Table 1 shows an overview of which genes survived the different stages of validation.

Table1: Verification, validation and diagnostic evaluation of the top15 candidates

Rank	Gene	Verification		Validation		Diagnostic evaluation
		Primer optimization (Yes/No)	Significant association (Yes/No)	Primer optimization (Yes/No)	Significant difference (Yes/No)	Significant difference (Yes/No)
1	OLIG2	Yes	Yes	No		
2	CR753863.1	Yes	Yes	No		
3	GFRA1	Yes*	Yes*	Yes	Yes	Yes
4	EVX2	Yes	No			
5	AL356961.2	Yes	No			
6	TBX5	Yes	Yes	Yes	No	
7 #	AC096537.2(218)	Yes	No			
8	SOX1	Yes*	Yes*	Yes	Yes	Yes
9	SYT6	Yes	No			
10 #	AC096537.2(219)	Yes	Yes	No		
11	FREM3	Yes	No			
12	TBX20	Yes	Yes	Yes	Yes	Yes
13	AC004963.1	Yes	Yes	No		
14	SLC6A5	Yes	Yes	Yes	Yes	Yes
15	SOX14	Yes	Yes	Yes	Yes	Yes

\# Same gene, different region

* These genes were verified by MSP, the remaining genes were verified by pyrosequencing

For these 10 markers (SOX1, GFRA1, SLC6A5, TBX5, OLIG2, AC004963.1, TBX20, AC096537.2(219), CR753863.1, SOX14), MSP primers were designed. Four genes showed high methylation levels in leukocytes and WGA and were therefore excluded from further validation. Methylation patters of the remaining 6 genes were analyzed on DNA from an independent series comprising normal cervix and cervical cancers (ADC as well as SCC). Except for TBX5, 5 genes (SOX1, SOX14, GFRA1, SLC6A5, TBX20) showed a significant difference of methylation positivity between normal and cancer ($P<0.05$), with a methylation frequency in both ADC and SCC >50% (Table 2).

Table 2: Methylation positivity in external validation cohort of FFPF samples

Genes	Positive rate	Positive rate		
		Cancer total	ADC	SCC
SOX1*	0% (0/15)	91.67% (11/12)	100% (5/5)	85.7% (6/7)
GFRA1*	0% (0/11)	83%(10/12)	83.33% (5/6)	83.33% (5/6)
SOX14*	25% (4/16)	83%(11/13)	83.33% (5/6)	85.7% (6/7)
SLC6A5*	6.67% (1/15)	83.33% (10/12)	100% (5/5)	71.4% (5/7)
TBX20*	5.88% (1/17)	83.33% (10/12)	100% (5/5)	71.4% (5/7)
TBX5	56.25% (9/16)	67% (8/12)	60% (3/5)	71.4% (5/7)

* Comparison of positive rate in normal cervices vs. cancer by Fisher exact test ($P<0.05$)

Diagnostic evaluation by QMSP for normal versus cancer scrapings

To determine their diagnostic performance, QMSP was set up for 5 genes (SOX1, SOX14, GFRA1, SLC6A5, TBX20) and evaluated on scrapings from a large series of cervical cancer patients (n=125, (ADC: n=57) and SCC: n=68) and healthy, age-matched controls (n=89). QMSP analysis indicated that the level of DNA methylation for all five genes was significantly higher in cancer scrapings compared to normal scrapings ($P<0.001$), but as expected similar between ADC and SCC (Fig5).

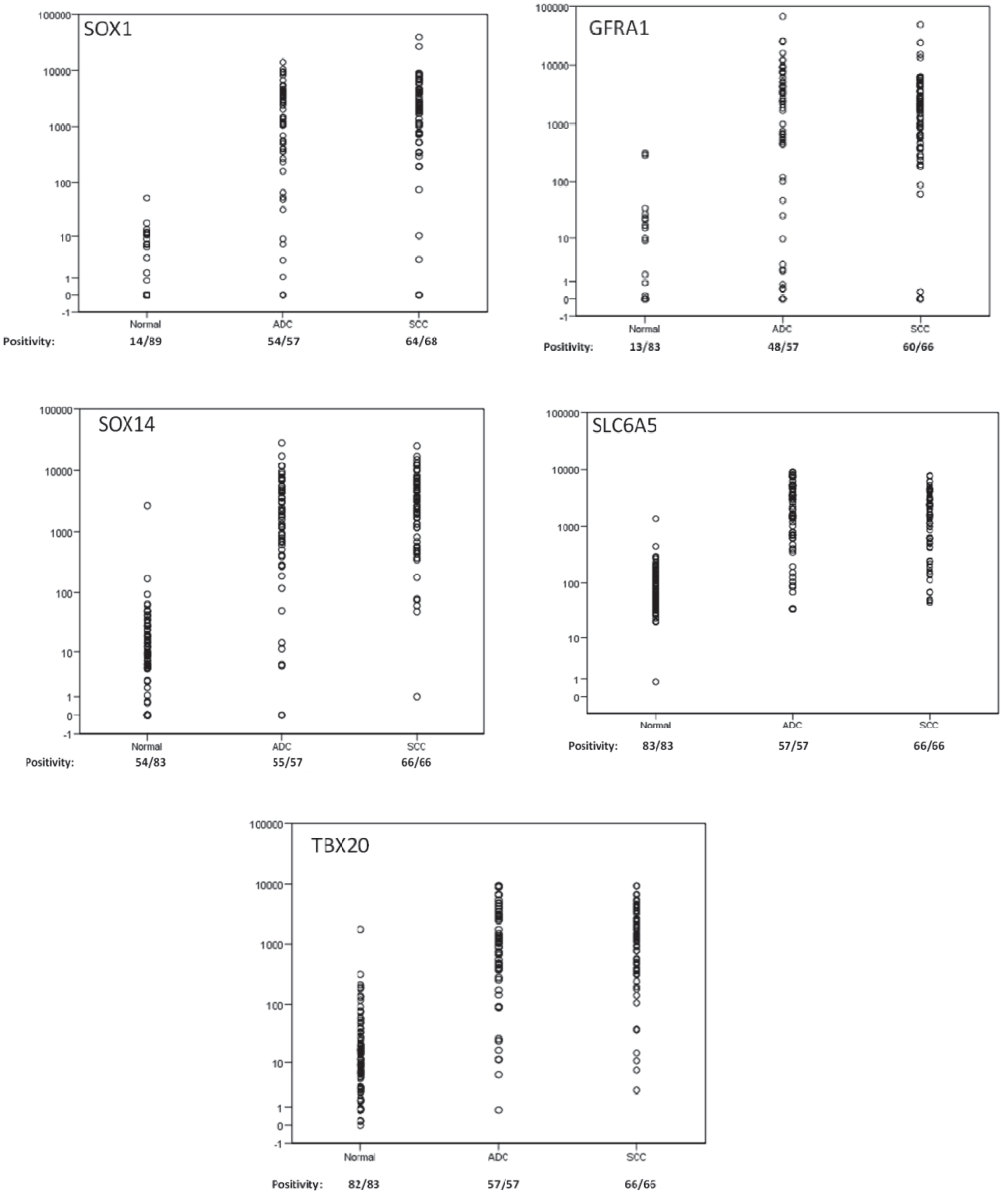

Fig5：Scatter plot in cervical scrapings of women with normal cervix or cervical cancer patients.
Mann-Whitney U test shows significant difference between normal and cancer ($P<0.001$) for all 5 genes ;
Mann-Whitney U test shows no difference between SCC and ADC ($P>0.05$) for all 5 genes.

Because many markers were also methylated in normal scrapings, albeit at lower levels as observed in cancer scrapings, a threshold was calculated at the highest Youden´s index based on the ROC analyses of the individual genes. Subsequently, sensitivity and specificity to detect ADC and/or SCC for all individual genes were determined (Table 4). The sensitivity for the 5 genes ranged from 80.5%-91.9%, while the specificity ranged from 94% -98.9% (Table 3a). For four genes (SOX1, SOX14, SLC6A5, TBX20), the sensitivity to detect either ADC or SCC was comparable, albeit slightly lower for ADC than for SCC.

In order to discover an optimal methylation marker panel with the highest sensitivity and specificity different gene combinations were evaluated. For a combination, a sample was considered positive if either one of the genes was positive. As expected, the combination of 2-3 genes decreased the specificity with the highest specificity calculated at 97.6% (Table 3b). However, simultaneously the sensitivity for cervical cancer increased by most of the combinations. The combination of SOX1, GFRA1 and SLC6A5, showed the highest AUC (0.959) with a sensitivity of 94.3% and a specificity of 97.6%. The methylation positivity for ADC and SCC was 89.5% and 98.5%, respectively (Table 3b). Addition of other genes did increase neither sensitivity nor specificity.

Discussion

Currently, an effective early detection method for ADC and its precursors is lacking and therefore ADC is mainly diagnosed in advanced stages. Screening for cervical cancer by Pap smear analysis is associated with significant false positive and false negative rates[39] and especially ADC are easily missed.[40,41] Compared to cytology, hrHPV screening will detect more (pre)malignant cervical cancers irrespective of histology. However, in population-based screening programs hrHPV-positive women will require triage analysis, which are until now still more specific for SCC and its precursors than for ADC[42,13]. Therefore, it is important to find better methods to improve the detection of the ADC. Recently, many studies have shown that inactivation of tumor suppressor genes by promoter hypermethylation is an early event in (cervical) carcinogenesis,[43,44] which may simultaneously also serve as suitable markers to allow early diagnosis[45].

Table3a: Sensitivity and specificity of individual genes in cervical scrapings of women with normal cervix or cervical cancer patients (with cut-off)

Gene Name	AUC#	Threshold	Specificity	Sensitivity			
				Cancer total	ADC	SCC	χ^2 for ADC vs SCC
SOX1	0.962	25.085	98.9%	89.6%***	87.7%	91.2%	P=0.528
SOX14	0.960	70.309	96.4%	91.9%***	87.7 %	95.5%	P=0.118
GFRA1	0.919	40.925	97.6%	80.5%***	70.2%	89.4%	P=0.007
TBX20	0.935	140.383	94.0%	86.2%***	82.5%	89.4%	P=0.266
SLC6A5	0.928	315.308	97.6%	81.3%***	82.5%	80.3%	P=0.760

AUC without threshold
*** Comparison of positive rate in women with a normal cervix or cervical cancer by chi-square (P<0.001)

Table3b: Sensitivity and specificity of combination of genes[1] in cervical scrapings of women with normal cervix or cervical cancer patients

Gene names	AUC#	Specificity	Sensitivity		
			Cancer total	ADC	SCC
Combination 2 genes					
SOX1/GFRA1	0.955	97.6%	93.5%	87.7%	98.5%
SOX1/SOX14	0.949	96.4 %	93.5%	89.5%	97.0%
GFRA1/ SOX14	0.945	96.4%	92.7%	87.7 %	97.0%
GFRA1/ SLC6A5	0.935	97.6%	89.4%	84.2%	93.9%
SOX1/ TBX20	0.931	92.8 %	93.5%	87.7 %	98.5%
SOX14/ SLC6A5	0.921	96.4%	92.7%	89.5%	95.5%
SOX1/SLC6A5	0.947	97.6%	91.9%	89.5%	93.9%
SOX14/ TBX20	0.921	91.6 %	92.7%	87.7%	97.0%
GFRA1/ TBX20	0.911	92.8 %	89.4%	82.5%	95.5%
SLC6A5/ TBX20	0.915	92.8 %	90.2%	86.0%	93.9%
Combination 3 genes					
SOX1/GFRA1/ SOX14	0.953	96.4%	94.3%	89.5 %	98.5%
SOX1/GFRA1/ SLC6A5	0.959	97.6%	94.3%	89.5 %	98.5%
GFRA1/ SOX14/ SLC6A5	0.949	96.4%	93.5%	89.5%	97.0%
SOX1/SOX14/ SLC6A5	0.949	96.4%	93.5%	89.5 %	97%
SOX1/SOX14/ TBX20	0.929	91.6 %	94.3%	89.5%	98.5%
SOX1/GFRA1/ TBX20	0.931	92.8 %	93.5%	87.7 %	98.5%
SOX1/SLC6A5/ TBX20	0.935	92.8 %	94.3%	89.5%	98.5%
GFRA1/ SOX14/ TBX20	0.921	91.6 %	92.7%	87.7%	97.0%
GFRA1/ SLC6A5/ TBX20	0.923	92.8 %	91.9%	86.0%	97.0%
SOX14/ SLC6A5/ TBX20	0.925	91.6%	93.5%	89.5%	97.0%

1: Combination of genes is made after a threshold was set (threshold is shown in table 3a). For a combination, a sample was considered positive if either one of the genes was positive.
AUC based on combination of genes.

In this study, we performed an unbiased genome-wide DNA methylation analysis, comparing cervical cancer, both SCC as well as ADC with and healthy cervical epithelium. This study is one of the first providing an overview of the DNA methylome for ADC with simultaneous identification of new methylation markers for ADC.

Finally, 5 new methylation markers were systematically validated for the early detection of both ADC and SCC in cervical scrapings. A DNA methylation marker panel with a specificity of 97.6% and sensitivity of 94.9% for cervical cancer was identified with methylation positivity for ADC of 85.7% and SCC of 98.5%. In carcinogenesis, cancer cells often exhibit imbalanced expression of oncogenes and tumor-suppressor genes, thus acquiring preferential growth ability[46]. It is well established that aberrant DNA methylation may lead to overexpression of oncogenes and/or repression of tumor-suppressor genes. Analyseis of those aberrant methylation patterns eg, by (Q)MSP indicate that alterations in DNA methylation patterns may be used as cancer biomarkers eg, for early diagnosis.[44,47-49] In our study, many differentially methylated markers, either hypo- or hypermethylated, were observed when normal cervices were compared with ADC and SCC, respectively. ADC develops from mucus-producing glandular cells, while SCC most often occurs at the squamous cells[4,12]. There is some evidence to suggest that ADC and SCC may also be associated with different epidemiologic co-factors. Smoking and high parity are risk factors for SCC[50], while obesity is a risk factor for ADC[51]. Since epigenetic marks reflect both an individual's genetic background and exposure to different environmental factors[52], we performed gene ontology (GO) analysis on the differentially hypermethylated markers in normal vs ADC and normal vs SCC to identify possible common disrupted pathways. Most of the pathways disrupted by hypermethylation in SCC were also disrupted in ADC, indicating highly similar disruption of pathways by hypermethylation during carcinogenesis independent of histological cancer subtype. Pathways identified were among others sequence-specific DNA binding, transcription factor activity and transcription regulator activity, all known to be involved in carcinogenesis [53-55]. Of the 53 differentially methylated candidates that were found in both ADC and SCC, 18 genes (Supplemental table 3) were described previously in literature as being more frequently methylated in cancer, and 4 genes were previously related to (squamous) cervical cancer (SOX1[56], SOX14,

ONECUT1[56], WT1).

In this study, we combined MethylCap-seq to draw detailed methylome maps. Bock et al[57] compared MeDIP, MethylCap and bisulfite-based methods. MethylCap and MeDIP provide broad coverage of the genome with higher coverage for MethylCap.. Both methods allow comparable distinction between methylated and unmethylated regions as bisulfite-based methods, but are less accurate for quantifying the DNA methylation levels in partially methylated genomic regions. However, as to DMRs, MethylCap was able to detect roughly twice as many in comparison to MeDIP at comparable sequencing depths. Nevertheless, every analytical technique has its limitations. Current methodology cannot differentiate between the two alleles of a gene. Presence of methylation does therefore not exclude that an active copy of the gene might be present.

In order to identify methylation markers for a diagnostic setting, we focused on the hypermethylated candidate genes. Using genome-wide MethylCap-seq it is difficult to draw an authentic map for the methylation of single CG sites and therefore a more sensitive and quantitative technique, i.e. pyrosequencing was applied for verification. A significant correlation was observed between MethylCap-seq and pyrosequencing. Subsequently, primers for (Q)MSP were designed as these assays are more suitable in a diagnostic setting. Common clinical indexes such as sensitivity, specificity, ROC and AUC were determined to allow a first impression of the diagnostic efficiency of each candidate marker. All five identified candidate markers appeared to be quite capable to discriminate between normal epithelium versus cancer. However, if no threshold was set a relatively low specificity was observed for SOX14, SLC6A5 and TBX20. Subsequently, an optimal threshold was set to improve the specificity. Youden's index is an easy method to set up an effective cut-off [38]. In our dataset, with the optimal cut-off, the specificity increased, while not losing too much sensitivity. In addition, marker combinations are also common choices to enhance the accuracy of clinical diagnosis[42]. From our analysis, the best result came from a 3 marker panel with a specificity of 97.6% and a sensitivity of 94.3%. So far only a limited number of methylated genes have been examined in ADC, especially using cervical scrapings in a large series. These studies revealed no adequate markers for clinical application,

as most of them had lower sensitivity for both ADC and SCC or either one.[23,24] Two genes of the Wnt pathway, DKK3 and SFRP2, showed more methylation in ADC tissue compared to SCC tissue (82% vs. 18% and 84% vs. 39%) and combined they allowed detection of all AdCIS and ADC. However, this analysis was limited to a very small number of scrapings (n=8)[23]. Recently, PAX1, PTPRR, SOX1 and ZNF582, previously reported to be frequently methylated in SCC scrapings, were also analyzed in ADC scrapings and showed a sensitivity of the single genes of 81.8%~93.3% with a specificity of 81.0%~95.2% in a Taiwanese population[24]. However, the sample size of scrapings again is small and a threshold was used to score samples methylation negative or positive. Combined this might easily affect the specificity when a larger number of normal scrapings is analyzed. [24].

Some of the 5 genes, that we found to be potential biomarkers, have previously been reported to be methylated in cancer. SOX1 and SOX14 belong to the SOX family. SOX proteins are the best examples of transcription factors having similar DNA binding specificities yet with divergent functions [58]. SOX1 encodes a transcription factor implicated in the regulation of embryonic development and in the determination of cell fate. DNA methylation of SOX1 in cervical cancer has been reported by Lai et al[24,56]. Furthermore, SOX1 was identified as a tumor suppressor gene, because it interfered with Wnt/β-catenin signaling in cervical cancer[59] and hepatocellular cancer[60]. Hypermethylated SOX1 was also found in ovarian cancer cells that are chronically exposed to cisplatin[61]. SOX1 methylation, at least in part, is responsible for cisplatin resistance in human non-small cell lung cancer (NSCLC)[62]. SOX14, in contrast to our data, has been reported to be a good marker to differentiate between ADC and SCC, with more methylation in SCC as determined by an array-based technique[63].

GFRA1, GDNF (glial cell line-derived neurotrophic factor) family receptor alpha 1, is a member of the GDNF receptor family, and mediates activation of the RET tyrosine kinase receptor. This gene is a candidate gene for Hirschsprung disease[64] and in lung adenocarcinoma its methylation status determines tumor aggressiveness and outcome[65]. Furthermore, in a recent study of our group GFRA1 methylation was also identified as a methylation marker for cervical intraepithelial neoplasia CIN2/3 (see Chapter 4).

T-box (TBX) transcription factors belong to an ancient gene family with critical roles in embryogenesis, in early cell fate decisions and in control of differentiation and organogenesis. TBX20 and TBX5 act synergistically to control vertebrate heart morphogenesis. Currently, TBX20 and TBX5 are TBX genes defined to have multiple protein isoforms created by alternative splicing and characterized by expression and functional studies. These proteins are important for development as mutations lead to severe developmental disorders in humans. Regulation of TBX transcription factor activity has been characterized through protein interactions and DNA binding affinities[66]. Only TBX20 methylation has previously been related to the recurrence of lung adenocarcinoma[65].

SLC6A5 (GLYT2) encodes a sodium- and chloride-dependent glycine neurotransmitter transporter important for the clearance of extracellular glycine during glycine-mediated neurotransmission. Mutations in this gene cause hyperekplexia, a heterogeneous neurological disorder. However, so far, there is no report showing the relationship between SLC6A5 and cancer.

Conclusions:

Overall, our approach resulted in new cervical cancer methylation markers with high specificity and high sensitivity for ADC as well as for SCC. Preliminary results showed that especially SOX1, GFRA1 and SLC6A5 combined might be promising markers. However, further research is needed to validate the clinical significance and reproducibility for these markers. For instance, validation of these markers in a population-based screening setting, particular for ADC precursor lesions such as cervical AdCIS has not been tested yet. Knowledge of the pathogenesis-associated epigenetic alterations based on the methylome analysis of ADC and SCC may result in new targets for both therapeutic as well as diagnostic approaches.

Reference

1. Ferlay J, Shin HR, Bray F, et al: Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 127:2893-917, 2010

2. Vizcaino AP, Moreno V, Bosch FX, et al: International trends in incidence of cervical cancer: II. Squamous-cell carcinoma. Int J Cancer 86:429-35, 2000

3. Vizcaino AP, Moreno V, Bosch FX, et al: International trends in the incidence of cervical cancer: I. Adenocarcinoma and adenosquamous cell carcinomas. Int J Cancer 75:536-45, 1998

4. Seoud M, Tjalma WA, Ronsse V: Cervical adenocarcinoma: moving towards better prevention. Vaccine 29:9148-58, 2011

5. Organization. WH: WHO/ICO information on centre on human papillomavirus virus and cervical cancer. aviailable at:http://apps.who.int/hpvcentre/statistics/dynamic/ico/DataQuery Select.cfm, 2010

6. Vinh-Hung V, Bourgain C, Vlastos G, et al: Prognostic value of histopathology and trends in cervical cancer: a SEER population study. BMC Cancer 7:164, 2007

7. Bulk S, Visser O, Rozendaal L, et al: Cervical cancer in the Netherlands 1989-1998: Decrease of squamous cell carcinoma in older women, increase of adenocarcinoma in younger women. Int J Cancer 113:1005-9, 2005

8. Lai CH, Hsueh S, Hong JH, et al: Are adenocarcinomas and adenosquamous carcinomas different from squamous carcinomas in stage IB and II cervical cancer patients undergoing primary radical surgery? Int J Gynecol Cancer 9:28-36, 1999

9. Eifel PJ, Burke TW, Morris M, et al: Adenocarcinoma as an independent risk factor for disease recurrence in patients with stage IB cervical carcinoma. Gynecol Oncol 59:38-44, 1995

10. Hopkins MP, Morley GW: A comparison of adenocarcinoma and squamous cell carcinoma of the cervix. Obstet Gynecol 77:912-7, 1991

11. Chen RJ, Lin YH, Chen CA, et al: Influence of histologic type and age on survival rates for invasive cervical carcinoma in Taiwan. Gynecol Oncol 73:184-90, 1999

12. Gien LT, Beauchemin MC, Thomas G: Adenocarcinoma: a unique cervical cancer. Gynecol Oncol 116:140-6, 2010

13. Kinney W, Fetterman B, Cox JT, et al: Characteristics of 44 cervical cancers diagnosed following Pap-negative, high risk HPV-positive screening in routine clinical practice. Gynecol Oncol 121:309-13, 2011

14. Hoque MO, Kim MS, Ostrow KL, et al: Genome-wide promoter analysis uncovers portions of the cancer methylome. Cancer Res 68:2661-70, 2008

15. Baylin SB, Ohm JE: Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer 6:107-16, 2006

16. Esteller M, Corn PG, Baylin SB, et al: A gene hypermethylation profile of human cancer. Cancer Res 61:3225-9, 2001

17. Yang N, Nijhuis ER, Volders HH, et al: Gene promoter methylation patterns throughout the process of cervical carcinogenesis. Cell Oncol 32:131-43, 2010

18.Eijsink JJ, Lendvai A, Deregowski V, et al: A four-gene methylation marker panel as triage test in high-risk human papillomavirus positive patients. Int J Cancer 130:1861-9, 2012

19.Wentzensen N, Sherman ME, Schiffman M, et al: Utility of methylation markers in cervical cancer early detection: appraisal of the state-of-the-science. Gynecol Oncol 112:293-9, 2009

20.Reesink-Peters N, Wisman GB, Jeronimo C, et al: Detecting cervical cancer by quantitative promoter hypermethylation assay on cervical scrapings: a feasibility study. Mol Cancer Res 2:289-95, 2004

21.Nijhuis ER, Reesink-Peters N, Wisman GB, et al: An overview of innovative techniques to improve cervical cancer screening. Cell Oncol 28:233-46, 2006

22.Kang S, Kim JW, Kang GH, et al: Comparison of DNA hypermethylation patterns in different types of uterine cancer: cervical squamous cell carcinoma, cervical adenocarcinoma and endometrial adenocarcinoma. Int J Cancer 118:2168-71, 2006

23.van der Meide WF, Snellenberg S, Meijer CJ, et al: Promoter methylation analysis of WNT/beta-catenin signaling pathway regulators to detect adenocarcinoma or its precursor lesion of the cervix. Gynecol Oncol 123:116-22, 2011

24.Chang CC, Huang RL, Wang HC, et al: High methylation rate of LMX1A, NKX6-1, PAX1, PTPRR, SOX1, and ZNF582 genes in cervical adenocarcinoma. Int J Gynecol Cancer 24:201-9, 2014

25.Ongenaert M, Wisman GB, Volders HH, et al: Discovery of DNA methylation markers in cervical cancer using relaxation ranking. BMC Med Genomics 1:57, 2008

26.Weber M, Davies JJ, Wittig D, et al: Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37:853-62, 2005

27.Wilson IM, Davies JJ, Weber M, et al: Epigenomics: mapping the methylome. Cell Cycle 5:155-8, 2006

28.Hurd PJ, Nelson CJ: Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief Funct Genomic Proteomic 8:174-83, 2009

29.Huang RL, Gu F, Kirma NB, et al: Comprehensive methylome analysis of ovarian tumors reveals hedgehog signaling pathway regulators as prognostic DNA methylation biomarkers. Epigenetics 8:624-34, 2013

30.Huang TT, Gonzales CB, Gu F, et al: Epigenetic deregulation of the anaplastic lymphoma kinase gene modulates mesenchymal characteristics of oral squamous cell carcinomas. Carcinogenesis 34:1717-27, 2013

31.Carvalho RH, Haberle V, Hou J, et al: Genome-wide DNA methylation profiling of non-small cell lung carcinomas. Epigenetics Chromatin 5:9, 2012

32.van Dongen JJ, Langerak AW, Bruggemann M, et al: Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17:2257-317, 2003

33.Langmead B, Trapnell C, Pop M, et al: Ultrafast and memory-efficient alignment of short DNA

sequences to the human genome. Genome Biol 10:R25, 2009

34. De Meyer T, Mampaey E, Vlemmix M, et al: Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. PLoS One 8:e59068, 2013

35. Wisman GB, Nijhuis ER, Hoque MO, et al: Assessment of gene promoter hypermethylation for detection of cervical neoplasia. Int J Cancer 119:1908-14, 2006

36. Wang R, Wang G, Zhang N, et al: Clinical evaluation and cost-effectiveness analysis of serum tumor markers in lung cancer. Biomed Res Int 2013:195692, 2013

37. Youden WJ: Index for rating diagnostic tests. Cancer 3:32-5, 1950

38. Fluss R, Faraggi D, Reiser B: Estimation of the Youden Index and its associated cutoff point. Biom J 47:458-72, 2005

39. Koss LG: Cervical (Pap) smear. New directions. Cancer 71:1406-12, 1993

40. Zappa M, Visioli CB, Ciatto S, et al: Lower protection of cytological screening for adenocarcinomas and shorter protection for younger women: the results of a case-control study in Florence. Br J Cancer 90:1784-6, 2004

41. Smith HO, Tiffany MF, Qualls CR, et al: The rising incidence of adenocarcinoma relative to squamous cell carcinoma of the uterine cervix in the United States--a 24-year population-based study. Gynecol Oncol 78:97-105, 2000

42. Steenbergen RD, Snijders PJ, Heideman DA, et al: Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. Nat Rev Cancer 14:395-405, 2014

43. Jones PA, Laird PW: Cancer epigenetics comes of age. Nat Genet 21:163-7, 1999

44. Feinberg AP, Tycko B: The history of cancer epigenetics. Nat Rev Cancer 4:143-53, 2004

45. Baylin SB, Esteller M, Rountree MR, et al: Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet 10:687-92, 2001

46. Hanahan D, Weinberg RA: The hallmarks of cancer. Cell 100:57-70, 2000

47. Ehrlich M: DNA methylation in cancer: too much, but also too little. Oncogene 21:5400-13, 2002

48. Esteller M: Epigenetics in cancer. N Engl J Med 358:1148-59, 2008

49. Tsai HC, Baylin SB: Cancer epigenetics: linking basic biology to clinical medicine. Cell Res 21:502-17, 2011

50. Lacey JV, Jr., Frisch M, Brinton LA, et al: Associations between smoking and adenocarcinomas and squamous cell carcinomas of the uterine cervix (United States). Cancer Causes Control 12:153-61, 2001

51. Lacey JV, Jr., Swanson CA, Brinton LA, et al: Obesity as a potential risk factor for adenocarcinomas and squamous cell carcinomas of the uterine cervix. Cancer 98:814-21, 2003

52.Wild CP, Scalbert A, Herceg Z: Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. Environ Mol Mutagen 54:480-99, 2013

53.Robertson KD, Jones PA: DNA methylation: past, present and future directions. Carcinogenesis 21:461-7, 2000

54.Herman JG, Baylin SB: Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med 349:2042-54, 2003

55.Baylin SB, Herman JG: DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet 16:168-74, 2000

56.Lai HC, Lin YW, Huang TH, et al: Identification of novel DNA methylation markers in cervical cancer. Int J Cancer 123:161-7, 2008

57.Bock C, Tomazou EM, Brinkman AB, et al: Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28:1106-14, 2010

58.Malas S, Duthie S, Deloukas P, et al: The isolation and high-resolution chromosomal mapping of human SOX14 and SOX21; two members of the SOX gene family related to SOX1, SOX2, and SOX3. Mamm Genome 10:934-7, 1999

59.Lin YW, Tsao CM, Yu PN, et al: SOX1 suppresses cell growth and invasion in cervical cancer. Gynecol Oncol 131:174-81, 2013

60.Tsao CM, Yan MD, Shih YL, et al: SOX1 functions as a tumor suppressor by antagonizing the WNT/beta-catenin signaling pathway in hepatocellular carcinoma. Hepatology 56:2277-87, 2012

61.Yu W, Jin C, Lou X, et al: Global analysis of DNA methylation by Methyl-Capture sequencing reveals epigenetic control of cisplatin resistance in ovarian cancer cell. PLoS One 6:e29450, 2011

62.Li N, Li X, Li S, et al: Cisplatin-induced downregulation of SOX1 increases drug resistance by activating autophagy in non-small cell lung cancer cell. Biochem Biophys Res Commun 439:187-90, 2013

63.Senchenko VN, Kisseljova NP, Ivanova TA, et al: Novel tumor suppressor candidates on chromosome 3 revealed by NotI-microarrays in cervical cancer. Epigenetics 8, 2013

64.Lui VC, Samy ET, Sham MH, et al: Glial cell line-derived neurotrophic factor family receptors are abnormally expressed in aganglionic bowel of a subpopulation of patients with Hirschsprung's disease. Lab Invest 82:703-12, 2002

65.Sato T, Arai E, Kohno T, et al: DNA methylation profiles at precancerous stages associated with recurrence of lung adenocarcinoma. PLoS One 8:e59444, 2013

66.Yu J, Ma X, Cheung KF, et al: Epigenetic inactivation of T-box transcription factor 5, a novel tumor suppressor gene, is associated with colon cancer. Oncogene 29:6464-74, 2010

Table S1. Pyrosequence primers of the genes

Gene name	Forward Primers	Reverse Primers	Amplicons(bp)	Sequence primers
SLC6A5	ATGTAAATTGGGAAATTGGTTAGAA	AACACCCCCCAAACTCTTAACATCTTA	278	TGGGAAATTGGTTAGAAT
OLIG2	GGGAGATATATTTTAGAATTAGGGGTAA	ACTAAATTTCTCACACCCATACA	298	CAAATCTCCAATACTAACACTTAAC
CR753863.1	TTTGTTTTTAGGTTTTTGGGTAGTTAGT	CCCTACACTTACCAAAACTAAAACTTA	166	GATTTTGTGTAATGG
TBX5	GTTGTAAGGAAAAGGGGAAA	CTATAAAACTAAACTTACAAAATCCTTTAA	131	AGGAAAAGGGGAAAT
AL356961.2	TGTTTTGGGTTTGGTTGTG	ATTAATCCAACCCCAAATTTCCAAATA	200	TGGGTTTGGTTGTGT
AC096537.2 (218)	TGTTTAGGGGTAGGAGGTAGTT	ACCCACCACAAATACTAACAT	292	GGTGTTTTGGTTGGAG
AC096537.2 (219)	TGTGGTGGGTTGTATTTAGATATAA	TTCCTCCATAACCTTAATAATACCTAC	211	ATTTAGATATAATTGGAGATAAAGT
SOX14	GTTTAAGAATTTGTTTAAGAAGGATAGG	AAAACACTTCCCCCATCTTCTAAAT	224	GTTTAAGGAGGTTGGTTTG
EVX2	GGGGAAAGTAGAGTGGGTATT	CCTAAACCAAAATATCCCATTCAATC	300	CCAAACATAACAAAATTAAACATC
SYT6	GTTATAGGGGAGAGAGGGGGTGG	AACACCCTTAAAACTTCAATTTCC	386	GGAGAGATATTAGGATATAAAGAT
FREM3	TAGGGGTATTAGGAGTTTAGGGT	ACCTACCCCTCAACAAATCCCTCTTTA	296	GGTTTTAAGTGAGGATGT
TBX20	GGGGAAGTAGTGAATTTTATGGTTTTTAGTA	CTCCTCTCCCACTACCTTCCAAATCTCTAC	249	AAAAAATAAAAACCAATATACAAC
AC004963.1	TGGTTTTATTTTGGAAATATGTTAGTATAG	AACTCTAAAACTATTTCCTCCTACAC	164	AAATATGTTAGTATAGTTGGTT

Table S2. (Q)MSP primer and probe sequences of the genes used validation of the top15 genes

Genes	M/U	Forward primer 5'->3'	Reverse primer 5'->3'	Amplicon (bp)	6FAM /ZEN/ IBFQ Probe 5'->3'
SOX1_1	M	CGAATTTTAGTGTACGTCGC	AAAACGCTCCCTTAAACG	124	TCGAGAACGGATGTAGGGCGG
	U	TGGTGAATTTTAGTGTATGTTGT	AAAAAACACTCCCTTAAACAC	129	
GFRA1	M	TAGGGGGAATCGATGTTTC	GAATCCTAAACACCGAACGA	129	TTTATTCGTCGCGCGTTTTCGG
	U	AGTTAGGGGGAATTGATGTTTT	CCAAATCCTAAACACCAAACAAA	134	
SOX14	M	GTTCGTGGGGGTTTTCGAC	CAAAAAATAAAACGCCGAAACCG	85	TGAGCGCGTTCGAGAAAGTTCGGG
	U	GTTTGTTTGTGGGGGTTTTTGATG	TCCAACAAAAAATAAAACACCAAA ACCA	94	
SLC6A5	M	CGTGATTGTTTTTCGGGGAAGTCG	CAAACTCGAACTCAAAAAACGATC	113	CGTAGAGGGAGGTTTATATCGATTTTT GGGTTC
	U	TGTGATTGTTTTTTGGGGAAGTTGG	CAAACAAACTCAAACTCAAAAAAC AATC	117	
TBX20	M	G GGGTCGTTCGAATTGTC	TAACCCCGCTATAACTACTCTC	91	GATTTTTTAAGGGAGATAAAGATTCGA AATATAGTTT
	U	GGGATGGGGTTGTTTGAATTG	CTAACCCCACTATAACTACTCTCA	97	
TBX5	M	GGAAATTTCGAGCGTAGCG	CTACAAAACCCGCAATCCG	70	
	U	GGGGAAATTTTGAGTGTAGTG	AACTACAAAACCCACAATCC	74	
OLIG2	M	TAATAACGGAAGTAATTCGACG	CGCGCGAACACACACG	66	
	U	GTAAAAGTTTTTTAATAATGGAAGTAAT TTGATG	CTATAAACACACACACAAACACAC ACA	89	
AC004963.1	M	TTCGGTTTTCGCGTGC	TCCACTAATTTCTACGCTCG	68	
	U	GTTGGTTTTTGGTTTTTGTGTGT	TCCACTAATTTCTACACTCA	75	
AC096537.2(219)	M	TAGGGAGAAGTCGAGTTTACG	TTCTCGCCTTTCCCCCACC	81	
	U	GAGGTAGGGAGAAGTTGAGTTTATG	TTCTCACCTTTCCCCCACCAC	85	
CR753863.1	M	CGGTTGTAGTCGAGTTGGTTTTAGC	CGACAAAAACTCGACGTCCAAAAC	80	
	U	GTGGTTGTAGTTGAGTTGGTTTTAGTG	CAACAACAAAAACTCAACATCCAA AACAC	84	

TableS3. The top 15 markers Of the 53 identified candidate DMRs

Rank	Ensembl_gene_id	Gene_label	Gene_description	Count if normals ⩽1	Count if ADC⩾3	Count if SCC⩾3	Remark
1	ENSG00000089225	TBX5	T-box transcription factor TBX5 (T-box protein 5)	16	6	5	*
2	ENSG00000168875	SOX14	Transcription factor SOX-14 (Protein SOX-28)	15	5	6	#
3	ENSG00000229335	CR753863.1		19	5	5	
4	ENSG00000224243	AL356961.2		17	5	5	
5	ENSG00000233384	AC096537.2		16	5	5	
6	ENSG00000174279	EVX2	Homeobox even-skipped homolog protein 2 (EVX-2)	18	5	4	*
7	ENSG00000182968	SOX1	Transcription factor SOX-1	16	5	4	#
8	ENSG00000134207	SYT6	Synaptotagmin-6 (Synaptotagmin VI)(SytVI)	16	5	4	*
9	ENSG00000233384	AC096537.2		16	5	4	
10	ENSG00000183090	FREM3	FRAS1-related extracellular matrix protein 3 Precursor	16	5	4	
11	ENSG00000164532	TBX20	T-box transcription factor TBX20 (T-box protein 20)	16	5	4	*
12	ENSG00000228880	AC004963.1	cDNA FLJ39672 fis, clone SMINT2009233	16	5	4	
13	ENSG00000165970	SLC6A5	Sodium- and chloride-dependent glycine transporter 2 (GlyT2)(GlyT-2)(Solute carrier family 6 member 5)	16	5	4	
14	ENSG00000205927	OLIG2	Oligodendrocyte transcription factor 2 (Oligo2)(Class B basic helix-loop-helix protein 1)(bHLHB1)(Protein kinase C-binding protein RACK17)(Protein kinase C-binding protein 2)	20	5	3	*

15	ENSG00000151892	GFRA1	GDNF family receptor alpha-1 Precursor (GFR-alpha-1)(GDNF receptor alpha)(GDNFR-alpha)(TGF-beta-related neurotrophic factor receptor 1)(RET ligand 1)	19	5	3	*
16	ENSG00000132130	LHX1	LIM/homeobox protein Lhx1 (LIM homeobox protein 1)(Homeobox protein Lim-1)(hLim-1)	19	5	3	*
17	ENSG00000217236	SP9	Sp9 transcription factor homolog	19	4	5	
18	ENSG00000181195	PENK	Proenkephalin-A Precursor [Contains Synenkephalin;Met-enkephalin(Opioid growth factor)(OGF);PENK(114-133);PENK(143-183);Met-enkephalin-Arg-Gly-Leu; Leu-enkephalin;PENK(237-258);Met-enkephalin-Arg-Phe]	16	4	5	*
19	ENSG00000104435	STMN2	Stathmin-2 (Superior cervical ganglion-10 protein)(Protein SCG10)	16	4	5	
20	ENSG00000169856	ONECUT1	Hepatocyte nuclear factor 6 (HNF-6)(One cut domain family member 1)	19	4	4	
21	ENSG00000184937	WT1	Wilms tumor protein (WT33)	18	4	4	#
22	ENSG00000155269	GPR78	Probable G-protein coupled receptor 78	18	4	4	
23	ENSG00000159409	TNRC4	CUG-BP- and ETR-3-like factor 3 (CELF-3)(Bruno-like protein 1)(RNA-binding protein BRUNOL-1)(ELAV-type RNA-binding protein 1)(ETR-1)(Trinucleotide repeat-containing gene 4 protein)(Expanded repeat domain protein CAG/CTG 4)(CAG repeat protein 4)	18	4	4	
24	ENSG00000227757	AP000282.1		18	4	4	
25	ENSG00000169856	ONECUT1	Hepatocyte nuclear factor 6 (HNF-6)(One cut domain family member 1)	17	4	4	#
26	ENSG00000196109	ZNF676	Zinc finger protein 676	17	4	4	
27	ENSG00000215023	AC114730.3		17	4	4	
28	ENSG00000168621	GDNF	Glial cell line-derived neurotrophic factor Precursor (Astrocyte-derived trophic factor)(ATF)(hGDNF)	17	4	4	*

29	ENSG00000225554	AL359764.1		17	4	4	
30	ENSG00000150275	PCDH15	Protocadherin-15 Precursor	16	4	4	
31	ENSG00000187811	AP000867.1	Putative uncharacterized protein FLJ42569	16	4	4	
32	ENSG00000139915	MDGA2	MAM domain-containing glycosylphosphatidylinositol anchor protein 2 Precursor (MAM domain-containing protein 1)	16	4	4	*
33	ENSG00000237751	AC007040.7		16	4	4	
34	ENSG00000112562	SMOC2	SPARC-related modular calcium-binding protein 2 Precursor (Secreted modular calcium-binding protein 2)(SMOC-2)(Smooth muscle-associated protein 2)(SMAP-2)	16	4	4	
35	ENSG00000180053	NKX2-6	Homeobox protein Nkx-2.6 (Homeobox protein NK-2 homolog F)	16	4	4	*
36	ENSG00000186369	AL390816.1	UPF0730 protein CS0DE013YM09	20	4	3	
37	ENSG00000139910	NOVA1	RNA-binding protein Nova-1 (Neuro-oncological ventral antigen 1)(Onconeural ventral antigen 1)(Paraneoplastic Ri antigen)(Ventral neuron-specific protein 1)	19	4	3	
38	ENSG00000167614	TTYH1	Protein tweety homolog 1 (hTTY1)	19	4	3	
39	ENSG00000117600	RP4-788L13.1	Lipid phosphate phosphatase-related protein type 4 (EC 3.1.3.4)(Plasticity-related gene 1 protein)(PRG-1)(Brain-specific phosphatidic acid phosphatase-like protein 1)	19	4	3	
40	ENSG00000143340	FAM163A	UPF0417 protein FAM163A (Neuroblastoma-derived secretory protein)	19	4	3	
41	ENSG00000074964	ARHGEF10L	Rho guanine nucleotide exchange factor 10-like protein (GrinchGEF)	19	4	3	
42	ENSG00000230898	AL596087.1		19	4	3	
43	ENSG00000158258	CLSTN2	Calsyntenin-2 Precursor (Alcadein-gamma)(Alc-gamma)	19	4	3	
44	ENSG00000106853	PTGR1	Prostaglandin reductase 1 (PRG-1)(EC 1.3.1.-)(NADP-dependent leukotriene B4 12-hydroxydehydrogenase)(EC 1.3.1.74)(15-oxoprostaglandin 13-reductase)(EC 1.3.1.48)	19	3	5	

45	ENSG00000120937	NPPB	Natriuretic peptides B Precursor (Gamma-brain natriuretic peptide) [Contains Brain natriuretic peptide 32(BNP-32)(BNP(1-32));BNP(1-30);BNP(1-29);BNP(1-28);BNP(2-31);BNP(3-32);BNP(3-30);BNP(3-29);BNP(4-32);BNP(4-31);BNP(4-30);BNP(4-29);BNP(4-27);BNP(5-32);BNP(5-31);BNP(5-29)]	19	3	4	
46	ENSG00000179088	C12orf42	Uncharacterized protein C12orf42	20	3	3	
47	ENSG00000174059	CD34	Hematopoietic progenitor cell antigen CD34 Precursor (CD34 antigen)	20	3	3	*
48	ENSG00000070748	CHAT	Choline O-acetyltransferase (Choline acetylase)(CHOACTase)(ChAT)(EC 2.3.1.6)	20	3	3	*
49	ENSG00000197757	HOXC6	Homeobox protein Hox-C6 (Hox-3C)(HHO.C8)(CP25)	19	3	3	
50	ENSG00000217236	SP9	Sp9 transcription factor homolog	19	3	3	
51	ENSG00000081087	OSTM1	Osteopetrosis-associated transmembrane protein 1 Precursor	19	3	3	
52	ENSG00000169427	KCNK9	Potassium channel subfamily K member 9 (Acid-sensitive potassium channel protein TASK-3)(TWIK-related acid-sensitive K(+) channel 3)(Two pore potassium channel KT3.2)	19	3	3	*
53	ENSG00000188620	HMX3	Homeobox protein HMX3 (Homeobox protein H6 family member 3)(Homeobox protein Nkx-5.1)	19	3	3	

Sorted by the methylation reads in ADC from large to small

Sorted by the methylation reads in SCC from large to small

Sorted by the unmethylation reads in normal from large to small

Since GFRA1 are overlap the identification for our CIN2/3 marker selection, therefore it is included.

* Methylation related to cancer

Methylation related to cervical cancer