RESEARCH ARTICLE

Editorial Process: Submission:06/15/2022 Acceptance:10/24/2022

Identification of Malignancy in PAP Smear Samples Using the *CGB3* and *NOP56* Genes as Methylation Markers

Palak Singh[1], Nakarin Kitkumthorn[2], Pattamawadee Yanatatsaneejit[3]*

Abstract

Background: Although various improvements have been made in the reporting of the Papanicolaou (PAP) test in recent years, there remain several challenges that have yet to be addressed in terms of determining a standardized methodology for categorizing atypical squamous cells of undetermined significance (ASC US). **Methods:** The present study focuses on evaluating the performance of the methylation status of two genes (*CGB3* and *NOP56*) using a total of 200 PAP samples, which were divided into the "determined" group, with 78 samples based on cytology, and the "undetermined" group (ASC US), with 122 samples. The promoter methylation status of the *CGB3* and *NOP56* genes was detected for the 200 PAP samples using methylation specific PCR (MSP). The diagnostic abilities of the *CGB3* and *NOP56* genes in PAP samples were measured, and receiver operating characteristic (ROC) curves were generated using Python programming language. **Results:** Based on the validation of *CGB3* and *NOP56* methylation in the 200 PAP samples, both genes exhibited higher methylation percentages in abnormal samples compared with normal samples. In addition, on the basis of diagnostic performance analysis, the *CGB3* gene exhibited the highest sensitivity and specificity in both histology based ASC US and cytology based 'determined' PAP samples, with significant diagnostic abilities [area under the curve (AUC) values of 0.83 and 0.74, respectively, where AUC ≥0.5 was determined to be significant] to distinguish between the "normal" and "abnormal" samples. **Conclusion:** The findings of the present study will contribute toward identifying a DNA methylation marker for the early detection of abnormal samples before they reach the initial stages of cervical cancer, and should prove to be helpful for clinicians in terms of diagnosing patients whose cells are ASC US.

Keywords: promoter methylation- ASC-US- chorionic gonadotropin subunit 3- nucleolar protein 56- diagnostic value

Asian Pac J Cancer Prev, **23 (10)**, 3541-3551

Introduction

Cervical cancer is one of the most frequently occurring types of cancer in women, with an estimate of >600,000 cases reported in 2020, where in excess of 90% of those cases were shown to occur among countries with lower level and middle level incomes (Sung et al., 2021). Specifically, in Asia there were in excess of 315,000 reported cases, and >168,000 reported deaths, due to cervical cancer (Bray et al. 2018). Focusing on Thailand, it was shown that cervical cancer is the second most common cause of cancer associated death (11.7%) after breast cancer (21.8%) amongst women. Furthermore, the number of cervical cancers associated deaths has been steadily increasing in the recent past (Imsamran et al., 2018).

In terms of the national perspective, a lack of screening and defective treatment of abnormal cells during the pre cancerous stages have been identified as major clinical deficiencies associated with cervical cancer, even though these could both easily be prevented (Bray et al., 2018; Arbyn et al., 2013). The low diagnostic accuracy of PAP tests makes it difficult to differentiate benign from malignant lesions due to post radiation cellular changes (Desai et al., 2021). Pre cancerous lesions caused by human papilloma virus (HPV) may be detected using early screening, which helps to prevent cancers from developing further (Pimple and Mishra, 2019). The Papanicolaou test (PAP smear), visual inspection of the cervix with acetic acid, and HPV testing are a few of the screening tests that may be used for cervical cancer (Ngo-Metzger and Adsul 2019; Burd, 2003). The "Bethesda System" guidelines 2001 (Solomon et al., 2002) involving colposcopy and biopsy can also be used for the further examination of women with cytology proven severe lesions (Koh et al., 2019; Wright Jr et al., 2002). However, there are opportunities for further explorations, especially when it comes to the management of treatment for women with atypical squamous cells of undetermined significance

*[1]Program in Biotechnology, Faculty of Science, Chulalongkorn University, Bangkok 10330 Thailand. [2]Department of Oral Biology, Faculty of Dentistry, Mahidol University, Bangkok 10400 Thailand. [3]Human Genetics Research Group, Department of Botany, Faculty of Science, Chulalongkorn University, Center of Excellence in Molecular Genetics of Cancer and Human Diseases, Chulalongkorn University, Bangkok 10330, Thailand. *For Correspondence: pattamawadee.y@chula.ac.th*

(ASC US) where additional explanations are needed (Lahue et al., 2015), and this points towards further analyses in terms of differentiating between a negative result and a confirmed squamous intraepithelial lesion (SIL) (Wright et al., 2015). Based on further analyses of the ASC US results, the SIL results may also be categorized into low SIL (LSIL) and high SIL (HSIL) (Iavazzo et al., 2012).

HPV infection alone, however, as a causative agent does not fully explain the progression of cervical cancer, and additional genetic and epigenetic modifications in the host genome may serve critical roles in the oncogenic process (Soto et al., 2017; Steenbergen et al., 2014). Additionally, distinct patterns of specific promoter hypermethylation were found to be consistent with the underlying mechanism of HPV E6 and E7 oncoprotein induced DNA methyltransferase activity following gene silencing (Sen et al., 2018). Here, silencing of tumor suppressor (TS) genes by DNA methylation of promoter regions is a significant predictor of progressive oncogenesis. High risk (HR) HPV (involving HPV types 16 and 18) is induced by two majors viral oncoproteins, E6 and E7 (Burgers et al., 2007). In addition, promoter methylation of CCNA1, acting as a TS gene, which results in its silencing, may be induced by HPV E6 and E7 oncoproteins (Chalertpet et al., 2015). The gene selection criteria for the present study were based on previous work (Singh et al., 2022), which determined that the *CGB3* and *NOP56* promoter sequences share the same sequence as that of CCNA1, regulated by HPV16 E7. That study demonstrated silencing of the *CGB3* and *NOP56* genes by DNA methylation of promoter regions in cervical carcinogenesis; therefore, DNA methylation of the *CGB3* and *NOP56* genes may provide a credible target for the analysis of clinical samples.

The 'U' in ASC US stands for 'undetermined' significance, which indicates that women with the cytology of ASC US may have negative results, or that they may have low or high grade SIL, which immediately needs colposcopy test for diagnosing the ASC-US cytology to increase the detection rate of high-grade cervical lesions or invasive cancer. (Abdulaziz et al., 2020). Followed up by cytological analysis, abnormal ASC US cells are further examined by clinicians to determine which stage the cells are at by performing a colposcopy or biopsy. However, by using the *CGB3* and *NOP56* genes as methylation markers, it would be possible for us to identify the stage of the ASC US samples without the patient being required to go through the painful biopsy process. Therefore, the aim of the present study was to identify novel methylation markers to detect the malignancy of PAP samples at the early cancer stages. On the basis of the study of the ALTS group (Group, 2003) from the U.S.A., cervical screening programs have shown that HR HPV is responsible for ~50% of the cases of women with ASCs. Therefore, based on HPV typing, ASC US samples were categorized into the HPV types HPV16 and HPV18, 'other HPV' types (see below for further details) and no HPV categories using a Cobas® 4800 genotyping assay with specific identification of the HPV16 and HPV18 types (Rao et al., 2013). To identify malignancy in the collected liquid based PAP test samples of ASC US, these were classified into four categories, namely LSIL, HSIL and normal samples (with or without HPV). Note that there are no cancer samples in the ASC US, as these are supposed to contain only normal or pre cancerous samples by definition. By using methylation specific PCR (MSP), the *CGB3* and *NOP56* genes were used to differentiate between the normal and abnormal samples through detecting promoter methylation in both types of samples. The findings of the present study should help to evaluate PAP samples at an early stage, and this will be of importance in terms of the clinical diagnosis of ASC US samples.

Materials and Methods

The present study was conducted to evaluate the performance of methylation of the *CGB3* and *NOP56* genes in PAP samples and in different types of HPV to distinguish between the normal and abnormal samples. The diagnostic abilities of both the genes were also measured with respect to a total of 200 PAP samples.

Collection of clinical samples

The total of 200 PAP test samples was identified from the pathology archive of the Chulalongkorn Medical Hospital, and from the National Research Institute of Cancer (NCI), Thailand. HPV typing had already been performed using the Cobas® 4800 HPV genotyping assay that targets 14 different types of HPV genotypes, including HPV types 16 and 18 with specific primers of identification, and "other" types of HPV (consisting of 12 HR HPV types, namely 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68), referred to subsequently as "other HPV" (Heideman et al., 2011). The 200 PAP samples were divided into two groups, i.e., 'determined' and 'undetermined' PAP samples. Based on cytology (microscopic observation), 78 samples were categorized as 'determined' PAP samples, whereas the remaining 122 samples were categorized as 'undetermined'. The stage of each of the ASC US samples was categorized into the LSIL, HSIL and normal (with and without HPV) stages based on histology, followed up by biopsy. All the samples were approved by the ethical committee (IRB no. 477/61) of the Faculty of Medicine, Chulalongkorn University.

Detection of the methylation of the *CGB3* and *NOP56* genes' promoters in 200 PAP samples by MSP PCR extraction of DNA from the clinical samples.

DNA extraction from the total of 200 individual PAP samples was performed using 10% sodium dodecyl sulfate (SDS) (Sigma Aldrich; Merck KGaA), and subsequently lysis buffer II (0.75 M NaCl/0.024 M EDTA, pH 8.0) was added, together with 20 mg/ml proteinase K (Invitrogen; Thermo Fisher Scientific, Inc.) for digestion of the cells. The incubation process was performed in a water bath at 50°C overnight. The next day, phenol:choloroform:isoamyl alcohol in a 25:24:1 ratio was added to the same samples, and the upper layer was carefully separated after the centrifugation step. In the next step, 100% ethanol was added into the separated part of the upper layer of the samples for precipitation. The DNA pellet was washed with 75% ethanol, and then air dried by inverting the

tube at room temperature. Subsequently, the pellet was resuspended by adding distilled H2O (Gilbert and Vance, 1998). The DNA concentration was determined using a Nanodrop™ 2000c spectrophotometer (Thermo Fisher Scientific, Inc.). Genomic DNA from SiHa cells was also extracted and used as a positive methylated control (MC), whereas for the negative control, C33A genomic DNA was extracted and used as a negative unmethylated control (UC).

Sodium bisulfite treatment and MSP

Aliquots (750 ng) of the extracted DNA for each sample were subjected to bisulfite treatment using an EZ DNA Methylation Gold kit (Zymo Research Corp.) according to the manufacturer's protocol. The eluted DNA of each of the samples obtained from the bisulfite kit was used to perform the detection of methylation using methylated and unmethylated specific primers of *CGB3* and *NOP56*. The thermocycling conditions for the *CGB3* gene were as follows: an initial incubation at 95°C for 15 min, 27 cycles of 95°C for 45 sec, 54°C for 45 sec, 72°C for 45 sec, and a final incubation at 72°C for 7 min (116 bp in size), and the sequences of the primers were as follows [forward (FW) primer, CGGGTTGAATTTTTCGTTGGC; and reverse (RV) primer, CCCAAAAAAAACGCGACTTCG]. The thermocycling conditions for the *NOP56* gene were as follows: an initial incubation at 95°C for 15 min, 27 cycles of 95°C for 45 sec, 38°C for 45 sec, 72°C for 45 sec, and a final incubation at 72°C for 7 min (116 bp in size) (FW primer, TATTTTTTATTATATTTTGGAATC; RV primer, ATTAAATTATTTTAACCGTCG). Subsequently, 10 µl of the PCR product was loaded onto an 8% acrylamide gel for gel electrophoresis. The gels were stained with SYBR reagent (Lonza Group, Ltd.), and visualization of the methylated and unmethylated band intensities in each sample was achieved using a Storm 840 (Cytiva). The positive controls for methylation and unmethylation were contained in an EpiTect DNA set (Qiagen GmbH).

Statistical analysis

The categorical data sets of abnormal or pre cancerous samples were compared with the normal samples, and methylation in the various HPV types was analyzed using GraphPad Prism software, version 5 (GraphPad Software, Inc.) χ^2 tests, using the two tailed test, were used for statistical analysis, where P≤0.05 was considered to indicate a statistically significant value. The diagnostic abilities of the *CGB3* and *NOP56* genes in the PAP samples were measured, and receiver operating characteristic (ROC) curves were generated using Python programing language (version 3.8) and the Jupyter Notebook Integrated Development Environment (IDE) (Kluyver et al. 2016; McKinney 2010). The area under the curve (AUC) values were subsequently calculated (AUC values ≥0.5 were considered to be significant).

Results

Classification of the 200 PAP samples

The total number of samples on which the experiments were performed was 200, amongst which, 122 were ASC US samples based on histology, and the remaining 78 samples were 'determined' samples, based on cytology. Furthermore, the 78 'determined' PAP samples were categorized as cancerous samples (20/78 samples), HSIL (24/78 samples), LSIL (12/78 samples), and normal samples (with HPV, 8/78 samples; and without HPV, 14/78 samples).

On the basis of biopsy, the histology based (ASC US) samples were further categorized as HSIL (15/122 samples), LSIL (33/122 samples) and normal (with HPV, 66/122 samples; and without HPV, 8/122 samples).

Distribution of CGB3 and NOP56 genes' promoter methylation percentages in PAP samples

To determine the presence of *CGB3* and *NOP56* gene methylation in PAP samples, MSP PCR was performed using the specific primers of the genes. On the basis of these experiments, methylation in the promoter region of the *CGB3* and *NOP56* genes can be used to differentiate between normal cervical cells and abnormal cervical cells. Considering the ASC US PAP samples specifically, for the *CGB3* gene, 43 abnormal samples exhibited promoter methylation out of the total number of 48 samples (89.5%), whereas, out of the total of 74 normal samples, 17 samples (22%) exhibited *CGB3* gene promoter methylation, with a prediction of close to 80% in terms of the accuracy of the results.

The 48 abnormal ASC US PAP samples were further broken down into HSIL (which showed 80% of *CGB3* promoter methylation from a total sample size of 15 samples) and LSIL (which showed 93% of *CGB3* methylation from a total sample size of 33 samples). For the remaining 74 normal samples, 25.7% of the normal

Table 1A. Statistically Significant Difference of *CGB3* in ASC-US Samples, when stages of Normal compared with abnormal with level of [LSIL, HSIL and (LSIL+HSIL)]

Stage compared with normal (without HPV)	Gene	Chi-Square p-value
LSIL	*CGB3*	0.0001 (***)
HSIL	*CGB3*	0.0003 (***)
LSIL+HSIL	*CGB3*	0.0001 (***)
Stage compared with normal (with and without HPV)	Gene	Chi-Square p-value
LSIL	*CGB3*	0.0001 (***)
HSIL	*CGB3*	0.0001 (***)
LSIL+HSIL	*CGB3*	0.0001 (***)

CGB3, chorionic gonadotropin subunit 3; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion.

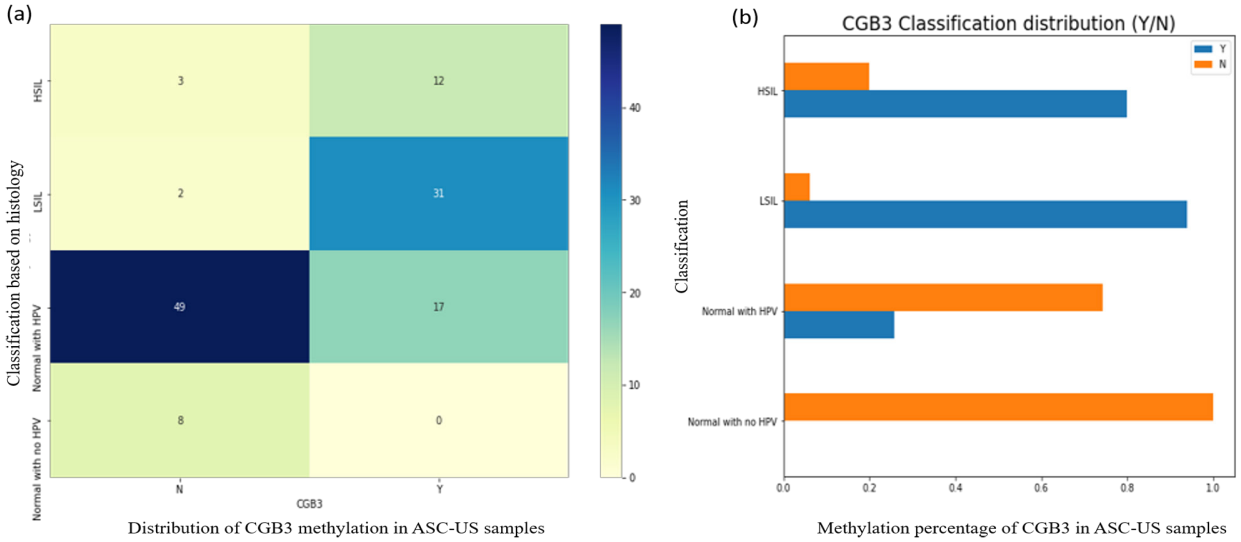

Figure 1. Classification and Distribution of Presence (Y) and absence (N) of Methylation Percentage of *CGB3* Gene in ASC-US Samples. (a) CGB3 promoter methylation distribution in HSIL samples (N-3, Y-12), LSIL samples (N-2, Y-31), Normal with HPV samples (N-49, Y-17, and Normal with no HPV samples (N-8, Y-0). (b) Presence of *CGB3* methylation percentage in HSIL (80%), LSIL (93%), Normal with HPV (25%) and Normal with no HPV (0%). *CGB3,* chorionic gonadotropin subunit 3; ASC-US, Atypical squamous cells of undetermined significance; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion.

Table 1B. Statistically Significant Difference of *NOP56* in ASC-US Samples, when stages of Normal compared with abnormal with level of [LSIL, HSIL and (LSIL+HSIL)]

Stage compared with normal (without HPV)	Gene	Chi-Square p-value
LSIL	*NOP56*	0.0009 (***)
HSIL	*NOP56*	0.0038 (**)
LSIL+HSIL	*NOP56*	0.0005 (***)
Stage compared with normal (with and without HPV)	Gene	Chi-Square p-value
LSIL	*NOP56*	0.0001 (***)
HSIL	*NOP56*	0.0031 (**)
LSIL+HSIL	*NOP56*	0.0001 (***)

NOP56, nucleolar protein 56; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion

samples with HPV showed *CGB3* methylation from a total sample size of 66, whereas the normal samples without HPV showed 0% *CGB3* methylation from a total sample size of 8, as shown in Figure 1. The 25% methylation level in normal samples with HPV may have been due to the presence of HPV types which could have caused the methylation.

Similarly, regarding the results with the *NOP56* gene, out of the total of 48 abnormal samples, 36 of them (75%) showed *NOP56* gene methylation, whereas out of the total 74 normal samples, 24 (32%) samples showed *NOP56* gene methylation, with a prediction of close to

Table 2A. Statistically Significant Difference of *CGB3* in Determined PAP Samples, when stages of Normal compared with pre-cancerous with level of [LSIL, HSIL and (LSIL+HSIL)]

Stage compared with normal (without HPV)	Gene	Chi-Square p-value
LSIL	*CGB3*	0.001 (**)
HSIL	*CGB3*	0.0001 (***)
Cancer	*CGB3*	0.0007 (***)
LSIL+HSIL+CANCER	*CGB3*	0.0001(***)
Stage compared with normal (with and without HPV)	Gene	Chi-Square p-value
LSIL	*CGB3*	0.008 (**)
HSIL	*CGB3*	0.0001 (**)
Cancer	*CGB3*	0.004 (**)
LSIL+HSIL+CANCER	*CGB3*	0.0001 (***)

CGB3, chorionic gonadotropin subunit 3; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion.

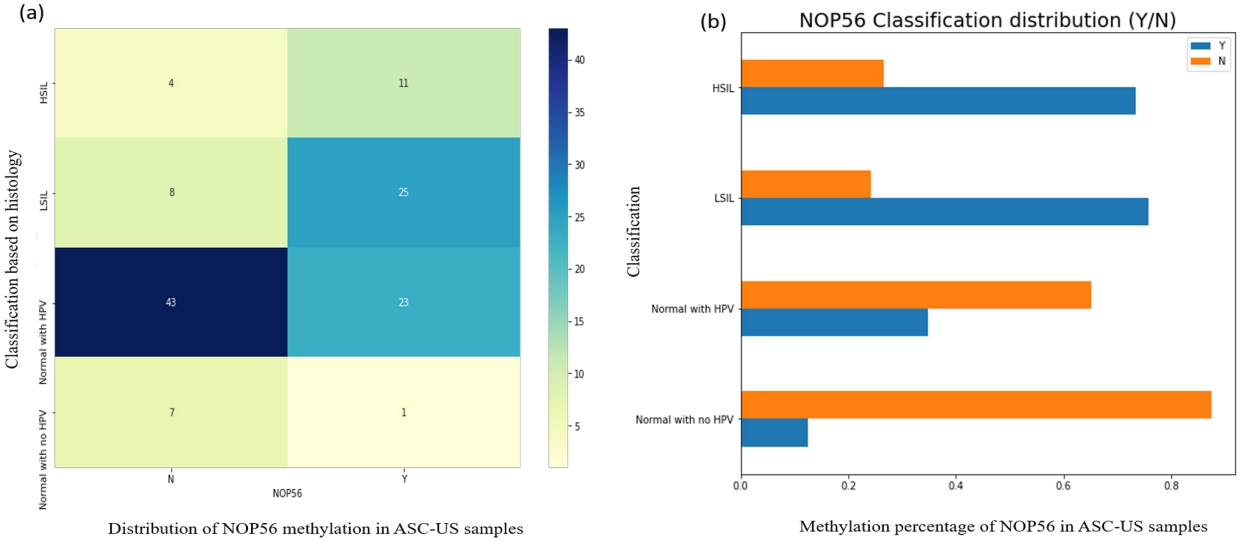

Figure 2. Classification and Distribution of Presence (Y) and Absence (N) of Methylation Percentage of *NOP56* in ASC-US samples. (a) *NOP56* promoter methylation distribution in HSIL samples (N-4, Y-11), LSIL samples (N-8, Y-25), Normal with HPV samples (N-43, Y-23, and Normal with no HPV samples (N-7, Y-1). (b) Presence of *NOP56* methylation percentage in HSIL (73%), LSIL (75%), Normal with HPV (34%) and Normal with no HPV (12%). *NOP56*, nucleolar protein 56; ASC-US, Atypical squamous cells of undetermined significance; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion.

70% accuracy in terms of the results.

The 48 abnormal ASC US PAP samples were further broken down into HSIL (where 73% of the samples exhibited *NOP56* promoter methylation from a total sample size of 15 samples) and LSIL (where 76% of the samples showed *NOP56* methylation from a total sample size of 33 samples). For the remaining 74 normal samples, the normal samples with HPV exhibited a 35% level of *NOP56* methylation (from a total sample size of 66), and the normal samples without HPV showed 12% *NOP56* methylation from a total sample of 8, as shown in Figure 2. The 35% methylation level identified in normal samples with HPV could have been due to the presence of HPV types, which may have led to the methylation. Taken together, these results indicated that the abnormal samples (LSIL+HSIL) exhibited a higher percentage level of *CGB3* and *NOP56* methylation in the ASC US PAP samples compared with the normal samples.

In parallel with these experiments, the effects of the presence or absence of *CGB3* and *NOP56* gene promoter methylation were also analyzed in cytology based 'determined' PAP samples categorized as LSIL, HSIL,

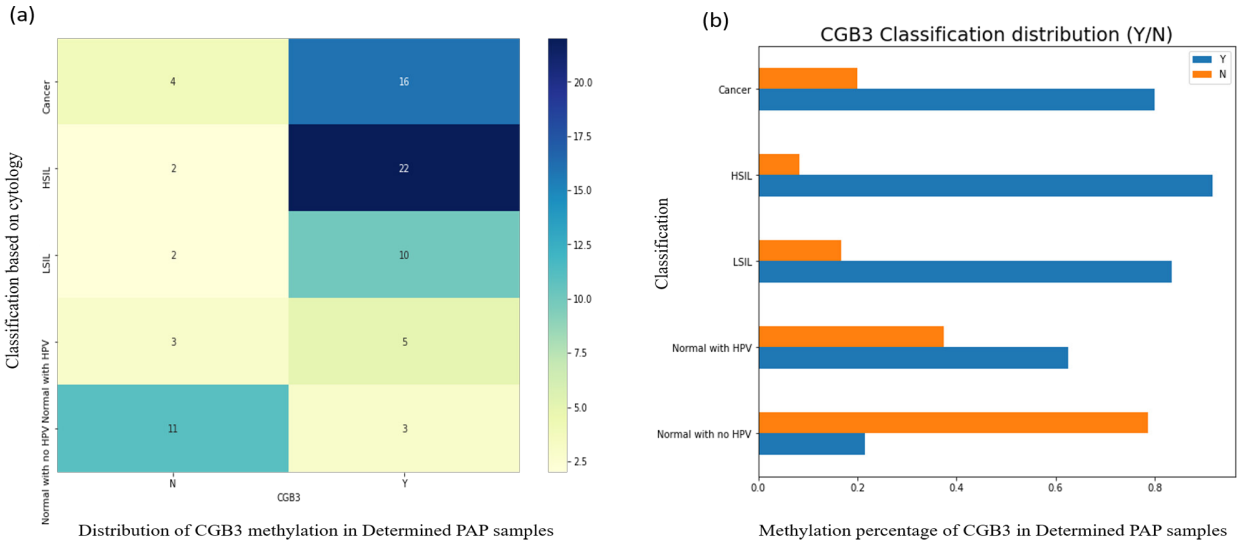

Figure 3. Classification and Distribution of Presence (Y) and absence (N) of methylation percentage of *CGB3* in Determined PAP samples. (a) *CGB3* promoter methylation distribution in Cancer samples (N-4, Y-16), HSIL samples (N-2, Y-22), LSIL samples (N-2, Y-10), Normal with HPV samples (N-3, Y-5, and Normal with no HPV samples (N-11, Y-3). (b) Presence of CGB3 methylation percentage in Cancer samples (80%), HSIL (92%), LSIL (83.3%), Normal with HPV (62.5%) and Normal with no HPV (21%). *CGB3,* chorionic gonadotropin subunit 3; PAP, Papanicolaou test; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion.

Table 2B. Statistically Significant Difference of *NOP56* in Determined PAP Samples, when stages of Normal compared with pre-cancerous with level of [LSIL, HSIL and (LSIL+HSIL)]

Stage compared with normal (without HPV)	Gene	Chi-Square p-value
LSIL	*NOP56*	0.006(**)
HSIL	*NOP56*	0.01(*)
Cancer	*NOP56*	0.005 (**)
LSIL+HSIL+CANCER	*NOP56*	0.001 (**)
Stage compared with normal (with and without HPV)	**Gene**	**Chi-Square p-value**
LSIL	*NOP56*	0.007 (**)
HSIL	*NOP56*	0.01 (*)
Cancer	*NOP56*	0.005 (**)
LSIL+HSIL+CANCER	*NOP56*	0.001(**)

NOP56, nucleolar protein 56; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion

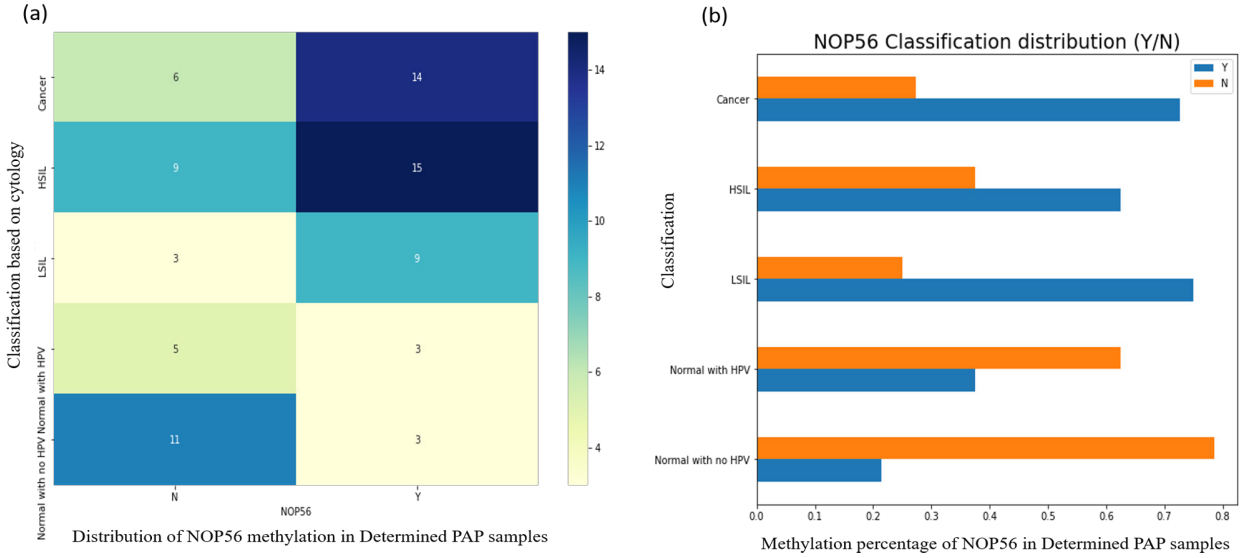

Figure 4. Classification and Distribution of Presence (Y) and absence (N) of methylation percentage of *NOP56* in Determined PAP samples. (a) *NOP56* promoter methylation distribution in Cancer samples (N-6, Y-14), HSIL samples (N-9, Y-15), LSIL samples (N-3, Y-9), Normal with HPV samples (N-5, Y-3, and Normal with no HPV samples (N-11, Y-3). (b) Presence of *NOP56* methylation percentage in Cancer samples (70%), HSIL (62.5%), LSIL (67%), Normal with HPV (37.5%) and Normal with no HPV (21%). *NOP56,* nucleolar protein 56; PAP, Papanicolaou test; LSIL, Low-grade squamous intraepithelial lesion; HSIL, High-grade squamous intraepithelial lesion.

cancer and normal (with or without HPV) samples by using the same MSP technique. According to the results of *CGB3* gene methylation, for the abnormal samples, 48 samples exhibited *CGB3* gene methylation out of the total size of 56 samples (85%), whereas for the normal samples, out of a total of 22 normal samples, 8 samples (36.3%) showed *CGB3* gene methylation, with a prediction of close to 65% in terms of the accuracy of the results.

The 56 abnormal 'determined' PAP samples were further broken down into cancer (which showed 80% *CGB3* promoter methylation from a total size of 20 samples), HSIL (which exhibited 92% *CGB3* promoter methylation from a total sample size of 24 samples) and LSIL (which showed 83.3% *CGB3* promoter methylation from a total sample size of 12 samples). For the remaining 22 normal samples, the normal samples with HPV showed a 62.5% level of *CGB3* promoter methylation from a total sample size of 8, whereas the normal samples without HPV showed 21% *CGB3* promoter methylation from a total sample of 14, as shown in Figure 3.

Similarly, based on the results of *NOP56* methylation, we found that, out of the total number of 56 abnormal

Table 3A. Screening Test Efficiency in Term of Sensitivity, Specificity, PPV and NPV of *CGB3* and *NOP56* Genes in Histology-based ASC-US PAP Samples.

Genes	Histology based PAP Samples	Sensitivity (%)	Specificity (%)	Positive predictive value (PPV%)	Negative predictive value (NPV%)	Accuracy of methylation marker (%)
CGB3	ASC-US PAP SAMPLES	90	77	72	92	82
NOP56	ASC-US PAP SAMPLES	73	68	60	79	70

CGB3, chorionic gonadotropin subunit 3; NOP56, nucleolar protein 56; ASC-US, atypical squamous cells of undetermined significance, PPV, positive predictive test, NPV, negative predictive test

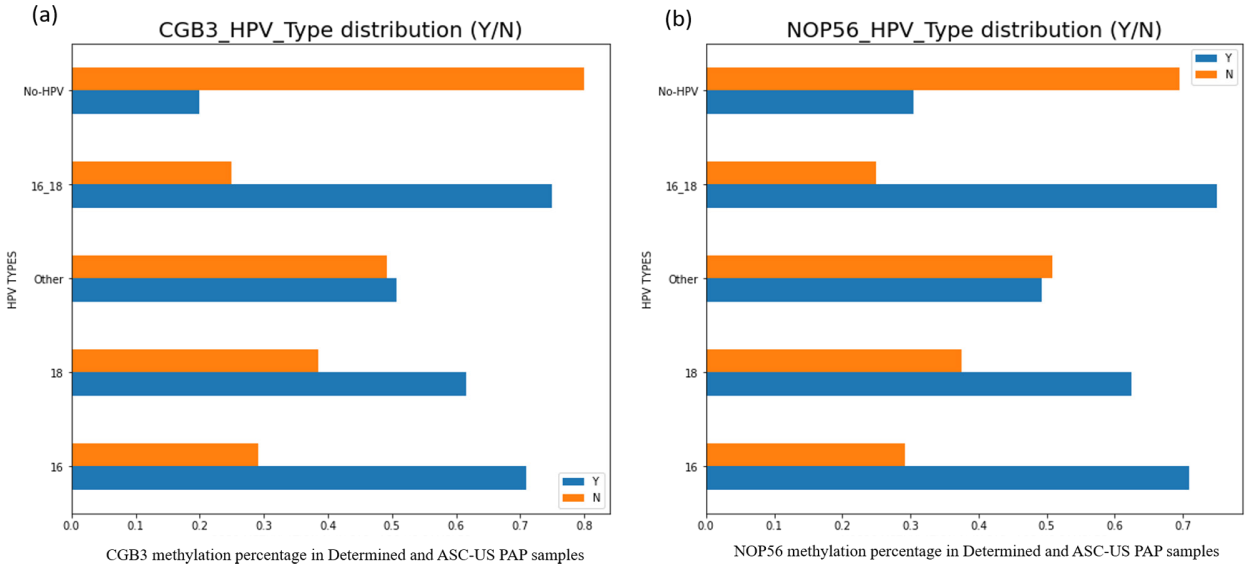

Figure 5. Distribution of Methylation Percentage of *CGB3* and *NOP56* Genes on the basis of of HPV types present in the samples. (a) Presence of *CGB3* methylation percentage in No-HPV types (20%), HPV16_18 type (75%), Other HPV type (50%), HPV 18 (61%) and HPV 16 (70%). (b) Presence of *NOP56* methylation percentage in No-HPV types (30%), HPV16_18 type (75%), Other HPV type (49%), HPV 18 (62%) and HPV 16 (70%). *CGB3, chorionic gonadotropin subunit* 3; NOP56, nucleolar protein 56; HPV, Human papillomavirus; Atypical squamous cells of undetermined significance; PAP, Papanicolaou test

Table 3B. Screening Test Efficiency in Term of Sensitivity, Specificity, PPV and NPV of *CGB3* and *NOP56* genes in cytology-based determined PAP samples.

Genes	Cytology based PAP Samples	Sensitivity (%)	Specificity (%)	Positive predictive value (PPV%)	Negative predictive value (NPV%)	Accuracy of methylation marker(%)
CGB3	Determined PAP SAMPLES	86	64	86	64	80
NOP56	Determined PAP SAMPLES	68	73	86	47	70

CGB3, chorionic gonadotropin subunit 3; *NOP56*, nucleolar protein 56; ASC-US, atypical squamous cells of undetermined significance, PPV, positive predictive test, NPV, negative predictive test

samples, 38 (68%) samples showed *NOP56* gene methylation, whereas out of the total of 22 normal samples, 6 samples (27.2%) showed *NOP56* gene with methylation, with a prediction of close to 73% in terms of the accuracy of the results.

The 56 abnormal 'determined' PAP samples were further broken down into cancer (which showed 70% samples with *NOP56* promoter methylation from a total size of 20 samples), HSIL (which exhibited 62.5% *NOP56* promoter methylation from a total sample size of 24 samples) and LSIL (which showed 67% *NOP56* promoter methylation from a total sample size of 12 samples). For the remaining 22 normal samples, the normal samples with HPV showed a 37.5% level of *NOP56* gene methylation from a total sample size of 8, whereas the normal samples without HPV showed a 21% level of *NOP56* gene methylation from a total sample of 14, as shown in Figure 4.

The distribution of the results of the methylation of the *CGB3* and *NOP56* genes shows that there was a higher level of methylation in the abnormal samples both for the histology based ASC US samples and for the cytology based 'determined' PAP samples. On the other hand, a lower level of methylation was detected in the normal samples for both histology and cytology based samples (with and without HPV), indicating the effectiveness of the *CGB3* and *NOP56* genes as methylation markers for the early detection of cervical cancer.

HPV typing and methylation distribution of CGB3 and NOP56 in PAP samples

The data were further analyzed according to the methylation levels shown for selected HPV strains that were known to be causal agents of cervical cancer (Burd 2003) and also proposed as an alternative screening test for cervical cancer (Koliopoulos et al. 2017). All the samples based on the HPV strains were classified into no HPV, HPV16, HPV18, combined HPV (containing both HPV16 and HPV18) and other HPV types [i.e., HPV genotypes (HR 12) including HPV 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68]. In this study, the different types of HPVs were analyzed using the 200 PAP samples for both the ASC US and 'determined' PAP sample groups.

There was a total of 31 ASC US samples with the HPV16 type, a total of 11 samples had the HPV18 type, 4 samples had the combined HPV type, 65 samples had 'other HPV' types, and the sample size of the 'no HPV' group was 11. Similarly, as far as the classification of the other remaining 'determined' PAP samples, which included the LSIL, HSIL, cancer and normal (with or without HPV) sample groups, was concerned, there were 49 samples with HPV16, 15 samples with HPV18, and

the no HPV sample group size was 14.

The distribution of the methylation percentages of the *CGB3* and *NOP56* genes in all the 200 PAP samples were subsequently calculated in the different types of HPV strains. Based on the results, significant methylation differences in the *CGB3* and *NOP56* genes were observed in the distribution of HPV genotyping, with P values of 0.0003 and 0.019, respectively. The methylation percentages of *CGB3* were distributed as follows: the HPV16 strain showed 70% methylation, the HPV18 strain showed 61% methylation, the combined HPV (HPV16/18) strain showed 75% methylation, the 'other HPV strains' group showed 50% methylation, and the no HPV group showed 20% methylation, as shown in Fig. 5A. Similarly, for the *NOP56* gene, the HPV16 strain showed 70% methylation, the HPV18 strain showed 62% methylation, the combined HPV strain showed 75% methylation, the 'other HPV strains' group showed 49% methylation, and the no HPV strain showed 30% methylation, as shown in Fig. 5B. Based on the collective results in Fig. 13A and 13B, it may be observed that the methylation percentages of samples containing either or both the HPV16 and HPV18 strains were relatively higher compared with those lacking the HPV16 and HPV18 strains.

Examining the statistically significant differences in *CGB3* and *NOP56* methylation in the histology based ASC US PAP samples (122/200 samples).

A statistical comparison of methylation in the promoter regions of the *CGB3* and *NOP56* genes in the ASC US samples was then made, and differences between the stages of normal cervical cells and abnormal cells were then calculated. Our experiments showed that, when comparisons between different combinations of normal samples (with and without HPV) and pre cancerous samples ("LSIL", "HSIL" and "LSIL+HSIL") were made separately for the *CGB3* and *NOP56* genes, it was possible to distinguish between the normal samples and the pre cancerous samples statistically (with P values ≤0.05), as shown in Table 1A and 1B, respectively.

Examining statistically significant differences of *CGB3* and *NOP56* methylation in cytology based 'determined' PAP samples (78/200 samples).

Similarly, to the ASC US samples with reference to histology, a statistical comparison of methylation in the promoter regions of the *CGB3* and *NOP56* genes in the cytology based 'determined' PAP samples was then made, and differences between the stages of normal cervical cells and pre cancerous cells were calculated. Our experiments showed that, when comparisons between different combinations of normal samples (with and without HPV) and pre cancerous samples ("LSIL", "HSIL" and "LSIL+HSIL") were made separately for the *CGB3* and *NOP56* genes, it was also possible to distinguish between the normal samples and the pre cancerous samples statistically (with P values ≤0.05), as shown in Table 2A and 2B, respectively.

Screening the test efficiencies and diagnostic abilities of the *CGB3* and *NOP56* methylation markers in histology based ASC US PAP samples (122/200 samples).

The effectiveness of methylation in the promoter regions of the *CGB3* and *NOP56* genes to be used to isolate and screen cervical cells was then assessed according to measurement methods. The performance calculations were made between "true label" × "predicted label", where "true positives (TP)", "false positives (FP)", "true negatives (TN)" and "false negatives (FN)" of the samples were calculated. On the basis of these four indices, "sensitivity", "specificity", "positive predictive value (PPV)" and "negative predictive value (NPV)" were then calculated. Based on the results of the histology based ASC US PAP samples, the following sample numbers were identified for the *CGB3* gene: TP, 43; FP, 17; TN, 57; and FN, 5. Based on screening with the methylation testing method in the promoter region of the *CGB3* gene, the following percentages of samples were identified: sensitivity, 90%; specificity, 77%; PPV, 72%; NPV, 92%; and the predictive accuracy was found to be 82%. Similarly, the efficiency of the *NOP56* gene in the ASC US PAP samples was then assessed, giving rise to the following results (numbers of samples) for the *NOP56* gene: TP, 50; FP, 24; TN, 36; and FP, 12. Based on screening with the methylation testing method in the promoter region of the *NOP56* gene, the following percentages of samples were identified: sensitivity, 75%; specificity, 68%, PPV, 60%; and NPV, 81%. The method of testing for methylation in the promoter of *NOP56* gene was found to be 70% effective, as shown in Table 3A.

Based on these findings, the sensitivity values (as percentages of methylation) for the *CGB3* and *NOP56* genes in the ASC US (histology based) samples were shown to be high (90 and 75%, respectively), indicating a higher level of efficiency in terms of the precise and correct identification of abnormal samples.

The other method we explored of measuring the diagnostic abilities of the *CGB3* and *NOP56* methylation marker was in terms of measuring the AUC (which represents the measure of separability) and ROC (which represents the diagnostic ability of a binary classifier system, which in this case was methylated vs. non methylated) values. AUC ROC values >0.5 were considered to be significant values, through which the classifier was able to distinguish between the normal and the abnormal samples (Carrington et al., 2022). Based on the analysis of the *CGB3* and *NOP56* methylation markers in the ASC US PAP samples, the AUC values were observed to be 0.83 and 0.71 for *CGB3* and *NOP56* respectively, indicating that *CGB3* has an 83%, and *NOP56* has a 71%, ability to detect between the normal and abnormal PAP samples correctly.

Screening test efficiency and diagnostic ability of *CGB3* and *NOP56* methylation marker in cytology based determined PAP samples (78/200 samples).

Similarly, to the analysis of the ASC US samples, the remaining cytology based 'determined' PAP samples were also assessed in terms of the calculation of these four indices, i.e., sensitivity, specificity, PPV and NPV. For the efficiency of the *CGB3* gene, the numbers of samples identified were as follows: TP, 48; FP, 8; TN, 14; and FN, 8. It was found that screening with the methylation testing method in the promoter region of the *CGB3* gene gave rise to the following percentage results: sensitivity, 86%; specificity, 64%; PPV, 86%; and NPV, 64%. The

method of testing for methylation in the promoter of *CGB3* gene gave a score of 79.4% (~80%) accuracy. Similarly, for the efficiency of the *NOP56* gene, the numbers of samples identified were as follows: TP, 38; FP, 6; TN, 16; and FN,18. Screening with the methylation testing method in the promoter region of the *NOP56* gene gave rise to the following percentage results: sensitivity, 68%; specificity, 73%; PPV, 86%; and NPV, 47%. Finally, the method of testing for methylation in the promoter of the *NOP56* gene gave a score of 69.2% (~70%) accuracy, as shown in Table 3B.

The diagnostic abilities of the *CGB3* and *NOP56* methylation markers observed in 'determined' PAP samples were found to be 0.74 and 0.63, respectively. As these values were >0.5, this signifies that the analysis was able to significantly distinguish between the positive and negative samples.

Discussion

ASC US is the most frequent type of abnormal cell found during cervical PAP screening, their grey zone of cytology can either be a sign of benignity (not cancer) or potential malignancy (Badea et al., 2019). Thus, those women who were found to be positive for HR HPV (with HR HPV types or had PAP results of ASC US, or higher stages) were considered to have positive screening test results, indicating exposure to the risk of cervical cancer, and these patients were subsequently referred for colposcopy and biopsy (Abdulaziz et al., 2020; Stany et al. 2006). In the present study, the 200 PAP samples were divided on the basis of histology [target biopsy with confirmed diagnosis; i.e., ASC US ('undetermined') samples], and cytology (under microscopic observation, for which the diagnosis was not 100% confirmed, known as 'determined' PAP samples). Methylation of the promoters of the *CGB3* and *NOP56* genes was analyzed in all the 200 PAP samples, with the aim of showing methylation in abnormal samples (including the LSIL, HSIL and cancer groups) and no methylation in normal (with or without HPV) samples. Based on the results of the histology based ASC US PAP samples, where the distribution percentages of the *CGB3* and *NOP56* genes were assessed, it was revealed that there was a higher level of methylation in abnormal PAP samples compared with normal samples, and moreover, the differences were found to be statistically significant.

Based on further analysis, the *CGB3* gene exhibited high sensitivity/specificity in both the histology and cytology based samples, with the percentages of methylation determined to be 90% and 77%, and 86% and 64%, respectively, which indicated that the *CGB3* gene had the ability to detect the presence of disease with high levels of accuracy (82% and 80%, respectively). Similarly, the *NOP56* gene showed high levels of sensitivity/specificity, with values for the percentages of methylation of 75% and 68%, and 68% and 73%, with respect to the histology and cytology based samples, respectively. These results indicated that the *CGB3* gene was able to act as a more effective marker compared with the *NOP56* gene, via analyzing the methylation process for both the histology (confirmed stage) and cytology (not confirmed stage) based samples. Furthermore, the study compared the diagnostic accuracies of the *CGB3* and *NOP56* genes in histology based ASC US PAP samples and cytology based 'determined' PAP samples by determining the ROC and AUC values. Usually, the ROC curve analyzes the probability of classifying the TP and FP rates by setting the thresholds based on the AUC value. However, for the majority of biomarkers, a cut off value needs to be set above a certain threshold of the methylation level in order to detect high specificity (Boers et al., 2016). The advantage of our methylation markers was that no cut off threshold value was used. In this case, if the PCR product was negative (i.e., no amplification of specific product), the samples were categorized as "negative", and any ratio above zero was categorized as "positive", as has been reported in Boers et al., 2016. Based on the analysis of the diagnostic ability results, the AUC values for the *CGB3* and *NOP56* genes in the histology based samples were found to be 0.83 and 0.71 respectively, which indicated that the *CGB3* and *NOP56* genes have an 83% and a 71% capability, respectively, to distinguish between the ASC US and normal samples. Similarly, in the cytology based samples, the *CGB3* and *NOP56* genes were shown to have AUC values of 0.74 and 0.63 respectively, which indicated that *CGB3* and *NOP56* respectively have a 74% and a 63% ability to distinguish between abnormal pre cancerous samples and normal samples.

A subsequent analysis was performed to validate the *CGB3* and *NOP56* methylation levels in different HPV types, which were categorized as HPV16, HPV18, combined HPV (HPV16/18), 'HPV others and no HPV. HPV infection in patients with abnormal cytology is an underlying cause of cervical cancer, which accounts for 28.8 61.3% of the reported cases (Çilingir et al., 2013; de Oliveira et al., 2018; Oranratanaphan et al., 2020). At present and in the future, the result of HPV genotyping being positive for HPV strains 16 and/or 18 indicates a high risk of cervical neoplasia, with either a high or low grade of squamous cells (Huh et al. 2015; de Oliveira et al., 2018; Einstein et al., 2011). Based on previously published data (Gultekin et al., 2018; Dursun et al., 2009; Stany et al., 2006), HPV type 16 accounts for approximately 60% of all cervical cancers as compared with HPV18 and the other HR HPV types, which are responsible for 15% and 35% of all cervical cancers, respectively. Based on the present study, HPV type infections overall were found to be associated with 87.5% (or 175/200) of the total PAP samples, which were further subdivided into instances of the various HR HPV types [HPV16, 40% (80 samples); HPV18, 13% (26 samples); the combined HPV16/18 strain, 2% (4 samples); and other types of HPV strains, 32.5% (65 samples)]. The remaining no HPV type of strain was represented by 12.5% (25/200) of the total samples. The overall methylation percentages were found to be the highest in the HR HPV types (HPV16 and HPV18), with methylation values of 68% and 87.5%, respectively. Therefore, the present study has demonstrated that the *CGB3* and *NOP56* genes exhibited higher methylation percentages in the HR HPV strains, which indicated a higher probability of cervical cancer, and therefore this

should be considered as a precautionary measure to further prevent the progression of cervical cancer infections to more dangerous levels.

In conclusion, in the present study we have identified two genes, *CGB3* and *NOP56*, for which *CGB3* showed the highest sensitivity and specificity to distinguish between abnormal and normal samples at a statistically significant level for both histologically (biopsy based) and cytologically (observation based) assessed samples. The findings in this study should be useful in the future to further test and develop large scale methylation markers at the commercial level for the purpose of detecting abnormal samples with the intention of preventing further spread of cervical cancers.

Author Contribution Statement

PS performed the experiments, analyzed the data and wrote the manuscript. NK analyzed the data and provided the ASC US samples. PY wrote the proposal for grants, designed the experiments, analyzed the data and wrote the manuscript. All authors read and approved the final manuscript. PY confirm the authenticity of all the raw data.

Acknowledgements

We are very grateful to Professor Dr. Yong Poovorawan, Center of Excellence in Clinical Virology, Chulalongkorn University for providing all the clinical PAP samples from Chulalongkorn Medical Hospital, and from the National Research Institute of Cancer (NCI), Thailand. We are also thankful to Dr. Apiwat Mutirangura, Center of Excellence in Molecular Genetics of Cancer and Human Diseases, Chulalongkorn University for providing the laboratory equipment and valuable suggestions.

Funding and support

The present study was financially supported by the Thailand Research Fund and Chulalongkorn University (grant no. RSA5880065), The National Science and Technology Development Agency, Thailand (grant no. P-15-50270).

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study protocol (IRB No. 477/61, COA No.868/2018) was approved by The Institutional Review Board of the Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand.

Conflict of interests

The authors declare that they have no conflict of interests.

References

Abdulaziz AM, Liu L, Sun Y, et al (2020). Clinicopathologic significance and treatment of ASC-US in cervical cytology. *Int J Clin Exp Pathol*, **13**, 307.

Arbyn M, Roelens J, Simoens C (2013). Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. *Cochrane Database Syst Rev*, **3**.

Badea M, Baroş AL, Bohîlţea RE et al (2019). Modern interdisciplinary monitoring of cervical cancer risk. *Rom J Morphol Embryol*, **60**, 469-78.

Boers A, Wang R, Van Leeuwen RW (2016). Discovery of new methylation markers to improve screening for cervical intraepithelial neoplasia grade 2/3. *Clin Epigenetics*, **8**, 1-6.

Bray F, Jacques F, Isabelle S, et al (2018). 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, **68**, 394-424.

Bray F, Ferlay J, Soerjomataram I, et al (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, **68**, 394-424.

Burd EM (2003). Human papillomavirus and cervical cancer. *Clin Microbiol Rev*, **16**, 1-7.

Burgers WA, Blanchon L, Pradhan S et al (2007). Viral oncoproteins target the DNA methyltransferases. *Oncogene*, **26**, 1650-5.

Carrington AM, Manuel DG, Fieguth P et al (2022). Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation IEEE *Trans Pattern Anal Mach Intell*, **25**, 1.

Chalertpet K, Pakdeechaidan W, Patel V, Mutirangura A, Yanatatsaneejit P (2015). Human papillomavirus type 16 E7 oncoprotein mediates CCNA1 promoter methylation. *Cancer Sci*, **106**, 1333-40.

Çilingir IU, Bengisu E, Ağaçfidan A (2013). Microarray detection of human papilloma virus genotypes among Turkish women with abnormal cytology at a colposcopy unit. *J Turk Ger Gynecol Assoc*, **14**, 23.

de Oliveira GG, Eleutério RM, Barbosa RD, et al (2018). Atypical squamous cells: cytopathological findings and correlation with HPV genotype and histopathology. *Acta Cytol*, **62**, 386-92.

Desai FS, Korant R, Gohil M, Singh LS (2021). The role of Ki67 and p16INK4a biomarkers on conventional cell blocks to differentiate post radiation dysplasia from cervical cancer in post therapeutic surveillance cytology. *Asian Pac J Cancer Biol*, **6**, 111-6.

Dursun P, Senger SS, Arslan H, Kuşçu E, Ayhan A (2009). Human papillomavirus (HPV) prevalence and types among Turkish women at a gynecology outpatient unit. *BMC Infect Dis*, **9**, 1-6.

Einstein MH, Garcia FA, Mitchell AL, Day SP (2011). Age-stratified performance of the cervista HPV 16/18 genotyping test in women with ASC-US cytology cervista HPV 16/18 genotyping across age strata. *Cancer Epidemiol Biomark Prev*, **20**, 1185-9.

Gilbert JR, Vance JM (1998). Isolation of genomic DNA from mammalian cells. *Curr Protoc Hum Genet*, **19**, A-3B.

Group AL (2003). Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. *Am J Obstet Gynecol*, **188**, 1383-92.

Gultekin M, Zayifoglu KM, Kucukyildiz I, et al (2018). Initial results of population based cervical cancer screening program using HPV testing in one million T urkish women. *Int J Cancer*, **142**, 1952-8.

Heideman DA, Hesselink AT, Berkhof J, et al (2011). Clinical validation of the cobas 4800 HPV test for cervical screening purposes. *J Clin Microbiol*, **49**, 3983-5.

Huh WK, Ault KA, Chelmow D, et al (2015). Use of primary high-risk human papillomavirus testing for cervical cancer screening: interim clinical guidance. *Gynecol Oncol*, **136**, 178-82.

Iavazzo C, Boutas I, Grigoriadis C, Vrachnis N, Salakos N (2012). Management of ASCUS findings in Papanicolaou smears. A retrospective study. *Eur J Gynaecol Oncol*, **33**, 605-9.

Imsamran W, Pattatang A, Supaattagorn P, et al (2018). Cancer in Thailand Volume IX: 2013-2015. J Natl Cancer Inst Ministry of Public Health: Bangkok, Thailand. 88.

Kluyver T, Ragan-Kelley B, Pérez F et al (2016). Jupyter Notebooks-a publishing format for reproducible computational workflows.

Koh WJ, Abu-Rustum NR, Bean S, et al (2019). Cervical cancer, version 3.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*, **17**, 64-84.

Koliopoulos G, Nyaga VN, Santesso N, et al (2017). Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database Syst Rev*, **8**.

Lahue BJ, Baginska E, Li SS, Parisi M (2015). Health technology assessment on cervical cancer screening, 2000–2014. *Int J Technol Assess Health Care*, **31**, 171-80.

McKinney W (2010). Data structures for statistical computing in python. InProceedings of the 9th Python in Science Conference. Vol. 445, Number. 1. pp 51-6,

Ngo-Metzger Q, Prajakta A (2019). Screening for cervical cancer. *Am Fam Physician*, **99**, 253-54.

Oranratanaphan S, Kobwitaya K, Termrungruanglert W, et al (2020). Value of CCNA1 promoter methylation in triaging ASC-US cytology. *Asian Pac J Cancer Prev*, **21**, 473.

Pimple SA, Mishra GA (2019). Global strategies for cervical cancer prevention and screening. *Minerva Ginecol*, **71**, 313-20.

Rao A, Young S, Erlich H, et al (2013). Development and characterization of the cobas human papillomavirus test. *J Clin Microbiol*, **51**, 1478-84.

Sen P, Ganguly P, Ganguly N (2018). Modulation of DNA methylation by human papillomavirus E6 and E7 oncoproteins in cervical cancer. *Oncol Lett*, **15**, 11-22.

Singh P, Chalertpet K, Sukbhattee J et al (2022). Association between promoter methylation and gene expression of *CGB3* and *NOP56* in HPV-infected cervical cancer cells. *Biomed Rep*, **16**, 1-2.

Solomon D, Davey D, Kurman R, et al (2002). The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA*, **287**, 2114-9.

Soto D, Song C, McLaughlin-Drubin ME (2017). Epigenetic alterations in human papillomavirus-associated cancers. *Viruses*, **9**, 248.

Stany MP, Bidus MA, Reed EJ, et al (2006). The prevalence of HR-HPV DNA in ASC-US Pap smears: A military population study. *Gynecol Oncol*, **101**, 82-5.

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

Sung H, Ferlay J, Siegel RL et al (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, **71**, 209-49.

Wright Jr TC, Cox JT, Massad LS, Twiggs LB, Wilkinson EJ (2002). ASCCP-Sponsored Consensus Conference. 2001 Consensus Guidelines for the management of women with cervical cytological abnormalities. *JAMA*, **287**, 2120-9.

Wright TC, Stoler MH, Behrens CM, et al (2015). Primary cervical cancer screening with human papillomavirus: end of study results from the ATHENA study using HPV as the first-line screening test. *Gynecol Oncol*, **136**, 189-97.