CpG methylation of the PAX1 promoter in exfoliated cervical cells: a potential cervical cancer screening model

Yu Chen1,†, Yuanyuan Zhang1,2,†, Xiaochun Chen1,†, Hong Chu1, Songqing Guan1, Qiuni Liang1, Yuting Xu1, Xinyue Lv1, Chunhua Li1, Yun Li3, Jin Liu1, Wenlan Liu4,*, Zhiying Yu1,*.†

1 Department of Gynecology, Shenzhen Second People’s Hospital/the First Affiliated Hospital of Shenzhen University Health Science Center, 518035 Shenzhen, Guangdong, China
2 Department of Obstetrics and Gynecology, the Affiliated Hospital of Guizhou Medical University, 550025 Guizhou, China
3 The Center for Medical Genetics and Molecular Diagnosis, Shenzhen Second People’s Hospital/the First Affiliated Hospital of Shenzhen University Health Science Center, 518035 Shenzhen, Guangdong, China

*Correspondence
2011191208@email.szu.edu.cn
(Zhiying Yu)
wlliu@szu.edu.cn
(Wenlan Liu)

† These authors contributed equally.

Abstract
Cervical cancer screening is an effective mean of preventing cervical cancer, however, there are limitations of the current screening methods. In this study, an automated high throughput paired box-1 (PAX1) methylation detection method was designed by employing matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), which analyzed 88 PAX1 methylation sites. In total, 188 subjects were enrolled in the study. The diagnostic efficacy of various PAX1 methylation sites for detecting cervical intraepithelial neoplasia II or more severe lesions was analyzed and compared with current cervical cancer screening methods. A predictive model was constructed using logistic regression. Five sites with the largest area under receiver operating characteristic curve values were GRCh38/hg38, chr20: X21705687, X21705946, X21706427, X21706285 and X21706637. PAX1 methylation method had higher diagnostic efficacy than high risk human papillomavirus (hr-HPV) and ThinPrep cytology tests. A predictive model combining hr-HPV and PAX1 methylation sites was constructed as follows: prediction index = −5.993 + 15.211 × X21705687 + 7.890 × X21706427 + 1.846 × HPV16/18 (1,0) + 1.821 × other hr-HPV (1,0). Area under the curve was 0.868 and sensitivity and specificity were 0.863 and 0.756, respectively. Analysis of PAX1 promoter methylation alone or in combination with hr-HPV was a promising approach for cervical cancer screening. The study was registered in Chinese Clinical Trial Registry (registration number: ChiCTR2000029231, http://www.chictr.org.cn/index.aspx).

Keywords
Cervical cancer screening; MALDI-TOF MS; Methylation; PAX1

1. Introduction
Cervical cancer is the frequently diagnosed gynecological cancer and leading cause of gynecological cancer-related deaths in China and worldwide [1, 2]. It originates in the transformation zone of cervical epithelial tissue and predominantly caused by human papillomavirus (HPV), making it preventable and detectable in pre and early cancer stages [3, 4]. In 2018, World Health Organization announced a global call for action to eliminate cervical cancer. HPV vaccines are effective, however cervical cancer screening is the most effective for preventing cervical cancer. In recent decades, the incidence and mortality of cervical cancer have decreased in affluent countries. In urban areas of China, the threat posed by cervical cancer has declined because of cervical cancer screening [5]. Conversely, never- and under-screenings are common in underprivileged than in privileged populations which has led to the increased incidence and mortality of invasive cervical cancer [6, 7]. Hence, making cervical cancer screening more accessible, reliable, and cost-effective remains a global challenge.

Currently, the common cervical cancer screening methods are liquid-based cytology and HPV DNA testing [8]. However, both methods have limitations. HPV DNA testing is sensitive but lacks specificity as ~80% women are infected with HPV at some point in their lives. Hence, HPV DNA testing requires economic and medical resources which burden the patients with extra hospital visits and psychological distress [9]. Cytological testing has higher specificity than HPV DNA testing, however, it depends on the expertise of pathologists, and accuracy varies from doctor to doctor, making it unsuitable for regions lacking skilled pathologists [10].

PAX1 gene is a tumor suppressor gene located on the short arm of chromosome 20 [11]. A previous study demonstrated that PAX1 expression inhibited the phosphorylation of multiple kinases including the dual specificity phosphatase 1, 5 and 6, and of epidermal growth factor/mitogen-activated protein kinase signaling [12]. Numerous studies have shown that PAX1...
methylation is related to cervical cancer, however, the clinical significance of \textit{PAX1} methylation detection as cervical cancer screening method remains controversial. \textit{PAX1} methylation testing accuracy fluctuates from 30\% to 85\% \cite{13}. Literature review revealed that previous studies used methylation-specific polymerase chain reaction (MSP) which restricted detection to few methylation sites. When a methylation site relevant to cervical cancer is tested, the diagnostic value increases, and vice versa. The limitations of detection method may contribute to variations in the accuracy.

MALDI-TOF MS can be used to quantify DNA methylation and simultaneously detect multiple methylation sites. Herein, an automated high throughput \textit{PAX1} methylation detection method was designed by using MALDI-TOF MS. The methylation status of 88 sites was analyzed for identifying the most relevant ones and compared the diagnostic efficacy with that of hr-HPV DNA test and ThinPrep cytology test (TCT). Finally, a predictive model was constructed by combining the automatic detection methods for hr-HPV DNA and \textit{PAX1} methylation. The diagnostic efficiency of this automated detection method was substantially better than the traditional cervical cancer screening methods.

2. Materials and methods

2.1 Patient selection

Patients included for this study had abnormal cervical cancer screening results and had undergone cervical biopsy during colposcopy in Department of Gynecology at Shenzhen Second People’s Hospital between October 2019 and April 2021. Each patient was asked to fill in a questionnaire for assessing whether they met the inclusion or exclusion criteria and to add other relevant characteristics. The inclusion criteria were as follows: (1) 18–60 years’ age; (2) abnormal findings detected for the first time by cervical cancer screening; and (3) patients who were about to undergo cervical biopsy due to abnormal colposcopy findings. Exclusion criteria included: (1) a history of uterine or cervical surgery; (2) patients with other malignant tumors; and (3) women undergoing or within 42 days of pregnancy termination.

Ultimately, 188 participants were enrolled in the study. The study conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki), published by British Medical Journal (1964).

Cervical biopsy is the gold standard for pathological diagnosis. Samples included 108 inflamed cervical epithelia (non-malignant), 78 cervical intraepithelial neoplasms (CINs), and two cervical cancers. The 78 CINs comprised of 29 CIN I (less than 1/3 abnormal cells), 21 CIN II (1/3 to 2/3 abnormal cells), and 28 CIN III (2/3 to full cervical surface layer thickness) lesions.

2.2 Sample collection and DNA extraction

Exfoliated cervical epithelial cells were obtained by following the colposcopic examination prior to cervical biopsy. After exposing the cervix, blood and vaginal discharge were gently wiped off. Cytology swabs were inserted into the cervical squamocolumnar junction and rotated clockwise for five to ten times. The cells were immediately stored in cell preservation solution (Cat# 51306; Qiagen, Hilden, Germany) at 4 °C until DNA extraction.

2.3 \textit{PAX1} methylation detection

DNA from cervical epithelial cells was extracted using QIAmp Tissue Kit (Qiagen). Cytosines were converted to uracils using EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA). Methylated cytosines remained unmodified during the process. According to the University of California Santa Cruz genome browser database (http://genome.ucsc.edu/cgi-bin/hgGateway), the methylation sites in \textit{PAX1} promoter contained 141 5-cytosine-phosphate-guanine-3 (CpG) islands (chr20: 21705562–21707051) (Fig. 1A). Methylation specific primers were designed using EpiDesigner (https://epidesigner.com/start3.html, Agena Bioscience, San Diego, CA, USA). Five primer pairs were used to cover 88 CpG islands: (1) left: ATTTTAAAAGTTAGAGGAGGAAGGG, right: CCAAAAACCCTAAAACCTTCATC, CpG coverage: 18; (2) left: ATGAAGTTATTTGGGGTGGGG, right: CCAAACTTACTCCTACCCCCCCTA, CpG coverage: 34; (3) left: TTTAGTCTCCTTCCTTCCTTTCAGAAGT, right: TTCTAAAAATATAAATCCCTAAATCCCC, CpG coverage: 10; (4) left: GATTGTTGCGGGGTTTCCAGTTGAAGGG, right: CAAACAAACTCCTCCTCTACT, CpG coverage: 30; and (5) left: ATTTTGATTTTGTGTGGGAGAT, right: CTTAACGATATACCCCTTCCCCCTA, CpG coverage: 38 (Fig. 1B). Exfoliated cervical cell DNA was extracted by enzymatic lysis, transformed with bisulfite, and amplified using PCR. Base-specific cleavage was performed using homogeneous Mass Cleave assay. The 88 sites were cut into 52 pieces. The base fragments were analyzed using MassARRAY time-of-flight MS (Agena Bioscience).

2.4 Liquid-based cytology

Exfoliated cervical epithelial cells were obtained as described in Section 2.2. Samples for cytology were prepared according to the standard ThinPrep (Hologic, Marlborough, MA, USA) protocol. The samples were evaluated by two experienced pathologists. The results were classified as negative intraepithelial lesion or malignancy, low-grade squamous intraepithelial lesion, or high-grade squamous intraepithelial lesion, when both pathologists agreed with the cell morphology. Atypical squamous cells of undetermined significance and atypical squamous cells categorized as high-grade, that could not be excluded based on suspected underlying lesion (low-grade squamous intraepithelial lesion vs. high-grade squamous intraepithelial lesion, respectively), were used when pathologists were uncertain or in disagreement with the results.

2.5 HPV DNA testing

HPV DNA was amplified using Hema 9600 gene amplification instrument (Hema Medical Instrument Co. Ltd., Zhuhai, China). YN-H18 automatic nucleic acid molecular hybridization instrument (Yaneng Biosciences, Guangdong, China) was employed for detecting 18 hr-HPV genotypes,
2.6 Tissue collection and pathological diagnosis

All abnormal lesions were evaluated according to the American Society for Colposcopy and Cervical Pathology and the Chinese Society for Colposcopy and Cervical Pathology guidelines. The lesions were biopsied by cervical disease specialists. Tissues were fixed in 4% formaldehyde. Paraffin embedded tissues were cut into 5-µm sections for hematoxylin and eosin staining.

Pathological slides were analyzed by the two experienced pathologists. If two pathologists disagreed, a third pathologist examined the slides and performed immunohistochemical staining as supporting method, if necessary. The tissues samples in this study were fairly typical and above scenario was not created.

2.7 Statistical analysis

The data were analyzed using SPSS Statistics 27.0.1 (SPSS Inc., Chicago, IL, USA). Figs were plotted using GraphPad Prism 9.4.0 (GraphPad Software, San Diego, CA, USA). Quantitative variables were expressed as mean ± standard deviation (SD) or median (interquartile range). Between-group differences were assessed using t-tests or one-way analysis of variance for normally distributed variables and non-parametric tests for non-normally distributed variables. The Chi-square test or Fisher’s exact test was used to analyze the qualitative data.

Receiver operating characteristic (ROC) curves were plotted, and cut-off values were determined using maximum Youden index. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated based on optimal cut-off values. The predictive model was constructed using multivariate logistic regression and backward elimination. The model was internally evaluated using ROC curves.

3. Results

3.1 Participant characteristics

A total of 188 patients were included in this study and placed in five groups based on their pathological diagnoses: 108 cases of inflamed cervical epithelia (non-malignant), 29 of CIN I, 21 of CIN II, 28 of CIN III, and 2 of cervical cancer (Fig. 2). The baseline characteristics of patients are listed in Table 1. The median (interquartile range) age of enrolled patients was 35 (29–42) years. Differences in parity, hr-HPV counts, and TCT results were observed among the patients (p = 0.027, p < 0.01, and p < 0.01, respectively). No significant differences were found in age, weight, height, the onset of menstruation, or gravity. HPV52 (19.7%) was the most prevalent HPV genotype, followed by HPV16 (17.0%) and HPV58 (13.3%). This was consistent with the overall distribution of HPV genotypes in China but inconsistent with the most prevalent genotypes worldwide (HPV16 and HPV18) [14, 15]. The detailed distribution of HPV genotypes in relation to pathological diagnoses is described in Supplementary Table 1.

3.2 Diagnostic value of hr-HPV and TCT

According to cervical cancer screening guidelines, HPV16/18 testing alone or combined with other genotypes can be used as an independent screening method. TCT can be employed as an independent screening method or as a secondary screening method for hr-HPV-positive patients [16]. The diagnostic efficacy of these methods was evaluated. At least one positive test for either HPV16 or HPV18 was considered positive for HPV16/18 test. Patients tested positive for at least one hr-HPV genotype were considered hr-HPV-positive. A negative intraepithelial lesion or malignancy was defined as TCT-negative. Patients with equal to or more severe than atypical squamous cells of undetermined significance were considered TCT-positive.

The participants were divided into control and patient groups as per their pathological diagnosis. CIN II is the commonly accepted threshold for treatment and its detection is the primary
Women requiring cervical biopsy were evaluated to determine whether they met the criteria of this study. Before biopsy, exfoliated cervical epithelial cells were obtained. Patients were diagnosed as having inflamed cervical epithelia (non-malignant), cervical intraepithelial neoplasia CIN I, CIN II, CIN III, or cervical cancer based on pathological analysis. Abbreviations: MALDI-TOF MS: matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; PAX1: paired box-1.

3.3 Diagnostic value of PAX1 promoter methylation

Eighty-eight PAX1 methylation sites were evaluated for each sample and their methylation percentages were recorded (Fig. 3). After classifying participants to control and patient groups, ROC curves were plotted for each methylation site (Fig. 4 and Supplementary Fig. 1). The area under ROC curve (AUC) and Spearman correlation between PAX1 methylation and pathological diagnosis were analyzed. GRCh38/hg38 chr20: X21705687 (X21705687) had the largest AUC (0.792). The optimal cut-off values for five sites with the largest AUCs and their corresponding sensitivity, specificity, PPV, NPV, and accuracy are shown in Table 3. PAX1 methylation status of these sites had higher diagnostic efficacy levels than conventional screening methods. The hr-HPV genotypes had higher sensitivity, however their specificity of 0.299 was considered unsatisfactory as a false-positive result may consume unnecessary resources to perform further colposcopy and biopsy. In contrast, HPV16/18 had relatively high specificity but its low sensitivity of 0.333 was ineffective for cervical cancer screening and prevention.

3.4 Predictive model construction and evaluation

Based on the high sensitivity of hr-HPV, high specificity of HPV16/18, and advantages of automated detection process in clinical practice, a predictive model was constructed using multivariate logistic regression with backward elimination that combined HPV testing and PAX1 methylation. Using Chi-square test, HPV16/18 (p = 0.003) and other hr-HPV genotypes (p = 0.034) were confirmed to be associated with pathological diagnosis. PAX1 methylation sites associated with pathological diagnosis, as well as HPV16/18 and other hr-HPV genotypes, were initially included in the regression model. By gradually eliminating the variables having little effect on regression model until all variables having significant effect on regression coefficient, it was determined that HPV16/18, other hr-HPV genotypes, X21705687, and X21706427 had significant impact on regression model (p < 0.05) which allowed the prediction index (PI) to reach its maximum predictive value. The logistic regression equation was as follows:

\[
PI = \ln(P)/(1-P) = -5.933 + 15.211 \times X21705687 + 7.890 \times X21706427 + 1.846 \times HPV16/18(1,0) + 1.821 \times other hr - HPV (1,0).
\]
<table>
<thead>
<tr>
<th>Pathology</th>
<th>Non-malignant (n = 108)</th>
<th>CIN I (n = 29)</th>
<th>CIN II (n = 21)</th>
<th>CIN III (n = 28)</th>
<th>Cervical cancer (n = 2)</th>
<th>Total (n = 188)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>36.0 (30.0–43.0)</td>
<td>30.0 (26.0–35.5)</td>
<td>32.0 (27.0–39.5)</td>
<td>35.5 (29.3–48.5)</td>
<td>44.5 (39.0–50.0)</td>
<td>35.0 (29.0–42.0)</td>
<td>0.118</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>160.0 (156.0–162.0)</td>
<td>163.0 (158.5–165.0)</td>
<td>160.0 (155.5–165.5)</td>
<td>160.0 (154.3–165.0)</td>
<td>162.0 (160.0–164.0)</td>
<td>160.0 (156.0–163.0)</td>
<td>0.104</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>53.0 (49.0–57.0)</td>
<td>54.0 (47.0–60.8)</td>
<td>57.8 (48.5–63.5)</td>
<td>56.5 (63.0–60.0)</td>
<td>65.5 (53.0–60.0)</td>
<td>53.0 (49.0–59.0)</td>
<td>0.604</td>
</tr>
<tr>
<td>Age at menarche (yr)</td>
<td>13.0 (12.0–14.0)</td>
<td>13.0 (12.0–14.0)</td>
<td>13.0 (13.0–14.0)</td>
<td>13.0 (12.0–14.0)</td>
<td>13.0 (13.0–13.0)</td>
<td>13.0 (12.0–14.0)</td>
<td>0.696</td>
</tr>
<tr>
<td>Gravidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>22 (20.4%)</td>
<td>10 (34.5%)</td>
<td>10 (47.6%)</td>
<td>7 (25.0%)</td>
<td>0 (0.0%)</td>
<td>49 (26.1%)</td>
<td>0.132</td>
</tr>
<tr>
<td>1</td>
<td>19 (17.6%)</td>
<td>5 (17.2%)</td>
<td>0 (0.0%)</td>
<td>4 (14.3%)</td>
<td>0 (0.0%)</td>
<td>28 (14.9%)</td>
<td></td>
</tr>
<tr>
<td>&gt;1</td>
<td>67 (62.1%)</td>
<td>14 (48.2%)</td>
<td>11 (52.5%)</td>
<td>17 (60.7%)</td>
<td>2 (100.0%)</td>
<td>111 (59.0%)</td>
<td></td>
</tr>
<tr>
<td>Parity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>31 (28.7%)</td>
<td>16 (55.2%)</td>
<td>11 (52.4%)</td>
<td>8 (28.6%)</td>
<td>0 (0.0%)</td>
<td>66 (35.1%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42 (38.9%)</td>
<td>8 (27.6%)</td>
<td>4 (19.0%)</td>
<td>7 (25.0%)</td>
<td>0 (0.0%)</td>
<td>61 (32.4%)</td>
<td>0.027*</td>
</tr>
<tr>
<td>&gt;1</td>
<td>35 (32.4%)</td>
<td>5 (17.2%)</td>
<td>6 (28.6%)</td>
<td>13 (46.5%)</td>
<td>2 (100.0%)</td>
<td>61 (32.4%)</td>
<td></td>
</tr>
<tr>
<td>Contraception</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>26 (24.1%)</td>
<td>7 (24.1%)</td>
<td>7 (33.3%)</td>
<td>7 (25.0%)</td>
<td>0 (0.0%)</td>
<td>47 (25.0%)</td>
<td></td>
</tr>
<tr>
<td>Condoms</td>
<td>45 (41.7%)</td>
<td>14 (48.3%)</td>
<td>11 (52.4%)</td>
<td>11 (39.3%)</td>
<td>1 (50.0%)</td>
<td>82 (43.6%)</td>
<td>0.378</td>
</tr>
<tr>
<td>Other</td>
<td>32 (29.6%)</td>
<td>7 (24.1%)</td>
<td>2 (9.6%)</td>
<td>6 (21.4%)</td>
<td>0 (0.0%)</td>
<td>47 (25.0%)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (4.6%)</td>
<td>1 (3.4%)</td>
<td>1 (4.8%)</td>
<td>4 (14.3%)</td>
<td>1 (50.0%)</td>
<td>12 (6.4%)</td>
<td></td>
</tr>
<tr>
<td>hr-HPV count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>39 (36.1%)</td>
<td>2 (6.9%)</td>
<td>0 (0.0%)</td>
<td>1 (3.6%)</td>
<td>0 (0.0%)</td>
<td>42 (22.3%)</td>
<td>&lt;0.010**</td>
</tr>
<tr>
<td>1</td>
<td>62 (57.4%)</td>
<td>18 (62.1%)</td>
<td>16 (76.2%)</td>
<td>22 (78.6%)</td>
<td>1 (50.0%)</td>
<td>119 (63.3%)</td>
<td></td>
</tr>
<tr>
<td>&gt;1</td>
<td>7 (5.7%)</td>
<td>9 (31.0%)</td>
<td>5 (23.9%)</td>
<td>5 (17.9%)</td>
<td>1 (50.0%)</td>
<td>27 (14.4%)</td>
<td></td>
</tr>
<tr>
<td>TCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NILM</td>
<td>76 (70.4%)</td>
<td>17 (58.6%)</td>
<td>10 (47.6%)</td>
<td>8 (28.6%)</td>
<td>1 (50.0%)</td>
<td>112 (59.6%)</td>
<td></td>
</tr>
<tr>
<td>ASC-US</td>
<td>18 (16.7%)</td>
<td>7 (24.1%)</td>
<td>6 (28.6%)</td>
<td>7 (25.0%)</td>
<td>0 (0.0%)</td>
<td>38 (20.2%)</td>
<td></td>
</tr>
<tr>
<td>L-SIL</td>
<td>12 (11.1%)</td>
<td>3 (10.3%)</td>
<td>3 (14.3%)</td>
<td>3 (10.7%)</td>
<td>1 (50.0%)</td>
<td>22 (11.7%)</td>
<td>&lt;0.010**</td>
</tr>
<tr>
<td>ASC-H</td>
<td>1 (0.9%)</td>
<td>1 (3.4%)</td>
<td>1 (4.8%)</td>
<td>4 (14.3%)</td>
<td>0 (0.0%)</td>
<td>7 (3.7%)</td>
<td></td>
</tr>
<tr>
<td>H-SIL</td>
<td>1 (0.9%)</td>
<td>1 (3.4%)</td>
<td>1 (4.8%)</td>
<td>6 (21.4%)</td>
<td>0 (0.0%)</td>
<td>9 (4.8%)</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01 (Kruskal-Wallis or Fisher’s exact test). Abbreviations: ASC-H: atypical squamous cells of high-grade; ASC-US: atypical squamous cells of undetermined significance; CIN: cervical intraepithelial neoplasm; hr-HPV: high-risk human papillomavirus; L/H-SIL: low/high-grade squamous intraepithelial lesion; NILM: negative intraepithelial lesion or malignancy; TCT: ThinPrep cytology test.
For HPV16/18 and other hr-HPV genotypes, 1 = positive for one or more HPV genotypes; and 0 = negative for one or more HPV genotypes. AUC for this predictive model was 0.868 (95% confidence interval: 0.817–0.920) (Fig. 5). The cut-off, sensitivity, specificity, PPV, NPV, and accuracy were 0.300, 0.843, 0.622, 0.457, 0.913, and 0.683, respectively (Table 4). The variables adopted by this model could be inserted automatically; they were quantifiable, stable, and independent of the pathologist’s interpretation. Internal validation of predictive model showed higher accuracy than conventional screening methods which proved its clinical value.

### 4. Discussion

DNA methylation is a major form of epigenetic regulation and regarded as a symbolic event in carcinogenesis. It occurs when methyl group is added to cytosine by DNA methyltransferase to form CpG dinucleotide [13]. DNA methylation plays role in the development of cervical cancer, for example, the methylation of PAX1, sex-determining region Y-box 1, hsa-miR-124, telomerase reverse transcriptase, LIM homeobox transcription factor 1A [18], septin 9 [19], Wilms’ tumor 1, NK6 transcription factor related locus 1, deleted in bladder cancer 1 [20], junctional adhesion molecule 3 [21], erythrocyte membrane protein band 4.1 like 3, cell adhesion molecule 1, and maturation associated protein [22]. Among these, PAX1 methylation is the most correlated with CIN progression and cervical carcinogenesis [11]. PAX1 gene family interacts with homeobox gene family and is critical for embryonic tissue development and cellular differentiation [23, 24]. The mechanism of PAX1 methylation in cancer development has yet not been elucidated. Compelling evidence has suggested that PAX1 methylation could have diagnostic potential in cancer screening, especially for cervical cancer. Previous studies have examined the methylation status of PAX1 in cervical samples, however the diagnostic value of this data was found inconsistent as the sensitivity ranged from 0.640–0.941 and the specificity from 0.600–0.910, for diagnosing CIN II and higher or CIN III and higher-grade lesions [11, 13, 25–32].

The approaches used to detect DNA methylation can be categorized based on their pretreatment methods like restriction enzyme digestion, bisulfite conversion, and affinity enrichment. Mainstream DNA methylation detection methods are based on bisulfite conversion including MSP, digital MSP, nucleic acid MS, methylation chips, and first-, second-, or third-generation sequencing [33]. MALDI-TOF MS is an analytical technique in which a laser irradiates the crystalline film formed between sample and matrix to ionize biomolecules in the sample. The charged biomolecules move under an electric field, and their mass-to-charge (m/z) ratios are calculated by measuring their movement speed [34]. Compared with PCR and methylation chips, MALDI-TOF MS has accurate absolute quantification, higher throughput, and higher detection stability. Previous studies have largely used MSP as the detection method. Most studies did not clearly state the exact methylation sites. Only a scant number of methylation sites were analyzed because of the limitations of approach. Time-of-flight MS was used to detect PAX1 methylation in 122 cervical samples, in which 20 CpG islands were analyzed [35]. The lack of absolute quantification and coverage of methylation sites may result in large

### Table 2. Diagnostic efficacy of hr-HPV and TCT.

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>hr-HPV</td>
<td>0.980</td>
<td>0.299</td>
<td>0.342</td>
<td>0.976</td>
<td>0.484</td>
</tr>
<tr>
<td>HPV16/18</td>
<td>0.333</td>
<td>0.861</td>
<td>0.472</td>
<td>0.776</td>
<td>0.718</td>
</tr>
<tr>
<td>TCT</td>
<td>0.627</td>
<td>0.679</td>
<td>0.421</td>
<td>0.830</td>
<td>0.665</td>
</tr>
</tbody>
</table>

**Abbreviations:** hr-HPV: high-risk human papillomavirus; NPV: negative predictive value; PPV: positive predictive value; TCT: ThinPrep cytology test.

### Table 3. Diagnostic value of PAX1 promoter methylation.

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>95% CI</th>
<th>Optimal cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X21705687</td>
<td>0.792</td>
<td>0.719–0.866</td>
<td>0.125</td>
<td>0.843</td>
<td>0.622</td>
<td>0.457</td>
<td>0.913</td>
<td>0.683</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>X21705946</td>
<td>0.792</td>
<td>0.718–0.865</td>
<td>0.145</td>
<td>0.725</td>
<td>0.730</td>
<td>0.500</td>
<td>0.877</td>
<td>0.729</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>X21706427</td>
<td>0.790</td>
<td>0.721–0.859</td>
<td>0.175</td>
<td>0.608</td>
<td>0.825</td>
<td>0.564</td>
<td>0.850</td>
<td>0.766</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>X21706285</td>
<td>0.783</td>
<td>0.705–0.861</td>
<td>0.155</td>
<td>0.647</td>
<td>0.781</td>
<td>0.524</td>
<td>0.856</td>
<td>0.745</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>X21706637</td>
<td>0.777</td>
<td>0.698–0.855</td>
<td>0.165</td>
<td>0.745</td>
<td>0.696</td>
<td>0.481</td>
<td>0.879</td>
<td>0.710</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**Abbreviations:** AUC: area under the curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

### Table 4. Diagnostic efficacy of predictive model.

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>95% CI</th>
<th>Optimal cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Accuracy</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0.868</td>
<td>0.817–0.920</td>
<td>0.300</td>
<td>0.843</td>
<td>0.756</td>
<td>0.566</td>
<td>0.927</td>
<td>0.780</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**Abbreviations:** AUC: area under the curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.
**Figure 3.** Representative mass spectra for PAX1 methylation sites. Mass spectra of (A) X21705687, (B) X21705946, (C) X21706427, (D) X21706285, and (E) X21706637. The abscissa is mass-to-charge ratio and ordinate is the ion abundance. Abbreviation: PAX1: paired box-1.

**Figure 4.** ROC curves of five methylation sites with the highest AUCs. Abbreviations: AUC: area under the curve; ROC: receiver operating characteristic.

**Figure 5.** ROC curve of the predictive model. Abbreviation: ROC: receiver operating characteristic.
differences of sensitivity and specificity in previous studies. In addition, owing to the lack of comprehensive coverage of PAX1 methylation sites and accurate methylation estimation, these results have less referential value for guiding the design of primers used to detect PAX1 promoter methylation with MSP or MALDI-TOF MS.

There were 141 methylation sites determined in PAX1 promoter. A total of 88 methylation sites were detected using this method as some sites were lost because of inconsistent coverage of CpG islands by each primer pair. The diagnostic efficacy of each site was evaluated as a cervical cancer screening method and determined that X21705687 had the highest AUC of 0.792. To the best of our knowledge, this is the first time that AUC values for specific PAX1 methylation sites have been reported. Although the sensitivity of PAX1 was marginally lower than that of hr-HPV, the false-positive rate of hr-HPV was as high as 0.701. When used as an independent screening method, hr-HPV can lead to unnecessary colposcopies and cervical biopsies which increases the pain and psychological burden on patients and wastes medical resources. Although HPV16/18 testing has high specificity, HPV16 and HPV18 are not the only prevalent genotypes among Chinese women which result in low sensitivity [14]. Consequently, if only HPV16/18 testing is used, many patients will not be diagnosed and may develop serious lesions. Thus, our results do not support independent HPV testing as the sole means of cervical cancer screening. In contrast to HPV and TCT testing, PAX1 methylation has high and stable sensitivity and specificity which makes it an ideal cervical cancer screening marker.

Despite the low accuracy of HPV testing as independent screening method, we have noted the high sensitivity of hr-HPV and high specificity of HPV16/18 testing. This information has been used to construct a predictive model for improving the diagnostic efficacy of cervical cancer screening. Both tests are combined using logistic regression to build a predictive model. This model improves the specificity and accuracy when compared with PAX1 methylation alone. All variables included in the model can be automatically detected. This provides more reliable method than TCT or other similar cervical cytology tests. Our method is independent of human factors and is thus suitable for regions lacking skilled pathologists. Another model with marginally higher diagnostic efficacy was also constructed as follows:

$$PI = -5.911 + 13.172 \times X21705687 + 3.498 \times X21705946 + 6.371 \times X21706427 + 1.834 \times HPV16/18(1, 0) + 1.791 \times \text{other hr-HPV (1, 0)}$$

(2)

AUC of this model was 0.869, the optimal cut-off value was 0.273, and sensitivity, specificity, PPV, NPV, and accuracy were 0.863, 0.756, 0.571, 0.936 and 0.785, respectively. Neither X21705946 ($p = 0.433$) nor X21706427 ($p = 0.145$) significantly contributed to this model, and thus was not adopted.

5. Conclusions

The diagnostic efficacy of PAX1 methylation was higher than those of HPV DNA testing and TCTs. X21705687 had the largest AUC value compared to other methylation sites. The predictive model combining HPV testing and PAX1 methylation improved the accuracy of cervical cancer predictions. Although further evaluations will be required before this or similar models can be practically applied in cervical cancer screening. HPV52 was the most prevalent genotype among the patients included in this study. This is consistent with the distribution of HPV genotypes among Han Chinese women (the major nationality in China), whereas it is in contrast to global data that has identified HPV16 and HPV18 as the most prevalent genotypes. Therefore, we cannot rule out the possibility of varying results in different populations, which is a major limitation of this study [14, 15]. Nonetheless, the presented approach offers a promising alternative to traditional cervical cancer screening methods.

ABBREVIATIONS

AUC, area under the curve; CIN, cervical intraepithelial neoplasm; CpG, 5-cytosine-phosphate-guanine-3; HPV, human papillomavirus; hr-HPV, high-risk HPV; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight MS; MS, mass spectrometry; MSP, methylation-specific PCR; NPV, negative predictive value; PAX1, paired box-1; PI, prediction index; PPV, positive predictive value; ROC, receiver operating characteristic; TCT, ThinPrep cytology test.

AVAILABILITY OF DATA AND MATERIALS

The datasets used and/or analyzed during the current study are available from our data administrator Yu Chen (chenyu20171@email.szu.edu.cn) upon reasonable request.

AUTHOR CONTRIBUTIONS

ZYY and WLL—designed and supervised the investigation. YC and YYZ—recruited participants and analyzed the data. YC and XCC—wrote the manuscript. YTX, CHL, XYL and JL—collected samples. All authors contributed to the editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the Ethics Committee of the Shenzhen Second People’s Hospital (approval number: KS20190521002-FS2019052903) and registered in the Chinese Clinical Trial Registry (registration number: ChiCTR2000029231).

ACKNOWLEDGMENT

We would like to thank all the participants that were enrolled in this study.
FUNDING
This research was funded by Shenzhen Science and Technology Program (JCYJ20180228163529609); Sanming Project of Medicine in Shenzhen (SZSM201812041); Clinical Research Funding from Shenzhen Second People’s Hospital/the First Affiliated Hospital of Shenzhen University Health Science Center (20193357011).

CONFLICT OF INTEREST
The authors declare no conflict of interest.

SUPPLEMENTARY MATERIAL
Supplementary material associated with this article can be found in the online version, at https://oss.ejgo.net/files/article/1801487262440931328/attachment/Supplementary%20material.docx.

REFERENCES
