Development and Validation of a New HPV Genotyping Assay Based on Next-Generation Sequencing

Xin Yi, PhD,1,2 Jing Zou, MS,1 Jiajia Xu, MS,2 Tao Liu,2 Tengfei Liu,1 Sang Hua,1 Feng Xi, MS,1 Xifang Nie,2 Lili Ye, MS,1 Yufen Luo,1 Lin Xu,1 Hui Du, MD,3 Ruifang Wu, MD,3 Ling Yang, MD,2 Ren Liu, MD,4 Bin Yang, MD, PhD,5 Jian Wang, PhD,1 and Jerome L. Belinson, MD6

From 1BGI Shenzhen, Shenzhen, China; 2BGI Tianjin, Tianjin, China; 3Peking University Shenzhen Hospital, Shenzhen, China; 4Shenzhen Second People’s Hospital, Shenzhen, China; and 5Department of Anatomic Pathology and 6Preventive Oncology International and Women’s Health Institute, Cleveland Clinic, Cleveland, OH.

Key Words: Cervical cancer screening; Self-sampling; HPV genotyping test; Next-generation sequencing (NGS); Ion Torrent PGM; MiSeq

ABSTRACT

Objectives: We developed a new human papillomavirus (HPV) genotyping assay based on multiplex polymerase chain reaction and next-generation sequencing (NGS) methods for large-scale cervical cancer screening.

Methods: We first trained the assay on 1,170 self-collected samples, balancing the cutoff points for high-risk types. Then using 4,262 separate self-collected specimens, we compared concordance, sensitivity, and specificity for cervical intraepithelial neoplasia type 2 (CIN2) or higher and CIN type 3 (CIN3) or higher of the HPV sequencing assay with that of Hybrid Capture 2 (HC2) direct samples and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry assay self-samples.

Results: All assays had a good agreement. The sensitivity for CIN2 or higher and CIN3 or higher of the self-sampling specimens tested with the sequencing assay run on both MiSeq and Ion Torrent Personal Genome Machine sequencer was similar to that of direct-sampling specimens tested with HC2 (P > .05), but the specificity of the sequencing assay for CIN2 or higher and CIN3 or higher was significantly higher than that of HC2 (P < .01).

Conclusions: This population-based study has demonstrated the applicability of a new NGS high-risk HPV assay for primary cervical cancer screening based on self-collection.

Cervical cancer is the third most common cancer in women around the world, accounting for 13% of all cancers. It is the most common cause of death from cancer among young women and women in the developing world. Each year nearly half a million cases of cervical cancer occur around the world, with more than 275,000 deaths. Persistent infection with high-risk types of the human papillomavirus (HR-HPV) is the most important etiologic factor of cervical cancer.

Over the past decade, many population-based cervical cancer screening studies in multiple developing countries have demonstrated that HPV testing is appropriate for primary screening in low-resource settings for women who are at least 30 years of age. In a study in rural India, HPV testing was not only the most effective and reproducible of the screening tests but was less demanding in terms of training and quality assurance.

Over the past 16 years, we have studied self-collection as a way to reach the medically underserved population for cervical cancer screening. Because 85% of the global burden of cervical cancer occurs in developing countries, designing screening programs that would not require an extensive healthcare infrastructure, including individual pelvic examinations for screening, would be an advantage. A vaginal self-collected specimen tested for HR-HPV solves many of these issues because it requires fewer resources than physician-collected samples and is less dependent on a complex healthcare infrastructure than the traditional cytology-based screening programs found in most developed countries. When vaginal self-collected specimens have been compared with endocervical clinician-collected specimens tested for HR-HPV, the self-collected samples have had a lower sensitivity and specificity for high-grade precancers and cancer.
In 2007, the Shanxi Province Cervical Cancer Screening III (SPOCCS III) study discovered that by testing patient samples using polymerase chain reaction (PCR)–based technology (Roche Linear Array, Roche, Basel, Switzerland), a self-collected sample was as sensitive as a physician-obtained specimen. At that time, the results had no clinical usefulness because PCR methods were primarily costly research tools with low throughput.

All of this changed with the Shenzhen Cervical Cancer Screening Study II (SHENCCAST II) completed in 2010 in Guangdong Province, China. In that study, we showed that when using the PCR-based multiplex genotyping assay developed by BGI Shenzhen for mass spectrometry (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry [MALDI-TOF-MS]), a self-collected sample was equal in sensitivity to a physician-obtained specimen. However, in contrast to prior PCR-based technologies, MALDI-TOF-MS had processing speeds 10 to 20 times those of current technologies, with low consumable costs. This translated into a far lower cost per case than even the new assays designed to reach the medically underserved.

From 2007 to 2009, PCR-based technologies made a quantum leap to being clinically relevant and in the center of a healthcare paradigm that was becoming ever clearer. Now we ask the question if sequencing technology can move this paradigm even further. With the new bench-top next-generation sequencing (NGS) instruments currently available, such as the Ion Torrent Personal Genome Machine (PGM; Life Technologies, Carlsbad CA) and MiSeq (Illumina, San Diego, CA), fast progress in DNA sequencing technology has allowed a substantial reduction in costs along with improved accuracy and throughput. It is possible that by using NGS technology, we can drive the costs lower and the applicability even wider because of the cost and transportability of the latest in NGS technology.

To achieve this goal, we designed a high-throughput HPV genotyping assay based on multiplex PCR and NGS technology. We used PGM and MiSeq, and optimized the HPV positive cutoff of this new assay and then validated the assay on population-based self-collected samples.

Materials and Methods

Sample Source for Training and Testing

The study was approved by the institutional review boards of the Cleveland Clinic (Cleveland, OH) and Peking University Shenzhen Hospital (Shenzhen, China). The samples used for training and testing the new HPV sequencing assay came from the stored specimens from the population-based SHENCCAST II study. We selected 9,501 self-collected samples stored at −80°C from patients with complete data for the direct-collected Hybrid Capture 2 assay (HC2, Qiagen, Gaithersburg, MD), the self-collected MALDI-TOF-MS assay, and if indicated, patients who returned for colposcopy and biopsy. In the SHENCCAST II study, colposcopy was performed using the quadrant-based Preventive Oncology International (POI) microbiopsy protocol of direct and random biopsies. Therefore, if patient results were positive on any of the tests done in SHENCCAST II, they had a minimum of five cervical biopsies (including endocervical curettage).

The 9,501 self-collected samples were then randomized into a training set (4,750 specimens) and a testing set (4,751 specimens). Of the 4,750 in the training set (based on samples with MALDI self-test results of those who returned for biopsy if test results were positive), when we added direct HC2 to the analysis, 4,267 met the additional criteria of having HC2 test results and biopsy results when indicated. After using 1,170 training samples, we were confident in the performance of the assay. At that point, the entire testing set (4,267 samples) was tested with the new assay.

Histology results from SHENCCAST II served as the criterion standard. All slides were interpreted by a gynecologic pathologist from Peking University Shenzhen Hospital with whom we have worked for many years, and adjudicated by a gynecologic pathologist from the Cleveland Clinic.

HPV Genotyping Sequencing Assay

We designed a series of unique primers to amplify about 150 base pairs (bp) DNA of the L1 gene, which is a consensus area of the HPV sequence, coding the major HPV virus capsid protein, and a pair of primers to amplify about 150 bp DNA of the human β-globin gene (HBB) (accession numberAY260740) as the internal quality control for identifying the false negatives caused by inadequate DNA or failed PCR.

To improve the throughput of the assay, the PCR primers not only amplify the target DNA, but also provide a unique primer index for each of the 96 samples in each plate. Multiple PCR (multiple index PCR). The result is 96 sets of 10-bp nucleic acid tags at the 5’ end of each HPV and HBB primer (Figure 1). The second index for each sample is the 8-bp nucleic acid tags in the library adapter sequence identifying each sample as coming from a specific 96-well plate (Figure 1). This index was attached to amplicons of the samples through the adapter preparation process (Figure 2, library preparation). Using this “double index system,” thousands of samples can be mixed together and detected in one sequencing chip at the same time.

After sequencing the samples, the raw sequence data were analyzed using in-house software. First, all sequence data were traced back to the specimens according to the sequence of primer indices and adapter indices. Second, the amplicon
sequences of every sample were aligned with the standard references for HPV types and HBB sequences from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). Third, the number of sequences mapped to each reference were counted and the HPV cutoff standardized by using the following formula: \( \text{Cutoff}_{\text{STD}} = \frac{\text{Cutoff}_{\text{origin}}}{(2000 \times S/N)} \), where the S is the total number of samples in the library, and the N is the number of sequences mapped.

**Figure 1** A diagram of the high-throughput human papillomavirus (HPV) genotyping sequencing assay. There are three kinds of DNA in one self-collected sample: human β-globin (HBB) gene DNA (light green), HPV DNA (light blue), and other human DNA (black). In a multiplex index polymerase chain reaction (PCR), HBB index primers that contain HBB gene-specific sequence (“primer,” red) and index sequence (“Index,” purple) bind to the target sequence of HBB DNA (“HBB,” light blue) and amplify the DNA; then HPV index primers consisting of HPV gene-specific sequence (“primer,” dark green) and the same index sequence (“Index,” purple) amplify the target DNA of HPV L1 gene (“HPV,” light green). Thus, the same index is combined to the end of amplicon of HBB and HPV by the PCR reaction. Subsequent library preparation is performed with amplicon containing primer index as well as the adapter (“Adapter,” brown), which was required for attachment to the sequencing chip to prepare the template for sequencing. The last two diagrams were the detailed maps of HPV sequence for massively parallel sequencing. There are one primer index and one adapter index (“primer index,” dark blue) at both sides of the sequence in the MiSeq sequence strategy, but just one adapter index is at one side of sequence in the personal genome machine (PGM) sequence strategy.
to each reference in the library. The original cutoff for HPV was achieved from a receiver operating characteristic (ROC) curve that was constructed using the HPV sequencing data and biopsy results. If the number of the amplicons of a certain HPV type was higher than the CutoffSTD, the result was interpreted as positive for that HPV type and vice versa.

All the HPV genotype sequencing tests were done in the Center of BGI Health clinical laboratory, Shenzhen, China.

HC2 HPV Test

HC2 is an in vitro nucleic acid hybridization assay with signal-amplification using microplate chemiluminescence for the qualitative detection of 13 HR-HPV DNA types, in aggregate in cervical specimens. The test was performed in SHENCCAST II using the manufacturer’s product insert.

MALDI-TOF-MS Genotyping Assay

The PCR-based MALDI-TOF-MS assay is an MS method that uses a multiplex primary PCR with several HPV primers (GP5+/6+) that target type-specific base pairs in the L1 region of the HPV genome for 14 HPV types, followed by a mass extension reaction with a single primer of distinct mass that is also specific for each genotype. The MALDI-TOF-MS–based HPV multiplex assay was designed by BGI Shenzhen and the MassEXTEND method used on the MassARRAY genetic analysis system is from Sequenom (San Diego, CA).

Statistics

The direct endocervical HC2 results and the self-sampling MALDI-TOF-MS results are from the original SHENCCAST II database. They are compared with the results obtained from the stored self-collected specimens with the new sequencing assay on the MiSeq and PGM sequencers. The final diagnoses for the test set were known only to the statisticians at the POI center for biostatistics and epidemiology. To determine concordance we had to account for the fact that the HPV sequencing and MS assays are able to detect 14 HR-HPV types, but the HC2 can only detect 13 HR-HPV types (not HPV66). Therefore we transformed the HPV genotyping results into positive and negative results by using the following criteria: if a sample was positive for at least 1 of the 13 HR-HPVs, ie, types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68, then it was considered positive; otherwise, it was considered negative.

All data analyses were performed using STATA 10.0 (StataCorp LP, College Station, TX). We calculated the concordance rates and k coefficients with 95% confidence intervals (CIs) to determine if the sequencing assays effectively detect all 13 HR-HPV types separately, compared with the MALDI-TOF-MS assay and collectively with the HC2 assay. Sensitivities and specificities with 95% CIs of the HC2 and MALDI-TOF-MS assays, and the two sequencing assays for cervical intraepithelial neoplasia type 2 (CIN2) or
greater and cervical intraepithelial neoplasia type 3 (CIN3) or greater specimens were calculated to estimate whether the sequencing assay is a suitable primary screening method. All CIs are exact and binomial CIs. The specificities for MALDI-TOF-MS and the two sequencing assays were compared with HC2 using the McNemar test on the subset of patients with a negative biopsy finding. Similarly, the sensitivities were compared using a subset of patients with a positive biopsy finding. Because the subset of positive biopsy specimens is so much smaller, the McNemar exact test was used.

The samples used in this study were from women who either received the reference standard (cervical biopsies) or were negative on HPV testing and cytologic examination, which has been shown in previous studies to predict histologically negative findings. Because of this, the aforementioned results were calculated directly in this study. Previously collected biopsy data were combined to generate variables using the highest-grade biopsy, of the multiple biopsies, as the final histologic diagnosis.

## Results

### Optimization of HPV-Positive Cutoff for Self-Collected Samples

After building the methodologic procedures, we used the HPV sequencing results and the corresponding biopsy results of 1,170 self-collected samples from SHENCCAST II to determine the optimal HPV-positive cutoff for self-collected samples. We defined biopsy results of CIN2 or greater (≥CIN2) as positive, and less than CIN2 (<CIN2) as negative. Using different HPV-positive cutoff values, we developed an ROC curve. Finally, we selected the optimal cutoff point that corresponded to a sensitivity and specificity for ≥CIN2 with a MiSeq assay of 94.2% and 87.6%, respectively, and the sensitivity and specificity for ≥CIN2 with the PGM assay of 93.1% and 87.5%, respectively. Figure 3A. Next, in reference to biopsy, HC2, and MS data from the 1,170 self-collected samples, we adjusted the cutoff value of each of the 14 HPV types based on the optimal cutoff point from the ROC curve. For the 1,170 samples, the final sensitivity and specificity for ≥CIN2 was 92.3% and 90.1%, respectively, with the MiSeq assay, and 92.3% and 90.9%, respectively with the PGM assay.

### Study Population and HPV-Positive Rates in Testing Set

In the population-based SHENCCAST II study, 4,267 self-sampled DNA specimens of the test set were analyzed with the new HPV genotyping sequencing assay using the optimal HPV-positive cutoff determined in the training phase. An additional 10 patients were eliminated because 8 had MiSeq/PGM test failures and two had insufficient samples; five of these patients remained in the 4,267, making a final dataset of 4,262 patients with complete data. As shown in Table 1, in 4,163 specimens, the biopsy findings correlated with a status of <CIN2, 34 specimens had CIN2, 60 specimens had CIN3, and five specimens had cervical cancer. The prevalence of ≥CIN2 was 2.32% (99/4,262) and the prevalence of CIN3 or greater (≥CIN3) was 1.53% (65/4,262).

The HPV-positive rates of HC2 direct sampling, MS self-sampling, MiSeq self-sampling, and PGM self-sampling, among the different biopsy results, are summarized in Table 1. For all four assays, HPV-positive rates in higher grades of CIN were greater than those in lower grades; the HPV-positive rate in cervical cancer for all four HPV assays was 100%.

**Figure 3A** Receiver operating characteristic curves for the two human papillomavirus genotyping sequencing assays. FPR, false positive rate (1 – specificity rate); TPR, true positive rate (sensitivity rate).
Concordance of Four HPV Assays

The concordance of the four HPV assays is displayed in Table 2, Table 3, and Table 4. As noted in the “Materials and Methods” section, with the typing data pooled for all the assays using 13 types, the four HPV assays were in agreement in 3,834 of 4,262 cases (89.96%). The κ values of HC2 with the three other assays were similar and all three κ values were greater than or equal to 0.68 (Table 2), thus showing good agreement. The concordance among the MS assay and the two sequencing assays was also similar (Table 3). The total concordance rate and κ value for the two sequencing assays was the highest, reaching 97.75% and κ = 0.89, respectively (Table 4).

Sensitivity and Specificity of Four HPV DNA Tests

The sensitivity and specificity of four HPV assays for ≥CIN2 and ≥CIN3 is displayed in Table 5 and Table 6, respectively.

As shown in Table 5, the sensitivities of the four assays for ≥CIN2 were not significantly different. However, the
specificity of the sequencing assay both on the MiSeq and PGM were significantly higher than that on direct HC2 test. Similar results were found when using ≥CIN3 as the positive cutoff (Table 6).

**Application Comparison of Two Sequencing Platforms**

The two sequencing platforms we used in this study (MiSeq and PGM) are both suitable to be used as an HPV genotyping test [Table 7].

**Discussion**

Two research groups have reported HPV genotyping methods using NGS technology, and both studies showed the

**Table 4**
Concordance Among the Sequencing Assays for 13 HR-HPV Types

<table>
<thead>
<tr>
<th>Comparative Assay</th>
<th>Results</th>
<th>Positive (No.)</th>
<th>Negative (No.)</th>
<th>Total (No.)</th>
<th>Positive Concordance Rate (%)</th>
<th>Negative Concordance Rate (%)</th>
<th>Concordance Rate (%)</th>
<th>κ (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGM-self</td>
<td>Positive</td>
<td>444</td>
<td>61</td>
<td>505</td>
<td>92.69</td>
<td>98.39</td>
<td>97.75</td>
<td>0.89 (0.87-0.91)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>35</td>
<td>3,722</td>
<td>3,757</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>479</td>
<td>3,783</td>
<td>4,262</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval; HPV, human papillomavirus; HR, high-risk; PGM, personal genome machine.

**Table 5**
Sensitivity and Specificity of Four HPV Assays for ≥CIN2

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity (%) (CI)</th>
<th>P Valuea</th>
<th>Specificity (%) (CI)</th>
<th>P Valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC2-direct</td>
<td>98.0 (97.6-98.4)</td>
<td>—</td>
<td>88.8 (87.9-89.8)</td>
<td>—</td>
</tr>
<tr>
<td>MS-self</td>
<td>95.0 (94.3-95.6)</td>
<td>.45</td>
<td>89.0 (88.0-89.9)</td>
<td>.77</td>
</tr>
<tr>
<td>MiSeq-self</td>
<td>92.9 (92.2-93.7)</td>
<td>.13</td>
<td>90.7 (89.8-91.6)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>PGM-self</td>
<td>91.9 (91.1-92.7)</td>
<td>.07</td>
<td>90.1 (89.2-91.0)</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

CI, confidence interval; ≥CIN2, cervical intraepithelial neoplasia 2 or higher; HC2, Hybrid Capture 2; HPV, human papillomavirus; MS, mass spectrometry; PGM, personal genome machine.
a All comparisons were with HC2 direct.

**Table 6**
Sensitivity and Specificity of Four HPV Assays for ≥CIN3

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity (%) (CI)</th>
<th>P Valuea</th>
<th>Specificity (%) (CI)</th>
<th>P Valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC2-direct</td>
<td>98.5 (98.1-98.8)</td>
<td>—</td>
<td>88.2 (87.2-89.1)</td>
<td>—</td>
</tr>
<tr>
<td>MS-self</td>
<td>95.4 (94.8-96.0)</td>
<td>0.63</td>
<td>88.3 (87.3-89.3)</td>
<td>0.72</td>
</tr>
<tr>
<td>MiSeq-self</td>
<td>96.9 (96.4-97.4)</td>
<td>0.99</td>
<td>90.1 (89.2-91.0)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>PGM-self</td>
<td>98.8 (98.1-98.8)</td>
<td>0.99</td>
<td>89.5 (88.6-90.4)</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

CI, confidence interval; ≥CIN3, cervical intraepithelial neoplasia 3 or higher; HC2, Hybrid Capture 2; HPV, human papillomavirus; MS, mass spectrometry; PGM, personal genome machine.
a All comparisons were with HC2 direct.

**Table 7**
Comparison of Two Sequencing Platforms for HPV Genotyping

<table>
<thead>
<tr>
<th>Platform</th>
<th>Read Lengtha</th>
<th>Sequence Throughput, bp/run</th>
<th>Sample/Run</th>
<th>Time/Runb</th>
<th>Labor Demandc</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiSeq (V1)</td>
<td>2 × 150 bp</td>
<td>800 million-900 million</td>
<td>1,248</td>
<td>27h</td>
<td>4-5</td>
</tr>
<tr>
<td>PGM (318 chip)</td>
<td>200 bp</td>
<td>800 million-1,000 million</td>
<td>1,440</td>
<td>45h</td>
<td>3-4</td>
</tr>
</tbody>
</table>

bp, base pair; HPV, human papillomavirus; PGM, personal genome machine.
a MiSeq platform can do the sequencing from both ends of single-strand DNA and the sequence length is 150 bp. PGM platform just does the sequencing from one end of single-strand DNA and the sequence length is 200 bp.
b The sequencers can work continuously. A single run is preceded by polymerase chain reaction and library preparation. The run is followed by data analysis and reporting of results with total laboratory time of 3 to 5 days. The current laboratory maximum throughput is about 6,000 per day.
c Not including DNA extraction.
practicality and advantages of NGS for HPV genotyping. However, the throughput of these two methods was low, so they were limited to research application.

In this study, we focused on the clinical application of a new sequencing technology for cervical cancer screening. We hoped to develop a technology that could overcome some of the major obstacles of throughput, cost, accuracy, and compatibility with self-collection, which is desirable for large-scale population-based screening programs. To increase the throughput of NGS for HPV genotyping, we developed a high-throughput HPV genotyping method based on the “double index system,” which consists of a PCR primer index and a library adapter index. Using this system, thousands of samples can be pooled together and sequenced in a single run. Both of the sequencing platforms that we used in this study, MiSeq and PGM, successfully ran the HPV genotyping test. The maximum throughputs of two platforms were more than 1,200 samples per run (1,248 samples per run with the MiSeq V1, 1,440 samples with the PGM 318 chip). PGM may be more flexible, because it can be used for large-scale sequencing (318 chip, 1,440 samples per run), and is also suitable for routine clinical HPV testing in most hospitals (314 chip, 90 samples per run). With very high volume throughput in the future, we expect to keep the charge for the assay under $5.00 per case. NGS technology, which is based on massively parallel sequencing for single DNA molecules, also resolves the problem that multiple HPV infections coexisting in the same specimen were difficult to detect with conventional Sanger sequencing technology. In addition, it avoids the false positivity of some existing hybridization-based typing systems because of cross reaction with low-risk HPV. These characteristics account for the overall performance of the new NGS HPV test.

The concordance results show a strong correlation between the two sequencing assays and MS assay (MiSeq vs MS: concordance rate = 95.35%, κ = 0.78; PGM vs MS: concordance rate = 95.40%, κ = 0.79) and better than that of HC2 (MiSeq vs HC2: concordance rate = 93.15%, κ = 0.68; PGM vs HC2: concordance rate = 93.20%, κ = 0.69). To ensure the detection accuracy and sensitivity of the HPV genotyping sequencing assay, we designed a series of primers to amplify the HPV DNA based on the common general GP5+/6+ HPV primer and the PCR primers of the MS assay we developed in 2009. Because the two sequencing assays used the same PCR process to amplify the HPV DNA from the same specimens before sequencing, the concordance between these two sequencing platforms is highest (concordance rate = 97.75%, κ = 0.89). This result implies that we can apply this sequencing assay using the same type of sequencer technology and achieve the same result with instruments such as the Illumina HiSeq 2000 and the Life Technologies Ion Proton, which are the high-throughput versions of MiSeq and PGM.

Our previous population-based cervical cancer screening studies (SHENCCAST II and SPOCCS III) indicated that a vaginal self-sampling specimen tested for HR-HPV would have increased sensitivity for ≥CIN3 if a PCR-based assay was used. In SHENCCAST II, the sensitivity for ≥CIN3 of a direct cervical sample tested with HC2 and MALDI-TOF-MS were similar. In addition, a self-collected sample tested with MALDI-TOF-MS was similar to the physician-obtained endocervical sample. In this study, an HR-HPV genotyping sequencing assay used the PCR method to amplify the HPV DNA. We chose self-sampling specimens instead of the direct-sampling specimens for two reasons. First, to validate and confirm that the performance of these new assays was as good as that of the MALDI-TOF-MS assay for self-sampling. The result of this study showed that the sensitivity for ≥CIN2 and ≥CIN3 of the self-sampling specimens tested with this PCR-based HPV genotyping sequencing assay was similar to that of self-collected MALDI-TOF-MS and direct sampling specimens tested with HC2 (P > .05); the specificity for ≥CIN2 and ≥CIN3 of NGS HR-HPV assay was significantly better than that of HC2 (P < .01). Second, we believed that the worldwide impact of this technology would be the greatest if it performed accurately when paired with self-collected specimens.

In the beginning of this study, we tried to adjust the cutoff for each HPV type using ROC curves, but some less oncogenic high-risk types, such as HPV 35, HPV 56, HPV 59, HPV 66, and HPV 68, were unassociated with ≥CIN2 in our sample set. Therefore, it is difficult to plot an ROC curve for each HPV type. So we used HPV-positive/negative data of the NGS assay and the biopsy data to plot the ROC curve and obtained an optimal basic cutoff. Finally, we adjusted the cutoff for each of the HPV types according to the HC2, MS, and biopsy data to improve the specificity. For instance, increasing the cutoff point for some less important HR-HPV types will reduce the positive rate of these HPV types and potentially raise the specificity for ≥CIN2. In addition to the biopsy data, we also referred to HC2 and MS data, which allowed the NGS assay to have a good concordance rate with HC2 and MS.

In conclusion, a sensitive, specific, high-throughput HR-HPV genotyping test based on the “multiplex PCR + double index system + NGS” method was designed. The results of this study have demonstrated its performance in detecting self-collected cervical samples. In addition, because the result of the test is not positive or negative, but type-specific HPV, this allows the selection of high-risk types as a secondary screen to further improve the specificity of the management algorithm.

With the majority of the world’s medically underserved population now living in middle-income countries, it is no longer necessary to think small and use screening assays that are slow, simple, and have poor quality control. With self-collection and sequencing, it is possible to achieve centralized, high-throughput, highly sensitive, low-cost-per-case processing, so
as to reach millions of women in a short period. Again, these concepts need not be restricted to the medically underserved. The clarity of this paradigm is a realistic solution for a critical healthcare problem in the developing world. In addition, as new genomic markers arrive for multiple other malignancies, the “world as a community,” not just the privileged few, will actually benefit from this emerging technology.

**References**


