Molecular Triage of Premalignant Lesions in Liquid-Based Cervical Cytology and Circulating Cell-Free DNA from Urine, Using a Panel of Methylated Human Papilloma Virus and Host Genes

Rafael Guerrero-Preston1,2, Blanca L. Valle1, Anne Jedlicka3, Nitesh Turaga1, Oluwasina Folawiyo1, Francesca Pirini1, Fahcina Lawson1, Angelo Vergura1, Maartje Noordhuis1,4, Amanda Dziedzic5, Gabriela Pérez2, Marisa Renehan1, Carolina Guerrero-Diaz2, Edgar De Jesus Rodríguez2, Teresa Diaz-Montes5, José Rodríguez Orego6, Keimari Méndez2, Josefina Ramaguer2, Bruce J. Trock7, Liliana Florea8, and David Sidransky1

Abstract

Clinically useful molecular tools to triage women for a biopsy upon referral to colposcopy are not available. We aimed to develop a molecular panel to detect cervical intraepithelial neoplasia (CIN) grade 2 or higher lesions (CIN2+) in women with abnormal cervical cytology and high-risk HPV (HPV+). We tested a biomarker panel in cervical epithelium DNA obtained from 211 women evaluated in a cervical cancer clinic in Chile from 2006 to 2008. Results were verified in a prospective cohort of 107 women evaluated in a high-risk clinic in Puerto Rico from 2013 to 2015. Promoter methylation of ZNF516, FKBP6, and INTS1 discriminated cervical brush samples with CIN2+ lesions from samples with no intraepithelial lesions or malignancy (NILM) with 90% sensitivity, 88.9% specificity, 0.94 area under the curve (AUC), 93.1% positive predictive value (PPV), and 84.2% negative predictive value (NPV). The panel results were verified in liquid-based cervical cytology samples from an independent cohort with 90.9% sensitivity, 60.9% specificity, 0.90 AUC, 52.6% PPV, and 93.3% NPV, after adding HPV16-L1 methylation to the panel. Next-generation sequencing results in HPV+ cultured cells, and urine circulating cell-free DNA (ccfDNA) were used to design assays that show clinical feasibility in a subset (n = 40) of paired plasma (AUC = 0.81) and urine (AUC = 0.86) ccfDNA samples obtained from the prospective cohort. Viral and host DNA methylation panels can be tested in liquid cytology and urine ccfDNA from women referred to colposcopy, to triage CIN2+ lesions for biopsy and inform personalized screening algorithms. Cancer Prev Res; 9(12); 915–24. ©2016 AACR.

Introduction

Rapid advances in genomic technologies and The Cancer Genome Atlas project (TCGA) have revealed extensive, previously unsuspected complexity and heterogeneity in human cancers. This heterogeneity has led to calls for "precision medicine" to address the complex genetic and epigenetic processes that underlie somatic cell evolution to malignancy. Precision medicine efforts are underway to understand how contextual, clinical, and molecular data can be integrated to develop personalized treatment, preventive, early detection, and screening programs. Personalized treatment in the post-genomic era is a new creative and scientific endeavor that aims to integrate large datasets and disparate tools and methodologies into a new continuum with clinical and public health relevance. In this article, we use a Phase II Biomarker Development Trial framework to integrate tools from the fields of molecular epidemiology, early detection, and genomics to engineer a precision medicine instrument for cervical cancer screening, based on DNA sequence changes in methylation and mutations in the somatic human and HPV genomes.
Cancer screening is undergoing a major transformation worldwide with the adoption of testing for the presence of oncogenic HPV types (1, 2). Testing for high-risk HPV genotypes (hrHPV) has recently been shown to be a better indicator of cervical cancer risk than the cytology test (3–5). Several countries have changed their cervical cancer screening guidelines, accordingly. In the United States, hrHPV co-testing with Pap is recommended for women 30 years and older (6). In the Netherlands, 5-yearly cytologic screening for women aged 30 to 60 years is being replaced by primary HPV screening. HPV-positive women will be referred to colposcopy in case of abnormal cytology and the screening interval for hrHPV-negative women older than 40 years will increase to 10 years, in an effort to achieve a balance between safety and screening burden (7). However, clinical management of women with hrHPV and abnormal cytology results is not firmly established (8–10) because it cannot predict who is more likely to develop cervical carcinoma (11, 12).

When women test positive for Pap and hrHPV, they are routinely sent to colposcopy, during which most are biopsied. The majority of these biopsy results are either negative or CIN1 lesions, which regress to normal epithelium in less than 24 months (13). Nevertheless, a small percentage of hrHPV infections is associated with progression from low-grade squamous intraepithelial lesions (LSIL) to cervical intraepithelial neoplasia grade 3 (CIN3) lesions and close to 30% of CIN3 lesions progress to cervical cancer (14, 15).

Methylation of host and HPV DNA genes has been examined as a marker of progression in cervical cancer. Several groups have shown links between host DNA methylation or methylated HPV DNA and cervical intraepithelial neoplasia progression, carcinoma in situ, or cervical cancer (CIN2+), or from women with hrHPV (16–25). Some studies have reported a positive association between CIN2+ and methylation of CpG sites in host and viral DNA isolated from liquid cytology samples (16, 25–29). Yet, there are no agreed upon clinical tests of progression (15, 30).

Recent reports have explored the use of urine-based hrHPV testing, as a complementary approach to liquid cytology for cervical cancer screening, in an attempt to identify less invasive screening technologies (31, 32). Most of the studies have failed to attain clinical usefulness, as they are limited by poor sensitivity; inappropriate protocols for DNA extraction from circulating cell-free DNA (ccfDNA) in urine; and the limit of detection of the HPV assays utilized (33–35). Most of the studies also fail to recognize the importance of using DNA isolation methods optimized to enrich for ccfDNA that crosses the kidneys and can be obtained as fragmented DNA in urine (36).

In the current study, we test whether a high-throughput panel of methylated viral and human genes can identify women with CIN2+ lesions in liquid-based cervical cytology and urine ccfDNA. Participants from 2 independent cohorts consented in high-risk cervical cancer clinics in Chile and Puerto Rico provided the samples for this study. The panel is composed of 3 genes we previously found to be frequently methylated in cervical neoplasia, associated with hrHPV status and ethnicity in an unbiased genome-wide study (19). The current retrospective study was performed in DNA extracted from cervical brush biopsies in a Chilean population. The results were confirmed in liquid-based cytology samples obtained from patients enrolled in a prospective study in Puerto Rico, adding HPV16-L1 gene methylation to increase sensitivity. The panel was also tested in ccfDNA isolated from paired liquid-based cytology, serum, and urine samples, obtained from the prospective cohort in Puerto Rico. Next-generation sequencing (NGS) platforms were used to produce personalized maps of the HPV genome and epigenome from urine ccfDNA and to populate cloud-based servers that can be used to track hrHPV DNA throughout the life course, as we move toward personalized cervical cancer screening solutions.

Materials and Methods

Patient samples

Cervical brush, liquid-based cytology, serum/plasma, and urine samples were obtained from collaborators in Chile and Puerto Rico, under the Johns Hopkins University School of Medicine Institutional Review Board (IRB)-approved protocol #NA_00020633. The IRB of the Doctor Hernán Henríquez Aravena Tertiary Care Regional Hospital, in Temuco, Chile, and IRB of the University of Puerto Rico School of Medicine also approved this protocol. Patients in both cohorts were treatment-naïve, had never participated in a clinical study, or had a history of cancer. Patient characteristics are listed in Supplementary Table S1.

Phase II biomarker development trial for molecular markers of CIN2+

We designed a phase II biomarker development trial to identify methylated biomarkers of CIN2+ in cervical epithelium and urine samples. We selected 3 genes FKBP6, INTS1, and ZNF516 that we previously showed were associated with cervical cancer and abnormal cytology in a phase I Biomarker Development Trial (Supplementary Fig. S1A; ref. 19). To support the use of these targeted genes in a clinical assay, we first performed immunohistochemistry (IHC) to identify the differential distribution and localization of the expressed proteins in normal cervical mucosa samples and cervical carcinoma tissue samples from Chile. The IHC results revealed the 3 genes, FKBP6, INTS1, and ZNF516, were differentially expressed in normal and cervical cancer epithelium (Supplementary Fig. S1B). We then used quantitative methylation-specific PCR (qMSP) to examine the association between CIN2+ biopsies and methylation of these 3 genes (ZNF516, FKBP6, and INTS1) in cervical brush epithelium from a retrospective cohort of women in Chile. We subsequently verified the association between CIN2+ biopsies and methylation of FKBP6, INTS1, ZNF516, and HPV16-L1 in liquid-based cytology samples obtained from an independent prospective cohort of women in Puerto Rico (Supplementary Fig. S1C). An overview of the workflow for tissue processing, including nucleic acid extraction, prior discovery using MeDIP-chip, IHC confirmation of targeted biomarkers, NGS, high-throughput PCR quantitation of genomic and bisulfite-treated DNA, and cloud-based alignment and visualization tools is provided in Supplementary Fig. S1D and described in detail in the Supplementary Experimental Procedures.

Sequencing the hrHPV genome in urine ccfDNA

We isolated urine ccfDNA using Q-Sepharose anion exchange resin followed by a silica-based elution with LCL, as described in the Supplementary Experimental Procedures. To examine the hrHPV genome in urine ccfDNA, we developed custom dual-sequence capture baits to enrich samples for hrHPV DNA from 12 high-risk types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59). We used liquid-based sequence capture baits to measure the hrHPV genome in urine ccfDNA from women with and without cervical premalignant lesions. DNA sequencing libraries were...
prepared following published protocols and enriched using custom capture reagents. HPV-specific, biotinylated, long oligonucleotide probes were designed, synthesized and pooled for target selection and enrichment utilizing the SeqCap EZ (Roche/NimbleGen) dual capture approach for gDNA enrichment. To assess the enrichment efficiency of the dual-sequence capture method on clinical samples, we compared hrHPV ccfDNA-qPCR assay results obtained on 8 pre- and postcapture HPV-enriched circulating DNA samples obtained from patients with CIN1 (n = 4) and CIN2-3 (n = 4), as described in Supplementary Experimental Procedures.

We used massively parallel NGS to quantify the different HPV genotypes present in urine ccfDNA from 7 patients with CIN1 (n = 3) and CIN2-3 (n = 4) lesions, using DNA from 2 cervical cancer cell lines (HeLa and CSCC7) as positive controls. The DNA was enriched for hrHPV DNA with the custom dual-sequence capture hrHPV assay prior to multiplexed sequencing on a GS Junior Plus 454 (Roche) system. The Roche GS Analysis Software Suite was used to perform signal processing, QC, and initial alignments of the gDNA sequencing reads. To determine the HPV genotype, composition reads were aligned to the PapillomaVirus Episteme (PAVE) genome database of HPV genomes (37) and to custom servers that can be used for personalized tracking of HPV genotypes longitudinally.

Sequencing and visualizing the hrHPV epigenome in urine ccfDNA
To examine the hrHPV genome in urine ccfDNA, we used the Sure Select Methyl-Seq Target Enrichment (Agilent) assay to enrich DNA samples from 2 HPV16-positive cervical cancer cell lines, CaSkii (ATCC CRL-1550, 600 integrated HPV16 copies) and SiHa (ATCC HTB35, 2 integrated HPV16 copies); 2 HPV16-positive head and neck squamous cell carcinoma (HNSCC) cell lines, SCC-47 and SCC-90; and urine ccfDNA from 2 clinical samples: one from a patient with ASCUS and CIN1 (TrdNA47) and another one from a patient with high-grade cervical SIL (HSIL) and CIN3 (TrdNA50). Samples were PCR-amplified using sample-specific indexed (“barcoding”) primers for multiplexed sequencing on a MiSeq (Illumina) system. The bisulfite-converted DNA reads were analyzed using the Bismark version 0.13.0 suite of tools (38) and Bowtie 2 version 2.1.0. Reads from the FASTQ files were aligned to all HPV types in the PAVE database, before performing the CpG methylation analysis of hrHPV genomes.

Identification of hrHPV and methylation biomarkers of CIN2+ in urine ccfDNA
We designed quantitative PCR amplification assays for β-actin and the HPV E1 region to detect hrHPV in plasma/serum and urine ccfDNA. This ccfDNA hrHPV qPCR assay amplifies an HPV E1 region common to 13 hrHPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68). The resulting amplicon is 93 bp. Primers were obtained from Invitrogen. We also designed qMSP primers to test DNA methylation of host (ZNF516, FKBP6, and INTS1) and viral DNA (HPV16-L1) in urine ccfDNA.

Bioinformatics and cloud-based tools for personalized cervical cancer screening
A suite of bioinformatics tools was developed to analyze the HPV genotype composition and quantify the HPV methylation in ccfDNA. The Supplementary Experimental Procedures describe how each clinical sample sequence was aligned against the non-redundant PAVE genome database. Cloud-based tools were developed to provide alignment and visualization tools that can be used by patients, providers, insurance personnel, and researchers.

We used A suite of bioinformatics tools was developed to analyze the HPV genotype composition and quantify the HPV methylation in ccfDNA. The Supplementary Experimental Procedures describe how each clinical sample sequence was aligned against the non-redundant PAVE genome database. Cloud-based tools were developed to provide alignment and visualization tools that can be used by patients, providers, insurance personnel, and researchers.

Results
Identification of methylation biomarkers of CIN2+ in cervical brush and liquid-based cytology samples
After removing samples without a definitive biopsy result, we examined the promoter methylation frequency of 3 genes (ZNF516, FKBP6, and INTS1) in cervical brush samples from women with normal (n = 34), CIN1 (n = 34), CIN2 (n = 33), CIN3 (n = 20), and cervical cancer (n = 90) pathology reports. We compared samples from women with no intraepithelial lesions or malignancy (NILM) and CIN1 lesions with samples from women with CIN2+ lesions and found that ZNF516 has 91.7% sensitivity, 27.4% specificity, and an AUC of 0.62; FKBP6 has 92.4% sensitivity, 46.8% specificity, and an AUC of 0.68; INTS1 has 93.8% sensitivity, 30.3% specificity, and an AUC of 0.57. We then compared samples from women with NILM to samples from women with CIN2+ lesions, we found that ZNF516 has 91.7% sensitivity, 38.9% specificity, and an AUC of 0.76; FKBP6 has 90.9% sensitivity, 67.6% specificity, and an AUC of 0.86; INTS1 has 90.7% sensitivity, 40.5% specificity, and an AUC of 0.62.

The panel of 3 classifiers (ZNF516, FKBP6, and INTS1) has 90% sensitivity, 88.9% specificity, an AUC of 0.93, a positive predictive value (PPV) of 93.1%, and a negative predictive value (NPV) of 84.2%, when comparing women with NILM with women with CIN2+ lesions (Fig. 1A and B and Table 1).

We then tested this panel of 3 classifiers in liquid-based cytology samples (n = 67) from women in Puerto Rico, a subset of which had been previously tested for concordance of hrHPV genotype between cervical cytology and urine DNA (31). We compared samples from women with NILM and CIN1 lesions with samples from women with CIN2+ lesions and found that...
ZNF516 has 72.7% sensitivity, 48.1% specificity, and an AUC of 0.63; FKBP6 has 63.6% sensitivity, 34.6% specificity, and an AUC of 0.50; INTS1 has 91% sensitivity, 35% specificity, and an AUC of 0.66. We also tested the performance of HPV16-L1 methylation, using previously designed primers and probes (40), and found it had 63.6% sensitivity, 100% specificity, and an AUC of 0.79.

The panel of 4 classifiers in cervical cytology samples, ZNF516, FKBP6, INTS1, and HPV16-L1 has 90.9% sensitivity, 60.9% specificity, an AUC of 0.90, a PPV of 52.6%, and a NPV of 93.3% when comparing NILM with CIN2+ lesions (Fig. 1D and Table 2).

Genotype profiling of the hrHPV genome in urine ccfDNA
Supplementary Figure S3A shows an amplification plot demonstrating the successful amplification and enrichment obtained with the dual-sequence capture assay. AC, values for HeLa and CSCC7 were 14.88 and 12.66, respectively. On the basis of an

Table 1. Sensitivity, specificity, AUC, PPV, and NPV of ZNF516, FKBP6, and INTS1 promoter methylation in cervical brush samples comparing CIN2+ versus NILM/CINI and CIN2+ versus NILM

<table>
<thead>
<tr>
<th>Lesion comparison</th>
<th>Marker</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>AUC, %</th>
<th>PPV, %</th>
<th>NPV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN2+ vs. NILM/CINI</td>
<td>ZNF516</td>
<td>91.7</td>
<td>27.4</td>
<td>61.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBP6</td>
<td>92.4</td>
<td>46.8</td>
<td>68.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTS1</td>
<td>93.8</td>
<td>30.3</td>
<td>56.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-gene panel</td>
<td>60</td>
<td>74.5</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIN2+ vs. NILM</td>
<td>ZNF516</td>
<td>91.7</td>
<td>38.9</td>
<td>76.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBP6</td>
<td>90.9</td>
<td>67.6</td>
<td>85.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTS1</td>
<td>90.7</td>
<td>40.5</td>
<td>62.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-gene panel</td>
<td>88.3</td>
<td>88.9</td>
<td>93.2</td>
<td>86.9</td>
<td>90.2</td>
<td></td>
</tr>
</tbody>
</table>
estimated efficiency for the assay, the approximate fold enrichment was greater than 1,700. As shown in Supplementary Fig. S3B with data representing 5 clinical samples, the average ΔCt value for precapture versus postcapture samples enriched for hrHPV ccfDNA was 11.07, for an average fold enrichment of 670 of hrHPV ccfDNA in clinical samples. We then examined whether the hrHPV ccfDNA-qPCR assay could discriminate urine ccfDNA isolated from women, with and without cervical dysplasia. Supplementary Figure S3C shows the amplification curves for the hrHPV ccfDNA-qPCR assay on ccfDNA from CIN2−3 (n = 14), CIN1 (n = 13), and 10 samples from women with NILM. The frequency of amplified premalignant lesion samples differed significantly from the NILM samples: NILM (30%), CIN1 (77%, P = 0.02), and CIN2−3 (71%, P = 0.04).

**Sequencing the hrHPV epigenome in urine ccfDNA**

The multiplexed massively parallel sequencing runs produced 230,385 reads with an average length of 138.3 bp. The reads in the clinical samples covered 82% to 100% of the reference HPV16 and 73% to 100% of the HPV18 genomes, with an average of 21% (HPV16) to 33% (HPV18) percentage of all reads mapping to the sequences. These results were comparable to the percentage of all reads that mapped to the reference sequences in the positive control samples: 92% of the CESC7 reads mapped on target covering 77% of the HPV16 genome and 89% of the HeLa reads mapped to 67% of HPV18, as expected (Supplementary Table S2). Samples TrDNA-445 and TrDNA-455 had more than 80% of their reads mapped to HPV16 and HPV18; however, the remaining 5 samples were split, mapping to both HPV16 and HPV18 genotypes, as well as other subtypes (Supplementary Table S2B).

To resolve the remaining reads, we searched the human genome using the program Bowtie (41) and then a local copy of the database of NCBI reference bacterial genomes, using BLAST (43). Following this tiered mapping approach, only a small number of reads were still unmapped, a small number of which (<50) were linker contaminants, whereas the others could potentially represent novel HPV genotypes or other viruses or bacteria that comprise the human microbiome. Supplementary Table S2C lists the reads mapping to 21 HPV types after profiling reads of the clinical sample TrDNA-456 by tiered read mapping. Supplementary Figure S4 shows the percentage of reads that map to 13 HPV types, human, bacteria, and unknown genomes for clinical sample TrDNA-456.

**Cloud-based visualization tools for personalized cervical cancer screening**

Our demonstration of cloud-based servers shows the alignment results of 11 hrHPV types, 9 low-risk HPV types, and 7 clinical samples (TrDNA-445, TrDNA-455, TrDNA-456, TrDNA-481, TrDNA-504, TrDNA-513, and TrDNA-571) against a reference genome (HPV16). The large-scale HPV DNA server (43) produces a graphical “large-scale” view of the pairwise alignments of 20 HPV genomes against a reference genome (HPV16 in this demo), together with annotations of genome rearrangement events (Supplementary Fig. S5A). The “close-up” HPV DNA server (44) computes and displays nucleotide-level (“close-up”) multiple alignments of sequences in a 1-kb region starting at a user-specified address or gene in the reference genome (Supplementary Fig. S5B).

**Sequencing the hrHPV epigenome in urine ccfDNA**

Reads from the FASTQ files were aligned to all HPV types in the PAVE database and we performed CpG methylation analysis using Bismark modified to analyze hrHPV genomes. The multiplexed massively parallel sequencing run produced 14,442,406 reads with an average length of 100 bp. The percentage of all reads of the HPV-positive cervical cancer and HNSCC cell lines that mapped uniquely to some of the reference genomes in the PAVE database can be seen in the top of Supplementary Table S3: CaSki (93%), SiHa (13%), SCC-047 (68%), and SCC-090 (87%).

**Personalized HPV methylation landscapes**

The percentage of methylation across the HPV genome in all 6 samples was obtained with the Methylator Extractor module in Bismark (See Supplementary Experimental Procedures for a detailed description). Scatterplots of the percentage of methylation by chromosomal location in the HPV genome for each of the 6 samples are shown in Fig. 2. The percentage of methylation is shown on the y-axis of the top panel. The x-axis shows the chromosomal location along the HPV genome for both panels, including the promoters at positions 97 and 670 of the HPV genome. The top panel of each plot represents the percentage of Cpg methylation, whereas the bottom panel of each plot represents the HPV genes and the upper regulatory region (45).

The different HPV genomes that aligned to the 6 samples are shown in different colors and shapes (HPV16, black dots; HPV35, red squares; HPV52, red triangles; and HPV71, green squares). The 2 cervical cancer cell lines had different patterns of Cpg methylation. CaSki had overall higher levels of methylation than SiHa for Cpgs across the genome. SiHa exhibits a bimodal distribution of methylation percentage. The majority of CpGs below the 3,500 position in the HPV genome have less than 60% methylation, whereas CpGs located between 3,500 and 7,200 positions on the HPV genome show more than 80% methylation. The methylation patterns for both HPV-positive HNSCC cell lines were very similar to those observed in CaSki. The clinical samples aligned to more...
Figure 2.
Results from sequence capture of methylated hrHPV. 

A, complete hrHPV methylome of cervical cancer cell lines CaSki (600 HPV16 copies) and SiHa (2 HPV16 copies).

B, complete hrHPV methylome of HNSCC cell lines SCC-047 (HPV16+) and SCC-090 (HPV16-).

C, complete hrHPV methylome of urine ccfDNA clinical samples from a patient with a cytologic diagnosis of atypical squamous cells of undetermined origin (ASCUS) and a CIN1 lesion diagnosed in biopsy (TrDNA-34) and a patient with a cytologic diagnosis of HSIL and a CIN3 lesion in biopsy (TrDNA-50). All the samples aligned to the HPV16 methylome. The 2 clinical samples also aligned to additional HPV types. TrDNA-34 aligned to HPV16 and HPV35. TrDNA-50 aligned to HPV16, HPV52, and HPV71. Note that the methylation pattern of HPV16 in CaSki is similar to all the HPV16 patterns with the exception of SiHa, which is lower and may be due to the fact that it only has 2 copies of HPV16.
than one HPV type: TrDNA-34, a sample obtained from a patient with ASCUS and CIN1, aligned to HPV16 and HPV35. TrDNA-50, a sample obtained from a patient with HSIL and CIN3, aligned to HPV16, HPV52, and HPV71. The methylation patterns of HPV16 in both clinical samples are very similar to the methylation patterns observed in CaSkI and both HNSCC cell lines, albeit with less abundant number of reads, as expected. Methylation of the remainder of HPV types was low overall.

To examine the HPV16 CpG methylation patterns, we aligned the reads from the 4 cell lines to the HPV16 reference genome. The mapping efficiency (the percentage of total reads that aligned uniquely to the reference genome) to the HPV16 reference genome was very similar to the percentage of all reads that mapped to the PAVE reference database for the 4 cell lines: CaSkI (91%), SiHa (13%), SCC-047 (66%), and SCC-090 (86%). The mapping efficiency of the reads from the HNSCC cell lines to the HPV16 reference genome was in the range between the mapping efficiency obtained with SiHa and CaSkI, namely, 86% for SCC-90 and 66% for SCC-47 (Supplementary Table S3, middle). Because we know the number of copies of HPV16 DNA in SiHa (2) and CaSkI (~600), these sequencing results may be a good indicator of the number of HPV copies present in the HNSCC cell lines.

We then examined the mapping efficiency of the HPV16-L1–specific CpG methylation patterns in the 4 cell lines and found that with the exception of SiHa, (3%), there was hardly a difference in the percentage of uniquely aligned reads: CaSkI (18%), TrDNA34 (15%), and TrDNA50 (15%) (Supplementary Table S3, bottom). The mapping efficiency to the PAVE database was orders of magnitude less in clinical samples when compared with the lowest mapping efficiency obtained in cell lines (SiHa): SiHa (240,913 reads), TrDNA34 (1,848 reads), and TrDNA50 (4,869 reads), as can be seen in the top panel of Supplementary Table S3. The mapping efficiency of the clinical samples when aligned to the HPV16 reference database was also orders of magnitude less than the one seen in SiHa: SiHa (236,809 reads), TrDNA34 (1,729 reads), and TrDNA50 (1,078 reads) as can be seen in the middle panel of Supplementary Table S3. This difference in mapping efficiency was also observed when we aligned the clinical samples to the HPV16-L1 region: SiHa (46,872 reads), TrDNA34 (419 reads), and TrDNA50 (210 reads), as can be seen in the bottom panel of Supplementary Table S3.

We wanted to assess whether methylation levels in the HPV16-L1 variable region could be used as a marker of progression in cervical cancer premalignant lesions. The mapping efficiency of the 4 cell lines to the HPV-16-L1 gene was high: CaSkI (81%), SiHa (80%), SCC-47 (78%), and SCC-90 (82%). The mapping efficiency of the clinical samples to the HPV16-L1 gene was as high as for the positive controls: 76% for TrDNA34 and 81% for TrDNA50 (Supplementary Table S4).

To further determine whether HPV16-L1 methylation levels can be used as a surrogate marker of methylation of the HPV16 genome in urine cfDNA, we examined the distribution of CpG methylation after aligning the urine cfDNA samples to the HPV16-L1 gene. Box plots show the distribution of CpG methylation levels per sample after aligning to HPV16 (Supplementary Fig. S6A) and HPV16-L1 region (Supplementary Fig. S6B). The CpG methylation median in the clinical samples is significantly higher than in the cell lines and higher in urine cfDNA from the CIN3 than from the CIN1 clinical sample (P < 0.05), as expected.

Quantification of viral and host DNA methylation in plasma and urine cfDNA

To enable the testing of this 4-gene panel in urine cfDNA, we optimized a previously published urine cfDNA isolation method and compared it with the gold standard, phenol–chloroform DNA extraction method (Supplementary Table S5). We then designed primers and TaqMan probes to quantify HPV16-L1 DNA methylation and ZNF516, INTS1 and FKBP6 methylation in fragmented urine cfDNA using qMSP. Primers were designed to amplify in urine cfDNA short amplicons (80 basepairs long) of the same genomic regions previously used to quantify HPV16-L1 methylation in HNSCC (40) and ZNF516, INTS1, and FKBP6 in cervical cancer (19).

In a feasibility study, we found that HPV16-L1 qMSP methylation can discriminate bisulfite-treated cervical epithelium DNA and urine cfDNA from women with normal cytology from women with dysplastic cytology and premalignant cervical lesions with high sensitivity and specificity (Supplementary Fig. S7).

We then quantified the methylation levels of the panel of viral and host DNA genes in plasma and urine cfDNA samples. In plasma, we found that the panel of 4 classifiers has 85.7% sensitivity, 60.9% specificity, an AUC of 0.807, PPV of 40%, and an NPV of 93.3%, when comparing women with NILM/CIN1 to CIN2+ lesions (Fig. 3A). In urine cfDNA, we found that the panel of 4 classifiers has 75% sensitivity, 83.3% specificity, an AUC of 0.86, PPV of 50%, and an NPV of 93.8% when comparing women with NILM/CIN1 to CIN2+ lesions (Fig. 3B).

Discussion

We used the formal criteria biomarker development created by the National Cancer Institute Early Detection Research Network (EDRN) to guide this research project. The goals and primary aims for the EDRN, that is, the 5 phases of Biomarker Development Trials (46), are listed in Supplementary Table S6. The main goal of Phase II Biomarker Development Trials is to identify a clinical assay that can detect the desired endpoint. We performed a Phase II Biomarker Development Trial to identify a panel of methylated HPV and human host genes that can discriminate between CIN2+ and normal/CIN1 lesions in a reflex test performed in cervical cytology samples and in cfDNA in plasma and urine. Recognizing the previously unsuspected complexity and heterogeneity in human cancers unveiled by rapid advances in genomic technologies, we adopted specific strategies from the fields of molecular epidemiology, early detection, and genomics to engineer a precision medicine approach for cervical cancer screening. The specific tools include DNA sequence–based assays, such as qMSP, bisulfite genomic sequencing, and NGS, all of which generate complex datasets reflecting the complexity of neoplastic evolution in human tissues. We also tested urine samples using methods that optimize DNA isolation to enrich for cell-free fragmented DNA that is excreted in the urine, which opens a novel approach to early detection.

We are the first to show that a panel of host and viral DNA methylation markers can discriminate between CIN2+ and NILM in multiple body compartments from the same individual: cervical cytology, serum, and urine. Our results suggest that a precision medicine panel can be used as a reflex test in cervical cytology to triage women referred to colposcopy. NGS reads from urine cfDNA can be aligned in custom cloud-
methylation in plasma ccfDNA samples from patients diagnosed with CIN2
NILM/CIN1 lesions.

Based servers for life-course personalized cervical cancer screening.

Women with low probability of having a CIN2- lesion can be triaged out of a biopsy after colposcopy. The 4-gene classifier best performed in liquid-prep with sensitivity of 90.9%, specificity of 60.9%, and NPV (93.3%). In urine ccfDNA, the 4-gene classifier had equal NPV (93.8%), a better specificity (83.3%), and similar AUC (0.861). The results obtained for this classifier in cervical cytology and urine ccfDNA warrant further study of this panel as a molecular biomarker to triage women referred to colposcopy after testing positive for hrHPV and being diagnosed with cervical dysplasia with cytology. Women with low methylation values in this panel would be asked to return for follow-up cytology and HPV co-testing in 6 to 12 months, if the colposcopists do not see a clear indication of a lesion that should be biopsied. This would decrease the number of blind biopsies that are currently being performed, decreasing screening costs and increasing health care quality.

There is growing evidence that circulating short human, viral, and bacterial DNA fragments from dying cells throughout the body, approximately 150 to 250 bases long on average, pass through the renal barrier and can be isolated as ccfDNA in urine. These fragments are known as transrenal DNA (TrDNA) (47–49). Recently, a capillary electrophoresis TrDNA test that targets the E1 region of the HPV genome for the detection of hrHPV demonstrated high sensitivity and modest specificity for urine-based detection of cervical precancerous lesions (50).

Urine samples tested by the TrDNA HPV capillary electrophoresis test had high concordance with corresponding cervical and urine samples tested by the widely used linear array HPV genotyping test (51). However, the TrDNA capillary electrophoresis HPV test is not quantitative, can only detect the presence or absence of hrHPV and, similar to previously published urine-based HPV tests, has limited specificity. None of the hrHPV urine-based reports use sequencing-based approaches to quantify multiple HPV types, nor include methylated markers in their workflow to improve the sensitivity and specificity of the TrDNA test.

We quantified hrHPV and the hrHPV methylome in urine ccfDNA using custom sequence capture, which allow for multiplexed massively parallel sequencing of clinical samples and qPCR verification. Our results also show that qMSP quantification of HPV16-L1 methylation is a surrogate of genome-wide HPV16-L1 methylation. Furthermore, urine ccfDNA assays discriminate women with CIN2+ lesions when compared with women who do not have cervical intraepithelial lesions or malignancy, which could lead to high-throughput testing of urine ccfDNA by qPCR, digital PCR, or multiplexed massively parallel sequencing.

Although privacy, cultural, and infrastructure issues challenge the implementation of HPV testing for cervical cancer screening, several countries have already implemented HPV testing in their screening protocols. Co-testing with cytology and HPV at 5-year intervals is now the preferred or acceptable strategy for cervical cancer screening for women aged 30 to 64 years in the United States. The Netherlands is implementing this year landmark changes in the cervical cancer screening algorithms. Clinical management for HPV-positive/Pap-negative women, however, is not firmly established among practicing physicians. There is resistance from the some in the medical community to accept HPV-PAP co-testing for cervical cancer, due to the complex risk patterns associated to positive, negative, and undetermined cytology, with positive and negative HPV results.

Clinical detection of HPV is typically performed by in vitro diagnostic assays that detect viral gDNA or RNA on the same cervical mucosa samples collected for the cervical cytology test. However, because HPV infections are very common and because most women will clear HPV infections within 6 to 12 months, the presence of hrHPV DNA does not mean that cervical dysplasia or cervical cancer is present or that the infection will persist and the patient will progress to cervical cancer (12). Furthermore, cervical cancer screening programs currently in use are inefficient at identifying individuals at risk for disease, requiring multiple visits over a women’s lifetime, which is costly and cumbersome (52). New methods for cervical cancer screening and triage, which provide accurate, efficient and cost-effective ways of identifying women at risk for cervical cancer, will improve screening and treatment efforts worldwide.

The primary aim of Phase II Biomarker Development Trials is to estimate sensitivity, specificity, and/or receiver operator characteristics (ROC) curves for the clinical biomarker assay and not to report the data as percent agreement in 2 x 2 format (marker ± by CIN2+/CIN2; ref. 46). ROC curves have 2 main advantages over frequencies and summary statistics for biomarker data: (i) ROC

...
curves do not depend on the scale of raw data measurements, which greatly facilitates comparison of the discriminatory capacities of different biomarkers; and (ii) ROC curves display true- and false-positive rates, quantities that are more relevant for screening purposes than the raw biomarker values themselves.

This Phase II Biomarker Development Trial has some limitations. The 4-gene panel assay provides acceptable discrimination in liquid cytology samples when excluding patients with CIN1 lesions (75% sensitivity and 93.8% NPV) and in urine (75% sensitivity and 93.8% NPV). These results suggest that more work is needed to improve the performance of the biomarkers and move this technology forward, before we are ready to perform a Phase III Biomarker Development Trial in urine. To improve this performance, we are redesigning our primers to shorten our amplicons and increase the number of CpGs queried in our primers and probes. We will also develop new primers to query a larger region of the promoter of ZNF516, FBKP6, and INTS1, with overlapping primers and probes. In addition, we can query other regions of the HPV16-L1 gene, as well as design primers for other HPV16 genes and for HPV18, HPV33, and HPV51 genes.

Moving forward, we will design personalized screening assays, on the basis of resequencing the human and the HPV methylomes, to customize precision screening markers applicable to population subsets and individual patients, as we move forward with the creation of personalized medicine tools for screening, early detection, diagnosis, and cancer prevention.

In sum, our results reveal that patients harbor differentially methylated loci in multiple HPV genotypes and host genes, which can be identified with PCR and NGS technologies in cervical cytology and urine ccfDNA. Because they are bound to change overtime, these molecular panels can be used to personalize cervical cancer screening algorithms that complement co-testing with hrHPV tests, cervical cytology, and colposcopy. The development of custom urine ccfDNA assays can lead to a new generation of personalized cervical cancer screening algorithms, which can be tracked using cloud-based technologies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: R. Guerrero-Preston, N. Turaga, O. Folawayo, J.R. Orengo, D. Sidransky

Development of methodology: R. Guerrero-Preston, A. Jedlicka, O. Folawayo, B.J. Trock

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Guerrero-Preston, B.L. Valle, A. Jedlicka, F. Pirini, F. Lawson, A. Dzedzic, G. Perez, M. Renehan, E. De Jesus Rodriguez, T. Diaz-Montes, K. Mendes, J. Romaguela

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Guerrero-Preston, B.L. Valle, N. Turaga, O. Folawayo, F. Lawson, A. Vergura, B.J. Trock, L. Florea, D. Sidransky

Writing, review, and/or revision of the manuscript: R. Guerrero-Preston, A. Jedlicka, F. Lawson, M. Noordhuis, G. Perez, M. Renehan, T. Diaz-Montes, J.R. Orengo, J. Romaguela, B.J. Trock, L. Florea

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R. Guerrero-Preston, O. Folawayo, F. Lawson, K. Mendes

Study supervision: R. Guerrero-Preston, O. Folawayo, J.R. Orengo

Other (conduction of experiments: PicoGreen; bisulfite conversion; and bisulfite sequencing on some samples): C. Guerrero-Diaz

Acknowledgments

The authors thank Drs. Priscilla Brebi and Juan C. Roa for providing the cervical brush and urine samples from patients in Temuco. We also appreciate the generous input of Casey Matthews and Dwanye Dexter from Roche/ NimbleGen.

Grant Support

This work was supported, in part, by EDRN Associate Member Supplement NCIU01 CA084986(to R. Guerrero-Preston); and NCIU01 CA086986(to D. Sidransky).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 19, 2016; revised August 23, 2016; accepted September 7, 2016; published OnlineFirst September 26, 2016.

References


Guerrero-Preston et al.


Molecular Triage of Premalignant Lesions in Liquid-Based Cervical Cytology and Circulating Cell-Free DNA from Urine, Using a Panel of Methylated Human Papilloma Virus and Host Genes

Rafael Guerrero-Preston, Blanca L. Valle, Anne Jedlicka, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-16-0138

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2016/09/24/1940-6207.CAPR-16-0138.DC1

Cited articles
This article cites 50 articles, 5 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/9/12/915.full#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cancerpreventionresearch.aacrjournals.org/content/9/12/915.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.