Gynecologic Oncology 138 (2015) 94–100

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

HPV16 CpG methyl-haplotypes are associated with cervix precancer and cancer in the Guanacaste natural history study

Marina Frimer [a,*], Chang Sun [b], Thomas McAndrew [c], Benjamin Smith [d], Ariana Harari [d], Zigui Chen [b], Lisa Mirabello [e], Nicolas Wentzensen [e], Gary L. Goldberg [a], Ana C. Rodriguez [f], Mark Schiffman [e], Robert D. Burk [b,c,d]

[a] Division of Gynecologic Oncology, Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine/Montefiore Medical Center, United States
[b] Department of Pediatrics, Albert Einstein College of Medicine, United States
[c] Department of Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, United States
[d] Department of Microbiology and Immunology, Albert Einstein College of Medicine, United States
[e] Division of Cancer Epidemiology and Genetics, National Cancer Institute, United States
[f] Proyecto Epidemiologico Guanacaste, Fundacion INCIENSA, San Jose, Costa Rica

HIGHLIGHTS

- Next Gen sequencing is an efficient method to quantitate HPV16 CpG methylation.
- Methyl-haplotypes of HPV16 methylation are associated with cervical precancer.
- L2 methylation increases over time in HPV16 infections leading to precancer.

ARTICLE INFO

Article history:
Received 27 February 2015
Accepted 2 May 2015
Available online 19 May 2015

Keywords:
HPV16
DNA methylation
Cervix cancer
Next generation sequencing

ABSTRACT

Objective. To evaluate HPV16 CpG methylation and methyl-haplotypes and their association with cervix precancer and cancer utilizing massively parallel single molecule next-generation sequencing (NGS).

Methods. A nested case–control study of HPV16 positive women was performed in a prospective cohort from Guanacaste, Costa Rica designed to study the natural history of HPV and cervical neoplasia. Controls encompassed 31 women with transient infections; there were 44 cases, including 31 women with CIN3 and 13 with cervical cancer. DNA samples from exfoliated cervical cells were treated with bisulfite and four regions (E6, E2, L2 and L1) were amplified with barcoded primers and tested by NGS. CpG methylation was quantified using a bioinformatics pipeline.

Results. Median methylation levels were significantly different between the CIN3+ cases versus controls in the E2, L2, and L1 regions. Methyl-haplotypes, specifically in 5 CpG sites included in the targeted L2 region, with the pattern "−−+−+" had the highest Area Under the Curve value (AUC = 88.40%) observed for CIN3 vs. controls. The most significant CpG site, L2 4277, determined by bisulfite NGS had an AUC = 78.62%.

Conclusions. This study demonstrates that NGS of bisulfite treated HPV DNA is a useful and efficient technique to survey methylation patterns in HPV16. This procedure provides quantitative information on both individual CpG sites and methyl-haplotypes that identify women with elevated present or subsequent risk for HPV16 CIN3 and cancer.

1. Introduction

Cervical cancer is the third most commonly diagnosed cancer and the fourth leading cause of cancer deaths in women worldwide [1]. There are approximately 530,000 new cases and 275,000 associated deaths annually [2]. Nearly all cases of cervical cancer are caused by a persistent infection with a high-risk (HR) type of human papillomavirus (HPV), which includes approximately 12 genotypes within the genus *Alphapapillomavirus* [3]. Of these, HPV16 and HPV18 are the two most important carcinogenic HPV types and together cause 70% of cervical cancer and 50% of the precancerous lesions, specifically cervical intraepithelial neoplasia (CIN) grade 3 (CIN3) [4–6]. Cervical cancer develops over decades from an acute infection with a carcinogenic HPV

* Corresponding author at: Albert Einstein College of Medicine, 1695 Eastchester Rd, Suite 601, Bronx, New York 10461, United States. Tel.: +1 718 405 8086; fax: +1 718 405 8087.
E-mail address: mfrimer@montefiore.org (M. Frimer).

http://dx.doi.org/10.1016/j.ygyno.2015.05.001
0090-8258/

type that is maintained as a persistent infection [7], whereas the majority of infections clear spontaneously [8].

Prevention of cervix cancer has been accomplished through Pap test screening and more recently co-testing that also includes Pap and HPV testing. Most recently, primary stand-alone HPV testing is also an option [9]. Management options for abnormal screening tests include colposcopy versus close-surveillance [10]. A positive HPV result, however, has insufficient specificity since it does not discriminate between HPV-associated cancer-relevant lesions (CIN2+) and transient, clinically benign infections [11,12]. HPV persistence represents a critical distinction between infections with substantial risk of progression to CIN3 or cancer and those that are benign or transient [13]. Referral rates to colposcopy are high in screening settings and diagnostic tests, which can readily distinguish between women with cervical precancer from those transiently infected are necessary to improve the utility of HR-HPV testing [11,12]. With the FDA approval of multiple HR-HPV tests and recent data indicating the advantages of HPV testing as part of the secondary prevention of cervical cancer, differentiation between benign cervical HR-HPV infections and those at risk to progress represents an urgent and significant clinical challenge [10].

The molecular mechanisms underlying persistence and progression to precancer and cancer are largely unknown [10,14]. As a result, there is a clinical need for additional biomarkers, particularly among HPV+/cytology-women; a substantial proportion of women tested for HPV in cervix cancer screening programs [10]. An area of promising study is the epigenetic modifications, i.e., DNA methylation of the viral double-stranded DNA genome as a biomarker for women at risk for cancer [15]. Moreover, these epigenetic changes may also provide insights into the molecular mechanisms of progression [16,17]. In recent years, multiple studies have focused on evaluating the association of HPV16 genome methylation with cervical precancer and cancer [18–25]. These studies provide a consistent model whereby the HPV16 genome shows increased DNA CpG methylation in women with precancerous lesions and cervical cancer, compared to women who are able to clear their infection [18–20,26].

In this study, we use an emerging technology for detection of HPV methylation by sequencing bisulfite-treated HPV16 DNA using massively parallel single molecule sequencing. Specifically, we evaluated the HPV16 methylome for methyl-haplotypes, the combination of methylated CpG sites on a single molecule as previously described [24], combined with quantifying CpG methylation sites. The methyl-haplotype is the combination of the methylation statuses of CpG sites (+, the C is methylated; −, the C is unmethylated) in cis determined from a single read. For example, a DNA amplicon with 3 CpG sites has the following eight possible methyl-haplotypes: −−−, +−−, −+−, −−+, ++−, +−+, −++, +++. With the anticipated widespread implementation of HPV DNA testing for cervix cancer prevention, the clinical objective of this study was to determine the significance of HPV16 methyl-haplotypes to identify HPV16-positive women with or at risk for CIN3+ in a prospective study previously evaluated for single site methylation [23].

2. Methods

2.1. Study population

Cervical specimens were selected from a random sample, population-based longitudinal study of the natural history of HPV infection and cervical neoplasia among 10,049 women in Costa Rica [27]. Women over the age of 18 were recruited for screening between June 1993 and December 1994 and were followed for up to 7 years. At each clinical visit, a cervical sample was taken from the cervix for HPV testing [27,28]. The study protocol was reviewed and reapproved annually by the National Cancer Institute and Costa Rican Institutional Review Boards [27].

2.2. Cervical samples, DNA isolation, and HPV16 detection

For the current nested case–control study, 75 women with single HPV16 type infections were selected for 3 main infection outcomes: (1) transient or cleared, 31 women; (2) CIN3, 31 women; and (3) cervical cancer, 13 women. In addition, there were 13 women who had multiple specimens prior to the diagnosis of CIN3 or cancer. DNA was isolated and previously tested for the presence of HPV16 by MY09/MY11 PCR and type specific oligonucleotide (dot blot) hybridization [29].

2.3. Bisulfite treatment

Cervical DNA samples were treated with bisulfite using the EZ DNA methylation kit (Zymo Research, Orange, CA), as recommended by the manufacturer. In brief, 25 μl of extracted DNA from cervical samples was mixed with dilution buffer and denatured at 37 °C for 15 min. Following denaturation, 100 μl of freshly prepared sodium bisulfite was added and the samples were incubated at 50 °C for 16 h. Post-treatment, samples were desulphonated, washed with 70% ethanol, eluted in 30 μl of elution buffer and stored at −20 °C. Following bisulfite conversion and PCR amplification, unmethylated C's are converted to T's, whereas methylated C's are retained as C's [30]. The ratio of C/C + T indicates the proportion of methylated cytosines at each CpG site in the assayed sample.

2.4. Primer design, PCR amplification and NG sequencing

Four regions within the HPV16 genome were selected for testing using NG sequencing assays for quantitating CpG methylation based on our previous data using pyrosequencing [20]. Methylation sensitive primers based on converted C to T changes at non-CpG sites after bisulfite treatment were designed using MethPrimer (http://www.urogene.org/methprimer/index1.html) [31]. The primers were designed to amplify fragments containing—3 CpG sites in the E6 Open Reading Frame (ORF) region (494, 502, 506); 9 CpG sites in the E2 ORF region (3412, 3415, 3417, 3433, 3436, 3448, 3462, 3473, 3496); 5 CpG sites in the L2 ORF (4240, 4249, 4261, 4270, 4277); and 4 CpG sites in the L1 ORF (7034, 7091, 7136, 7145). In total, 21 CpG sites across four open reading frames were surveyed. Each forward primer contained a specific 8 base pair barcode to identify each sample (see Bioinformatics below). Oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA) (see Table S1).

Each reaction mixture contained 1 μl of bisulfite-treated DNA, 400 nM dNTP, 4.0 μM $MgCl_2$, 2.5 μl 10× Buffer, 0.2 μM of the forward and reverse primers, and 0.08 units of HotStart-IT polymerase (United States Biochemicals, Cleveland, OH) in a total reaction volume of 25 μl. The PCR conditions and annealing temperatures for each assay are provided in Table S1. After confirming the signal intensity of each PCR product on a 3% agarose gel, the amplicons were pooled at similar molar quantities. The PCR products from each of the four assays were pooled separately and isolated by electro-elution, precipitated with isopropanol and 3 M NaOAc (Sodium Acetate) and resuspended in 50 μl of Elution Buffer as previously described [25]. The isolated DNA's were submitted for library construction and sequenced using paired-end reads on an Illumina HiSeq 2000 at the Albert Einstein Epigenomics Core Facility.

2.5. Bioinformatics and determination of CpG methylation status

Illumina sequencing data files were first filtered for low quality reads using a minimum average read PHRED score of 25 and minimum nucleotide PHRED score of 25 in a 5 bp sliding window [32,33]. The data was then demultiplexed based on the 8 bp DNA Hamming barcodes to assign sequencing reads to their sample of origin [34]. Quality control and demultiplexing were performed using the mubiomics package