THE ROLE OF HPV AND DNA METHYLATION IN THE DEVELOPMENT OF PRECANCEROUS CERVICAL LESIONS

Ayodele Gomih

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2017

Approved by:

Jennifer S. Smith

Cathrine Hoyo

Michael G. Hudgens

Kari E. North

Wendy R. Brewster

ABSTRACT

Ayodele Gomih: The Role of HPV and DNA Methylation in the Development of Precancerous Cervical Lesions
(Under the direction of Jennifer S. Smith)

Biological factors associated with cervical intraepithelial neoplasia (CIN) development may be utilized to improve the efficiency of cervical cancer screening. The present study examined human papillomavirus (HPV) genotypes and DNA methylation of imprinted genes in a survival analysis, using two endpoints: CIN1 regression and progression to CIN2+.

The CIN Cohort Study (CINCS) was conducted at 10 Duke University clinics in North Carolina. Women ages 18-65 years with enrollment colposcopies following an abnormal Pap test had follow-up cytology/histology for 3 years with up to 5 visits. DNA was extracted from exfoliated cervical cells for methylation at differentially methylated regions (DMR) of imprinted genes and for HPV genotyping. Hazard ratios (HR) and 95% confidence intervals (CI) were calculated using Cox regression models to estimate the association between aberrant methylation and CIN1 regression (normal epithelia) versus persistence/progression (CIN1+). HRs/95% CIs were estimated to quantify genotype-specific risk of progression to CIN2+.

Of 1303 participants, 472 normal/CIN1 cases with HPV data and 151 CIN1 cases with HPV/methylation data were included in the prospective analyses. In the analysis of genotype-specific risk of progression, hrHPV-66/51/16 were most prevalent in multiple-type infections among 364 HPV-unvaccinated women; 16/52/35 in single-type. Among 108 HPV-vaccinated women, hrHPV-66/39/51 were most prevalent in multiple-type infections; hrHPV-51/66/52/58 in single-type. Over 3 years, there were CIN2+ events (11%;38% normal;62% in CIN1). HrHPV-16/51 had a 3-fold risk of progression

(HR=3.2,95% CI=1.5-7.2;HR=3.2,95% CI=1.3-7.6) versus non-16/51 infections. HPV-16/51 were most predictive of progression to CIN2+.

In the CIN1 regression analysis, one-third regressed to normal epithelium (n=53;35.1%). Median time-to-regression was 12.6 months (range:4.5-24.0). The probability of CIN1 regression was negatively correlated with increased methylation at *IGF2AS* CpG 5 (HR=0.42,95% CI=0.23-0.77;*p*=0.005) and *PEG10* DMR (HR=0.78,95% CI=0.63-0.96;*p*=0.02).

HPV-51 was positively associated with increased risk of progression; aberrant methylation of *PEG10*/*IGF2AS* reduced the likelihood of CIN1 regression. *PEG10*/*IGF2AS* methylation may serve as potential biomarkers for screening, given further characterization of tumorigenesis pathways related to dysregulation of imprinted gene expression. HPV-51 may be useful for CIN2+ risk stratification. If confirmed in other populations, implementation of these novel biomarkers may improve LSIL management and patient care.

To the ancestors… I am because you were.

ACKNOWLEDGEMENTS

I would like to thank you my advisor and dissertation committee chair, Jennifer S. Smith, for her guidance in and dedication to researching HPV and cervical cancer. I thank my committee members who also enriched my doctoral research: Michael Hudgens for his statistical expertise, Kari North for the extensive knowledge of genetic epidemiology, Wendy Brewster for providing meaningful clinical direction and context for my research, and to Cathrine Hoyo for helping me navigate the blossoming field of epigenetics in epidemiology. To Cathrine and Susan Murphy, the Co-PIS of the CINCS Study, thank you for allowing me to be a part of this innovative project. A heartfelt thank you to CINCS participants for being willing to contribute their own personal experiences to helping others, as well as CINCS research team for facilitating the process. To all collaborators, thank you for your invaluable feedback on my manuscripts. I acknowledge the funding sources that supported this research—the National Institutes of Health (NIH) T32-5T32AI007001, R01CA142983, and R01CA142983-02S1. To Chris Wiesen and the Odum Institute: sincerely grateful to you for saving me from the SAS and R frustrations. I thank the EPID student services administrators Nancy, Carmen, Valerie, Jennifer, Chandra, Lisa, etc. fielding all questions, big and small. To the administrators at SPH Office of Student Affairs (Felicia, Charletta, Trinnette, etc.), and SPH IT (Melvin, Joseph, OJ, etc.), thank you! Thank you to all my EPID colleagues for sharing your brains during this process. Special thanks to Dr. Cavin McCaviness (computational biology support); Dr. Christian Douglas (biostatistics support); Dr. Mia Layne (clinical support). To my mentors, Dr. Funmi Olopade and Dr. Ron Aubert, sincerest of thank yous to you both for believing in me. And the deepest, deepest thanks to my Mental and Spiritual Tribe: Gomih Family (Dad, Mom, Dayo, Deji, Dami and all extended), OCD-ish Besties, PhDivas, Klubhouse Kidz, Vegunkies, Chi-town homies, Bull City homies, Jand and Naija homies, homies across the multiverse, all those who prayed/sent positive and encouraging vibes—we made it. Adupe lowo Olorun.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHR	Adjusted hazard ratio
AIC	Akaike Information Criteria
AIS	Adenocarcinoma in situ
ASCCP	American Society for Colposcopy and Cervical Pathology
ASC-US	Atypical squamous cells of undetermined significance
ASC-H	Atypical squamous cells, cannot exclude HSIL
AGUS	Atypical glandular cells of undetermined significance
AGUS-H	Atypical glandular cells of undetermined significance, cannot exclude HSIL
CI	Confidence interval
CIN	Cervical intraepithelial neoplasia
CINCS	Cervical Intraepithelial Neoplasia Cohort Study
CKC	cold knife conization
CPG	Cytosine-phosphate-guanine
DAG	Directed acyclic graph
DCRU	Duke Clinical Research Unity
DCTP	Deoxycytidine triphosphate
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DPC	Duke Primary Care
DTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
FDA	Food & Drug Administration
HPV	Human papillomavirus
HRHPV	High-risk human papillomavirus

HR	Hazard ratio
HSIL	High-grade squamous intraepithelial lesion
ICC	Invasive cervical cancer
IQR	Interquartile range
LBC	Liquid-based cytology
LEEP	Loop electrosurgical excision procedure
LRHPV	Low-risk human papillomavirus
LSIL	Low-grade squamous intraepithelial lesion
LSIL-H	Low-grade squamous intraepithelial lesion, cannot exclude HSIL
ML	Milliliter
MRNA	Messenger ribonucleic acid
NCE	normal epithelium specimen
PAP	Papanicolaou
PBR	Peripheral benzodiazepine receptor
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
SCC	Squamous cell carcinoma
STI	Sexually transmitted infection

CHAPTER ONE: STATEMENT OF SPECIFIC AIMS

Cervical cancer is a highly preventable cancer by screening. In the United States, current screening guidelines recommend the use of high-risk human papillomavirus (HPV) detection with liquid-based cytology testing for women between the ages of 30 and 65 years for detection of high-grade cervical intraepithelial neoplasia (CIN), the precursor of cervical cancer [1]. All low-grade cytology and HPV-16/18-positive/cytology-negative cases are referred for immediate colposcopy [1]. The role of other non-16/18, high-risk HPV genotypes in clinical management of cervical lesions has not been well characterized. The role of epigenetics in cervical carcinogenesis— specifically, DNA methylation of imprinted genes— may also serve as a potential biomarker for CIN progression. Novel biomarkers in screening can help further stratify risk of lesion progression and reduce the number of colposcopy referrals, improving patient care and quality. To identify potential biomarkers in the development of CIN, the following research aims were addressed:

AIM 1: To examine the association between HPV infection by genotype and the progression **of low-**grade cervical epithelial neoplasia (CIN1 to CIN2+) or persistence (CIN1 to CIN1) of precancerous cervical lesions in the Cervical Intraepithelial Neoplasia Cohort Study (CINCS).

AIM 2: To examine the impact of DNA methylation of imprinted genes on the regression of low-grade precancerous lesions (CIN1 to normal) in the Cervical Intraepithelial Neoplasia Cohort Study (CINCS).

CHAPTER TWO: BACKGROUND AND SIGNIFICANCE

Burden of human papillomavirus and cervical cancer

Human papillomavirus (HPV) is an extremely common sexually transmitted infection (STI), affecting nearly 79 million people in the United States (US), with approximately 14 million new infections each year [2]. Of 100 known HPV genotypes, 40 can affect the genital area [3]. HPV infection is more prevalent among Black and Hispanics compared to Whites, especially in low-income populations [4, 5].

HPV infection is a necessary cause in cervical cancer development [6]. Fourteen high-risk oncogenic HPV types have been identified as the main etiologic agents in cervical cancer, in which types 16 and 18 alone cause 70% of all cases [7]. There are over 12,000 new HPV-related cervical cancer cases in the US each year [8]. In North America, 3 women per 100,000 die from cancer of the cervix [8].

The highest rates of cervical cancer in the US occur in the South; the incidence in North Carolina is comparable to overall cancer incidence nationwide (age-adjusted rate: 7.4 per 100,000). Furthermore, cervical cancer disproportionately affects racial and ethnic minorities relative Whites in the US, with higher prevalence of both HPV infection and cervical cancer in Blacks than Whites [4, 5]. While there are comparable screening rates, both incidence and mortality is higher among Blacks and Hispanics [9-11]. In NC, cancer incidence is more than 60% higher in Blacks and Hispanic women compared to White women [9, 12].

HPV and Cervical Cancer Screening/Prevention

Methods for Cervical Cancer Screening

Early detection and treatment of precursor cervical intraepithelial neoplasia (CIN) is an effective strategy to prevent cervical cancer and related deaths. Screening methods to identify CIN and invasive

cervical cancer (ICC) have been available for more than three decades. The Papanicolaou test, best known as a Pap smear test, is the original cytology-based method for detection of abnormal cervical cells. Developed in 1941 by Dr. George Papanicolaou, the Pap test allowed for the identification of precancerous and cancerous cells in cytology samples from vaginal aspirates [13]. The American Cancer Society endorsed the use of Pap tests in formal cervical screening settings in 1957.

The implementation of cytology-based screening and national screening and management guidelines designed by the American Society for Colposcopy and Cervical Pathology (ASCCP) have led to substantial decreases in the incidence and mortality of ICC in the United States [13, 14]. However, conventional Pap tests have variable sensitivity to detected high grade cervical lesions due to variation in sample and slide preparation [13, 15, 16]. Liquid-based cytology (LBC) screening methods were developed to improve clinical performance within the past 17 years. In the US, 90% of Pap tests used to screen for cervical lesions are LBC-based methods such as ThinPrep (Cytyc Corp, Marlborough, MA), preferred over conventional Pap smear tests [13].

Overall, Pap tests have high specificity to distinguish non-diseased (>90%)[17]. However, detection by cytology alone requires a series of repeated Pap tests and/or unnecessary referral to colposcopy and potential surgical excision [15, 18-21]. While women with low-grade CIN— especially young women— are more likely to regress, women who have persistent infection with high-risk HPV (hrHPV) genotypes can lead to persistent dysplasia and progression to ICC [22]. Evidence demonstrates that high-risk HPV infection can better predict development of CIN3 and ICC [22, 23]. As a result, recent age-specific recommendations include the use of HPV testing in conjunction with cytology-based screening for women over 30 years of age [22]. The Food and Drug Administration (FDA) approved the use of the cobas assay (Roche Diagnostics, Indianapolis, IN) for routine screening [22, 24]. Other companies are amidst developing and testing of other high-risk genotyping assays, such as APTIMA (Hologic, Marlborough, MA) and BD (Franklin Lakes, NJ).

Primary screening with HPV is currently recommended in the US for clinical practice as a method of screening triage. The ASCCP currently recommend that women who test positive for HPV-

16/18 are referred for immediate colposcopy, whereas positive tests for the other 12 hrHPV genotypes are referred for cytology and HPV-negative women follow routine screening guidelines[1, 24]. While HPV co-testing has increased the ability to detect high-risk women, improving the efficacy of cervical screening methods can reduce the number of unnecessary screenings and associated cost, patient anxiety and issues with low adherence to follow-up recommendations. Therefore, it is important to further explore the impact of other oncogenic HPV types among women who have negative cytology and low-grade lesions that are more likely to progress to high-grade cervical lesions. With a better understanding of the sociodemographic determinants of CIN and ICC, cancer screening and management efforts may better address those groups at highest risk.

Additionally, recognizing the extent to which other oncogenic HPV (e.g. non-16/18) types lead to CIN progression is crucial to current prevention strategies. Further research in HPV typing in carcinogenesis can inform vaccine development and improve coverage in preventing future transmission. Hence, this dissertation research aimed to determine whether infection with a particular HPV type varies the rate of progression to high-grade CIN or cervical cancer, and determine whether these rates differ among women of different racial/ethnic backgrounds in a multiracial cohort.

Epigenetics and Cervical Tumorigenesis

Epigenetics and Cancer/Disease

Genetic research has ensued for years in order to understand potential targets for disease prevention and clinical interventions. However, epigenetics of cancer has become an increasingly popular topic of interest for researchers. Epigenetics— the control of gene expression by external chemical modifications— may alter biological pathways involved in tumorigenesis [25]. Such alterations can include DNA methylation and chromatin modification, both of which can control gene activity and affect gene expression and activation in cells [26]. DNA methylation has been well researched, especially in the context of human cancers, making them probable candidates for targeted therapies and screening markers.

DNA methylation occurs at cytosine-phosphate-guanine (CpG) sites, which are concentrated in regions throughout the genome in "islands" [27]. These CpG rich regions are not typically methylated, and are situated at the 5' end of regulatory regions for several genes, at promoters, that control gene expression [27]. Unmethylated CpG islands at gene promoters allow for the binding of transcription machinery and transcription of DNA into mRNA[28] (Figure 2.1). However, methylation of the promoter prevents transcription and subsequent gene expression[28] (Figure 2.1).

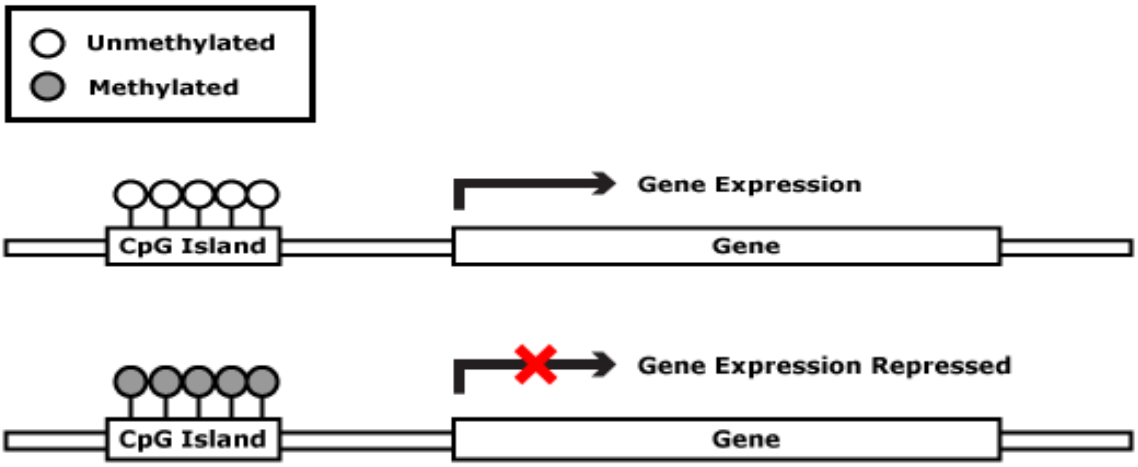


Figure 2.1 Methylation at CpG islands at promoter regions controls gene expression [29]

Genomic Imprinting Loss of Imprinting and Cancer

An alternative mechanism of interest for determining the extent to which epigenetic changes impacts tumorigenesis is the role of DNA methylation in imprinted gene networks. Unlike autosomal genes that are dually expressed from both the maternally- and paternally-derived chromosomal alleles, expression of genomically imprinted genes are determined by one parent and epigenetically marked to express only one allele while the other allele is transcriptionally silenced [26, 30]. Aberrant methylation at regulatory elements of imprinted genes can either silence the only active allele (loss of imprinting) or reactivate the normally silent allele (bilallelic expression), both nullifying normal gene function.

Imprinted genes typically exist in clusters, thus share common regulatory elements. These elements include differentially methylated regions (DMRs), which can be intergenic or intragenic and are critical in tissue-specific gene expression [31-33]. In mice, *Igf2, H19, Zac, Mest, Peg3, Dlk1, Meg3, Grb10, Ndn, Cdkn1c* and *SLC38a4* appear to exhibit coordinated repression with growth deceleration [34]. *In vitro* experiments also showed genetic induction of several imprinted genes on different chromosomes related to the overexpression of another gene as a result of aberrant methylation [35]. As imprinted genes are susceptible to deregulation, they should be strongly considered as key mediators in cancer development.

Their role in cancer was first supported by their role in Beckwith-Wiedemann syndrome (BWS) caused by defects in imprinting on chromosome 15 with affected individuals predisposed to Wilms tumor and hepatoblastoma [26, 36]. Loss of imprinting also has occurred colorectal cancer with insulin growth factor-2 or *IGF2* [37].

DNA Methylation Markers for Cervical Dysplasia

Because cancer progression occurs through the loss and/or gain of genetic function, genetic and epigenetic biomarkers for CIN2+ have received considerable attention in research aimed to distinguish women with benign infection from those requiring treatment over the past 15 years. Specifically, DNA methylation can influence transcription and expression of genetic factors involved in cervical cancer development without directly altering the underlying gene's DNA sequence [25].

One proposed mechanism for CIN progression involves epigenetic modifications at gene promoter CpG islands that impact transcription, inhibiting expression of tumor suppressor genes. However, aberrant methylation of promoter regions has not been strongly associated with CIN progression [38] and data is inconclusive on whether any specific methylation markers can be utilized as markers for screening[31, 38-40]. The role of methylation in HPV's carcinogenicity has been explored as a potential marker of cervical progression, DNA methylation that alter the expression of critical genes that can either inhibit or promote cervical carcinogenesis[31, 40]. Methylation of high-risk HPV genomes has

also been assessed as a potential pathway in the progression of cervical dysplasia [31, 40]. Specifically, integration of HPV-16 into the host genome has resulted in the expression of E6/E7 oncoprotein genes that in turn, results in overexpression of p53 and PBR tumor suppressing genes (citation). Hypermethylation of HPV-16 E6 and E7 oncogenes (which promote HPV virulence) may cause binding of DNA transcription factors that influence cancer development [31]. Further studies in humans have demonstrated differential changes in methylation at imprinting centers among women who had cervical intraepithelial and invasive cervical cancer. An analysis of publicly available microarray data found that expression of imprinted genes differed between specimens from ICCs compared to normal cervical epithelium (NCE) specimen (Table 2.1).

Table 2. 1 DNA Methylation of Imprinted Genes in Cancerous vs. Normal Cervical Tissue (Hoyo/Murphy Group)

Repressed Imprinted Genes in Cervical Cancer				
		CVCX (N=33) / NCE (N=9)		
Gene	**Probe**	**Ratio**	**P**	**Active allele**
SNRPN/SNURF	206042_x_at	0.49	0.0160	Paternal
HYMA1	215513_at	0.50	0.0012	Paternal
PEG3	209242_at	0.53	0.0034	Paternal
CPA4	205832_at	0.53	0.0040	Maternal
NDN	209550_at	0.53	0.0207	Paternal
PLAGL1 (ZAC)	209318_x_at	0.54	0.0052	Paternal
PLAGL1 (ZAC)	207943_x_at	0.57	0.0131	Paternal
PLAGL1 (ZAC)	207002_s_at	0.58	0.0598	Paternal
WT1	206067_s_at	0.64	0.4346	Paternal
Upregulated Imprinted Genes in Cervical Cancer				
		CVCX (N=33) / NCE (N=9)		
Gene	**Probe**	**Ratio**	**P**	**Active allele**
PEG10	212094_at	8.43	0.2945	Paternal
IGF2	202410_x_at	5.45	0.6017	Paternal
PEG10	212092_at	3.78	0.4341	Paternal
MEST	202016_at	3.78	0.0350	Paternal
CDKN1C	213182_x_at	3.69	0.0516	Maternal
CDKN1C	216894_x_at	3.43	0.0454	Maternal
IGF2	210881_s_at	3.42	0.5372	Paternal
CDKN1C	219534_x_at	2.30	0.0952	Maternal
GRB10	210999_s_at	2.22	0.0294	Isoform-dep.
CVCX, cervical cancer; NCE, normal cervical epithelium				

For example, paternally-expressed SNRPN/SNURF genes were suppressed compared to NCE samples; similarly, paternally-expressed PEG10 was overexpressed in comparison to normal cases. This finding supports that the expression profile of imprinted genes differs in cervical cancer cases compared to normal cases and thus these genes warrant further research in CIN progression.

Influence of HPV and External Factors on DNA Methylation and Cancer

Epigenetic changes to the host genome are typically attributed to external forces, such as the environment and lifestyle behaviors. Cofactors such as smoking, high parity, and oral contraceptive use may influence these epigenetic changes that lead to progression of cervical dysplasia. It is established that, in addition to persistent high-risk HPV infection, cigarette smoking, long term oral contraceptive use, and high parity are risk factors for cervical cancer and CIN [20]. These identified factors may also influence DNA methylation and therefore lead to loss of imprinting and potential functionality of both growth factors and tumor suppressor genes.

The relationship between DNA methylation and HPV infection, as previously described, has been explored in the context of mainly high-risk HPV types 16 and 18. To our knowledge, no studies have examined aberrant methylation patterns of imprinted genes within other high-risk HPV types. In addition, while smoking is known to be associated with cervical cancer and many other cancer outcomes, the mechanism for smoking's impact on CIN is not well understood. Smoking does not appear to have an effect on the clearance or persistence of HPV infection, suggesting that the effect of smoking may be independent of HPV infection in the process of cervical cancer progression [41-43]. Further, it is inconclusive whether smoking increases the risk of HPV infection by lowering immunity and antibody levels [44, 45]. Still, smoking appears to influence methylation of IGF2 or Insulin-like Growth Factor 2, leading to loss of imprinting of IGF2/H19 [46]. This finding is of great importance, as loss of imprinting of IGF2/H19 gene complex occurs in 58% of cervical cancers [47]. Hence, additional research is required to characterize the roles of both HPV types and smoking in the methylation of imprinted genes.

Thus, differentiation of CIN cases that are more likely to progress or regress using epigenetic biomarkers requires further investigation. Ultimately, the findings of this research can improve the efficiency of cervical screening and optimize clinical management of CIN, reducing unnecessary testing and treatment of affected women.

CHAPTER THREE: RESEARCH PLAN AND METHODS

Study Population

The Cervical Intraepithelial Neoplasia Cohort Study (CINCS) is a 3-year prospective study that enrolled participants between 2010 to 2014 to determine whether epigenetic deregulation of known imprinted genes can be used to distinguish women more likely to progress to CIN2+.

The cohort includes 1303 women aged 18 years and older with abnormal cytology attending 10 Duke University and Duke Primary Care (DPC) clinics in Durham County, North Carolina**.** North Carolina is located in southeastern region of the USA, where cervical cancer is most prevalent.

Recruitment and Study Follow-Up

Recruitment at the clinics

Potential participants included women who were referred for colposcopy following an abnormal cervical screening test. Prospective participants were identified by their physicians two weeks prior to a scheduled colposcopy appointment through the electronic clinic appointment logs of the Duke University and community clinics. The attending physicians provided signed letters of invitation for study participation to mail to potential study participants prior to the colposcopy. Study invitation letters explained the broad goals of the study and expectations of participants. The colposcopy patients were instructed to call a toll-free number to register their intent to decline participation. During the scheduled colposcopy appointment visit, the recruiting interviewer solicited participation directly from the participant and administered the informed consent process. After consent, participants were given a standardized questionnaire to ascertain demographic and clinical information. Further information was collected on cervical dysplasia diagnosis through histopathological confirmation typically two weeks after the colposcopy visit.

Follow up of women after abnormal cytology diagnosis

Study participants attended a clinic visit approximately every 6 months to 1 year for 3 years, following clinical practice guidelines for management of CIN (Figure 1). At each visit, clinic physicians conducted a cervical examination to ascertain information on any cytological/histological changes. A follow-up questionnaire was administered to ascertain information on any behavioral and clinical characteristics of the participant since enrollment.

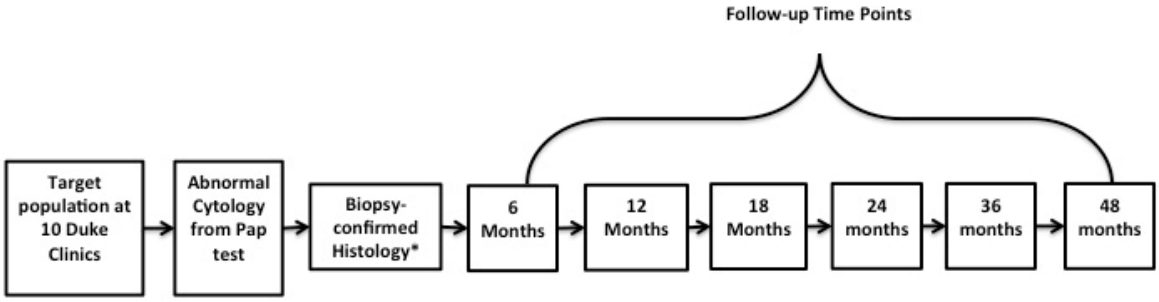


Figure 3. 1 CINCS Study Visit Flow Chart

Eligibility criteria

To be eligible to participate in CINCS overall, women were new visitors to the colposcopy clinic with histopathology results from a colposcopy-directed biopsy (inclusive of endocervical curettage) following an abnormal Pap test. Because most CIN is commonly transient in younger women (<21 years old) [22], enrollees ranged in age from 21 to 79 years. Participants were English or Spanish-speakers and mentally competent to give informed consent. Women who had treatment for cervical lesions (e.g. cold knife conization, loop electrosurgical excision procedure, hysterectomy) were excluded from the study. Women were further excluded if they did not intend to receive follow-up care in one of the 10 DPC clinics or moved out of the area for other reasons.

Eligibility criteria for Aim 1

Women were required to have histopathologically-confirmed diagnosis of normal epithelial/CIN1 at their enrollment visit. Participants must have attended at least one follow-up visit following enrollment, at which they received either a Pap test and/or colposcopy with cervical biopsy and/or endocervical curettage. Participants without data on HPV status were excluded.

Eligibility criteria for Aim 2

Participants with a histopathology-confirmed CIN1 diagnosis at the enrollment visit were included. at their enrollment visit. Participants must have attended at least one follow-up visit following enrollment, at which they received either a Pap test and/or colposcopy with cervical biopsy and/or endocervical curettage. Participants without data on HPV status or imprinted gene methylation were excluded.

Data Collection

Cervical Cytology & Histology Specimens

During the Pap test at baseline, the clinic physician utilized a spatula and cytobrush to obtain exfoliated cervical cells. Cervical specimens were suspended in a Thinprep vial containing proprietary fluid with at least 50% methanol (Cytyc®, Malborough, MA, USA) for cytology confirmation. Following abnormal cytology results, clinic physicians performed colposcopy-directed biopsy at baseline. Information on cytology and lesion morphology, including size and location, were abstracted from patient medical records. All specimens were tested for adequacy using the 2012 ASCCP guidelines. The specimens were stored at 4°C prior to HPV testing.

All Pap test results were classified as one of the following by a pathologist according to 2012 Bethesda system [48]: i) negative/normal cytology; ii) low-grade squamous intraepithelial lesion (LSIL); iii) atypical squamous cells of undetermined significance (ASCUS); iv) atypical glandular cells of

undetermined significance (AGUS); v) high-grade squamous intraepithelial lesion (HSIL); vi) low-grade squamous intraepithelial lesion, cannot exclude HSIL (LSIL-H); or vii) atypical squamous cells cannot exclude HSIL (ASC-H).

To ascertain the presence of CIN, cervical biopsies were adjudicated by a clinical pathologist. CIN diagnoses were classified as follows: i) CIN-1; ii) CIN-2; iii) CIN-2/3; iv) CIN-3; or v) ICC.

HPV DNA testing

HPV typology was assessed using excess cervical tissue obtained from the enrollment biopsy. ThinPrep® specimens and homogenized aliquoted biopsies were collected during the same baseline visit and shipped to the University of Hawaii Cancer Center. Following DNA extraction, PGMY09/PGMY11 primers [49, 50] were used in PCR to target a 450-bp region of the HPV L1 genome. Amplification of the human β-globin gene was included as an internal control for sample sufficiency. HPV-positive specimens were subsequently genotyped by using the HPV Linear Array® (Roche Diagnostics, Branchburg, NJ, USA). This assay is designed to detect 14 high-risk HPV types— 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68— and low-risk HPV types— 6, 11, 26, 40, 42, 53, 54, 61, 62, 64, 67, 69, 70, 71, 72, 73, 81, 82, 83, 84.

DNA Methylation

Nucleic acid extraction

At the enrollment of all participants, a 10ml blood sample in an EDTA tube was collected for DNA and plasma. The phlebotomy laboratory run by the Duke Clinical Research Unit (DCRU) collected peripheral blood for methylation assays. Peripheral blood DNA was extracted from buffy coat using PureGene Reagents (Qiagen) according to the manufacturer's instructions. RNA and DNA from the ThinPrep cell pellet were extracted using the protocol for simultaneous nucleic acid extraction provided by Teltest for their RNA Stat60 and DNA Stat60 reagents. Nucleic acids will be aliquoted, barcoded, and

stored at -80ºC until required. Methylation and host genetic analyses were performed by the Duke Epigenetics Group led by Drs. Murphy and Jirtle within their respective laboratories.

Methylation analysis

DNA methylation was measured at differentially methylated regions (DMRs) that regulate the imprinted genes *IGF2* AS*, IGF2/H19, PEG1/MEST, KV DMR, DLK1/MEG3, PLAG1/HYMA1, PEG10* and *PEG3* DMR using the Sequenom MassARRAY platform with EpiTYPER analysis software. Primers were designed with EpiDesigner software which designs primers complementary to bisulfite treated DNA in regions without CpG nucleotides; it then adds a T7 promoter site to all forward primers (Table 3.1).

Table 3. 1 Imprinted Gene Primers used in Sequenom analysis for DNA methylation

Gene	Primer Seq.	Chr.	Start	End
H19-IGF2	aggaagagagTATTTTGAGGTTTTGGGGGATATTA	11	2130112	2130388
	cagtaatacgactcactatagggagaaggctCTCCCTCAACAAAAACTAACAAATC			
MESTIT1, MEST	aggaagagagGGGTTTAGAGGTATAAGAAAGAGGG	7	130130648	130131063
	cagtaatacgactcactatagggagaaggctTTTCTAAAAACAACCAAACCCCTAC			
KvDMR	aggaagagagTTTGGTAGGATTTTGTTGAGGAGTTTT	11	2721161	2721464
	cagtaatacgactcactatagggagaaggctCTCACACCCAACCAATACCTCATAC			
MEG3	aggaagagagTTGTGATAAGGTTAGTGAGGGGTTA	14	101293947	101294390
	cagtaatacgactcactatagggagaaggctCCAACCAAAACCCACCTATAACTAC			
HYMAI, ZAC	aggaagagagGAAAAAGTTTGTTTTAAGTAATAATGGGAT	6	144328445	144328885
	cagtaatacgactcactatagggagaaggctAAAAAACCAAAACCTCAATAAAACC			
PEG10	aggaagagagAGGTGTGGGATTTTATTTTTTTTGT	7	94285845	94286061
	cagtaatacgactcactatagggagaaggctCAAACCTTTAAAACTTAATTTCCCC			
PEG3, ZIM2	aggaagagagTATTGGGTGTTATTTTTTATGAGGG	19	57350715	57351051
	cagtaatacgactcactatagggagaaggctTCTACTACCAACCAACCAAAACAAC			

After PCR amplification, unincorporated nucleotides were removed with shrimp alkaline phosphatase, and the PCR products were transcribed by T7 RNA & DNA polymerase with simultaneous cleavage by RNase A. Including either dCTP or dTTP in the transcription reaction restricts cleavage by RNase A to positions immediately 3' of thymine or cytosine residues, respectively. The fragmented transcripts were spotted onto SpectroCHIPS for mass spectrometry analysis on the MassARRAY instrument. The fragments were identified by matching the molecular mass of detected particles to that expected from

analysis of the reference sequence. The analysis software compares the relative amounts of methylated forms of each predicted fragment to quantify the methylated fraction.

Analytic Methods

Aim 1: To examine the association between HPV genotypes and the progression to CIN2+ among normal and low-grade CIN in the CINCS.

a. Determine whether progression to CIN2+ varies by individual HPV status (positivity) or HPV type-grouping (high-risk HPV types compared to infection with low-risk HPV types).
b. Assess whether the association between HPV infection (positive vs. negative; high-risk vs. low-risk) and the progression to CIN2+ is modified by race, smoking, parity, or oral contraceptive use.

Time-to-progression analysis with Cox proportional hazards were used to examine the relationship between HPV genotypes/risk group and progression to CIN2+. CIN outcome at follow-up was dichotomized as progression vs. persistence/regression.

We defined progression as a more severe CIN score (CIN-2+) at a follow-up visit; persistence as retention of the same diagnosis at a follow-up visit (CIN-1); and regression as a less severe CIN diagnosis. Cytology results were used if the participants had missing histology data as a result of conservative clinical practice. A description of the outcome classification is described in Tables 3.2.1-3.2.4.

Table 3.2. 1 Progression to CIN2+ outcome, given enrollment CIN1 histology and follow-up histology

Enrollment Histology	Histology Follow-Up	Status at Follow-Up
CIN1	Normal/Negative	Regress
	CIN1	Persist
	CIN2	Progress
	CIN2/3	Progress
	CIN3	Progress
	Cancer	Progress

Table 3.2. 2 Progression to CIN2+ outcome given enrollment normal histology and follow-up histology

Enrollment Histology	Follow-Up Histology	Status at Follow-Up
Normal	Normal/Negative	Persist
	CIN1	Non-Progress/ Persist*
	CIN2	Progress
	CIN2/3	Progress
	CIN3	Progress
	Cancer	Progress

*CIN2+ is main endpoint in analyses, as it warrants clinical intervention, e.g. (LEEP, etc.) [1]

Table 3.2. 3 Progression to CIN2+ outcome for enrollment CIN1 histology utilizing follow-up cytology, given missing follow-up histology missing

Enrollment Histology	Cytology Follow-Up	Status at Follow-Up
CIN1	Normal/Negative	Regress
	LSIL	Persist
	AGUS-L	Persist
	ASCUS-L	Persist
	HSIL	Progress
	AGUS-H	Progress
	ASCUS-H	Progress
	ICC	Progress

Table 3.2. 4 Progression to CIN2+ outcome for enrollment normal histology utilizing follow-up cytology, given missing follow-up histology

Enrollment Histology	Follow-Up Cytology	Status at Follow-Up
Normal	Normal/Negative	Regress
	LSIL	Non-Progress/Persist*
	AGUS-L	Non-Progress/Persist*
	ASCUS-L	Non-Progress/Persist*
	HSIL	Progress
	AGUS-H	Progress
	ASCUS-H	Progress
	Cancer (AIS/SCC)	Progress

* CIN2+/HSIL+ is main endpoint in analyses, as it warrants immediate clinical intervention, e.g. (colposcopy/biopsy, etc.) [1]

We defined progression as a cytology result of HSIL, LSIL-H, or ASC-H at follow-up; persistence as LSIL, ASCUS, or AGUS at follow-up; and regression as negative/normal cytology at follow-up (Table 3.2.3 and 3.2.4).

For women who received treatment (loop electrosurgical excision procedure, cold knife conization, hysterectomy, etc.) at a follow-up visit, the histological diagnosis from the treatment specimen will be used as their follow-up diagnosis. Women who were treated for CIN were excluded from further data analyses after their treatment.

The outcome was dichotomized as the time to the first incidence or first progression of CIN from enrollment status. Person-time accumulated until progression to CIN2+ event, or at the last recorded follow-up visit. Administrative censoring occurred at 3 years' post-study enrollment.

Univariate distribution of the 37 HPV types (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 56, 58, 59, 66, 61, 62, 64, 66, 67, 69, 70, 71, 72, 73, 81, 82, 83, 84) were examined individually, and categorized as HPV types by "high-risk," "low-risk," and "no HPV infection."

Correlates of CIN/cervical cancer examined as follows:

- Age was examined as a continuous variable as well as a categorical variable
- Race was categorized as a binary variable of either "Black" and "(Non-Hispanic) White"
- Current cigarette smoking was categorized as "Non-current smoker" vs. "Current smoker"
- Parity was examined as a continuous variable and categorical variable
- Current Oral contraceptive use was categorized as "Non-current Use" and "Current Use"
- History of HPV Vaccination was categorized as "never received HPV vaccine" vs. "prior HPV vaccination"

Potential confounders were identified *a priori* and placed into an elastic net Cox regression model [51][52]. Effect measure modification was also assessed using the Akaike Information Criterion (AIC) to determine the quality of model fit for interactions between covariates of interest and HPV type risk group.

Aim 2: To examine the association between aberrant methylation of imprinted genes and the regression (CIN1 to normal) of precancerous cervical lesions in the Cervical Intraepithelial Neoplasia Cohort Study (CINCS).

The exposure of aberrant methylation was treated as a continuous variable and transformed by the interquartile range (IQR) for CIN1 regression modeling. Median methylation scores of multiple CpG loci for each imprinted gene were also calculated.

We also explored dichotomizing the outcome by establishing specific cutpoint for hyper- or hypo- methylation, given an expected value of 50% DNA methylation for monoallelically-expressed genes such as imprinted genes (as one allele is fully methylated while the other is completely unmethylated). Methylation levels deviating by 5% from this expected value were considered. Based on the data, methylation was coded continuously for the analyses.

Regression of cervical lesions was defined as a diagnosis of negative/normal cytology at two consecutive follow-up visits (Table 3.3.1 and 3.3.2).

Table 3.3. 1 CIN1 regression outcome, given enrollment CIN1 histology and follow-up histology

Enrollment Histology	Follow-Up Histology	Outcome Status at Follow-Up
CIN1	Normal/Negative	Regress
	CIN1	Persist
	CIN2	Progress
	CIN2/3	Progress
	CIN3	Progress
	Cancer	Progress

Table 3.3. 2 CIN1 regression outcome for enrollment CIN1 histology utilizing follow-up cytology, given missing histology

Enrollment Histology	Follow-Up Cytology	Outcome Status at Follow-Up
CIN1	Normal/Negative	Regress
	LSIL	Persist
	AGUS-L	Persist
	ASCUS-L	Persist
	HSIL	Progress
	AGUS-H	Progress
	ASCUS-H	Progress
	Cancer (AIS/SCC)	Progress

Per conservative clinical practice, cytology results were utilized to determine regression status if the participants had missing histology data. Persistence was defined as a diagnosis of low-grade histology (CIN1) at follow-up or low-grade lesions during cytology testing (e.g. LSIL, ASC-US, AGUS-L). Progression was defined as a follow-up histological diagnosis of CIN2+, or as development of high-grade cytology, e.g. HSIL, LSIL-H, or ASC-H.

Women with a negative/normal screening at one follow-up time point were not considered to have regressed for the main study analyses. For women who received treatment (LEEP, CKC, cryotherapy, or hysterectomy) at a follow-up visit, the histological diagnosis from the pre-treatment specimen was utilized.

The covariates were examined as follows:

- Age will be examined as a continuous variable as well as a categorical variable (categories to be explored)
- HPV Status will be dichotomized as "HPV-Positive" and "HPV-Negative"
- HrHPV Status will be dichotomized as "hrHPV-Positive" and "hrHPV-Negative (low-risk HPV or no HPV infection)"
- Race was categorized as "Black", "(Non-Hispanic) White", and "Other"
- History of Oral contraceptive use will be categorized as "Never Use" vs. "Ever Use"
- Current cigarette smoking will be categorized as "Non-current smoker" vs. "Current smoker"
- Parity will be examined as a continuous variable as well as a categorical variable.

Associations between methylation at individual CpG loci and across the gene and time-to-CIN1 regression were estimated using Cox regression models to produce HRs and 95% confidence intervals.

CHAPTER FOUR: HIGH-RISK HUMAN PAPILLOMAVIRUS GENOTYPES IN THE PROGRESSION OF CERVICAL INTRAEPITHELIAL NEOPLASIA

Background

Over 250,000 women in the United States (US) are living with invasive cervical cancer (ICC), caused by oncogenic human papillomavirus HPV types [6, 53]. A total of 14 oncogenic or high-risk HPV (hrHPV) genotypes are causal factors for the development of ICC, which is preceded by high-grade cervical intraepithelial neoplasia (CIN2+)[54]. HPV genotypes 16 and 18 account for ~70% of cervical cancer cases, while hrHPV types 31, 33, 45, 52 and 58 account for approximately 15% of detected cases and 35, 59, 51, 56, 39, 68, 66 for the remaining 15% [7, 55, 56].

Cervical cancer is a highly preventable cancer by screening for cytology. In the US, annual cytology testing (Pap smear) has reduced morbidity and mortality from cervical cancer over the past several decades[57]. However, cytology testing alone has high specificity (93-98%), albeit relatively low sensitivity (11-28%) for CIN2+ detection[1, 17].

With the development of new screening technologies, the American Society of Colposcopy and Cervical Pathology (ASCCP) guidelines have evolved to improve detection of CIN2+ with the use of HPV testing [1]. Current ASCCP guidelines recommend liquid-based cytology (LBC) testing every 3-5 years (vs. annually), with HPV co-testing to improve the sensitivity of cytology to detect CIN2+ among women ages 30 and older[58]. HPV-16/18 positive women with negative cytology are referred for immediate colposcopy. For negative cytology cases who test positive for hrHPV excluding types 16/18 ("non-16/18 hrHPV-positive"), clinical management is less certain. Currently, follow-up recommendations for HPV-positive/cytology-negative cases include repeat testing after one year [1]. The increased number of follow-up visits results in increased burden of cost and psychological stress and decreased clinical visit adherence [1, 59, 60].

To minimize over-screening, new methods of HPV-testing at screening have been developed to include additional genotyping. As risk of CIN2+ due to non-16/18 types is not well characterized [1, 61], new screening assays may help in establishing risk stratification categories by hrHPV genotype for women who are hrHPV-positive with negative cytology. Additionally, low-grade cytological abnormalities are currently referred for immediate colposcopy, with the exception of low-grade/HPV-negative cases[1]. Risk stratification by non-16/18 hrHPV types may also possibly further triage women with low-grade cytology results, ultimately reducing the number and frequency of colposcopies in CIN management.

With the advancement of new screening methods for cervical precancer, one must also consider the influence of HPV vaccination (bivalent, quadrivalent and 9-valent) on the distribution of HPV genotypes in the population. Any changes in prevalence due to immunity conferred from vaccination are important for contextualizing risk of CIN progression.

Improved risk management of patients who present with negative cytology, as well as low grade cervical abnormalities is essential to reducing the incidence of cervical cancer for all women. The use of hrHPV genotyping as a means of risk stratification would advance triaging methods for women who have low-grade cytology or are cytology-negative. This study aimed to investigate HPV genotypes in the development of cervical dysplasia. The objective of this study was to examine individual HPV types as predictors of progression to CIN2+ among negative and low-grade cytology patients in the Cervical Epithelial Neoplasia Cohort Study (CINCS).

Methods

Study population

From June 2010 – April 2014**,** women attending 10 Duke University hospitals and clinics in Durham, North Carolina were invited to participate in CINCS, as previously described [62]. Briefly, all clinics used Duke-affiliated pathology laboratory for cytology and histological evaluation. CINCS

comprised of a cohort of 1303 women who were referred for a colposcopy following an abnormal LBC result from a pelvic exam**.** Participants were eligible if they provided written consent, were new visitors to the clinic, 18-79 years old, and were English or Spanish speakers. We excluded women who had received previous treatment for cervical lesions—cold knife conization (CKC), electrosurgical excision procedure (LEEP), cryotherapy, or hysterectomy; or had moved out of the study area or did not intend to receive follow-up care at one of the 10 Duke clinics. Women with no lesion or with a CIN1 diagnosis at enrollment who had at least one follow-up visit were included in the present analyses. Approval was granted by the Institutional Review Boards at Duke University (Durham, NC, USA), North Carolina State University (Raleigh, NC, USA) and University of North Carolina (Chapel Hill, NC, USA).

Data collection and laboratory analyses

At enrollment, participants had a physician-directed cervical examination with a colposcopy-directed biopsy. Women diagnosed with normal/negative or CIN1 by colposcopic impression at enrollment were also included in the study, as prevalence of hrHPV was comparable to the hrHPV prevalence among those who underwent a biopsy and all women underwent colposcopy because of their initial abnormal cytology result. Study participants attended a clinical visit approximately every 6 months for the first two years, and at the end of 12 months for final third year, following clinical practice guidelines for management of CIN. During each follow-up visit, all women underwent a LBC test. Given an abnormal cytology result, clinic physicians performed a colposcopic examination and obtained a biopsy sample at follow-up visits if clinically necessary, according to physician's best judgement and per clinical guidelines for management of precancerous cervical lesions [1]. Study staff administered a questionnaire at enrollment to ascertain information on any behavioral and clinical characteristics associated with CIN and cervical cancer, including age, race/ethnicity, current smoking status, history of HPV vaccination, current hormonal contraceptive use, and parity.

Ascertainment of Cervical Cytology and Histology

To conduct a LBC test, the clinic physician utilized a spatula and cytobrush to obtain exfoliated cervical cells. Cervical specimens were suspended in a ThinPrep® vial containing proprietary fluid with at least 50% methanol (Cytyc®, Malborough, MA, USA) for cytological assessment. All study pathologists evaluated cytology according to Bethesda criteria [48]. The residual cervical exfoliated cell specimens were stored at 4°C prior to HPV DNA testing.

Biopsy results were also reviewed and graded for severity by a pathologist at the study site. All histological biopsy specimens were tested for adequacy using the 2002 and 2012 ASCCP guidelines. Information on cytology and histology were abstracted from patient medical records.

HPV Testing and Typology

HPV typology was assessed using cervical exfoliated cells from the enrollment pelvic exam. ThinPrep® specimens were collected during the same enrollment visit and sent to Johns Hopkins University and the University of Hawaii Cancer Center for laboratory testing. Following DNA extraction, PGMY09/PGMY11 primers [49, 50] were used in PCR to target a 450-bp region of the HPV L1 genome. Amplification of the human β-globin gene was included as an internal control for sample sufficiency. HPV-positive specimens were subsequently genotyped by using the HPV Linear Array® (Roche Diagnostics, Branchburg, NJ, USA). This assay detects 37 HPV types— 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 56, 58, 59, 66, 61, 62, 64, 66, 67, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39 and CP6108.

Statistical analyses

HPV genotypes were evaluated individually and by high risk vs. low-risk status, where hrHPV types included 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 and lrHPV types included 6, 11, 26, 40, 42, 53, 54, 61, 62, 64, 67, 69, 70, 71, 72, 73, 81, 82, 83, and 84. We also examined incidence and progression of CIN2+ in categories of hrHPV genotypes associated with risk of CIN3+ (defined by Stoler

et al.): A) 16/31/18 (>5% risk of CIN3+); B) 33/58/52/45 (2-5% risk of CIN3+); and C) 39/68/35/51/59/56/66 (<2% risk of CIN3+) [63].

CIN progression was defined as a follow-up diagnosis of CIN2+ (CIN2, CIN2/3, CIN3, ICC) at a follow-up visit. Cytology results were utilized to determine progression status if the participants had missing histology data, per conservative clinical practice. A cytological diagnosis of high-grade squamous epithelial lesions (HSIL), LSIL-H (LSIL, cannot exclude HSIL) or ASC-H (ASC, cannot exclude HSIL) at follow-up was also considered a progression event. Persistence was defined as a diagnosis of low-grade histology (CIN1) at follow-up or low-grade lesions during cytology testing (e.g. low-grade squamous epithelial lesions or LSIL, atypical squamous cells of undetermined significance or ASC-US). Regression was defined as a follow-up diagnosis of negative/normal cytology or histology. For women who received treatment (LEEP, CKC, cryotherapy, or hysterectomy) at a follow-up visit, we utilized the histological diagnosis from the treatment specimen as their follow-up diagnosis.

We conducted bivariable analyses to assess the distribution of HPV genotypes and other risk factors of cervical dysplasia and cervical cancer among women with no CIN lesion and CIN1 cases and by HPV vaccination history (Tables 4.1 and 4.2). We compared differences by enrollment CIN diagnosis within vaccination status using the Chi-Square test and Fisher's Exact test for stratum with less than 10 observations. The cumulative probabilities of progression to CIN2+ over 3 years were compared within HPV vaccination history (vaccinated vs. no vaccine) by enrollment diagnosis (no lesion vs. CIN1) using the Kaplan-Meier product-limit method and Log-rank test. Because probability estimates were similar regardless of enrollment diagnosis, subsequent analyses combined women with no lesion or CIN1 by their HPV vaccination status.

Incidence rates and 95% confidence intervals were estimated for each HPV genotype, infection with any hrHPV/lrHPV type, and hrHPV risk categories among all infections and among single type infections (Tables 4.3 and 4.4). The Kaplan-Meier product-limit method was used to estimate the cumulative probability of progression to CIN2+ by hrHPV risk categories among single genotype infection (Figure 4.1). The Log-rank test was used to assess differences between progression probabilities

over time by infection status. In the time-to-event analyses, Cox proportional hazards regression models were utilized to estimate unadjusted and adjusted hazard ratios (HR) and 95% confidence intervals (95% CIs) for the association between HPV genotypes and progression to CIN2+ by vaccination status (Tables 4.5 and 4.6). Time-to-progression was measured from the date of enrollment to the date of CIN2+ histological or HSIL+ cytological diagnosis, estimated in woman-months. Participants contributed woman-months up to the occurrence of progression or the date of the last attended clinical study visit. Participants who received treatment during the study were right censored from further follow-up analyses at the date of procedure. Administrative censoring occurred at 3 years.

Confounders selected for the multivariable Cox regression model were determined *a priori* using conceptual models (directed acyclic graphs). Covariates considered for the analyses included continuous age at enrollment, HPV vaccination (yes vs. no), race/ethnicity (non-Hispanic White, Black/African-American, Other), current smoking status at enrollment (current vs. non-current), current oral contraceptive use at enrollment (current vs. non-current), and parity (continuous). A penalized Cox proportional hazards regression model was employed to determine the strongest predictors of progression to CIN2+ from a large set of variables [64]. Model parameter optimization for progression to CIN2+ was determined using the **c060** package for extended inference with elastic net Cox models, allowing for the inclusion of highly correlated variables [52, 65]. The optimal alpha-level selected by the algorithm was α=0.012. We assessed for modification using Akaike Information Criteria for model fit and found evidence of modification by race/ethnicity. Sensitivity analyses were conducted to assess for potential selection bias due to attrition. All statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC) and R version 3.3.3 (Vienna, Austria).

A total of 780 women of the 1303 enrolled CINCS participants had HPV DNA laboratory results. Of these, we excluded 36 women with an inconclusive CIN diagnosis, and 111 women with CIN2+ histological diagnosis. The remaining 631 women diagnosed with no lesion (n=389) or CIN1 (n=242) were considered for the analyses.

Results

One quarter of CINCS participants with no lesion or CIN1 dropped out of the study after enrollment (N=159, 25.2%). There were no differences in characteristics between participants who dropped out after enrollment compared to those who remained in the longitudinal study (data not shown). Median age at enrollment among 472 women was 29.3 years (range: 19.7-64.7). One-fifth of CINCS participants had initiated HPV vaccination by their date of enrollment (22.9%, Table 4.1). Unvaccinated participants were 8 years older than those who were vaccinated (median age: 32.3 vs. 24.4 years), whereas two-thirds of vaccinated women were Non-Hispanic white relative to those with no prior vaccination (57.4% vs. 43.5%). Most women were non-current smokers regardless of vaccination history (81.5% unvaccinated, 86.1% vaccinated). A higher proportion of vaccinated women were current users of oral contraceptives compared to unvaccinated women (46.7% vs. 25.1%).

Over 80% of CINCS participants had infection with any HPV type (88.6%) (Table 4.2). Most women tested positive for more than one genotype (53.6%). Approximately 65% of women had infection with any hrHPV type and 43% with any lrHPV type. The prevalence of any hrHPV type among unvaccinated women was slightly lower compared to women with HPV vaccination by enrollment (67.9% vs. 77.8%), however genotypes in the quadrivalent HPV vaccine (HPV-6, HPV-11, HPV-16, HPV-18) were less prevalent among vaccinated participants. Most participants were infected with multiple genotypes, regardless of vaccination status. The most frequently occurring hrHPV genotypes among single type infections were 16, 52, 35 in unvaccinated women, whereas genotypes 66, 51, and 16 occurred most frequently for multiple infections. Vaccinated participants were predominately infected with other single non-16/18 hrHPV genotypes, including 51, 66, 52. The most prevalent hrHPV genotypes in multiple infections among women with a history of HPV vaccination included 66, 39, 51. No women in the study had infection with lrHPV types 11 and 64.

Median study follow-up time was 22.4 months (range: 0.7-36 months). Only 11% (n=52 women) progressed to CIN2+ over 3 years of follow-up (n=20 with no lesion, n=32 with CIN1), compared to 418 (89%) who did not progress (307 with no lesion, 110 with CIN1). Most CIN2+ events occurred among

unvaccinated participants (73%). Four of the 52 CIN2+ events occurred among patients who were HPV-negative at enrollment, whereas the most frequently occurring types among those who progressed were HPV-16 (14 CIN2+ events) and HPV-51 (10 CIN2+ events). Only 3 of the 10 CIN2+ patients with HPV-51 were co-infected with HPV-16/18. Median time-to-progression was 12.2 months (range: 0.8-36.0 months). Unvaccinated women progressed to CIN2+ within a shorter duration compared to those with a history of HPV vaccine (9.1 months vs. 12.2 months).

Over a 3-year period, the incidence rate of CIN2+ among unvaccinated HPV-positive unvaccinated women was 4.3 per the 1000 woman-months (95% CI, 2.4-6.9; Table 4.3), whereas the rate of CIN2+ among vaccinated women was 5.7 per 1000 woman-months (95% CI: 2.1-12.6). Unvaccinated women infected with hrHPV genotypes 16, 33, 51, and 35 (inclusive of both single and multiple infections) had the highest rates of CIN2+ over 3 years of follow-up. For vaccinated women, though the incidence rates of CIN2+ was highest among women infected with genotypes 33, 31, 45 and 16, there were less than 10 women exposed for each type. Among single infections, there were sparse data and events across vaccination status; however, rates of CIN2+ were highest among genotypes 66, 33, 51 among unvaccinated women and 31, 39, and 16 among vaccinated women (Table 4.4). Unvaccinated participants infected with HPV-16 (single and multiple infection) experienced at least a 2-fold higher risk of progression to CIN2+ compared to those not infected with HPV-16 (adjusted HR or aHR, 2.5; 95% CI, 1.2-5.6; Table 4.5). The risk of progression to CIN2+ was highest among unvaccinated women infected with HPV-16 (aHR, 2.5; 95% CI: 1.2-5.6), followed by HPV-51 (aHR, 2.2; 95% CI, 0.9-4.9) and HPV-33 infection (aHR, 2.2; 95% CI: 0.5-9.3). High-risk genotypes HPV-16 and HPV-51 were most predictive progression to CIN2+ in CINCS women, with 3 times the risk of progression compared to women who did not have infection with either hrHPV type (HR: 3.2, 95% CI: 1.5-7.2). Using risk stratification groups developed by Stoler et al. [63], unvaccinated women infected with hrHPV types 16, 31 or 18 (single and multiple infection) were twice as likely to progress to CIN2+ compared to women infected with other HPV types (HR: 2.1; 95% CI: 1.1-4.1; adjusted HR: 1.9; 95% CI: 1.0, 3.7; Table 4.6). For women with a single type infection and no history of the HPV vaccine, the 3-year cumulative

probability of progression to CIN2+ was highest for infection with genotypes 16/31/18, followed by those infected with types 33/58/52/45 (Figure 4.1).

Discussion

This 3-year prospective study investigated progression to CIN2+ by HPV genotype among women in North Carolina with no CIN lesion and colposcopy-confirmed CIN1 at enrollment. The most prevalent hrHPV types among unvaccinated women were 66, 51, 16 in multiple infections and types 16, 52, 35 in single infections. The highest CIN2+ incidence rates among unvaccinated CINCS participants generally corresponded with prevalence estimates for any infection type, with the exception of high incidence in hrHPV-33 infection. However, incidence rates among single infections were highest among types 66, 33 and 51. In women with a history of HPV vaccination, there was decreased the prevalence of types 6, 11, 16, and 18, as expected. HrHPV types with the highest CIN2+ incidence rates (31, 39, 16 in single infections; 33, 31, and 45 in multiple infection) were not the most prevalent types (51, 66, 52/58 in single infections; 66, 39, 51 in multiple infections). HPV-16 and HPV-51 collectively were most predictive of progression to CIN2+ among unvaccinated participants, whereas hrHPV types 33 and 42 were most predictive in vaccinated women. These findings increase the understanding of HPV epidemiology and attribution of HPV genotypes in high-grade CIN progression, implicating potentially new screening criteria for clinical screening and management guidelines.

We assessed the predictability of hrHPV types for progression to CIN2+ individually and in three risk categories. Previous data suggest that collective assaying of hrHPV types could increase specificity of HPV testing while maintaining high sensitivity (>90%) to detect CIN2+ [66]. The present study findings, when stratified by absolute risk of CIN2/3+ [63] support that types 16, 31 and 18 may distinguish women at highest risk of progression to CIN2+ in the CINCS population. Results from the BD Onclarity™ HPV Clinical trial proposed stratification of patients according to CIN2/3+ risk attributed to HPV genotype [63]. In the BD trial, baseline absolute risk of CIN2+ was 30.1%, 17.6% and 12.3%, respectively, for hrHPV types 16, 31 and 18. The risk of CIN3+ was 20.0%, 9.9%, 6.6%, respectively for

16, 31, 18. Though the absolute risk of CIN2/3+ attributed to HPV-51 was 1.6%-9.0% in the BD Onclarity™ HPV Clinical Trial, our study may suggest that HPV-51, as well as established oncogenic genotype 16, are potentially predictive of progression to CIN2+ over a 3-year period.

Cross-sectional studies of HPV prevalence have attributed genotypes 16, 18 and 45 to CIN2+/ICC [61, 67]. By contrast, a US-based screening population study showed that 16, 31, and 18 may indicate the highest risk of CIN2+ (20%, 10%, 6.6%, respectively) compared to other hrHPV types [68]. Our study, though a smaller sample size, provides useful information regarding the risk of progression to CIN2+ over time among negative and low-grade cases relative to case-control or prevalence study design. The current clinical guidelines for HPV co-testing with LBC recommend genotyping for hrHPV, and specifically for HPV 16/18, given a negative cytology result [1]. However, the predictive use of other hrHPV types in progression has not been well characterized for CIN management. It is imperative in the case of negative cytology, ASCUS and potentially LSIL results that the risk of CIN2/3+ be adequately quantified to avoid excessive follow-up and to continue improving efficacy of overall precancer screening.

As new methods for screening are developed, one must also consider the impact of single and multiple type HPV infections in risk assessment for ICC. Previous research suggests there may be a clustering of homologous HPV types during HPV infection [69]. Infection with lrHPV types has been shown to decrease the potential for progression to more severe cervical lesions [70]. Infection with both lrHPV and hrHPV types may prolong the diagnosis of cancer in situ or squamous cervical cancer compared to cases infected with only hrHPV types [71]. Data on the risk of progression to CIN2+ due to any single HPV type infection in the CINCS cohort were inconclusive due to small sample size, thus the presence of multiple type infections should be factored into the interpretations of the results. However, as most CIN2+ events with HPV-51 infection were not co-infected with hrHPV genotypes 16 or 18, the rate of progression to CIN2+ for women with multiple infection with types inclusive of genotype 51 is noteworthy for future investigation in larger cohorts.

The potential for genotypic interactions in multiple HPV infections may be addressed with the introduction of the recently FDA approved 9-valent (9v) vaccine, which protects against genotypes 6,11,16,18, 31, 33, 35, 52 and 58 [72]. Clinical trial data on the quadrivalent vaccine demonstrated cross-protective efficacy against CIN2+ associated with types 33, 31, 45, and 51 [73]. Though we did not observe high prevalence or attribution of progression to CIN2+ for all types included in the 9v vaccine in the CINCS cohort regardless of vaccination status, the inclusion of other hrHPV types in the latest iteration of the vaccine may show increased protection against infection with multiple and related genotypes and ultimately reduce CIN2+ risk.

To our knowledge, this study is among the few to prospectively examine progression to CIN2+ by HPV genotype who had normal epithelia or CIN1 following an abnormal cytology test. The use of longitudinal data to explore genotype-specific CIN2+ risk is advantageous compared to cross-sectional studies to quantify risk over time. The data on predictive HPV types were strengthened by the use of the elastic net Cox regression model for high-dimensional variable selection. In order to determine which combination of the genotypes may be most predictive of progression to CIN2+, the elastic net method allowed for the inclusion of correlated variables, as would be the case with HPV genotypes that are phylogenically-related.

Among potential limitations, we did not assess HPV infection by genotype at study follow-up, which would have improved the ability to examine persistent HPV infection and observe any genotypic changes in infection in relation to progression to CIN2+. Future work would be strengthened with multiple HPV genotype measurements throughout the study period. There was potential for selection bias due to attrition, but no differences were observed between women who dropped out compared to women who remained in the study. It is also possible that there was an over-estimation of person-time contributed by each patient, as the person-time was dependent on study visit adherence and clinical recommendations for CIN follow-up. Small sample size impacted the ability to make strong inferences on CIN2+ risk attributed to a single HPV genotype, as well as among vaccinated CINCS participants, which comprised only 20% of the total analytic sample. There were fewer CIN2+ events and fewer women infected with

certain HPV genotypes in the vaccinated group, affecting the precision of HR estimates. However, it should be noted that fewer events among women with a history of HPV vaccination potentially demonstrates the efficacy of protection against development of CIN2+. The endpoint was defined as CIN2+, given the short duration of follow-up for the study. Observing CIN3+ events over a longer study duration would have strengthened our analyses, as a smaller proportion of CIN2+ cases progress to invasive cancer compared to CIN3+ cases [74]. Though the CIN3+ endpoint is more proximal to invasive cancer, there is clinical value in determining risk stratification earlier in the natural history of HPV-associated CIN.

In conclusion, the data support the need to further investigate the utility of non-16/18 hrHPV types for predicting progression to CIN2+ among women with negative or low-grade cytology/histology over time. Preventive strategies for CIN will benefit from increased knowledge of individual and composite risk attribution of HPV genotypes to CIN progression. With the increasing vaccine coverage in the population and subsequent shifts in HPV prevalence, it is important to identify hrHPV types that may require more immediate intervention and that can improve the risk stratification model for the clinical management of CIN. Successful characterization and implementation of hrHPV type-specific risk stratification would shift the screening paradigm, creating a new template for clinical management practice of CIN and cervical cancer prevention in the US.

Tables and Figures

Table 4. 1 Enrollment characteristics of 472 CINCS* colposcopy participants with normal histology or low-grade CIN (CIN1)*, stratified by HPV† vaccination status

	No prior HPV† Vaccination(n=364) N (%)				Prior HPV† Vaccination (n=108) N (%)			
Enrollment characteristic	**N (%)**	**Normal N (%)**	**CIN1* N (%)**	***p*-value‖**	**N (%)**	**Normal N (%)**	**CIN1* N (%)**	***p*-value‖**
Total	364 (77.1)	218 (59.9)	146 (40.1)		108 (22.9)	55 (50.9)	53 (49.1)	
Age (years)								
Median (Range)	*29.8 (20.1-64.7)*	*32.3 (20.5-64.7)*	*28.5 (20.1-64.4)*	*0.19*	*24.5 (19.7-39.9)*	*24.4 (20.9-35.2)*	*24.5 (19.7-39.9)*	*0.30*
18-24	84 (23.1)	32 (14.7)	52 (35.6)		61 (56.5)	31 (56.4)	30 (56.6)	
25-29	102 (28.0)	60 (27.5)	42 (28.8)		39 (36.1)	21 (38.2)	18 (34.0)	
30-34	52 (14.3)	39 (17.9)	13 (8.9)		6 (5.6)	2 (3.6)	4 (7.5)	
35+	126 (34.6)	87 (39.9)	39 (26.7)		2 (1.8)	1 (1.8)	1 (1.9)	
Race				0.39				0.80
Non-Hispanic White	158 (43.5)	98 (45.2)	60 (41.1)		62 (57.4)	34 (61.8)	28 (52.8)	
Black	166 (45.7)	101 (46.5)	65 (44.5)		31 (28.7)	14 (25.5)	17 (32.1)	
Other§	39 (10.8)	18 (8.3)	21 (14.4)		15 (13.9)	7 (12.7)	8 (15.1)	
Current Smoker‡				0.78				1.0
No	296 (81.5)	173 (79.7)	123 (84.3)		93 (86.1)	46 (83.6)	47 (88.7)	
Yes	67 (18.5)	44 (20.8)	23 (15.7)		15 (13.9)	9 (16.4)	6 (11.3)	
Current Oral Contraceptive Use‡				0.10				0.15
No	248 (74.9)	144 (75.8)	104 (73.8)		56 (53.3)	29 (52.7)	27 (54.0)	
Yes	83 (25.1)	46 (24.2)	37 (26.2)		49 (46.7)	26 (47.3)	23 (46.0)	
Parity‡				0.25				0.16
Nulliparous	155 (44.2)	90 (43.1)	65 (45.8)		85 (79.4)	47 (85.5)	38 (73.1)	
Primiparous (1)	96 (27.4)	61 (29.2)	35 (24.7)		16 (15.0)	5 (9.1)	11 (21.2)	

Multiparous (2+)	100 (28.5)	58 (27.8)	42 (29.6)		6 (5.6)	3 (5.5)	3 (5.8)	
High-risk HPV[†]				0.13				0.11
Negative	117 (32.1)	82 (37.6)	35 (24.0)		24 (22.2)	16 (29.1)	8 (15.1)	
Positive	247 (67.9)	136 (62.4)	111 (76.0)		84 (77.8)	39 (70.9)	45 (84.9)	

* CIN = cervical intraepithelial neoplasia; CINCS = Cervical Intraepithelial Neoplasia Cohort Study

[†] HPV = Human papillomavirus

[‡] Numbers do not add up to the total sample size due to missing data

[§] "Other" includes Hispanic/Asian/Pacific Islander/Native American/Multiracial

[‖] Chi-Square test p-value, comparing normal histology to low-grade CIN by vaccination status. Fisher's Exact test p-value used for strata where n<10.

Table 4. 2 Distribution of single/multiple HPV* genotypes in 472 CINCS† participants with normal histology/CIN1† at enrollment, stratified by HPV* vaccination status

	No Prior HPV* Vaccination (n=364)			Prior HPV Vaccination* (n=108)		
HPV* Genotype	**HPV*-Positive N (%)**	**Single N (%)**	**Multiple N (%)**	**HPV*-Positive N (%)**	**Single N (%)**	**Multiple n (%)**
Total Infected n (%)	317 (87.1)	152 (41.8)	165 (45.3)	101 (93.5)	42 (38.9)	59 (54.6)
High-Risk						
All types	247 (67.9)	109 (71.7)	138 (83.6)	84 (77.8)	34 (80.9)	50 (84.8)
16	40 (11.0)‡	16 (10.5)‡	24 (14.6)‡	10 (9.3)	3 (7.1)	7 (11.9)
18	22 (6.0)	0 (0.0)	6 (4.0)	1 (0.9)	0 (0.0)	1 (1.7)
31	22 (6.0)	8 (5.3)	14 (8.5)	2 (1.9)	1 (2.4)	1 (1.7)
33	7 (1.9)	1 (0.7)	6 (3.6)	2 (1.9)	0 (0.0)	2 (3.4)
35	20 (5.5)	11 (7.2)‡	9 (5.5)	4 (3.7)	1 (2.4)	3 (5.1)
39	29 (8.0)	7 (4.6)	22 (13.3)	12 (11.1)‡	2 (4.8)	10 (17.0)‡
45	15 (4.1)	7 (4.6)	8 (4.9)	2 (1.9)	0 (0.0)	2 (3.4)
51	36 (9.9)‡	8 (5.3)	28 (17.0)‡	20 (18.5)‡	11 (26.2)‡	9 (15.3)‡
52	33 (9.1	15 (9.9)‡	18 (10.9)	10 (9.3)	4 (9.5)‡	6 (10.2)
56	20 (5.5)	4 (2.6)	16 (9.7)	7 (6.5)	0 (0.0)	7 (11.9)
58	19 (5.0)	6 (4.0)	12 (7.3)	7 (6.5)	4 (9.5)‡	6 (10.2)
59	26 (7.1)	7 (4.6)	19 (11.5)	9 (8.3)	1 (2.4)	8 (13.6)
66	46 (12.6)‡	9 (5.9)	37 (22.4)‡	20 (18.5)‡	6 (14.3)‡	14 (23.7)‡
68	16 (4.4)	4 (2.6)	12 (7.3)	4 (3.7)	1 (2.4)	3 (5.1)
Low-Risk						
All types	169 (46.4)	40 (26.3)	129 (78.2)	54 (50.0)	8 (19.1)	46 (78.0)
6	13 (3.6)	5 (1.4)	8 (4.9)	0 (0.0)	0 (0.0)	0 (0.0)
11	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
26	3 (0.8)	1 (0.7)	2 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)
40	8 (2.2)	1 (0.7)	7 (4.2)	2 (0.4)	0 (0.0)	2 (3.4)
42	13 (3.6)	1 (0.7)	12 (7.3)	3 (2.8)	0 (0.0)	3 (5.1)
53	35 (9.6)	14 (9.2)‡	21 (12.7)	17 (15.7)	3 (7.1)	14 (23.7)
54	12 (3.3)	2 (1.3)	10 (6.1)	4 (3.7)	0 (0.0)	4 (6.8)

55	13 (3.6)	2 (1.3)	11 (6.7)	1 (0.9)	0 (0.0)	1 (1.7)
61	22 (6.0)	3 (2.0)	19 (11.5)	5 (4.6)	0 (0.0)	5 (8.5)
62	29 (8.0)	2 (1.3)	27 (16.4)	13 (12.0)	3 (7.1)	10 (17.0)
64	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
67	6 (1.7)	1 (0.7)	5 (3.0)	2 (1.9)	1 (2.4)	1 (1.7)
69	1 (0.3)	0 (0.0)	1 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)
70	16 (4.4)	5 (3.3)	11 (6.7)	4 (3.7)	0 (0.0)	4 (6.8)
71	2 (0.6)	0 (0.0)	2 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)
72	4 (1.1)	0 (0.0)	4 (2.4)	0 (0.0)	0 (0.0)	0 (0.0)
73	12 (3.3)	1 (0.7)	1 (6.7)	3 (2.8)	0 (0.0)	3 (5.1)
81	18 (5.0)	0 (0.0)	18 (10.9)	2 (1.9)	0 (0.0)	2 (3.4)
82	8 (2.2)	1 (0.7)	7 (4.2)	1 (0.9)	0 (0.0)	1 (1.7)
83	7 (1.9)	1 (0.7)	6 (3.6)	7 (6.5)	0 (0.0)	7 (11.9)
84	23 (6.3)	0 (0.0)	23 (13.9)	5 (4.6)	1 (2.4)	4 (6.8)

* HPV = Human papillomavirus

† CINCS = Cervical Intraepithelial Neoplasia Cohort Study; CIN = cervical intraepithelial neoplasia

‡ Top 3 most prevalent high-risk HPV genotypes in the among infection type stratum (all vs. single vs. multiple)

Table 4. 3 Incidence rates of CIN2+*, stratified by HPV† genotype‡ infection and vaccination history among 472 CINCS* participants over 3 years of follow-up

	No HPV† Vaccination (n=364)			HPV† Vaccination (n=108)		
HPV† Genotype	**CIN2+* Events**	**Woman-Months§**	**Incidence Rate (95% CI)‖**	**CIN2+* Events**	**Woman-Months‖**	**Incidence Rate (95% CI)‖**
Total	34	7860	4.3 (3.0, 6.0)	14	2230	6.3 (3.6, 10.3)
High-Risk						
All	30	5428	5.5 (3.8, 7.8)	13	1729	7.5 (4.2, 12.5)
16/31/18¶	14	1684	8.3 (4.7, 13.6)	4	201	2.0 (0.6, 4.8)
33/58/52/45¶	6	1597	3.8 (1.5, 7.8)	5	482	1.0 (0.4, 2.3)
39/68/35/51/59/56/66¶	10	3255	3.1 (1.6, 5.5)	4	1248	3.2 (1.0, 7.7)
16	10	891	11.2 (5.7, 20.0)	3	175	17.1 (4.4, 46.7)¶
18	2	535	3.7 (0.6, 12.4)¶	0	20	0.0 (0.0)
31	2	407	4.9 (0.8, 16.2)	1	26	38.5 (1.9, 189.7)¶
33	2	182	11.0 (1.8, 36.3)¶	1	18	55.6 (27.8, 274.0)¶
35	3	371	8.1 (2.1, 22.0)¶	1	64	15.6 (0.7, 77.1)¶
39	4	592	6.8 (2.1, 16.3)	2	267	7.5 (1.3, 24.8)
45	1	361	2.8 (0.1, 13.7)¶	1	41	24.4 (1.2, 120.3)¶
51	7	718	9.7 (4.3, 19.3)	2	342	5.8 (1.0, 19.3)¶
52	5	669	7.5 (2.7, 16.6)	1	218	4.6 (0.2, 22.6)¶
56	2	491	4,1 (0.7, 13.5)	2	186	10.8 (1.8, 35.5)¶
58	2	385	5.2 (0.9, 17.2)	3	222	13.5 (3.4, 36.8)¶
59	3	637	4.7 (1.2, 12.8)	0	154	0.0 (0.0)
66	4	954	4.2 (1.3, 10.1)	2	439	4.6 (0.8, 15.1)
68	1	376	2.7 (0.1, 13.1)	0	97	0.0 (0.0)
Low-Risk						
All	17	3732	4.6 (2.7, 7.1)	6	1130	5.3 (2.2, 11.0)
6	2	275	7.3 (1.2, 24.0)¶	0	--	0.0 (0.0)
11	--	--	--	--	--	--
26	1	71	14.1 (0.7, 69.5)¶	--	--	--
40	0	154	0.0 (0.0)	0	53	0.0 (0.0)
42	2	313	6.4 (1.1, 21.1)	3	75	40.0 (10.1, 108.9)¶
53	4	824	4.9 (1.5, 11.7)	3	294	10.2 (2.6, 27.8)
54	0	237	0.0 (0.0)	1	74	13.5 (0.7, 66.7)¶
55	3	267	11.2 (2.9, 30.6)	0	36	0.0 (0.0)

61	5	589	8.5 (3.1, 18.8)	1	74	13.5 (0.7, 66.7)¶
62	2	607	3.3 (0.6, 10.9)	0	221	0.0 (0.0)
64	--	--	--	--	--	--
67	1	149	6.7 (0.3, 33.1)¶	0	53	0.0 (0.0)
69	0	24	0.0 (0.0)	--	--	--
70	1	434	2.3 (0.1, 11.4)	0	92	0.0 (0.0))
71	0	72	0.0 (0.0)	--	--	--
72	0	83	0.0 (0.0)	--	--	--
73	1	217	4.6 (0.2, 22.7)	1	48	20.8 (1.0, 102.7)¶
81	0	475	0.0 (0.0)	0	56	0.0 (0.0)
82	1	231	4.3 (0.2, 21.4)¶	0	11	0.0 (0.0)
83	1	170	5.9 (0.3, 29.0)¶	2	154	13.0 (2.1, 42.9)¶
84	3	552	5.4 (1.4, 14.8)	0	135	0.0 (0.0)

* CINCS = Cervical Intraepithelial Neoplasia Cohort Study; CIN = cervical intraepithelial neoplasia

† HPV = Human papillomavirus

‡ Includes single and multiple genotype infections

§ Sum of woman-months for all women infected with noted genotype at enrollment at risk of developing CIN2+

‖ IR = Incidence rate per 100 woman-months; 95% CI = 95% Confidence interval

¶ N<10 women "exposed" or infected with HPV type

** Risk stratification groups adapted from Stoler et al. 2017 [63]

Table 4. 4 Incidence rates of CIN2+* by single HPV† infection, stratified by vaccination status among 472 CINCS* participants over 3 years of follow-up

	No HPV† Vaccination (n=152)			HPV† Vaccination (n=42)		
HPV† Genotype	**CIN2+* Events**	**Woman-Months**	**Incidence Rate (95% CI)‡**	**CIN2+* Events**	**Woman-Months**	**Incidence Rate (95% CI) ‡**
Total	14	3323	4.2 (2.4, 6.9)	5	877	5.7 (2.1, 12.6)
High-Risk						
16/31/18‖	2	678	3.0 (0.5, 9.7)	2	73	27.4 (4.6, 90.5)¶
33/58/52/45‖	3	736	4.1 (1.0, 11.1)	2	216	9.3 (1.6, 30.6)¶
39/68/35/51/59/56/66‖	5	1055	4.7 (1.7, 10.5)¶	1	422	2.4 (0.1, 11.7)
16	2	393	5.1 (0.9, 16.8)	1	67	14.9 (0.7, 73.6) §
18	0	148	0.0 (0.0)	--	--	--
31	0	137	0.0 (0.0)	1	6	166 (8.0, 822.0) §
33	1	30	33.3 (1.7, 164.4) §	--	--	--
35	2	197	10.1 (1.7, 33.5)	0	9	0.0 (0.0)
39	0	128	0.0 (0.0)	1	24	41.7 (2.1, 205.5) §
45	0	208	0.0 (0.0)	--	--	--
51	2	144	13.9 (2.3, 45.9) §	0	226	0.0 (0.0)
52	1	369	2.7 (0.1, 13.4)	1	109	9.2 (0.5, 45.3) §
56	0	100	0.0 (0.0)	--	--	--
58	1	129	7.8 (0.4, 38.2) §	1	107	9.3 (0.5, 46.1) §
59	0	181	0.0 (0.0)	0	32	0.0 (0.0)
66	1	162	61.7 (3.1, 304.4) §	0	112	0.0 (0.0)
68	0	144	0.0 (0.0)	0	19	0.0 (0.0)
Low-Risk						
All	4	791	5.1 (1.6, 12.2)	0	165	0.0 (0.0)
6	1	71	14.1 (0.7, 69.5) §	--	--	--
11	--	--	--	--	--	--
26	0	36	0.0 (0.0)	--	--	--
40	0	12	0.0 (0.0)	--	--	--
42	0	12	0.0 (0.0)	--	--	--
53	1	284	3.5 (0.2, 17.4)	0	64	0.0 (0.0)
54	0	24	0.0 (0.0)	--	--	--
55	0	48	0.0 (0.0)	--	--	--
61	1	57	17.5 (0.8, 86.5) §	--	--	--

62	1	16	62.5 (3.1, 308.2)[§]	0	59	0.0 (0.0)
64	--	--	--	--	--	--
67	0	36	0.0 (0.0)	0	17	0.0 (0.0)
69	--	--	--	--	--	--
70	0	120	0.0 (0.0)	--	--	--
71	--	--	--	--	--	--
72	--	--	--	--	--	--
73	0	6	0.0 (0.0)	--	--	--
81	--	--	--	--	--	--
82	0	36	0.0 (0.0)	--	--	--
83	0	34	0.0 (0.0)	--	--	--
84	--	--	--	0	25	0.0 (0.0)

* CINCS = Cervical Intraepithelial Neoplasia Cohort Study; CIN = cervical intraepithelial neoplasia

† HPV = Human papillomavirus

‡ IR = Incidence rate per 100 woman-months; 95% CI = 95% Confidence interval

§ N<10 women "exposed" or infected with HPV type

‖ Risk stratification groups adapted from Stoler et al. 2017 [63]

¶ $p<0.05$

Table 4. 5 Risk of progression to CIN2+*, stratified by HPV† genotypes‡ among 472 CINCS* participants by HPV† vaccination history

	No HPV† Vaccination (n=364)			HPV† Vaccination (n=108)		
HPV† Genotype	**Unadjusted HR (95% CI)§**	**Adjusted HR (95% CI)§**	**Predictive Model HR (95% CI)§‖**	**Unadjusted HR (95% CI)§**	**Adjusted HR (95% CI)§**	**Predictive Model HR (95% CI)§‖**
High-Risk						
All	1.7 (0.8, 3.6)	1.6 (0.7, 3.6)	--	4.1 (0.5, 31.6)	5.7 (0.7, 46.3)	--
16	2.7 (1.3, 5.6)**	2.5 (1.2, 5.6)**	3.2 (1.5, 7.2)**	4.3 (1.1, 15.8)¶	10.0 (2.0, 50.0)¶	--
18	0.7 (0.2, 3.1)	0.7 (0.2, 2.9)	--	--	--	--
31	1.0 (0.2, 4.3)	1.2 (0.3, 5.1)	--	9.9 (1.2, 79.5)¶	18.4 (1.9, 180.0)¶	--
33	2.2 (0.5, 9.3)	2.4 (0.5, 10.8)	--	8.5 (1.1, 66.6)¶	6.5 (0.5, 93.1)	11.1 (1.4, 88.8)¶
35	1.8 (0.6, 5.9)	1.6 (0.5, 5.7)	--	3.7 (0.5, 29.0)	6.3 (0.7, 55.0)	--
39	1.5 (0.5, 4.2)	1.9 (0.6, 5.7)	--	1.3 (0.3, 6.0)	1.6 (0.3, 7.6)	--
45	0.6 (0.1, 4.2)	0.7 (0.1, 5.3)	--	4.3 (0.6, 33.1)	7.6 (0.8, 71.4)	--
51	2.2 (1.0, 4.9)	2.1 (0.9, 4.9)	3.2 (1.3, 7.6)**	1.0 (0.2, 4.7)	1.1 (0.2, 5.1)	--
52	1.6 (0.6, 4.1)	1.8 (0.7, 4.6)	--	0.7 (0.1, 5.0)	0.5 (0.1, 4.1)	--
56	0.9 (0.2, 3.6)	0.9 (0.2, 3.7)	--	2.0 (0.4, 8.9)	3.5 (0.7, 18.7)	--
58	1.0 (0.2, 4.3)	0.9 (0.2, 3.9)	--	2.3 (0.6, 8.3)	2.4 (0.6, 10.4)	--
59	1.0 (0.3, 3.4)	1.0 (0.3, 3.4)	--	--	--	--
66	0.9 (0.3, 2.4)	1.0 (0.4, 2.9)	--	0.6 (0.1, 2.9)	0.6 (0.1, 3.0)	--
68	0.6 (0.1, 4.2)	0.6 (0.1, 4.2)	--	--	--	--
Low-Risk						--
All	0.9 (0.5, 1.7)	0.8 (0.4, 1.6)	--	0.8 (0.3, 2.4)	0.9 (0.3, 2.7)	--
6	1.6 (0.4, 6.6)	1.5 (0.4, 6.5)	--	--	--	--
11	--	--	--	--	--	--
26	2.9 (0.4, 21.4)	3.4 (0.4, 26.3)	--	--	--	--
40	--	--	--	--	--	--
42	1.3 (0.3, 5.4)	1.7 (0.4, 7.2)	--	7.8 (2.1, 29.2)¶	23.8 (1.6, 360.9)¶	8.3 (2.2, 31.8)¶
53	1.0 (0.4, 2.9)	1.0 (0.3, 2.7)	--	1.9 (0.5, 6.8)	3.3 (0.8, 13.5)	--
54	--	--	--	3.2 (0.4, 24.6)	4.6 (0.5, 40.5)	--
55	2.6 (0.8, 8.6)	2.3 (0.7, 7.6)	--	--	--	--

61	1.9 (0.7, 4.9)	2.2 (0.8, 5.8)	--	2.9 (0.4, 22.3)	4.2 (0.4, 44.6)	--
62	0.7 (0.2, 2.7)	0.7 (0.2, 2.8)	--	--	--	--
64	--	--	--	--	--	--
67	1.3 (0.2, 9.3)	1.2 (0.2, 9.3)	--	--	--	--
69	--	--	--	--	--	--
70	0.5 (0.1, 3.4)	0.4 (0.1, 3.2)	--	--	--	--
71	--	--	--	--	--	--
72	--	--	--	--	--	--
73	1.01 (0.1, 7.4)	1.3 (0.2, 10.0)	--	4.6 (0.6, 35.3)	6.4 (0.7, 55.5)	--
81	--	--	--	--	--	--
82	0.8 (0.1, 6.0)	0.7 (0.1, 5.4)	--	--	--	--
83	0.7 (0.1, 5.4)	1.4 (0.2, 10.6)	--	2.5 (0.5, 11.1)	1.0 (0.2, 5.7)	--
84	1.0 (0.3, 3.4)	0.7 (0.2, 2.9)	--	--	--	--

* CIN = cervical intraepithelial neoplasia; CINCS= Cervical Intraepithelial Neoplasia Cohort Study

† HPV = human papillomavirus

‡ Includes single and multiple genotype infections

§ HR = hazard ratio; 95% CI = 95% confidence interval; adjusted HR controls for continuous age, race, current smoking status, history of HPV vaccine, continuous parity, and current oral contraceptive use

‖ Predictive elastic net Cox regression model. Model parameter specifications: α=0.012, λ = 1.7. Adjusted for other variables in model.

¶ $p<0.05$

** $p<0.001$

Table 4. 6 Risk of progression to CIN2+*, stratified by risk groups for high-risk HPV† infection‡

	No HPV† Vaccination (n=364)		HPV† Vaccination (n=108)	
HPV† Genotype	**Unadjusted HR (95% CI)§**	**Adjusted HR (95% CI)§**	**Unadjusted HR (95% CI)§**	**Adjusted HR (95% CI)§**
16/31/18‖	2.1 (1.1, 4.1)**	1.9 (1.0, 3.7)	5.7 (1.7, 18.9)¶	17.2 (3.4, 87.6)¶
33/58/52/45‖	1.4 (0.7, 2.8)	1.5 (0.7, 3.1)	1.9 (0.6, 5.6)	1.8 (0.5, 6.2)
39/68/35/51/59/56/66‖	1.5 (0.8, 2.8)	1.6 (0.8, 3.1)	0.9 (0.3, 2.6)	1.1 (0.4, 3.4)

* CIN = cervical intraepithelial neoplasia; CINCS= Cervical Intraepithelial Neoplasia Cohort Study

† HPV = human papillomavirus

‡ Includes single and multiple infections

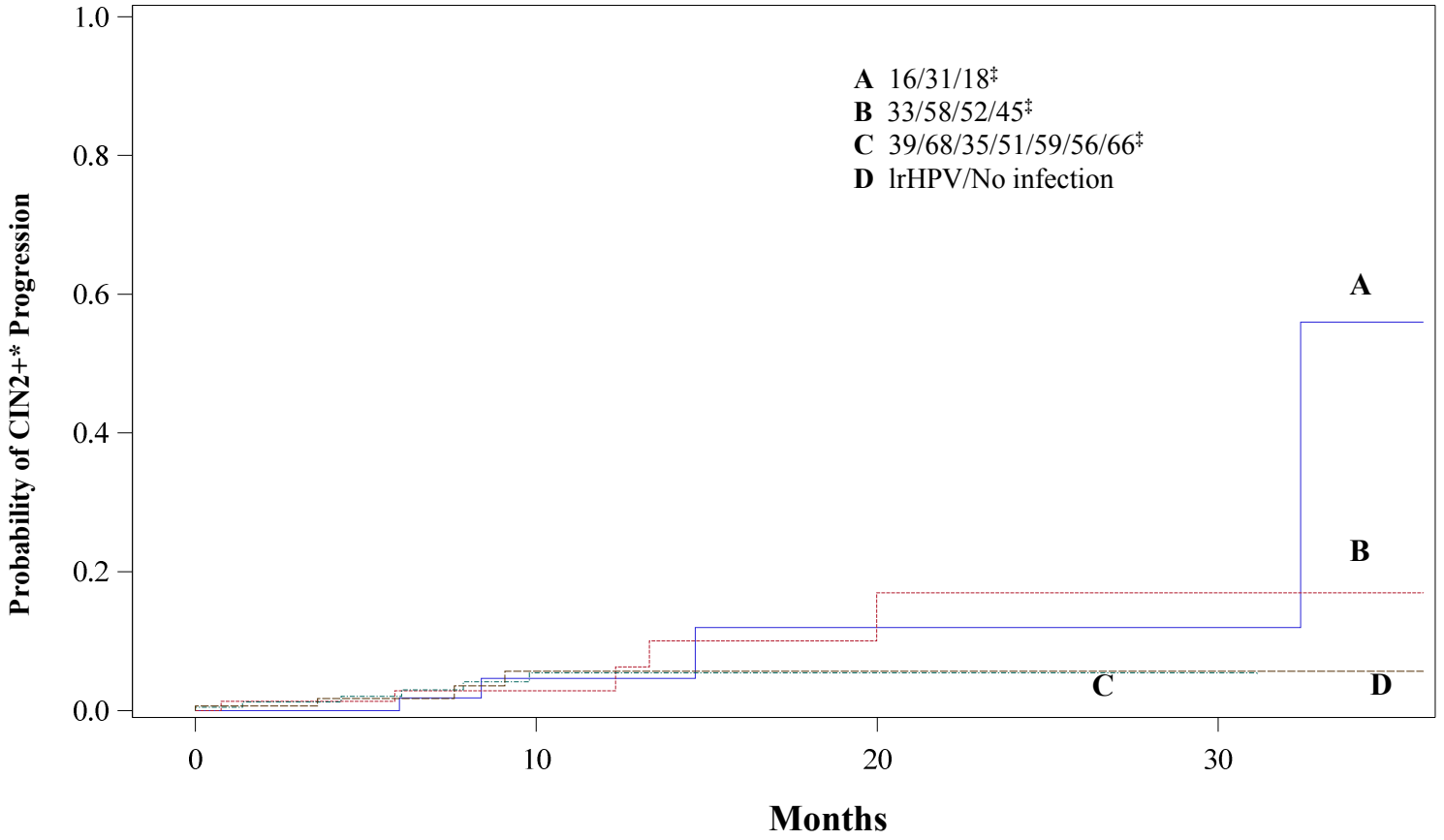
§ HR = hazard ratio; 95% CI = 95% confidence interval; adjusted HR controls for continuous age, race, current smoking status, history of HPV vaccine, continuous parity, and current oral contraceptive use

‖ Risk stratification groups adapted from Stoler et al. 2017 [63]

¶ $p<0.05$

** $p<0.001$

Figure 4. 1 3-year cumulative probability of CIN2+* progression in 194 unvaccinated and vaccinated CINCS women with single HPV† type infection, stratified by risk group



* CIN = Cervical intraepithelial neoplasia
† HPV = Human papillomavirus
‡ Risk stratification groups adapted from Stoler et al. 2017 [63]

CHAPTER FIVE: DNA METHYLATION OF IMPRINTED GENE CONTROL REGIONS IN THE REGRESSION OF LOW-GRADE CERVICAL LESIONS

Background

As of 2014, an estimated 250,000 women were living with cancer of the cervix in the United States (US) [53]. At current incidence and mortality rates, approximately 13,000 women will be diagnosed with cervical cancer in the US in 2017, resulting in over 4,000 subsequent deaths [75]. Though overall rates of cervical cancer in the US have decreased over time, the highest rates of cervical cancer incidence and mortality occur in the Southern states [75].

Nearly all invasive cervical cancers are caused by the human papillomavirus (HPV), a sexually transmitted infection that affects over 79 million people in the US [2, 6, 76]. While most HPV infections clear spontaneously, oncogenic or high-risk HPV (hrHPV) types often lead to persistent HPV infection and subsequent high-grade cervical intraepithelial neoplasia (CIN2+), a risk factor for progression to invasive cervical cancer [54].

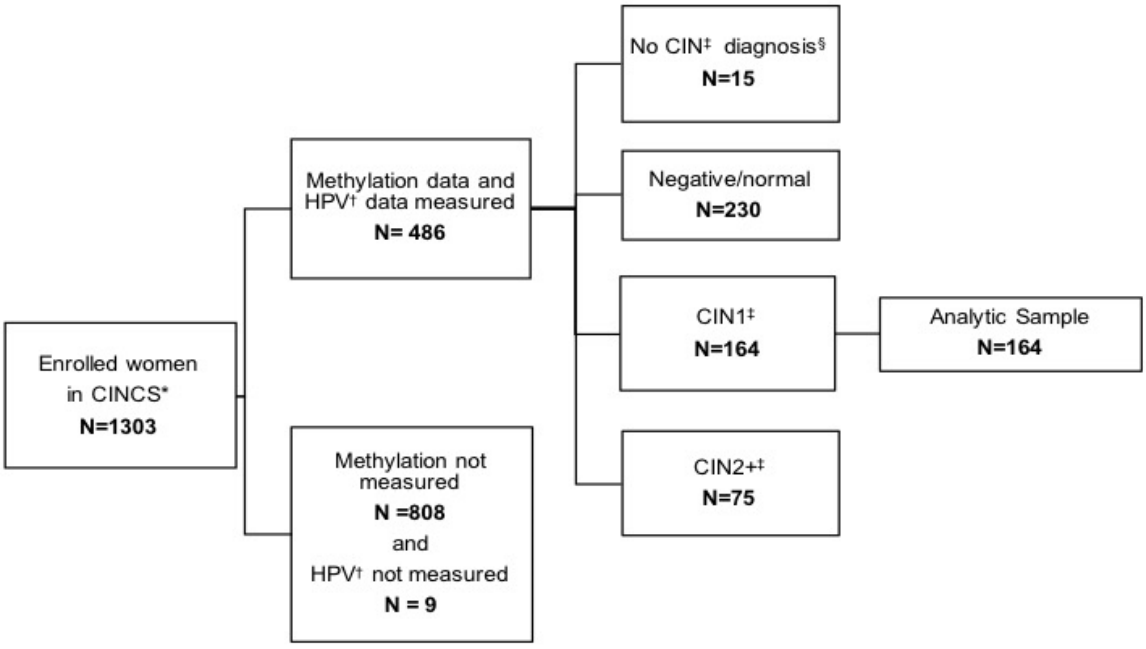
Current cervical cancer prevention strategies include the use of cytology-based testing (Pap testing) as a primary screening tool, with the addition of HPV testing to increase screening sensitivity for the detection of CIN2+ among women 30 years and older, as well as primary hrHPV screening. Although hrHPV testing is more sensitive than liquid-based cytology to detect high-grade (CIN2+), it is not as predictive of disease course among women with low-grade CIN (CIN1)[77]. A relatively small proportion of CIN1 cases progress to CIN2+, while most CIN1 cases regress to normal epithelia[78]. Follow-up of low-grade CIN is still recommended in the US until regression to normal colposcopic impression or negative cytology, leading to a high burden of cost and decreased clinical visit adherence[1, 60]. Therefore, it would be advantageous to identify novel biomarkers that can differentiate CIN1 cases which progress from CIN1 cases which regress.

Epigenetic profiles have been hypothesized as potential diagnostic biomarkers for susceptibility to cervical cancer [79, 80]. Modifications of the epigenome include DNA methylation at cytosine-guanine dinucleotide sequences (CpG sites) which can affect the expression of genes involved in cancer tumorigenesis [26]. Genomic imprinting involves inheritance of parent-of-origin specific epigenetic modifications controlling allele specific gene expression[26, 80]. Imprinted genes often exist in clusters and are regulated by imprinting centers, which can include differentially methylated regions (DMRs) that are rich in CpG sites[26].

Loss of imprinting (LOI) due to aberrant methylation at DMRs has been linked to various growth and developmental disorders[81], including Beckwith-Wiedemann Syndrome (BWS)[82]. In case-control studies, differential methylation of targeted imprinted genes has been associated with cancer outcomes, such as Wilms' Tumor of the kidney[83], breast cancer[84, 85], colorectal cancer[79, 86, 87], and prostate cancer[88]. Preliminary analyses have also found dysregulated expression of imprinted genes involved in tumor suppression (e.g. *HYMAI, PEG3, PLAGL1, MEST, CDKN1C*) in cervical cancer specimens compared to normal cervical tissue [89]. Studies have examined the influence of methylation patterns on the expression of HPV E6/E7 oncogenic proteins which deactivate host cell tumor suppressor p53 and thus may promote cervical carcinogenesis [31, 46]. The influence of host aberrant methylation at imprinted gene control regions on the natural history of low-grade CIN has not been assessed.

It is important to establish molecular-based methods of differentiating CIN1 cases which progress versus regress in order to improve clinical management. The current study examines whether aberrant DNA methylation patterns of imprinted genes influence regression of low-grade CIN in the Cervical Intraepithelial Neoplasia Cohort Study (CINCS).

Figure 5. 1 CINCS Study Flowchart



Participants with measured HPV genotypes and DMR methylation data were comparable to those of the entire cohort not included in this analysis.

Methods

Study population

From June 2010 – April 2014**,** women attending ten Duke University hospitals and clinics in Durham, North Carolina were invited to participate in CINCS, as previously described [62]. Briefly, all clinics used Duke-affiliated pathology laboratories for cytology and histological evaluation. The CINCS cohort is comprised of 1303 women who were referred for a colposcopy following an abnormal liquid-based cytology result. Participants were eligible if they provided written consent, were new visitors to the clinic, 21-79 years old, English or Spanish speakers, and able to give informed consent. We excluded women who had received previous treatment for cervical lesions—cold knife conization (CKC), electrosurgical excision procedure (LEEP), cryotherapy, or hysterectomy; had moved out of the study area; or did not intend to receive follow-up care at one of the 10 Duke clinics. Women who were

diagnosed with CIN1 at enrollment and had at least one follow-up visit were included in the present statistical analyses. Approval for this study was granted by the Institutional Review Boards at Duke University (Durham, NC, USA), North Carolina State University (Raleigh, NC, USA) and University of North Carolina (Chapel Hill, NC, USA).

Data collection and laboratory analyses

At enrollment, participants had a physician-directed cervical examination with a colposcopy-directed biopsy. Women diagnosed with CIN1 by colposcopic impression without biopsy (n=29) at enrollment were also included in the study, as prevalence of hrHPV was comparable to the hrHPV prevalence among those who underwent a biopsy. Study participants attended a clinical visit approximately every 6 months for the first two years, and every 12 months for final third year, following clinical practice guidelines for management of CIN. During each follow-up visit, all women underwent a liquid-based cytology (LBC) test. For participants with abnormal cytology results, clinic physicians performed colposcopy examination if necessary. Directed biopsies at follow-up colposcopy visits occurred only if clinically necessary, according to physician's best judgement and per clinical guidelines for management of precancerous cervical lesions[1]. Study staff administered a questionnaire to ascertain information on any behavioral and clinical characteristics at enrollment and follow-up visits, including age, race/ethnicity, current smoking status, history of oral contraceptive use and parity - risk factors for CIN and cervical cancer.

Ascertainment of Cervical Cytology and Histology

To conduct a LBC test, the clinic physician utilized a spatula and cytobrush to obtain exfoliated cervical cells. Cervical specimens were suspended in a ThinPrep® vial containing proprietary fluid with at least 50% methanol (Cytyc®, Malborough, MA, USA) for cytological assessment. All study clinic pathologists evaluated LBC cytology according to Bethesda criteria[48]. The residual LBC cervical exfoliated cell specimens were stored at 4°C prior to HPV DNA testing.

Biopsy results were also reviewed and graded for severity by a pathologist at DUPL. All histological biopsy specimens were tested for adequacy using the 2012 American Society for Colposcopy and Cervical Pathology (ASCCP) guidelines. Information on cytology and histology were abstracted from patient medical records.

HPV Testing and Typology

HPV typology was assessed using cervical exfoliated cells from the enrollment pelvic exam. ThinPrep® specimens were collected during the same enrollment visit and sent to the Johns Hopkins University and the University of Hawaii Cancer Center for laboratory testing, as previously described [49, 50]. Following DNA extraction, HPV status was determined by targeted amplification of a 450bp region of the HPV L1 genome using PGMY09/PGMY11 primers [49, 50] amplify. Amplification of the human β-globin gene was included as an internal control for sample sufficiency. Specimens identified as HPV-positive were subsequently genotyped using the HPV Linear Array® (Roche Diagnostics, Branchburg, NJ, USA). This assay is designed to detect 13 high-risk HPV types and 27 low-risk HPV types.

Assessment of DNA Methylation in Imprinted Differentially Methylated Regions (DMRs)

Nucleic acid extraction

DNA and RNA was extracted from the LBC cell pellet using a protocol for simultaneous nucleic acid extraction provided by Teltest (Friendswood, TX) for their RNA Stat60 and DNA Stat60 reagents. Nucleic acids were aliquoted, barcoded, and stored at -80ºC until required.

DNA methylation analysis

DNA methylation was measured using genomic DNA at differentially methylated regions (DMRs) regulating genomic imprinting of *IGF2/H19*, *IGF2*, *PEG1/MEST*, *Kv* DMR, *DLK1/MEG3*, *PLAGL1/HYMAI*, *PEG10* and *PEG3* imprinted domains, using Sequenom (San Diego, CA) MassARRAY EpiTYPER assays. Bisulfite-treated DNA was processed using the EZ-96 DNA Methylation Kit (Zymo

Research Corporation, Irvine, CA) to convert unmethylated DNA cytosine bases to uracil bases, leaving methylated cytosines unchanged using per manufacturer's protocol. We used Sequenom (San Diego, CA) EpiDesigner software to design primers complementary to bisulfite-converted DNA in regions without CpG nucleotides, adding a T7 promoter site to all forward primers. Polymerase chain reaction (PCR) assays were performed on the treated DNA samples using HotStarTaq DNA Polymerase (Qiagen, Hilden, Germany). PCR products were treated with Shrimp alkaline phosphatase (SAP, Sequenom, San Diego, CA) followed by transcription and T cleavage reactions according to the protocol. Cleanup and sequencing were performed according to the EpiTYPER user guide. Matrix-assisted laser desorption/ionization and time-of-flight mass (MALDI-TOF) spectrometry analysis was performed on resulting transcripts using the MassARRAY system (Sequenom). Fragments generated from the PCR assay differed by size and mass, allowing for quantification of methylated forms of each targeted fragment.

Statistical analyses

Methylation percentage was calculated at each CpG site of each imprinted gene DMR. Because imprinted gene DMRs are characterized by having one fully methylated allele and one unmethylated allele, the expected value of methylation for an imprinted gene in a diploid cell is approximately 50%. A total of 8 imprinted DMRs were analyzed with 5 CpG sites for the *IGF2/H19* DMR; 10 CpGs for the *IGF2AS* DMR; 31 CpGs for the *MEST/MESTIT1* DMR; 27 CpG sites for the *Kv* DMR; 31 CpG sites for the *DLK1/MEG3* DMR; 8 CpG sites for the *PLAGL1/HYMAI* DMR; 11 CpG sites for the *PEG10* DMR; and 12 CpG sites for the *PEG3* DMR. Genomic coordinates for each DMR have been previously published [90]. CpG sites with greater than 10% missing data at single CpG sites were excluded. As a result, DMRs in the analysis included *IGF2AS*, *DLK1/MEG3*, *PEG10* and the *Kv* DMR. Median percentages were calculated across DMRs.

Regression of cervical lesions was defined as a diagnosis of negative/normal cytology (or histology if applicable) at two consecutive follow-up visits. Per conservative clinical practice, cytology

results were utilized to determine regression status if the participants had missing histology data. Persistence was defined as a diagnosis of low-grade histology (CIN1) at follow-up or low-grade lesions during cytology testing (e.g. low-grade squamous epithelial lesions (LSIL), or atypical squamous cells of undetermined significance (ASC-US)). Progression was defined as a follow-up histologic diagnosis of CIN2+, or as a cytological diagnosis of high-grade squamous epithelial lesions (HSIL), LSIL-H (LSIL, cannot exclude HSIL), or ASC-H (ASC, cannot exclude HSIL). Women with a negative/normal screening at one follow-up time point were not considered to have regressed for the main study analyses. For women who received treatment (LEEP, CKC, cryotherapy, or hysterectomy) at a follow-up visit, the histological diagnosis from the pre-treatment specimen was utilized.

A univariate analysis was performed to assess the distribution of methylation biomarkers and covariates. Kaplan-Meier product-limit method was used to estimate the cumulative proportion of CIN1 regression, stratified by median methylation percentage at each DMR. The Log-rank test was used to assess differences between regression probabilities over time at methylation percentages below and above the median for each DMR (Figures 5.2.1-5.2.4). Cox proportional hazards regression models were employed to estimate unadjusted hazard ratios (HR) and 95% confidence intervals (95% CI) for the association between methylation at a specific CpG site and CIN1 regression. Time-to-regression was measured from the date of enrollment to the date of the second consecutive negative/normal histological or cytological diagnosis. Participants contributed person-time to the longitudinal analyses up to the occurrence of regression or the date of the last attended clinical study visit. Participants who received treatment during the study were right censored at the date of procedure. Administrative censoring occurred at 3 years. Woman-months was calculated as the sum of person-time for all women at risk among the specific methylation exposure group.

We calculated a median methylation percentage to represent a summary measure of methylation across each candidate region (when applicable) to estimate HRs and 95% CIs in the univariate and multivariate Cox regression models. In the Cox models, methylation levels (treated as a continuous variable) were rescaled using the interquartile range (IQR) for each CpG site or the IQR for median

methylation across the gene DMR. Confounders selected for the multivariable Cox regression model were determined *a priori* using conceptual models (directed acyclic graphs). Covariates considered for the analyses included continuous age at enrollment, hrHPV infection at enrollment, race/ethnicity, current smoking status at enrollment (current vs. non-current), history of oral contraceptive use (ever vs. never), and parity (continuous). Covariate modification was assessed using Akaike Information Criteria (AIC) for model fit. A sensitivity analysis was conducted to determine the change in estimate given a regression event at only one time point (Table 5.4). Further sensitivity analyses were also conducted to determine the impact of drop outs; CIN1 regression at a single follow-up visit; exclusion of women who were hrHPV negative at enrollment; and the exclusion of women who had high grade cytology (HSIL or higher) at their enrollment pap (preceding the enrollment colposcopy). All statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC).

A total of 486 women of the 1303 enrolled CINCS participants had both methylation and HPV DNA laboratory results. Of these, 15 women with no CIN diagnosis, 230 women with a negative/normal histological diagnosis at enrollment, and 75 women with CIN2+ histological diagnosis were excluded. The remaining 164 CIN1 cases at enrollment were included in analyses.

Results

Median age of CIN1 cases at enrollment (n=164) was 26.6 (range: 21-64.4 years) (Table 5.1). Nearly half of participants were non-Hispanic White (47%), while over one-third were Black (39%). Only 11% of participating women were current smokers, whereas most had a history of oral contraceptive use (78%). Over 80% of participants had infection with any hrHPV type; 16% were infected with HPV16 or HPV18, and <1% had infection with both HPV16 and HPV18.

Median DMR methylation levels were 58% for *IGF2AS*; 29% for *Kv* DMR, 52% for *MEG3*, and 32% for *PEG10* (Table 5.2). Within the *Kv* and *PEG10* DMRs, there was ≤30% methylation at most CpG sites. There was little variation in *Kv* DMR methylation (DMR IQR = 0.07) and in *PEG10* methylation (DMR IQR = 0.07) among women with CIN1.

Median study follow-up time was 10.5 months (SD: 6.9; range: 0.9-30.8). Thirteen women (8.5%) dropped out after enrollment. No differences were observed between those who dropped out (n=13) compared to those who had at least one follow-up visit (n=151). Over the 3-year study duration, a total of 53 (35.1%) women regressed from CIN1, compared to 98 (64.9%) who did not regress (37 women who regressed at only one visit, 41 women with persistent CIN1 and 20 women progressed to CIN2+). Median time to CIN1 regression was 12.6 months (range: 4.5-24.0 months). Fewer than half of follow-up diagnoses assessed for the longitudinal analysis were defined by cytology, as cases with normal or low-grade cytology at their follow-up visit did not warrant a biopsy (43% cytology-based follow-up diagnoses vs. 57% colposcopy/biopsy-based follow-up diagnoses).

The unadjusted estimated cumulative probability of regression plots showed that women with methylation percentages above the median at *IGF2AS* CpG 5 and at the *PEG10* DMR had lower incidence of regression over a 3-year period compared to women with methylation percentages below the median (Figures 5.2.1-5.2.4). No notable differences were observed in CIN1 regression probability at the *Kv* and *MEG3* DMRs (Figures 5.2.2-5.2.3).

There was at least a 40% decrease in the probability of CIN1 regression for women with higher methylation compared to women with lower methylation at *IGF2AS* CpG 5 (unadjusted HR, 0.57; 95% CI: 0.34, 0.794; p=0.03; Table 5.3). At the *PEG10* DMR, women with higher methylation had a 21% decrease in likelihood of CIN1 regression (unadjusted HR, 0.79; 95% CI: 0.65-0.97; p=0.02). After adjusting for continuous age, hrHPV status, race, current smoking status, continuous parity and history of oral contraceptive use, the probability of CIN1 regression decreased by 58% for women with higher methylation at *IGF2AS* CpG 5 compared to those with lower methylation (adjusted HR or aHR, 0.42; 95% CI, 0.23-0.77; p=0.005). Within the *PEG10* DMR, the likelihood of CIN1 regression decreased by 22% for women who had higher methylation compared to women with lower methylation (aHR: 0.78, 95% CI, 0.63-0.96, p=0.02). Evidence for modification by a covariate was not observed in data analyses. A sensitivity analysis was conducted to consider time to CIN1 regression at a single follow-up visit and found that increased methylation at both *PEG10* and *IGF2AS* DMRs remained associated with a

decreased probability of CIN1 regression (Table 5.4). Additionally, an increase in methylation at *Kv* DMR resulted in a 28% decrease in the probability of first CIN1 regression (aHR: 0.72; 95% CI, 0.56-0.93; p=0.01). Exclusion of either hrHPV-negative participants or participants with high grade cervical cytology at their enrollment Pap test did not significantly change adjusted hazard ratio estimates (data not shown)

Discussion

This longitudinal study of 164 CIN1 patients is among the first to prospectively examine aberrant methylation patterns of regulatory regions of imprinted genes and their association with low-grade CIN regression. The 3-year cumulative probability of CIN1 regression was lower among women with higher levels of methylation compared to women with lower levels of methylation at the *PEG10* DMR and the *IGF2AS* DMR CpG 5. Higher methylation levels at the *IGF2AS* and *PEG10* DMRs decreased the likelihood of CIN1 regression. A decrease in the probability of CIN1 regression due to increased methylation at *Kv* DMR was also observed over one follow-up visit. These findings may implicate these DMRs as potential epigenetic biomarkers of regression in low-grade CIN cases.

Abnormal methylation of *IGF2* may be associated with mechanism involved in cervical tumorigenesis. Though a notable association was found between aberrant methylation at *IGF2AS* CpG 5, conclusions from this data regarding the entire regulatory *IGF2AS* region could not be made. Aberrant DNA methylation of the *IGF2* and *PEG10* DMRs and other sequences regulating imprinted genes have been previously associated with higher risk of cervical dysplasia and invasive cancer in cross-sectional studies [91-94]. Hypomethylation at the *IGF2* DMR was associated with an increased risk of invasive cervical carcinoma in a case-control study conducted in Tanzania [93]. *IGF2AS* is a paternally expressed component of a downstream imprinted center, IC1 (located on human chromosome 11p15.5) that promotes cell proliferation[95].

Paternally-expressed *PEG10* also appears to have a role in cell proliferation [96, 97]. Overexpression of *PEG10* was associated with poor outcomes in hepatocellular carcinoma [96] and

increased proliferation of breast cancer cells in culture [97]. These findings warrant additional research into the relationship between *PEG10* methylation and expression in CIN/ICC tissues in a larger cohort, as well as further investigation into other epigenetic changes (e.g. chromatin modifications) that may be associated with imprinted gene deregulation in cervical cancer development.

Interestingly, when estimating time-to-regression at a single follow-up visit, increased methylation at the *Kv* DMR also decreased the probability of regression relative to women who had lower levels of methylation. The *Kv* DMR is maternally methylated, compromised of the imprint control region IC2, located at Chr11p15.5, which regulates at least 11 imprinted genes [98]. Imprinting at this region controls transcription of the long non-coding RNA *KCNQ1OT1*, which regulates the expression of *cyclin-dependent kinase inhibitor 1C* gene (*CDKN1C)*, an inhibitor protein involved in cell proliferation and growth regulation [99]. While studies on *Kv* DMR in the context of cervical dysplasia and ICC are limited, changes in methylation at *Kv* DMR/IC2 have been positively associated with colorectal cancer [87] and breast cancer [84].

Infection with hrHPV may influence methylation patterns in imprinted genes. Aberrant methylation of imprinted genes has been associated with a higher risk of hrHPV infection and may serve as an intermediate in the natural history of CIN1, as seen in previous studies [93, 94]. Differential methylation of DMRs associated with *PEG3, PEG1/MEST* and *IGF2* expression was strongly associated with HPV infection among women with ICC [93, 94]. Imprinted tumor suppressor *CDKN1C* was upregulated during E2 (HPV viral regulatory protein)-mediated HeLa cell senescence and concomitant repression of E6/E7 HPV viral oncogenes [100]. These findings implicate downregulation of *CDKN1C,* leading to upregulated cervical cell proliferation and subsequent cervical tumor development [100]. Inhibition of cell apoptosis due to loss of E2 expression in cervical carcinogenesis may be mediated in part by aberrant methylation and subsequent deregulation of pivotal genes, some of which are imprinted and implicated in cervical cancer development pathways [93, 100].

A major advantage of this study was the ability to prospectively assess the association between methylation markers and CIN. Accounting for time in estimating the probability of CIN1 regression

improves the strength of the association. The findings here, comparable to previous studies, further support the consideration of imprinted gene biomarkers as a screening tool for LSIL/CIN1 cases [93, 94].

Among potential limitations, this study did not assess HPV infection at study follow-up, which would have allowed for a more stringent definition of cervical lesion regression, including HPV negative status in addition to histological/cytological outcomes. However, the decision to define the regression as two consecutive negative screening results, rather than one, strengthened the robustness of this study. Future work should include capturing HPV infection status and HPV typing at all follow-up visits to determine the extent to which persistent hrHPV infection plays a role in the association between methylation patterns and the course of cervical dysplasia. Previous research has suggested that HPV-infection and cervical cell proliferation may be related to downregulation of *CDKN1C*, a tumor suppressor gene involved in cell proliferation and growth regulation controlled by IC2/*Kv* DMR [100]. The current study was also limited by the possibility that women with CIN1 at enrollment may have been misclassified. To address this, a sensitivity analysis was conducted by excluding women with high-grade cervical cytology at enrollment, producing similar results. The 2012 ASCCP guideline update [1] likely influenced loss to follow-up as well as created variability in the number of visits and duration between follow-up visits for each participant. Obtaining data on changes in smoking habits and other time-dependent behavioral/lifestyle factors also would broaden future research in order to investigate their influence on methylation patterns among women with cervical disease [101]. The incorporation of RNA/gene expression data would further characterize the impact of aberrant methylation among women with and without disease.

These study findings indicate further investigation into *IGF2* and *PEG10* DMRs as a diagnostic biomarker in women with low-grade CIN is warranted. It is critical to understand risk factors that determine the natural course of CIN in order to improve the effectiveness of current cervical cancer screening methods. Characterization of potential cervical tumorigenesis pathways related to the dysregulation of imprinted gene networks would help to establish novel epigenetic biomarkers in CIN management to reduce cervical cancer incidence.

Tables and Figures

Table 5. 1 Characteristics of 164 Women with CIN1 at enrollment in the CINCS Study*

Enrollment characteristic	N	(Range)%
Age (years)		
Median	26.6	(21.0-64.4)
18-24	65	39.6
25-29	56	34.2
30-34	13	7.9
35+	30	18.3
High-Risk HPV†		
Negative	32	19.5
Positive	132	80.5
HPV 16/18		
Negative	138	84.2
Positive	26	15.9
Race		
Non-Hispanic White	77	47.0
Black/African-American	64	39.0
Other‡	23	14.0
Current Smoker		
No	146	89.0
Yes	18	11.0
Ever Use of Oral Contraceptives§		
No	33	21.7
Yes	119	78.3
Parity§		
Nulliparous	93	57.4
Primiparous (1)	29	17.9
Multiparous (2+)	40	24.7

* CIN = Cervical Intraepithelial Neoplasia; CINCS = Cervical Intraepithelial Neoplasia Cohort Study
† Includes high-risk HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68
‡ Other includes Hispanic/Asian/Pacific Islander/Native American/Multiracial
§ Numbers do not add up to the total sample size due to missing data

Table 5. 2 Imprinted gene DMR[§] methylation distributions in 164 CIN1 patients at enrollment*

Gene Name	CpG site	N[†]	Mean	Median	SD[‡]	Interquartile Range	Range
IGF2AS	5	157	0.47	0.58	0.32	0.57	(0.00, 0.96)
Kv							
	DMR[§]	144	0.30	0.29	0.05	0.07	(0.16, 0.42)
	1	143	0.25	0.26	0.06	0.07	(0.12, 0.39)
	6	144	0.25	0.26	0.06	0.08	(0.08, 0.41)
	8,9	144	0.27	0.27	0.05	0.07	(0.03, 0.40)
	10,11,12	142	0.23	0.23	0.06	0.07	(0.03, 0.38)
	15	144	0.30	0.31	0.06	0.08	(0.14, 0.43)
	17,18	144	0.30	0.31	0.05	0.07	(0.14, 0.49)
	20	144	0.31	0.31	0.06	0.06	(0.10, 0.47)
	21	144	0.30	0.30	0.06	0.05	(0.12, 0.47)
	22	144	0.24	0.24	0.04	0.04	(0.12, 0.43)
	24	144	0.29	0.30	0.08	0.10	(0.11, 0.44)
	26,27	144	0.55	0.55	0.06	0.07	(0.35, 0.70)
MEG3[‖]							
	DMR[§]	152	0.49	0.52	0.11	0.19	(0.26, 0.82)
	3	152	0.57	0.59	0.16	0.17	(0.11, 1.00)
	6	150	0.54	0.54	0.09	0.28	(0.31, 0.96)
	15	152	0.40	0.39	0.10	0.14	(0.00, 0.80)
	16	150	0.38	0.38	0.10	0.15	(0.00, 0.82)
	20,21	150	0.43	0.42	0.09	0.14	(0.05, 0.75)
	22	150	0.54	0.54	0.09	0.17	(0.31, 0.96)
	23	152	0.46	0.46	0.10	0.14	(0.21, 0.95)
	26,27,28	152	0.55	0.55	0.11	0.21	(0.03, 0.89)
	29,30,31	152	0.57	0.56	0.12	0.19	(0.12, 1.00)
PEG10							
	DMR[§]	140	0.30	0.32	0.08	0.07	(0.00, 0.45)
	5	137	0.31	0.32	0.10	0.07	(0.02, 0.83)
	6	138	0.26	0.27	0.12	0.14	(0.00, 0.47)
	7	139	0.32	0.34	0.10	0.07	(0.00, 0.85)

***** CIN = Cervical Intraepithelial Neoplasia; CINCS = Cervical Intraepithelial Neoplasia Cohort Study

[†] Numbers do not add up to the total sample size due to missing data

[‡] SD = Standard Deviation

[§] DMR = Differentially methylated region; median across all CpG loci for candidate gene

[‖] MEG3 intronic differentially methylated region

Table 5. 3 Analysis of CIN1 regression by imprinted gene DMR‖/CpG site among CINCS participants*

Gene Name	CpG Site	Woman-Months	Regression Events (N)†	Unadjusted HR‡ (95% CI)	p-value	Adjusted HR (95% CI) ‡§	*p*-value
IGF2AS	5	2742	53	0.57 (0.34, 0.94)	0.03	0.42 (0.23, 0.77)	0.005
Kv							
	DMR‖	2514	49	0.81 (0.55, 1.19)	0.28	0.84 (0.57, 1.28)	0.42
	1	2485	48	0.78 (0.54, 1.14)	0.20	--	--
	6	2514	49	0.75 (0.50, 1.12)	0.15	--	--
	8,9	2514	49	1.06 (0.74, 1.50)	0.75	--	--
	10,11,12	2473	48	0.92 (0.64, 1.30)	0.63	--	--
	15	2514	49	0.73 (0.51, 1.05)	0.09	--	--
	17,18	2514	49	0.87 (0.59, 1.27)	0.46	--	--
	20	2514	49	0.89 (0.67, 1.17)	0.39	--	--
	21	2514	49	0.89 (0.66, 1.19)	0.62	--	--
	22	2514	49	1.02 (0.76, 1.37)	0.90	--	--
	24	2514	49	0.70 (0.45, 0.99)	0.04	--	--
	26,27	2514	49	1.10 (0.80, 1.53)	0.55	--	--
MEG3¶							
	DMR‖	2762	53	0.80 (0.52, 1.21)	0.28	0.82 (0.53, 1.27)	0.37
	3	2762	53	0.88 (0.58, 1.33)	0.54	--	--
	6	2720	53	0.90 (0.61, 1.35)	0.61	--	--
	15	2762	53	0.67 (0.49, 0.91)	0.01	--	--
	16	2741	52	0.76 (0.54, 1.06)	0.10	--	--
	20,21	2725	52	0.75 (0.54, 1.03)	0.07	--	--
	22	2720	53	0.90 (0.60, 1.35)	0.61	--	--
	23	2762	53	0.92 (0.68, 1.25)	0.60	--	--
	26,27,28	2762	53	0.78 (0.53, 1.15)	0.21	--	--
	29,30,31	2762	53	0.85 (0.57, 1.28)	0.44	--	--
PEG10							
	DMR‖	2517	50	0.79 (0.65, 0.97)	0.02	0.78 (0.63, 0.96)	0.02
	5	2461	49	0.57 (0.42, 0.78)	<0.001	--	--
	6	2475	50	0.66 (0.47, 0.92)	0.02	--	--
	7	2510	50	0.85 (0.69, 1.05)	0.12	--	--

* CIN = cervical intraepithelial neoplasia; CINCS = Cervical Intraepithelial Neoplasia Cohort Study; cases with at least one follow-up visit

† Regression event defined as histology- or cytology-negative diagnosis over two consecutive follow-up visits
‡ HR= Hazard ratio; 95% CI= 95% confidence intervals; continuous methylation levels rescaled using interquartile range for each CpG site
§ Adjusted for continuous age, high risk-HPV, race, smoking, continuous parity and history of oral contraceptive use
‖ DMR = differentially methylated region (median methylation of all CpG sites)
¶ MEG3 intronic differentially methylated region

Table 5. 4 Sensitivity analysis: CIN1 regression by imprinted gene using one regression time-point*

Gene Name	CpG Site	Woman-Months	Number of Events	Unadjusted HR† (95% CI)	p-value	Adjusted HR (95% CI) †‡	*p*-value
IGF2AS							
	5	1406	92	0.63 (0.42, 0.95)	0.03	0.46 (0.28, 0.75)	0.002
Kv							
	DMR§	1323	85	0.71 (0.53, 0.95)	0.02	0.65 (0.47, 0.92)	0.01
MEG3¶							
	DMR§	1410	93	0.95 (0.68, 1.32)	0.74	1.04 (0.72, 1.49)	0.88
PEG10							
	DMR§	1253	84	0.84 (0.71, 0.99)	0.03	0.82 (0.69, 0.97)	0.02

* CIN=cervical intraepithelial neoplasia; 151 cases with at least one follow-up visit

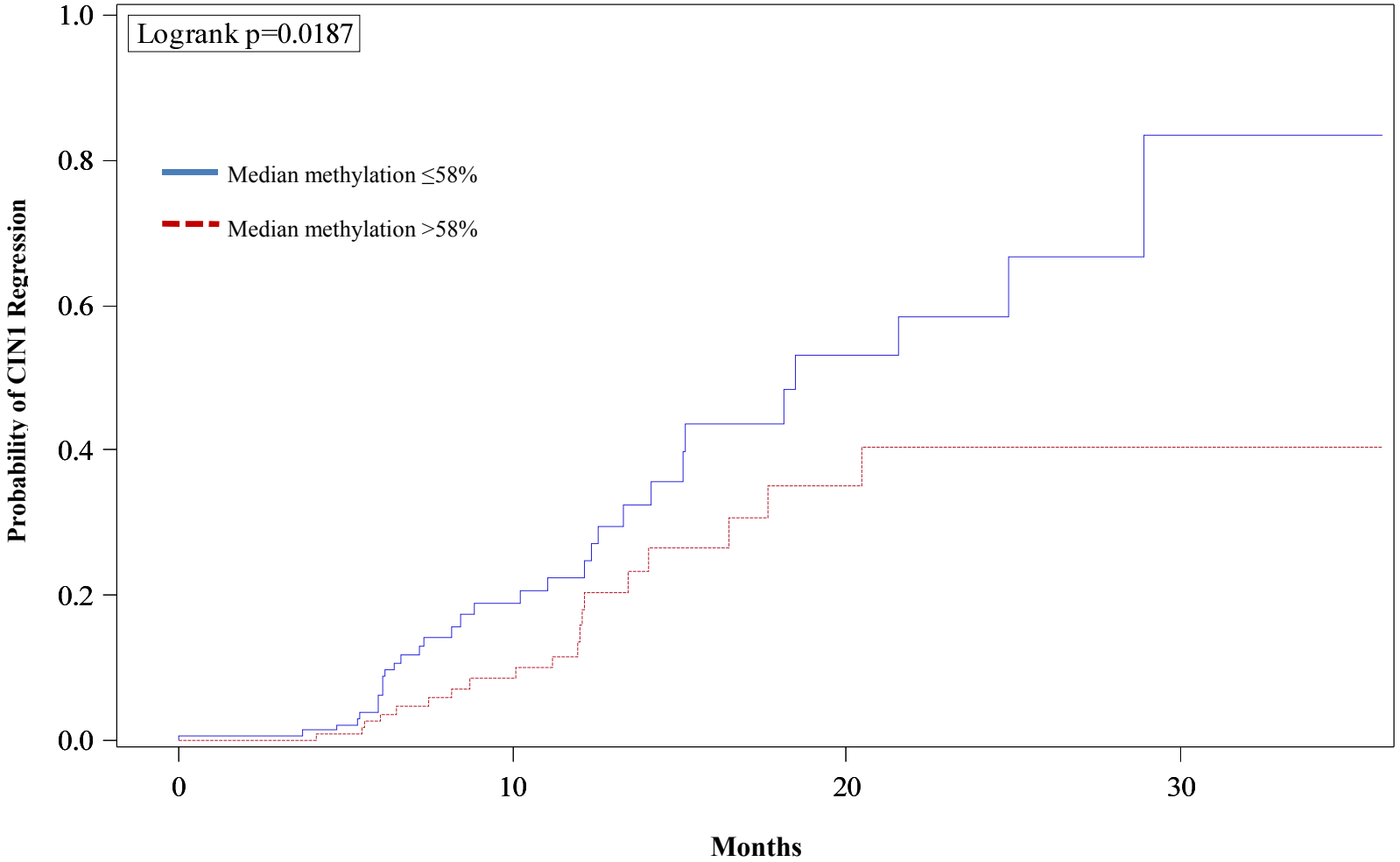
† HR=Hazard ratio; 95% CI= 95% confidence intervals; continuous methylation levels rescaled using interquartile range for each CpG site

‡ Adjusted for high-risk type human papillomavirus infection, race, current smoking, continuous parity and history of oral contraceptive use

§ DMR = differentially methylated region (median of all CpG sites)

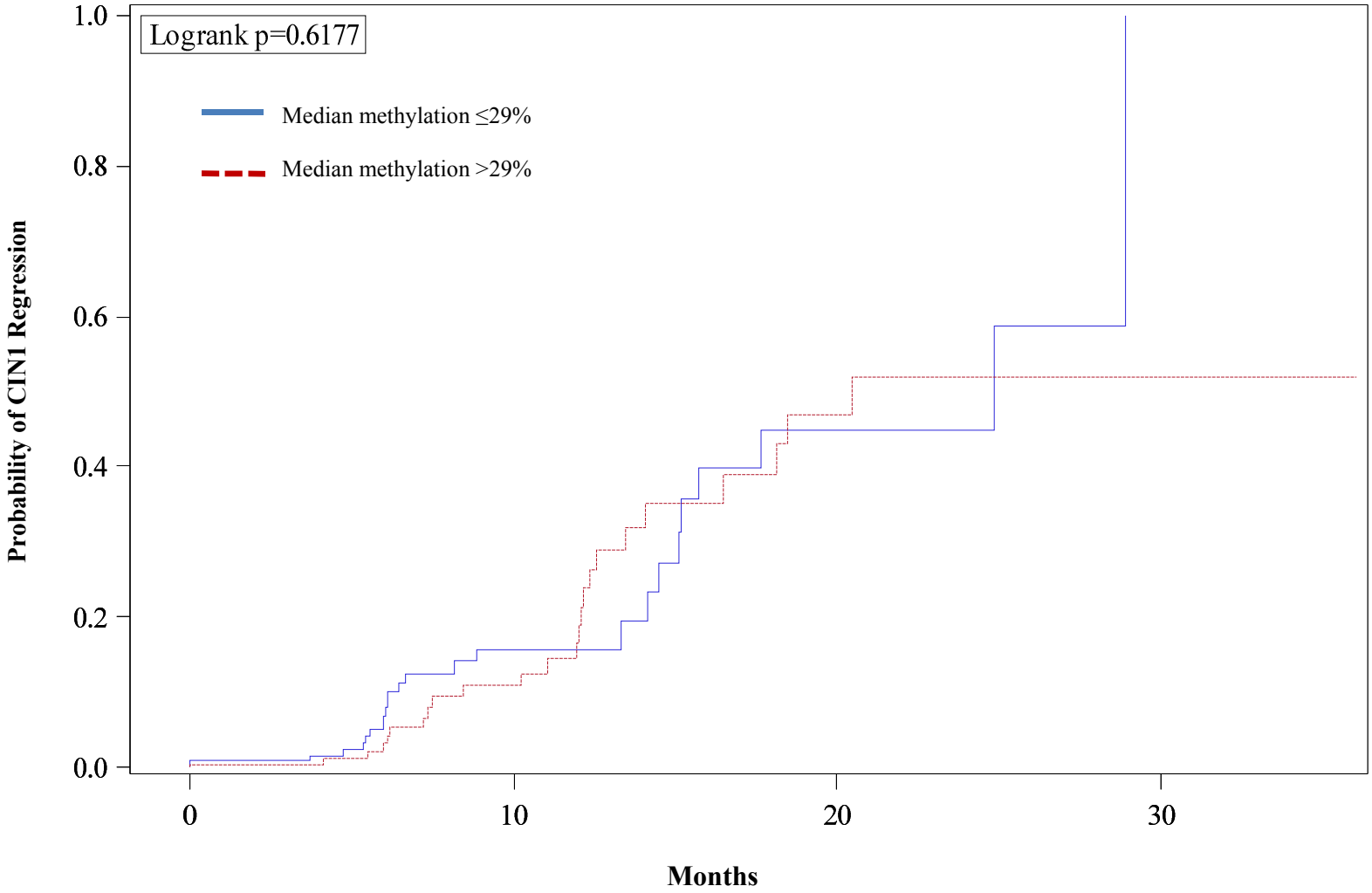
‖ MEG3 intronic differentially methylated region

Figure 5.2. 1 Time to CIN1 regression for *IGF2AS* (at CpG 5), stratified at median methylation percentage*



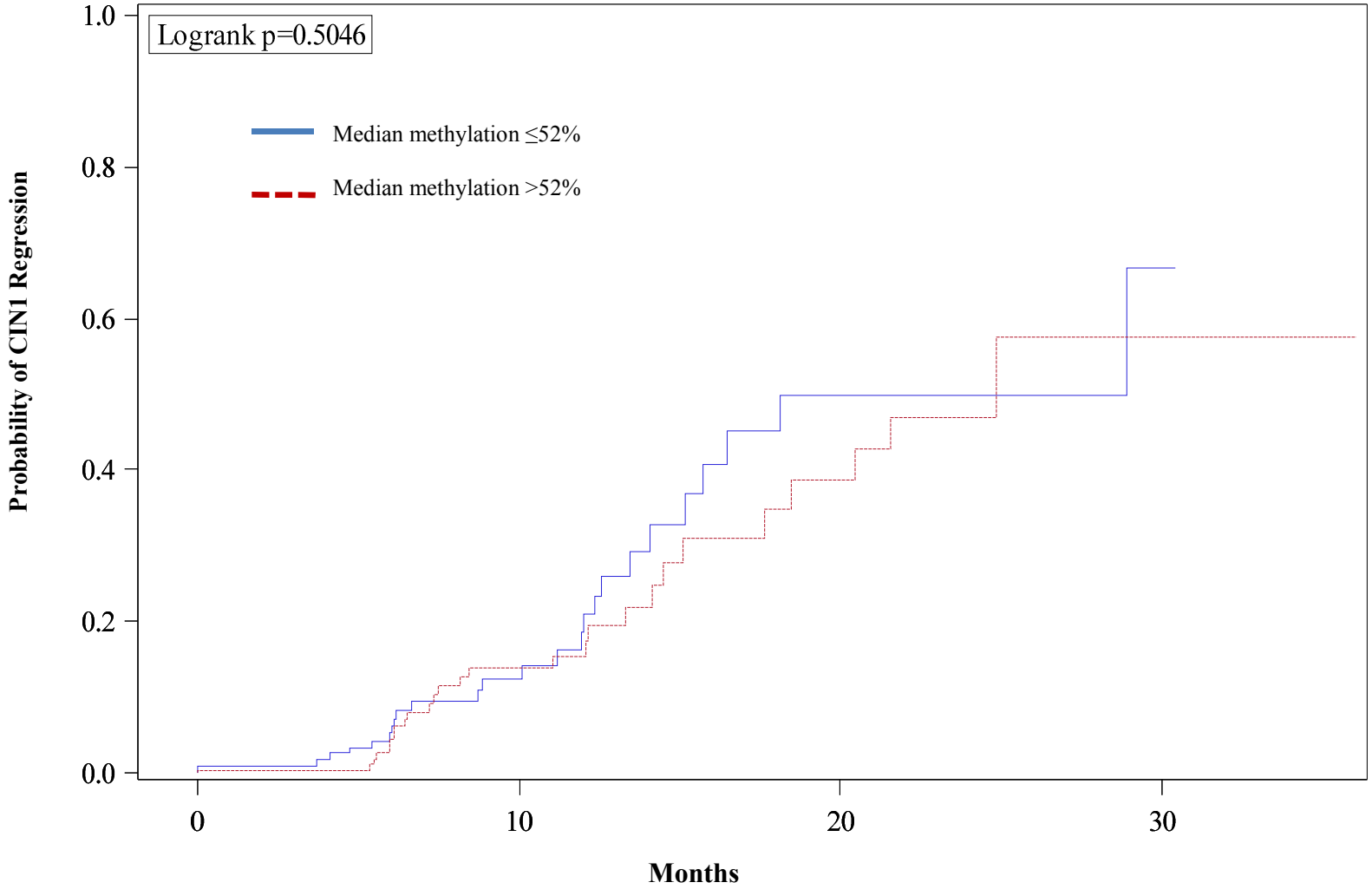
*CIN = cervical intraepithelial neoplasia

Figure 5.2. 2 Time to CIN1 regression for *Kv* DMR, stratified at median methylation percentage*



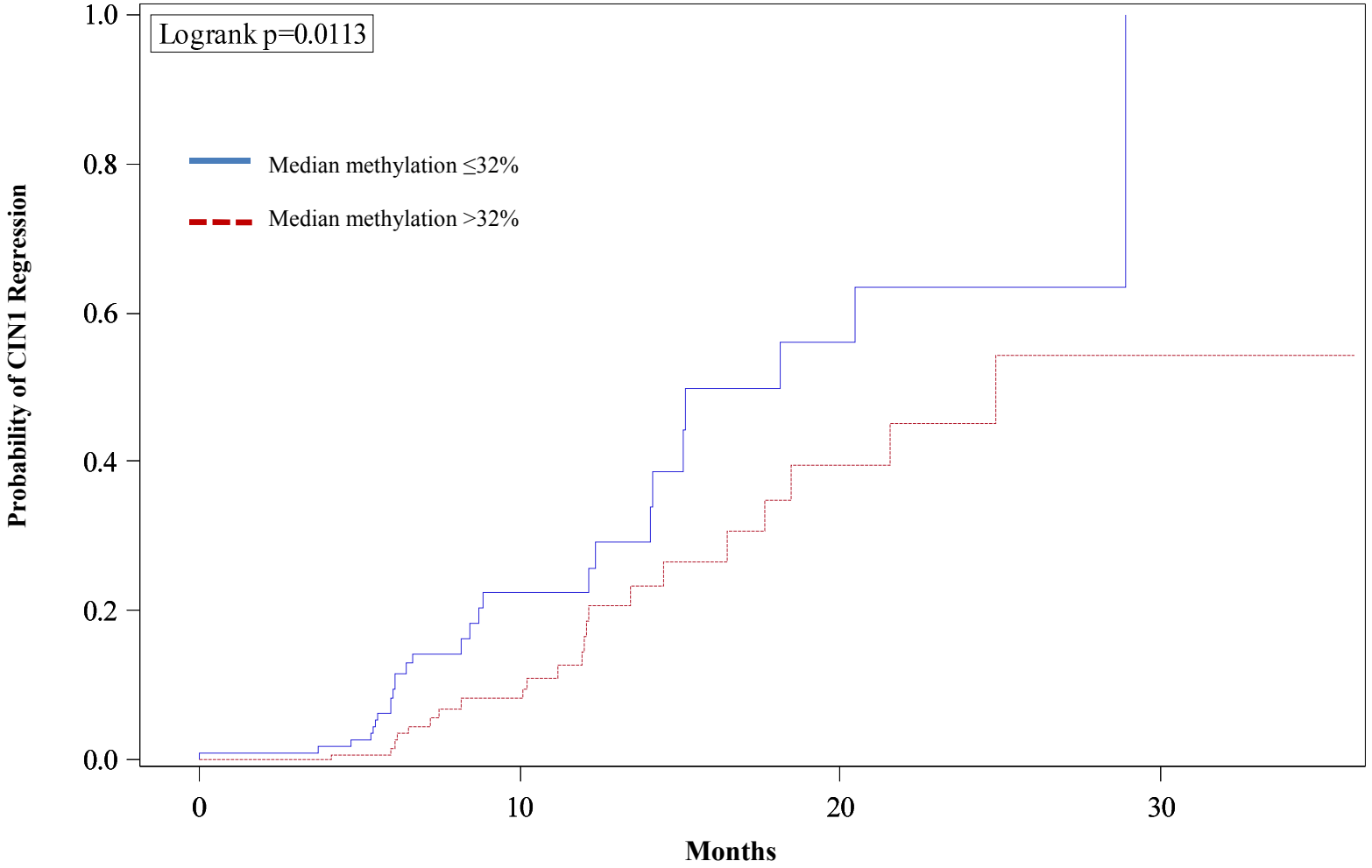
* CIN = cervical intraepithelial neoplasia; DMR = differentially methylated region (median methylation of all CpG sites)

Figure 5.2. 3 Time to CIN1 regression, stratified at median MEG3 DMR methylation percentage*



* CIN = cervical intraepithelial neoplasia; DMR = differentially methylated region (median methylation of all CpG sites)

Figure 5.2. 4 Time to CIN1 regression, stratified at median *PEG10* DMR methylation percentage*



* CIN = cervical intraepithelial neoplasia; DMR = differentially methylated region (median methylation of all CpG sites)

CHAPTER SIX: DISCUSSION

Overview

This dissertation aimed to examine biological markers associated with the natural history of CIN in a cohort of women in North Carolina in order to better characterize CIN for clinical management and improve cervical cancer prevention strategies. Detection of hrHPV, as well as aberrant methylation, may be used for triage to colposcopy among HPV-screen positive women for CIN.

Summary of Findings

The first aim of this dissertation was to examine the epidemiology of HPV infection among CINCS women and to assess the association between individual and collective HPV genotypes and progression of CIN2+ over a 3-year period. We found that hrHPV prevalence among this population was highest among types 66, 51 and 16 among unvaccinated women with multiple infections and in types 16, 52 and 35 in single infection. These prevalence estimates differed from vaccinated CINCS participants, where hrHPV genotypes 66, 39, and 51 occurred most frequently in multiple infections and genotypes 51, 66, 52/58 in single infections. Considering these prevalence estimates, we found that the highest incidence rates of CIN2+ among with women without prior vaccination occurred in hrHPV genotypes 16/33/51 in multiple, and genotypes 66, 31, and 51 in single infections. Among vaccinated women, hrHPV genotypes 33, 31, and 45 had the highest CIN2+ incidence rates in multiple infections, with genotypes 31, 39 and 16 for single infections. As expected, we found that hrHPV-16, was associated with a 3-fold increased risk of progression over 3 years. Interestingly, hrHPV-51 was also associated with a 3-fold increase of progression in the CINCS study. The implication of genotype 51 was unexpected, as

other studies on hrHPV risk attribution of CIN2/3+ have identified other alpha-9 viruses such as 18, 31 and 45.

The second aim of this dissertation was to examine the association between aberrant methylation of imprinted genes and CIN1 regression. The 3-year cumulative probability of CIN1 regression was lower among women with higher levels of methylation compared to women with lower levels of methylation at the *PEG10* DMR and the *IGF2AS* DMR CpG 5. Higher methylation levels at the *IGF2AS* and *PEG10* DMRs decreased the likelihood of CIN1 regression. A decrease in the probability of CIN1 regression due to increased methylation at *Kv* DMR was also observed over one follow-up visit. These findings may implicate *PEG10* and *IGF2AS* DMRs as potential epigenetic biomarkers of regression in low-grade CIN cases.

Public Health Significance

The findings of this research addresses gaps in the literature related to cervical precancer screening and CIN management. The data provide an increased understanding of the epidemiology of HPV genotypes in high-grade CIN progression. With the addition of hrHPV types in screening, women who are hrHPV-positive but cytology-negative may be further differentiated for triage to colposcopy. Preventive strategies for CIN, such as the HPV vaccine, will benefit from increased knowledge of individual and composite risk attribution of HPV genotypes to CIN progression, especially given the effect of the current vaccines on HPV prevalence in the population. The use of DNA methylation in relation to imprinted genes and loss of imprinting, though a novel area of study, may provide an additional biomarker (similar to p16) in identifying high-risk cases and low-risk cases during screening. With both HPV and DNA methylation of imprinted genes, there are great implications for cervical screening practices and clinical management in the US. These findings, if replicated, could encourage the continual evolution of ASCCP guidelines and new screening criteria for clinical screening and management guidelines. Successful characterization and implementation of these two biological factors in the natural history of CIN would create a new template for clinical management practice of CIN and

cervical cancer prevention in the US, ultimately promoting the reduction of cervical cancer incidence globally.

APPENDIX

A1. Risk of progression to CIN2+* by race/ethnicity by HPV† type‡ in 364 unvaccinated CINCS* patients

	Adjusted HR (95% CI)§		
HPV† Genotype	**Non-Hispanic White**	**Black**	**Other**
High-Risk			
16	4.3 (1.5, 12.7) ¶	1.8 (0.4, 8.3)	0.8 (0.1, 7.8)
51	1.7 (0.4, 7.9)	2.9 (1.02, 8.1) ¶	--

***** CINCS = Cervical Intraepithelial Neoplasia Cohort Study; CIN = cervical intraepithelial neoplasia; CIN2+ event = CIN2+ diagnosis at follow-up given colposcopy-confirmed negative or CIN1 diagnosis at enrollment

† HPV = Human papillomavirus

‡ Includes single and multiple type infections

§ HR = hazard ratio; 95% CI = 95% confidence interval. Elastic net Cox regression model parameter specifications: α=0.012, λ = 1.7. Adjusted for the other covariates in the model.

¶ $p<0.05$

A2. Risk of progression to CIN2+* by race/ethnicity among single HPV† type infections in 152 unvaccinated CINCS* patients

	Adjusted HR (95% CI)‡		
HPV† Genotype	**Non-Hispanic White**	**Black**	**Other**
High-Risk			
16	4.3 (1.5, 12.7)§	1.8 (0.4, 8.3)	0.8 (0.1, 7.8)
51	1.7 (0.4, 7.9)	2.9 (1.02, 8.1)§	--

***** CINCS = Cervical Intraepithelial Neoplasia Cohort Study; CIN = cervical intraepithelial neoplasia

† HPV = Human papillomavirus

‡ HR = hazard ratio; 95% CI = 95% confidence interval. Elastic net Cox regression model parameter specifications: α=0.012, λ = 1.7. Adjusted for the other covariates in the model.

§ $p<0.05$

REFERENCES

1. Massad, L.S., et al., *2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors.* Obstetrics & Gynecology, 2013. 121(4): p. 829-846.

2. National Center for HIV/AIDS Viral Hepatitis STD and TB Prevention: Division of STD Prevention, *Genital HPV Infection–CDC Fact Sheet*, Center for Disease Control and Prevention, Editor. 2014.

3. Workowski, K.A., et al., *Sexually transmitted diseases treatment guidelines, 2015.* MMWR Recomm Rep, 2015. 64(RR-03): p. 1-137.

4. Dunne, E.F., et al., *Prevalence of HPV infection among females in the United States.* Jama, 2007. 297(8): p. 813-819.

5. Lin, L., et al., *Racial and ethnic differences in human papillomavirus positivity and risk factors among low-income women in Federally Qualified Health Centers in the United States.* Preventive medicine, 2015. 81: p. 258-261.

6. Walboomers, J.M.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* The Journal of pathology, 1999. 189(1): p. 12-19.

7. Smith, J.S., et al., *Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: A meta-analysis update.* International journal of cancer, 2007. 121(3): p. 621-632.

8. U.S. Cancer Statistics Working Group, *United States Cancer Statistics: 1999–2012 incidence and mortality web-based report.* Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2015.

9. Porterfield, D.S., M.G. Dutton, and M.Z. Gizlice, *Cervical Cancer in North Carolina Incidence, Mortality and Risk Factors.* 2003.

10. Smith, J.S., *Ethnic disparities in cervical cancer illness burden and subsequent care: a prospective view in managed care.* Am J Manag Care, 2008. 14(6 Suppl 1): p. S193-S199.

11. Rauscher, G.H., et al., *Accuracy of self-reported cancer-screening histories: a meta-analysis.* Cancer Epidemiology Biomarkers & Prevention, 2008. 17(4): p. 748-757.

12. American Cancer Society, *Cancer Facts & Figures for African Americans 2013-2014.* 2014: Atlanta: American Cancer Society.

13. Gibbs, R.M., MG, *The impact of liquid-based cytology in decreasing the incidence of cervical cancer.* Rev Obstet Gynecol, 2011. 4(1): p. S2-S11.

14. International Agency for Research on Cancer (IARC), *Cervix cancer screening.* 2005.

15. Nanda, K., et al., *Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review.* Annals of internal medicine, 2000. 132(10): p. 810-819.

16. Coste, J., et al., *Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening.* Bmj, 2003. 326(7392): p. 733.

17. Ratnam, S., E.L. Franco, and A. Ferenczy, *Human papillomavirus testing for primary screening of cervical cancer precursors.* Cancer Epidemiology and Prevention Biomarkers, 2000. 9(9): p. 945-951.

18. Wright, T.C., et al., *2006 consensus guidelines for the management of women with abnormal cervical cancer screening tests.* American journal of obstetrics and gynecology, 2007. 197(4): p. 346-355.

19. Wright, T.C., et al., *2006 consensus guidelines for the management of women with cervical intraepithelial neoplasia or adenocarcinoma in situ.* American journal of obstetrics and gynecology, 2007. 197(4): p. 340-345.

20. Schiffman, M., et al., *Human papillomavirus testing in the prevention of cervical cancer.* Journal of the National Cancer Institute, 2011. 103(5): p. 368-383.

21. Soost, H.J., et al., *The validation of cervical cytology. Sensitivity, specificity and predictive values.* Acta cytologica, 1990. 35(1): p. 8-14.

22. Saslow, D., et al., *American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer.* CA: a cancer journal for clinicians, 2012. 62(3): p. 147-172.

23. Ronco, G., et al., *Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials.* The lancet, 2014. 383(9916): p. 524-532.

24. Wentzensen, N., et al., *Triage of HPV positive women in cervical cancer screening.* Journal of Clinical Virology, 2016. 76: p. S49-S55.

25. Egger, G., et al., *Epigenetics in human disease and prospects for epigenetic therapy.* Nature, 2004. 429(6990): p. 457-463.

26. Feinberg, A.P. and B. Tycko, *The history of cancer epigenetics.* Nature Reviews Cancer, 2004. 4(2): p. 143-153.

27. Esteller, M., *Epigenetics in cancer.* New England Journal of Medicine, 2008. 358(11): p. 1148-1159.

28. Razin, A. and A.D. Riggs, *DNA methylation and gene function.* Science, 1980. 210(4470): p. 604-610.

29. University of California at San Francisco, *DNA Methylation*. 2007.

30. Reik, W. and J. Walter, *Genomic imprinting: parental influence on the genome.* Nature Reviews Genetics, 2001. 2(1): p. 21-32.

31. Ding, D.-C., et al., *Methylation of the long control region of HPV16 is related to the severity of cervical neoplasia.* European Journal of Obstetrics & Gynecology and Reproductive Biology, 2009. 147(2): p. 215-220.

32. Fazzari, M.J. and J.M. Greally, *Epigenomics: beyond CpG islands.* Nature Reviews Genetics, 2004. 5(6): p. 446-455.

33. Khulan, B., et al., *Comparative isoschizomer profiling of cytosine methylation: the HELP assay.* Genome research, 2006. 16(8): p. 1046-1055.

34. Lui, J.C., et al., *An imprinted gene network that controls mammalian somatic growth is down-regulated during postnatal growth deceleration in multiple organs.* American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 2008. 295(1): p. R189-R196.

35. Varrault, A., et al., *Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth.* Developmental cell, 2006. 11(5): p. 711-722.

36. Enklaar, T., B.U. Zabel, and D. Prawitt, *Beckwith–Wiedemann syndrome: multiple molecular mechanisms.* Expert reviews in molecular medicine, 2006. 8(17): p. 1-19.

37. Cruz-Correa, M., et al., *Loss of imprinting of insulin growth factor II gene: a potential heritable biomarker for colon neoplasia predisposition.* Gastroenterology, 2004. 126(4): p. 964-970.

38. Wentzensen, N., et al., *Utility of methylation markers in cervical cancer early detection: appraisal of the state-of-the-science.* Gynecologic oncology, 2009. 112(2): p. 293-299.

39. Badal, V., et al., *CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression.* Journal of virology, 2003. 77(11): p. 6227-6234.

40. Clarke, M.A., et al., *Human papillomavirus DNA methylation as a potential biomarker for cervical cancer.* Cancer Epidemiology Biomarkers & Prevention, 2012. 21(12): p. 2125-2137.

41. Koshiol, J., et al., *Smoking and time to clearance of human papillomavirus infection in HIV-seropositive and HIV-seronegative women.* American journal of epidemiology, 2006. 164(2): p. 176-183.

42. Harris, T.G., et al., *Cigarette smoking, oncogenic human papillomavirus, Ki-67 antigen, and cervical intraepithelial neoplasia.* American journal of epidemiology, 2004. 159(9): p. 834-842.

43. Syrjänen, K., et al., *Smoking is an independent risk factor for oncogenic human papillomavirus (HPV) infections but not for high-grade CIN.* European journal of epidemiology, 2007. 22(10): p. 723-735.

44. Eldridge, R.C., et al., *Smoking and HPV antibodies, a mediation analysis of HPV re-infection.* Cancer Research, 2015. 75(15 Supplement): p. 852-852.

45. Schabath, M.B., et al., *Smoking and human papillomavirus (HPV) infection in the HPV in Men (HIM) study.* Cancer Epidemiology Biomarkers & Prevention, 2012. 21(1): p. 102-110.

46. Dueñas-González, A., et al., *Epigenetics of cervical cancer. An overview and therapeutic perspectives.* Molecular cancer, 2005. 4(1): p. 38.

47. Douc-Rasy, S., et al., *High incidence of loss of heterozygosity and abnormal imprinting of H19 and IGF2 genes in invasive cervical carcinomas. Uncoupling of H19 and IGF2 expression and biallelic hypomethylation of H19.* Oncogene, 1996. 12(2): p. 423-430.

48. Solomon, D., et al., *The 2001 Bethesda System: terminology for reporting results of cervical cytology.* Jama, 2002. 287(16): p. 2114-2119.

49. Gravitt, P.E., et al., *Genotyping of 27 human papillomavirus types by using L1 consensus PCR products by a single-hybridization, reverse line blot detection method.* Journal of clinical microbiology, 1998. 36(10): p. 3020-3027.

50. Gravitt, P.E., et al., *Improved amplification of genital human papillomaviruses.* Journal of clinical microbiology, 2000. 38(1): p. 357-361.

51. Friedman, J., T. Hastie, and R. Tibshirani, *glmnet: Lasso and elastic-net regularized generalized linear models.* R package version, 2009. 1(4).

52. Sill, M., et al., *c060: Extended inference with lasso and elastic-net regularized Cox and generalized linear models.* Journal of Statistical Software, 2014. 62(1): p. 1-22.

53. United States National Cancer Institute, *SEER Cancer Stat Facts: Cervix Uteri Cancer*. 2016, National Cancer Institute. Bethesda, MD.

54. Schlecht, N.F., et al., *Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia.* Jama, 2001. 286(24): p. 3106-3114.

55. Saraiya, M., *Population-based HPV genotype attribution in HPV-associated cancers.*, in *Anal Intraepithelial Neoplasia Society Conference*. 2015: Atlanta, GA.

56. Li, N., et al., *Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication.* International journal of cancer, 2011. 128(4): p. 927-935.

57. International Agency for Research on Cancer, *IARC handbooks of cancer prevention*. Vol. 10. 2005: IARC.

58. Cuzick, J., et al., *Overview of the European and North American studies on HPV testing in primary cervical cancer screening.* International Journal of Cancer, 2006. 119(5): p. 1095-1101.

59. Cox, J.T., et al., *Comparison of cervical cancer screening strategies incorporating different combinations of cytology, HPV testing, and genotyping for HPV 16/18: results from the ATHENA HPV study.* American journal of obstetrics and gynecology, 2013. 208(3): p. 184. e1-184. e11.

60. Paskett, E.D., et al., *Psychosocial predictors of adherence to risk-appropriate cervical cancer screening guidelines: a cross sectional study of women in Ohio Appalachia participating in the Community Awareness Resources and Education (CARE) project.* Preventive medicine, 2010. 50(1): p. 74-80.

61. Schiffman, M., et al., *A study of genotyping for management of human papillomavirus-positive, cytology-negative cervical screening results.* Journal of clinical microbiology, 2015. 53(1): p. 52-59.

62. Vidal, A.C., et al., *HPV genotypes and cervical intraepithelial neoplasia in a multiethnic cohort in the southeastern USA.* Cancer Causes & Control, 2014. 25(8): p. 1055-1062.

63. Stoler, M., et al. *Balancing Prevalence vs. Risk for CIN3+ by HPV Genotype:*

Bayesian Perspectives from the BD OnClarity™ HPV Clinical Trial. in *International Federation of Cervical Pathology and Colposcopy World Congress*. 2017. Orlando, FL.

64. Tibshirani, R., *The lasso method for variable selection in the Cox model.* Statistics in medicine, 1997. 16(4): p. 385-395.

65. Muñoz, N., et al., *Epidemiologic classification of human papillomavirus types associated with cervical cancer.* New England Journal of Medicine, 2003. 348(6): p. 518-527.

66. Cuzick, J., et al., *Individual detection of 14 high risk human papilloma virus genotypes by the PapType test for the prediction of high grade cervical lesions.* Journal of Clinical Virology, 2014. 60(1): p. 44-49.

67. Guan, P., et al., *Human papillomavirus types in 115,789 HPV-positive women: A meta-analysis from cervical infection to cancer.* International journal of cancer, 2012. 131(10): p. 2349-2359.

68. Bergeron, C., et al., *The clinical impact of using p16INK4a immunochemistry in cervical histopathology and cytology: An update of recent developments.* International journal of cancer, 2015. 136(12): p. 2741-2751.

69. Vaccarella, S., et al., *Concurrent infection with multiple human papillomavirus types: pooled analysis of the IARC HPV Prevalence Surveys.* Cancer Epidemiology and Prevention Biomarkers, 2010. 19(2): p. 503-510.

70. Maver, P.J., et al., *Genomic diversity of low-risk human papillomavirus genotypes HPV 40, HPV 42, HPV 43, and HPV 44.* Journal of medical virology, 2014. 86(2): p. 272-282.

71. Sundström, K., et al., *Interactions between high-and low-risk HPV types reduce the risk of squamous cervical cancer.* JNCI: Journal of the National Cancer Institute, 2015. 107(10).

72. Petrosky, E., et al., *Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices.* MMWR Morb Mortal Wkly Rep, 2015. 64(11): p. 300-304.

73. Wheeler, C.M., et al., *Cross-protective efficacy of HPV-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by non-vaccine oncogenic HPV types: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial.* The lancet oncology, 2012. 13(1): p. 100-110.

74. Hariri, S., et al., *HPV type attribution in high-grade cervical lesions: assessing the potential benefits of vaccines in a population-based evaluation in the United States.* Cancer Epidemiology and Prevention Biomarkers, 2015. 24(2): p. 393-399.

75. U.S. Cancer Statistics Working Group, *United States Cancer Statistics: 1999–2013 Incidence and Mortality Web-based Report*, in *Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute*. 2016.

76. Satterwhite, C.L., et al., *Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008.* Sexually transmitted diseases, 2013. 40(3): p. 187-193.

77. Molden, T., et al., *Predicting CIN2+ when detecting HPV mRNA and DNA by PreTect HPV-proofer and consensus PCR: A 2-year follow-up of women with ASCUS or LSIL pap smear.* International Journal of Cancer, 2005. 114(6): p. 973-976.

78. Bansal, N., et al., *Natural history of established low grade cervical intraepithelial (CIN 1) lesions.* Anticancer research, 2008. 28(3B): p. 1763-1766.

79. Jirtle, R.L. and M.K. Skinner, *Environmental epigenomics and disease susceptibility.* Nature reviews genetics, 2007. 8(4): p. 253-262.

80. Skaar, D.A., et al., *The human imprintome: regulatory mechanisms, methods of ascertainment, and roles in disease susceptibility.* ILAR journal, 2012. 53(3-4): p. 341-358.

81. Rainier, S., et al., *Relaxation of imprinted genes in human cancer.* 1993.

82. Choufani, S., C. Shuman, and R. Weksberg. *Beckwith–Wiedemann syndrome*. in *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 2010. Wiley Online Library.

83. Weksberg, R., C. Shuman, and A.C. Smith. *Beckwith–Wiedemann syndrome*. in *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 2005. Wiley Online Library.

84. Rodriguez, B.A., et al., *Estrogen-mediated epigenetic repression of the imprinted gene cyclin dependent kinase inhibitor 1C in breast cancer cells.* Carcinogenesis, 2011: p. bgr017.

85. Larson, P.S., et al., *CDKN1C/p57 kip2 is a candidate tumor suppressor gene in human breast cancer.* BMC cancer, 2008. 8(1): p. 68.

86. Ito, Y., et al., *Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer.* Human molecular genetics, 2008. 17(17): p. 2633-2643.

87. Nakano, S., et al., *Expression profile of LIT1/KCNQ1OT1 and epigenetic status at the KvDMR1 in colorectal cancers.* Cancer science, 2006. 97(11): p. 1147-1154.

88. Monk, D., *Deciphering the cancer imprintome.* Briefings in functional genomics, 2010: p. elq013.

89. Song, J., et al., *Microarray analysis of normal cervix, carcinoma in situ, and invasive cervical cancer: identification of candidate genes in pathogenesis of invasion in cervical cancer.* International Journal of Gynecological Cancer, 2008. 18(5): p. 1051-1059.

90. Li, Y., et al., *Lead exposure during early human development and DNA methylation of imprinted gene regulatory elements in adulthood.* Environmental health perspectives, 2016. 124(5): p. 666.

91. Dowdy, S.C., et al., *Biallelic methylation and silencing of paternally expressed gene 3 (PEG3) in gynecologic cancer cell lines.* Gynecologic oncology, 2005. 99(1): p. 126-134.

92. Douc-Rasy, S., et al., *High incidence of loss of heterozygosity and abnormal imprinting of H19 and IGF2 genes in invasive cervical carcinomas. Uncoupling of H19 and IGF2 expression and biallelic hypomethylation of H19.* Oncogene, 1996. 12(2): p. 423-430.

93. Vidal, A.C., et al., *PEG1/MEST and IGF2 DNA methylation in CIN and in cervical cancer.* Clinical and Translational Oncology, 2014. 16(3): p. 266-272.

94. Nye, M.D., et al., *Associations between methylation of paternally expressed gene 3 (PEG3), cervical intraepithelial neoplasia and invasive cervical cancer.* PLoS One, 2013. 8(2): p. e56325.

95. Du, M., et al., *Insulator and silencer sequences in the imprinted region of human chromosome 11p15. 5.* Human molecular genetics, 2003. 12(15): p. 1927-1939.

96. Bang, H., et al., *Expression of PEG10 is associated with poor survival and tumor recurrence in hepatocellular carcinoma.* Cancer Research and Treatment, 2015. 47(4): p. 844-852.

97. Li, X., et al., *PEG10 promotes human breast cancer cell proliferation, migration and invasion.* International journal of oncology, 2016. 48(5): p. 1933-1942.

98. Cerrato, F., et al., *The two-domain hypothesis in Beckwith–Wiedemann syndrome: autonomous imprinting of the telomeric domain of the distal chromosome 7 cluster.* Human Molecular Genetics, 2005. 14(4): p. 503-511.

99. Matsuoka, S., et al., *Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, p57KIP2, on chromosome 11p15.* Proceedings of the National Academy of Sciences, 1996. 93(7): p. 3026-3030.

100. Wells, S.I., et al., *Transcriptome signature of irreversible senescence in human papillomavirus-positive cervical cancer cells.* Proceedings of the National Academy of Sciences, 2003. 100(12): p. 7093-7098.

101. Zeilinger, S., et al., *Tobacco smoking leads to extensive genome-wide changes in DNA methylation.* PloS one, 2013. 8(5): p. e63812.