Investigation of DNA methylation and microbial biomarkers to improve cervical cancer triage and early diagnosis

Barts and the London School of Medicine and Dentistry

Queen Mary University of London

Thesis submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

February 2020

Author:

Cristiana Banila

STATEMENT OF ORIGINALITY

I, Cristiana Banila, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any United Kingdom law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: 22th of February 2021

Details of collaboration and publications:

Collection and DNA extraction of samples originating from Bhutan, Colombia, Georgia, India, Philippines, South Africa and Spain were performed by Dr Gary Clifford and Dr Vanessa Tenet at the International Agency for Research on Cancer, World Health Organisation.

Collection of samples originating from the United Kingdom was performed by Dr Kate Cuschieri at the Department of Laboratory Medicine, University of Edinburgh.

Collection of samples originating from Ethiopia was performed by Dr Birhanu Kumbi at the College of Natural and Computational Sciences, Addis Ababa University.

Collection of samples originating from the United States of America was performed by Prof Cossette Wheeler at the Centre for HPV Prevention, University of New Mexico.

Work presented in Chapters 4 & 5 of this thesis has been submitted for publication and is currently under review.

ACKNOWLEDGEMENTS

I believe that at any crossroads moment one needs just another person to believe in them and guide them onto a path of growth. I thank my supervisor, Belinda Nedjai, for being that one person for me in 2017, for the giving me the opportunity to be the first PhD student in the lab and for providing all the training and tools that have allowed me to learn so much over the last years. Additionally, I thank my mentor, Attila Lorincz, for all the guidance and stimulating scientific discussions that helped me focus on my path. From our very first meeting, they showed remarkable belief in me, and I could not be more grateful for the balance of freedom and support they provided to enable me to become the scientist/entrepreneur that I am today.

Next, I thank my colleagues and friends in the lab for their constant day-to-day guidance, training and support both inside and outside the lab. A particular thank you for Caroline Reuter, Dorota Scibior-Bentkowska and Rawinder Banwait for answering all my questions, teaching me new techniques, ordering reagents and stimulating my creativity. I thank Dionisis Grigoriadis and Connor Knight for the endless scientific and entertaining discussions as well as the moral support provided. You kept me going, I thank you for your friendship.

Many thanks for all my collaborators, Gary Clifford, Kate Cuschieri, Cossette Wheeler and Birhanu Kumbi, for providing high-quality samples for the work carried in this thesis. Also, my thanks go to Rachael Adcock for all the statistical analysis training provided.

I am very grateful to have received the support of my friends throughout my PhD journey. I thank Andrei for endlessly believing in me and for being a constant pillar of support in my journey, Javi for holding me accountable, Daniela for inspiring me to be unapologetically myself, Ararat for the stimulating conversations and history, Dragoş and Guillem for listening to my worries. A special thank you goes to my housemates who have been amazing companions during lockdown and beyond. Thanks Peter for helping me detangle Python and Phil for motivating me to write. Last but not least, thank you to Shakiba for all her encouragement and support in finishing my thesis and starting a venture.

În final, sunt recunoscătoare părinţilor şi bunicii pentru toate sacrificiile pe care le-au depus ca să îmi ofere cele mai bune condiţii ca să ajung unde sunt astăzi, dar şi pentru toată susţinerea şi încrederea acordată. Îi mulţumesc surorii mele, Tavi, pentru prietenia, distracţia si toate momentele amuzante care m-au energizat în ultimii ani.

ABSTRACT

The shift towards primary HPV-based screening roused the search for a secondary triage test that provides sufficient sensitivity to detect high-grade cervical intraepithelial neoplasia (CIN) and cancer, but also brings high specificity to avoid unnecessary clinical work and colposcopy referrals. To date, molecular biomarkers like DNA methylation and microbial signatures were investigated for molecular triage purposes as they could offer an objective, cost-effective alternative to cytology.

The S5 DNA-methylation classifier based on target CpG sites of the host gene *EPB41L3*, and viral gene regions of HPV16/18/31/33 demonstrated improved performance for detecting CIN2+ compared to either HPV16/18 genotyping, cytology or combination. We tested the performance of the S5 in detecting CIN3 and cancer from diverse geographic settings using the cut-off of 0.80 and the exploratory cut-off of 3.70 for use in low- and middle-income countries. Assays were performed using exfoliated cervical specimens and formalin-fixed biopsies from women with cytology negative results, CIN3 and cervical cancer diagnoses. We observed that S5 can accurately detect high-grade CIN and malignancy irrespective of geographic context and setting. We also show that adjustment of the S5 cut-off can be performed considering the relative importance given to sensitivity versus specificity, thus reflecting local triage modality needed.

In parallel, we investigated biomarkers from the cervicovaginal microbiota and their association with CIN3 development and the S5 DNA methylation signatures. In a pilot longitudinal study, we identified *S. amnii* as a consistent microbial biomarker for CIN3 development. The increase in *S. amnii* abundance was directly proportional to the increase of S5 classifier scores and disease severity. *S. amnii* abundance might play a role in sustaining the epigenetic landscape of the cervicovaginal space. We also found that higher proportions of *L.helveticus*, *L.suntoryeus* and *L.vaginalis* might have a potential protective role against CIN3 development in women with persistent hrHPV infections.

TABLE OF CONTENTS

Statement of originality 2
Acknowledgements 4
Abstract 5
List of figures 9
List of tables 10
Abbreviations 11
Chapter 1 - Introduction 13
1.1 HPV life-cycle: Molecular implications in carcinogenesis 13
1.1.1 Viral replication 14
1.1.2 HPV-induced transformation 17
1.2 Cervical Cancer Progression 20
1.2.1 Types of cervical cancer 20
1.2.2 Progression mechanisms 20
1.2.3 Genetic alterations associated with cervical cancer 23
1.2.4 Single nucleotide polymorphisms associated with cervical cancer risk 23
1.3 Epigenetic alterations in cervical cancer 25
1.3.1 DNA methylation 25
1.3.2 HPV-induced host DNA hypermethylation 27
1.3.3 Viral DNA hypermethylation 28
1.4 The cervicovaginal microbiota 29
1.4.1 The cervicovaginal microbiota: HPV infection acquisition and persistence 30
1.4.2 The cervicovaginal microbiota: cervical cancer progression 31
1.5 Disease Control 35
1.5.1 Implementation of HPV vaccination 35
1.5.2 Cervical cancer screening: implementation of hrHPV as the primary screening test 36
1.5.3 Triaging hrHPV positive women 40
1.5.4 Alternative triaging opportunities 41
1.5.5 The S5 DNA methylation classifier 43
1.5.6 Microbial markers for cervical cancer progression 47
1.6 Project Aims 48
Chapter 2 – Materials and Methods 49
2.1 Study design and clinical specimens 49
2.1.1 Bisulfite conversion kit determination 49
2.1.2 Worldwide quantification of DNA methylation using the S5 classifier 51

2.1.3 Investigation of the cervicovaginal microbiota 54
2.2 DNA extraction and quality assessment 56
2.2.1 LBC DNA extraction 56
2.2.2 FFPE DNA extraction 56
2.3 DNA bisulfite conversion 56
2.3.1 Bisulfite conversion optimisation 56
2.3.2 Quality assessment 57
2.3.3 Real-time PCR primer design 57
2.3.4 Bisulfite conversion efficiency quantification 58
2.4 Targeted methylation sequencing 59
2.4.1 PCR amplification 59
2.4.2 Pyrosequencing 60
2.5 The S5 DNA methylation classifier 62
2.6 HPV genotyping 62
2.7 16S rRNA metagenomic sequencing 63
2.7.1 16S rRNA library preparation 63
2.7.2 Ion S5™ metagenomics sequencing 64
2.7.3 Bioinformatics analysis of the 16S rRNA metagenomics sequencing data 65
2.7.4 Markov chain analysis on transition probabilities and simulations 65
2.8 Statistical analysis 65
2.8.1 Bisulfite conversion kit determination 65
2.8.2 Worldwide quantification of DNA methylation using the S5 classifier 66
2.8.3 Investigation of the cervicovaginal microbiota 67
2.8.4 Associations between microbial abundances and the S5 classifier score 68
Chapter 3 - Bisulfite conversion kit determination for formalin-fixed paraffin-embedded material 69
3.1 Introduction 69
3.2 Investigation of bisulfite conversion rates in liquid-based cytology and formalin-fixed paraffin-embedded samples 72
3.3 Alternative measurement of bisulfite conversion rate using non-CpG DPYS pyrosequencing assay 74
3.4 *EPB41L3* methylation assay: differentiation between <CIN2 and CIN2+ 76
3.5 Comparison of S5 methylation scores in FFPE material treated with Epitect Fast and Zymo Standard bisulfite conversion kits 78
3.6 Discussion 80

Chapter 4 –Clinical performance methylation as a biomarker for cervical carcinoma-in-situ and cancer diagnosis 82
4.1 Introduction 82
4.2 Characteristics and selection criteria 84
4.3 hrHPV prevalence in the sampled groups 84
4.4 Increasing trend in the S5 methylation scores with disease severity 88
4.5 S5 classifier sensitivty in cervical cancers at the 0.80 predefined cut-off 88
4.6 S5 classifier cut-off adjusted per country to optimise triage capacity 92
4.7 Diagnostic potential of S5 classifier in LMIC with limited resources 92
4.8 S5 classifier performance in detecting cervical cancers at the 3.70 cut-off 93
4.9 S5 classifier performance: sensitivity and specificity 95
4.10 S5 classifier component breakdown 97
4.11 Discussion 100
Chapter 5 – Microbial biomarkers in women who developed CIN3 – a pilot study 104
5.1 Introduction 104
5.2 The cervicovaginal microbiota composition according to disease and HPV status 106
5.3 The cervicovaginal microbial diversity associated with disease and HPV status 111
5.4 The cervicovaginal microbiota dynamics leading towards CIN3 development and recovery after LEEP treatment 112
5.5 Identification of cervicovaginal microbiota markers for CIN3 development 114
5.6 Correlations between the S5 classifier scores and bacterial abundances 117
5.7 Discussion 120
Chapter 6 – Conclusions and Future Directions 123
6.1 Discussion & Future Directions 123
6.1.1 Strategies for cervical cancer prevention worldwide 123
6.1.2 The S5 classifier – universal biomarker for CIN3 and cervical cancer detection 126
6.1.3 Microbial biomarkers predicting CIN3 progression – a pilot longitudinal study 129
6.1.4 Integrating DNA methylation and microbial biomarkers on Nanopore sequencing technology 132
6.2 Concluding remarks 134
Bibliography 135

LIST OF FIGURES

1.1 Stages of the HPV life cycle in the squamous epithelium
1.2 Viral E7 and E6 cellular targets promoting carcinogenesis
1.3 Cervical cancer progression upon hrHPV infection
1.4 DNA methylation patterns in cancer
1.5 Summary of cervicovaginal microbiota changes related to HPV persistence and cervical cancer progression
1.6 Cytology-based screening flowchart
1.7 HPV Genotyping-based screening flowchart
2.1 LBC vs FFPE study design flowchart
2.2 S5 Classifier world-wide validation study design flowchart
2.3 Longitudinal microbiome study design flowchart
2.4 Pyrosequencing reaction flowchart
3.1 DNA bisulfite conversion reaction outline
3.2 Bisulfite conversion kit performance and genomic quality score in LBC and FFPE samples
3.3 *DPYS* non-CpG methylation assay performance
3.4 *EPB41L3* methylation assay performance
3.5 Performance of Zymo Standard and Epitect Fast kits in LBC and FFPE samples
4.1 Global and country-stratified hrHPV prevalence
4.2 S5 classifier performance in the study group
4.3 S5 classifier score distribution per country and disease endpoint
4.4 Distribution of EPB41L3, HPV16, HPV18, HPV31L1 and HPV33L2 methylation scores based on the histopathological diagnostic of the patient
4.5 S5 classifier component breakdown
5.1 Bacterial species beta diversity in the longitudinal study
5.2 Cervicovaginal microbiota composition based on relative bacterial abundance cording to disease and HPV status
5.3 Cervicovaginal microbiota richness and diversity with attributed community state types
5.4 Transition probabilities and stationary distribution of bacterial community state types
5.5 CIN3 progression microbial biomarkers by LEfSe analysis
5.6 HPV persistence microbial biomarkers by LEfSe analysis
5.7 S5 classifier performance in the longitudinal study
5.8 Associations between the S5 classifier score and microbial abundances

LIST OF TABLES

1.1 Methylation Biomarker tests: Component genes and performance characteristics for CIN2 endpoint identification
1.2 History of the S5 classifier development
2.1 LBC vs FFPE sample characteristics
2.2 Composition and procedures of the four bisulfite conversion kits investigated
2.3 Conversion rate real-time PCR primers
2.4 S5 classifier PCR primers and PCR conditions
2.5 Pyrosequencing primers and target CpG positions investigated
4.1 Baseline characteristics of the S5-validation study participants
4.2 S5 Classifier sensitivity at the 0.80 cut-off in a cervical cancer referral group
4.3 Univariate and bivariate odds ratios (OR with 95%CI confidence intervals) for the associations between the different clinical outcomes and HPV16/18 and/or S5 cut-offs in a HPV(+)/Cyt(-) baseline.
4.4 S5 Classifier sensitivity at the 3.70 cut-off in a cervical cancer referral group
4.5 Sensitivity and specificity of the S5 methylation classifier in identifying CIN3 and invasive cancer based on different classifier cut-offs
5.1 Characteristics of the women included in the microbiota characterisation longitudinal study
5.2 Rates of CST in accordance to disease severity and HPV status/genotype at a species level

ABBREVIATIONS

16S rRNA	16S Ribosomal RNA Gene
ADC	Adenocarcinomas
ADS	Adenosquamous Cell Carcinomas
APOBEC	Apolipoprotein B mRNA Editing Catalytic Polypeptide-Like
ASCUS	Atypical Squamous Cells Of Undetermined Significance
AUC	Area Under The Curve
CIN	Cervical Intraepithelial Neoplasia
CIN2+	Cervical Intraepithelial Neoplasia Grade 2 And Above
CpG island	Cytosine And Guanine Rich Sequences
CS	Cancer Stage
CST	Community State Type
DC	Dendritic Cell
DNMT	DNA Methyltransferase
E6AP	E6-Associated Protein
EGFR	Epidermal Growth Factor Receptor
FDR	False Discovery Rate
GWAS	Genome-Wide Association Studies
HPV	Human Papilloma Virus
HPV(-)/Cyt(-)	HPV Negative Women With Negative Cytology Results
HPV(+)/Cyt(-)	HPV Positive Women With Negative Cytology Results
HSIL	High Grade Squamous Intraepithelial Lesions
IARC	International Agency For Research On Cancer
IQR	Inter-quartile Range
KOL	Key Opinion Leader
LBC	Liquid-Based Cytology
LDA	Linear Discriminant Analysis
LEEP	Loop Electrosurgical Excision Procedure
LEfSe	LDA Effect Size

LMIC	Low-Middle Income Country
LSIL	Low Grade Squamous Intraepithelial Lesions
NGS	Next Generation Sequencing
NK	Natural Killer Cell
OR	Odds Ratio
OTU	Operational Taxonomic Units
PC	Principal Component
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PRMT	Protein Arginine N-Methyltransferase
ROC	Receiver Operating Characteristic
SCC	Squamous Cell Carcinomas
SNEC	Neuroendocrine Small Cell Carcinoma
SNP	Single Nucleotide Polymorphism
VIA	Visual Inspection With Acetic Acid

CHAPTER 1 - INTRODUCTION

1.1 HPV LIFE-CYCLE: MOLECULAR IMPLICATIONS IN CARCINOGENESIS

Cancer transformation of a mammalian cell has been defined as the acquisition of permanent disturbances of growth associated with immune evasion. This enables the cancer cell to grow uncontrollably and independent of extracellular cues, escape immune responses and potentially migrate to other locations in the body, causing metastasis. In 2000 Hanahan & Weinberg have coined six fundamental biological hallmarks acquired by the cells to maintain their transformed phenotype. They include sustained proliferative signalling, evading growth suppressors, evading apoptosis, replicative immortality, sustained angiogenesis and metastasis (Hanahan, Weinberg, & Francisco, 2000). In 2011 a further four hallmarks were added to the list: deregulation of the energy metabolism, increased tumour promoting inflammation, immune evasion and genetic instability (Hanahan & Weinberg, 2011). These come as a consequence of DNA mutations, which generate genetic diversity and foster the cancer hallmarks (Snijders, Steenbergen, Heideman, & Meijer, 2006). In the case of cervical cancer, infection with the human papilloma virus (HPV) is a significant contributing factor but alone is not sufficient to trigger initiation and sustain the cancer hallmarks. However, approximatively 5% of cervical cancer cases develop independent of HPV infection (Banister et al., 2017). Other important co-factors include some sexually transmittable infections (i.e. HIV and *Chlamidia trachomatis*), smoking or long-term use of contraceptives (Moody & Laimins, 2010).

Cervical cancer is currently the fourth most common cancer death in women worldwide with 604,000 new cases and 342,000 deaths according to the 2020 Globocan report (Sung et al., 2021). There has been an approximatively 18% increase in cases registered since the last 2018 report (Bray et al , 2018). Although cervical cancer may be almost completely preventable due to the effective primary (HPV vaccine) and secondary (screening) prevention measures, the implementation of such measures has not been equal across countries. More than 80% of high-income countries (i.e. United Kingdom) have implemented national HPV vaccination programmes compared to less than 30% low- and middle-income countries (LMICs) (Brisson et al., 2020). Also, screening covers more than 60% women in high-income countries compared to only 44% of women in LMICs (Lemp et al., 2021). In consequence, cervical cancer incidence rates are disproportionally high in LMICs compared to developed countries (18.8 vs 11.3 cases per 100,000) (Sung et al., 2021). The same applies with mortality as LMICs register more than 85% of the worldwide cervical cancer deaths (Allen-Leigh et al., 2017).

1.1.1 Viral replication

HPV is a small double stranded DNA virus with a circular genome of 8kb. Out of the ~120 HPV types identified, approximately 40 have specific tropism for squamous epithelium (Leventakos et al., 2017), especially targeting primary basal keratinocytes (basal cell layer) in the genital tract of women (Moody & Laimins, 2010). HPV is mainly transmitted through sexual contact and most people are infected with the virus shortly after onset of sexual activity. Based on their oncogenic potential, 13 types are classified as high-risk (hrHPV). These are types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68 with hrHPV16 and hHPV18 causing 70% of pre-cancerous and invasive cervical lesions (Snijders et al., 2006).

The HPV genome codes for six early genes (E1, E2, E4, E5, E6 and E7) responsible for viral replication, and two late genes (L1 and L2) with structural functions and antigen recognition sites (Snijders et al., 2006). As the virus genome does not encode for its own polymerase, the replication process relies on host polymerases and depends on epithelial differentiation mechanisms. The HPV life-cycle is tailored to the differentiation program of target cells, whereby the virus infects the basal cell layer through micro-wounds in the epithelia (Figure 1.1). The basal layer cells consist mainly of stem cells and transit-amplifying cells, and it is the epithelial stem cells that must be infected for a lesion to be maintained (Steenbergen et al., 2014). In healthy basal cells, DNA replication occurs early in the cell cycle of the basal cell, to create two copies of the DNA for the resulting daughter cells. Post-mitosis, one daughter cell remains attached to the basal layer, while the other one is detached from the basal layer and begins to migrate up through the suprabasal layers, where the process of terminal differentiation begins (Steenbergen et al., 2014). This involves shutting down the cell cycling and a cross-linkage of the keratin intermediate filament to form confined cellular envelopes (Doorbar, 2005). All together, these changes promote the formation of the epithelial surface which acts as a physical barrier against the environment.

Upon HPV infection, the viral DNA is replicated in synchrony with the cellular DNA during cell division. Translation of viral replication proteins (E1 and E2) leads to the initiation of the productive phase of the infection process which involves viral DNA amplification in the form of extrachromosomal plasmids known as episomes (Cheung et al., 2006). Transfer of the episomal HPV genome in dividing basal cells is achieved through E2, which orchestrates interactions with mitotic factors, allowing episomal tethering to host chromosomes. This results in equal viral segregation into newly formed cells (Reinson et al., 2015). Early viral gene expression is maintained at a low level to minimize antigen exposure (Steinbach & Riemer, 2018). However,

the limited expression of E6 and E7 abolishes cell cycle arrest and normal terminal differentiation is moderately retarded (Moody & Laimins, 2010). The viral oncoproteins' function is to maintain the cellular environment suitable for viral replication.

A characteristic complication of the HPV infections is the phenomenon of latency (Doorbar, 2005). This involves only a limited number of cells supporting productive viral cycle during cell differentiation. During latency, HPV DNA is thought to remain in the basal epithelial cells and until immune-surveilance levels decline and viral replication is reactivated (Reinson et al., 2015). During the latent phase of the infection, low levels of E1 and E2 are involved in episomal maintenance, until host cell division triggers viral gene expression (Doorbar et al., 2012).

Once the viral genome amplification has been completed, the L1 and L2 genes are expressed, and their products induce viral capsid assembly. This triggers the release of the newly formed virion by desquamation (Figure 1.1). Studies with both papillomaviruses indicate that, once expressed, capsid proteins assemble into icosahedral capsids *via* assistance from chaperone proteins (Harper et al., 2004). It is unknown if encapsidation of the viral genome takes place during capsid assembly or after; however, encapsidation is assisted by L2 and may be facilitated by E2 proteins (Gu et al., 2004; Heino, Zhou, & Lambert, 2000). In the newly formed envelope, E1 and E4 proteins are thought to interact with the cellular keratin networks, causing their collapse, thus allowing mature virions to escape from the cells (Brown et al., 2006; Bryan & Brown, 2000).

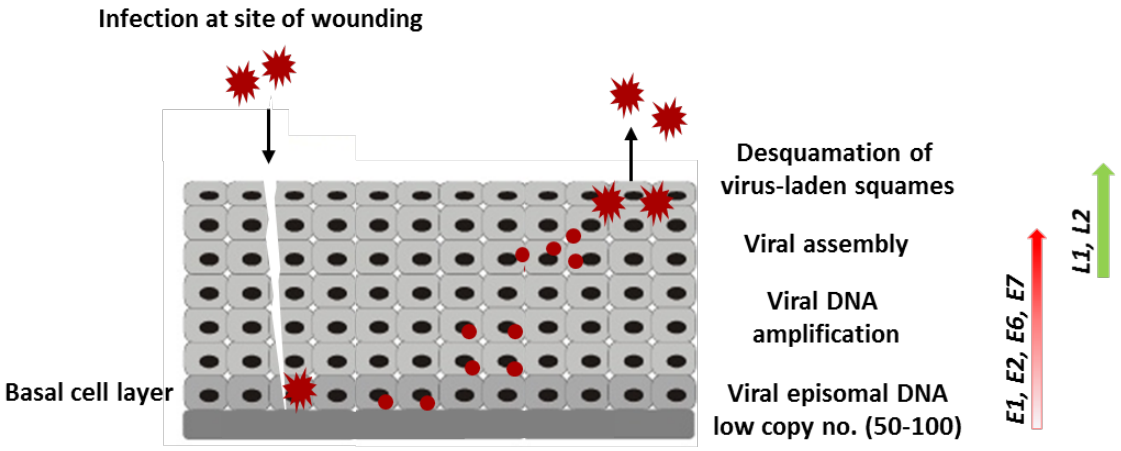

Figure 1.1 – Stages of the HPV life cycle in the squamous epithelium. The viral infection reaches the basal cell layer through micro-wounds in the squamous epithelium. The viral infection is latent until host cell division takes place. Until then, the viral DNA is kept in episomal form, at low copy number, to avoid immune detection. During cell division, viral episomal DNA is replicated in synchrony with the cellular DNA and the E1 and E2 proteins ensure that the viral DNA is equally partitioned to the daughter cells. Upon host cell division, the daughter cells relocate to the suprabasal cell layer for terminal differentiation. To sustain viral replication, E6 and E7 expression is increased preventing abolition of the cell cycle. Once the viral genome amplification has been completed, the L1 and L2 genes are expressed, and their products induce viral capsule assembly. When the virus capsule is formed, the newly formed virus is released as part of the desquamation process of the epithelium.

1.1.2 HPV-induced transformation

The link between HPV infections and cervical cancer was demonstrate in the early 1980s by Harold zur Hausen (Gravitt, 2011). The transforming potential of HPV is driven and maintained by the onco-proteins E6 and E7 and it comes as a consequence of maintaining a suitable environment for viral replication (Mirabello et al., 2017). The onco-proteins are able to affect the normal behaviour of the cell-cycle, promoting cellular immortalisation. E7 sequesters proteins from the retinoblastoma tumour suppressor (pRb) family, stimulating ubiquitin-mediated pRb degradation (Figure 1.2 A). The E7-pRb interaction releases the pRb-controlled E2F transcription factor, allowing it to freely stimulate the unregulated expression of S-phase genes (cyclin dependant kinases; CDKs) thus abrogating cell cycle control (Moody & Laimins, 2010). However, the sole disruption of the pRb/E2F complex is not enough for cellular transformation and interaction with an array of other cellular proteins such as p27 and subsequent inactivation of the CDK inhibitor have been reported (Reinson et al., 2015). Additionally, E7 has been shown to induce host genomic instability via aberrant centrosome duplication while interfering with cytokine and interferon responses, thus, allowing for immune evasion (Mirabello et al., 2017; Snijders et al., 2006).

Due to E7-mediated inhibition of pRb function, the expression of p53 tumour suppressor protein increases, priming E7 positive cells for apoptosis (Moody & Laimins, 2010). To counteract this effect and hijack normal cellular growth pathways, E6 sequesters p53 via the E6-associated protein (E6AP), forming an E6-E6AP-p53 complex. This targets p53 ubiquitin-mediated degradation and prevents p53-mediated apoptosis (Ganti et al., 2015). E6 was also reported to induce p53-independent activation of telomerase by inhibiting the action of telomerase repressors such as NFX1-91, sustaining cellular immortalization (Klingelhutz, Foster, & Mcdougall, 1996; Moody & Laimins, 2010) (Figure 1.2 B). Further, *in vitro* studies have also shown that the functions of E6 and E7 can be augmented by the expression of E5. This is achieved by the E5-mediated activation of the epidermal growth factor receptor (EGFR) which increases the signalling flux through the MAPK pathway and hence allows transformation (Moody & Laimins, 2010; Snijders et al., 2006).

Persistent presence of HPV DNA can lead to the viral genome integration in host chromosomes near regions of genomic instability, accelerating the neoplastic transformation. Many studies have noticed HPV DNA in an episomal state within precancerous lesions while in many higher grade lesions or cancer, the viral genome is found integrated in the host genome (Chen et al, 2014; Karbalaie et al., 2018; Y. Liu, et al, 2015). In the productive phase of the infection, E2 is

known to repress transcription of E6 and E7 onco-proteins (Moody & Laimins, 2010). Integration of viral DNA is accompanied by the deletion of the genes coding for E2 (Cheung et al., 2006), therefore fostering HPV-induced carcinogenesis via the constitutive activation of the viral oncogenes. Supporting evidence demonstrates that reintroducing E2 in HeLa cervical carcinoma cells is a driver for senescence rather than uncontrolled growth (Goodwin & Dimaio, 2000). This demonstrates the transformation-switch mechanism accomplished by loss of E2 expression that comes with viral integration as well as the importance of E6 and E7 in maintaining the transformed phenotype (Goodwin & Dimaio, 2000).

Lei *et al.* have postulated that about 5% of invasive cervical tumors are HPV negative (Lei et al., 2018). They also observed that patients with HPV-positive tumors show a 39% lower mortality rate than women with HPV-negative tumors (Lei et al., 2018). The inability to detect HPV could be explained by a low viral load in the tissue, below the sensitivity of detection of the test, or associated with loss of the L1 region, which can occur in a fraction of cancers (Tjalma & Depuydt, 2013). This is an interesting discovery as most HPV-tests to date target the L1 region, questioning the overall efficiency of current HPV testing. Moreover, Bannister C. *et al.* have shown that cervical cells can undergo cervical carcinogenesis without the E7-induced transforming mechanisms (Banister et al., 2017), strengthening the idea that the genetic background of the individual is a component of cervical cancer development.

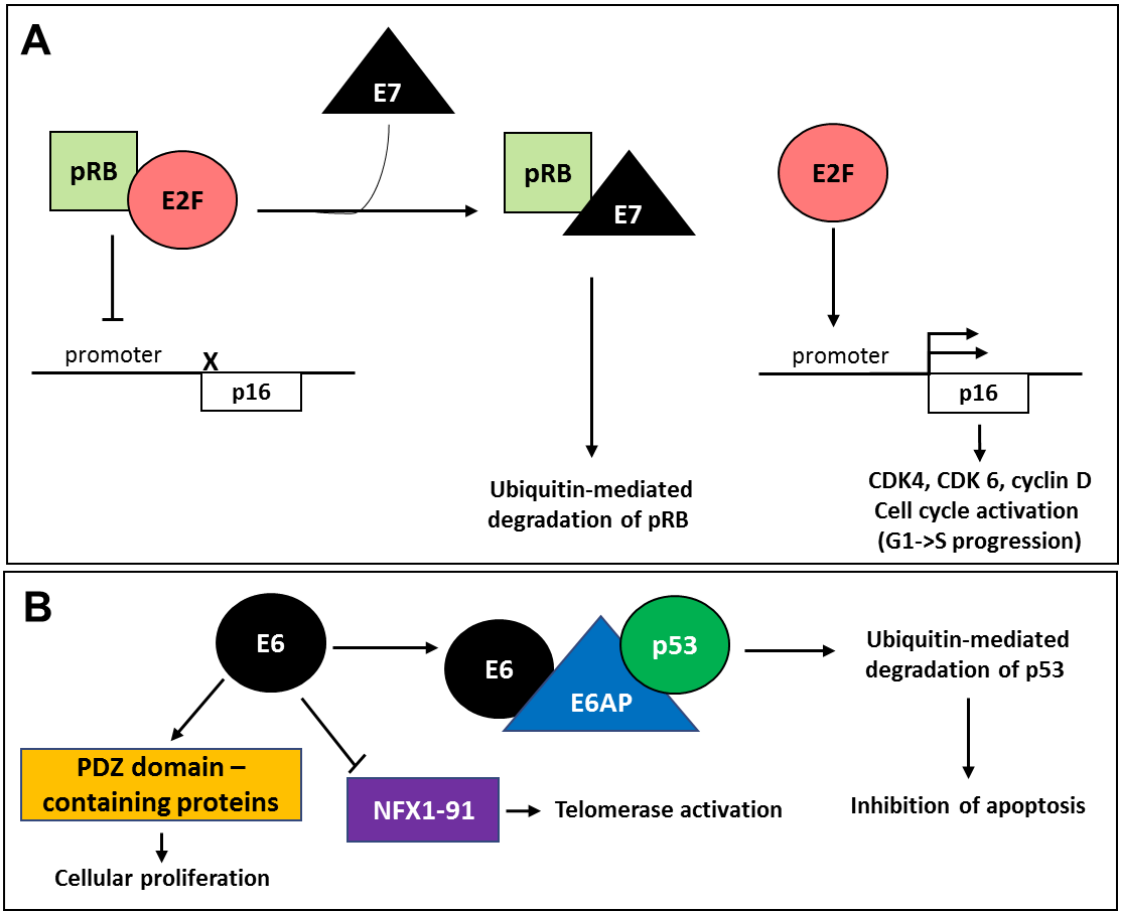

Figure 1.2 –Viral E7 and E6 cellular targets promoting carcinogenesis. A) Infection with HPV leads to the E7-mediated disruption of cell cycle control. Normally, the pRB-E2F complex regulates the expression of p16, an inhibitor of cyclin-dependent kinases (CDKs). In the presence of HPV E7, the pRB is sequestered in the E7-pRB complex, rendering E2F free to activate the transcription of p16 and, hence, stimulate CDKs expression. The E7-pRB complex eventually leads to ubiquitin-mediated degradation of pRB, leaving the cell cycle activated. **B**) The levels of p53 in a healthy cell are normally low. The HPV E6 protein promotes the degradation of p53 and prevents cell cycle arrest and apoptosis. E6 interacts with NFX1-91, PZD domain-containing proteins and several other cellular factors, facilitating the hallmarks of cancer.

1.2 CERVICAL CANCER PROGRESSION

1.2.1 Types of cervical cancer

Cervical cancer is divided into three main groups based on the type of cell that becomes cancerous. The first group are squamous cell carcinomas (SCC) originating from the squamous cells on the surface of the cervix or ectocervix. SCC represents the most common histological type of cervical cancer, with a prevalence of 70-80% (Alexandrov et al., 2013) of the clinically diagnosed cases. The second group are adenocarcinomas (ADC) starting from the glandular cells (Wilting & Steenbergen, 2016) and account for 10-20% of cervical cancers. The glandular cells produce mucus and are located along the endocervical canal which runs from the cervix to the womb (Jerant *et al*., 1997). Compared to SCCs, ADCs are not universally associated with HPV infections (Hodgson et al., 2019). Almost 100% SCC have been associated with HPV infections, only 86% of ADC were attributed to HPV (Holl et al. , 2015). The third group is a collection of rare histotypes of cervical cancers such as adenosquamous carcinomas which can occur as a consequence of both squamous cells and glandular cells becoming cancerous, although it is less common, accounting for only about 5% of cervical cancer cases (Alexandrov et al., 2013). Other less common histotypes include: neuroendocrine small cell and clear cell carcinomas. All cervical cancers are currently diagnosed through histology and subdivided in stages I-IV as per International Federation of Gynaecology and Obstetrics staging system (FIGO staging I-IV) to depict the severity of the lesion (Bhatla et al., 2019).

1.2.2 Progression mechanisms

Most HPV infections are productive but non-persistent and cleared within 2 years (Alizon et al., 2017). Such infections lead to the generation of new viral progeny but do not show any signs of cellular transformation. Productive infections may lead to moderate cellular abnormalities, histologically characterized as productive cervical intraepithelial neoplasia grade 1 or 2 (CIN1 or CIN2). However, lifetime risk of a woman to be infected with HPV is 80% and at least 80% of these are likely to be cleared without causing damage to the cervical tissue (Gravitt, 2011). It is not yet known why some individuals clear the infection and other do not. Host genetic evidence suggests that alleles of the polymorphic human histocompatibility leucocyte antigen (HLA; i.e. HLA-DRB1 and DQB1) and killer immunoglobulin-like receptor (KIR; i.e. KIR2DL2 and KIR2DL3) genes can have a protective effect towards CIN development (Carrington et al., 2005; Beskow et al., 2002). It is worth noting that carriers of the protective HLA II alleles show lower viral load and shorter HPV infections (Beskow et al., 2005; “Beskow et al., 2002). Overall, due to the inter-individual variation in effective immune responses, it is estimated that less than 5% of the

persistent HPV infections (Snijders et al., 2006) are actually transforming, priming the development of CIN3 and transition into malignancy (Figure 1.3).

As it is the case with most oncogenic viruses, the exact mechanisms of progression to cancer are still unknown (Chen et al., 2014). One hypothesis states that cervical cancer evolves in a sequential process from HPV-infected normal epithelia cells in the cervix, to a long-lasting evolution phase through all the CIN stages until sufficient mutations have been accumulated to cause the switch to cancer. It was estimated that it takes 12 to 15 years before persistently HPV infected cells undergo transformation, however quickly progressing cases have also been reported in young women (Snijders et al., 2006). The sequential model involves the development of low-grade abnormalities (i.e. low grade squamous intraepithelial lesions (LSIL) or CIN1) (Steenbergen et al., 2014) post-HPV infection. These lesions are usually transient and resolve without intervention within 1-2 years as a result of the immune system clearing the virus. Failure in the immunological control of the infection results in viral persistence which increases the risk of the development of high-grade abnormalities (i.e. high grade squamous intraepithelial lesions (HSIL) or CIN2/3) (Snijders et al., 2006; Steenbergen et al., 2014) and subsequently cancer.

An alternative hypothesis is the 'molecular switch' model. This argues that CIN3 lesions may develop directly from normal HPV-infected cells witch underwent viral integration without passing through other CIN stages (Winer et al., 2005; Nedjai et al., 2018). The 'molecular switch' model assumes that distinct epigenotypes trigger distinct morphologic changes such as CIN1 or CIN3 independently, describing partially the versatile nature of these lesions (Nedjai et al., 2018). Further, Wallin *et al.* and Zielinski *et al.* suggested that it takes an additional 10 to 12 years for a CIN3 to develop into cancer (Wallin et al. 1999; Zielinski et al., 2001), through mechanisms still not fully understood.

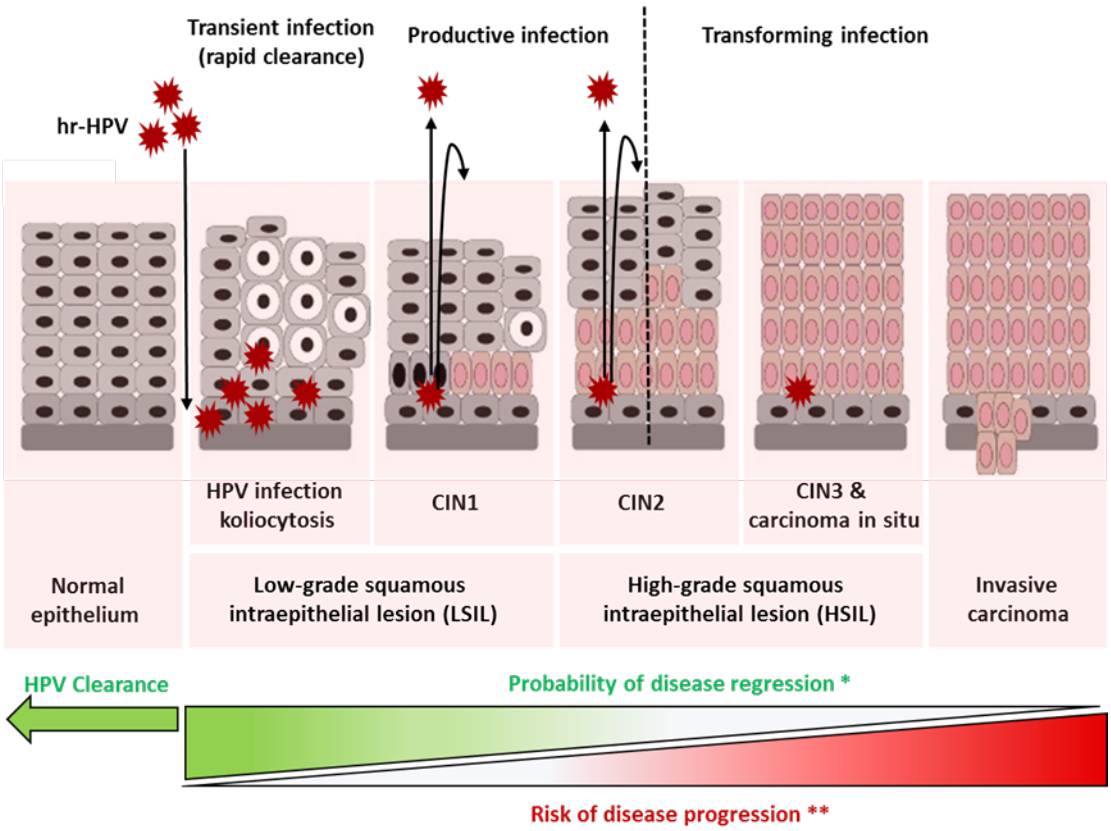

Figure 1.3 – Cervical cancer progression upon hrHPV infection. HPV accesses the basal cells though microlesions in the cervical epithelia. Upon division, the infected cells of the basal layer begin to migrate towards the surface of the epithelium. These cells are morphologically different from normal epithelial cells and show an irregular size, shape and color (koliocytosis). The outcomes of the exposure of epithelial cells to HPV are: transient infection (rapid HPV clearance), productive infection involving virion secretion by the infected cells (productive cervical intraepithelial neoplasia (CIN); mainly representing CIN1 and a subset of CIN2) and transforming infection (the remaining subset of CIN2 and CIN3). Productive CIN2 cannot be morphologically differentiated from transforming CIN2. All stages of CIN can either regress or progress to cancer. The probability of disease regression decreases with disease lesion. The risk of progressing to cancer is dependent on the molecular host cell alterations. Abbreviations: hrHPV: high risk human papillomavirus; CIN 1/2/3: cervical intraepithelial neoplasia grade 1/2/3; LSIL: low-grade squamous intraepithelial lesion; HSIL: high-grade squamous intraepithelial neoplasia.

1.2.3 Genetic alterations associated with cervical cancer

It is well known that progression to CIN3 and cervical cancer is driven by a great deal of host genetic alterations within cervical cells. However, the role of somatic mutations during HPV-induced carcinogenesis is still not fully understood (Alexandrov et al., 2013; Chang et al., 2013). To date, *PIK3CA, EPB300, FBXW7, HLA-B, PTEN, NFE2L2, ARID1A, KRAS* and *MAPK1* were found to be frequently mutated in cervical cancers (Cuschieri et al., 2018). A recent profiling study of 228 sequenced HPV-positive cervical cancer cases in the TGCA database have revealed further somatic mutations in *SHKBP1*, *ERBB3, CASP8, HLA-A, TGFBR2* (Cancer Genome Atlas Network, 2017) as well as in apolipoprotein B mRNA editing catalytic polypeptide-like (*APOBEC*) (Litwin et al., 2017) to be associated with cervical cancer development. The same analysis has been performed on 119 HPV-negative tumours and revealed high frequencies of *KRAS, ARID1A* and *PTEN* mutations (Cancer Genome Atlas Network, 2017; Vogelstein et al., 2013).

Approximatively 70% of cervical cancers exhibit genetic alterations in *PI3KCA* or *MAPK* pathways, with *PIK3CA* showing consistent somatic mutations in CINs, SCCs and ADCs (Verlaat et al., 2015). Interestingly, subsequent analysis of 209 CINs has revealed that *PIK3CA* exon 9 is a mutation locus in only 2.4% of CINs, suggesting that mutations in this gene might represent a late event during carcinogenesis (Litwin et al., 2017). Conversely, *HLA-A* and *HLA-B* mutations are considered an early event in carcinogenesis, being more frequent in CINs (Litwin et al., 2017). Evidence suggests that somatic mutations in *HLA-A* or *HLA-B* are found in 8% and 6-9%, respectively, of SCC (Ojesina et al., 2014). Additionally, *HLA-A/B* mutations were found as more common in HPV-positive (11%) than HPV-negative (7%) cancers (Cancer Genome Atlas Network, 2017; Ojesina et al., 2014).

1.2.4 Single nucleotide polymorphisms associated with cervical cancer risk

Development of cervical cancer varies among individuals. This disparity is thought to be partly caused by single nucleotide polymorphisms (SNPs), which are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) is altered (McKay et al., 2017). SNPs are the most common genetic variations in humans. The genetic variation caused by SNPs translates into non-synonymous changes, occurring in protein coding regions. These alter the encoded amino-acid mutated at the site and may cause functional changes in the mutated protein (Chen et al., 2013). With regard to cervical cancer predisposition, SNPs can be divided into susceptibility SNPs (germline genetic variations; indicate genetic predisposition to develop the disease) and progression SNPs (somatic variations; indicate the likelihood of viral infection progressing into high-grade CIN or cancer).

A literature review of case-control genome-wide association studies (GWASs) revealed susceptibility SNPs in 47 genes. The most abundant number of SNPs were identified in the Janus kinase/signal transducer and activator of transcription (*JAK/STAT)* genes. Their protein products are involved in signalling pathways of growth and division. Impaired functions of the members of the JAK-STAT signalling pathway alter the response of the cell to cytokines, thus increasing the risk of immunodeficiency disorders and various malignancies (Snijders et al., 2006). For instance, the presence of the SNPs rs10815144 in *JAK1* and rs17748980 in *STAT1* in were associated with an approximatively six and fivefold (*JAK1* SNP: Odds Ratio-OR 6.53; 95%CI 2.78–15.33, p<0.00001 and *STAT1* SNP: OR 5.23; 95%CI 3.11–14.21, p<0.00001) increase in the risk of an individual for cervical cancer. The same applies for rs4737999 (OR 6.61; 95%CI 2.92–15.45, p<0.00001) in *SULF1* or rs11177074 (OR 5.52; 95%CI 1.92–15.89, p<0.00001) in *IFNG*. In the context of GWAS, associations are considered significant for $p<10^{-7}$ and p<0.00001 (Panagiotou et al., 2012). Interestingly, most susceptibility SNPs have been found in key genes coding for proteins involved in signal transduction for growth, immune responses or tumorigenesis. This highlights a potential genetic background of selected individuals to acquire cervical cancer over their life-time, even in the absence of HPV infection.

Similarly, 63 studies identified progression SNPs in 150 genes (Van Tassell et al., 2008). A statistically significant association with cancer progression has been reported in rs284451147 located in the major histocompatibility region of chromosome 6 (OR 5.14; 95%CI 1.89–12.25, p<0.00001). It has been shown that patients with either the GA or GG genotype have an increased risk to progress to cervical cancer upon HPV-infection (McKay et al., 2017). Additionally, a study has shown a significant increase in the risk of occurrence of cervical cancer with the presence of rs187084 minor C allele in *TLR9* (Li et al., 2017) *(*OR 4.33; 95%CI 1.28–13.89, p<0.00001). Finally, individuals with the A>G change in the *CYP1A1* rs1048943 (OR = 7.02; 95% CI = 2.99–16.45, p<0.00001) are at greater risk for HPV infection progression to cervical disease (Wang et al., 2015).

1.3 EPIGENETIC ALTERATIONS IN CERVICAL CANCER

Cancer has long been known as a disease caused by alterations in the genetic blueprint of cells. However, in the past decades, it has become evident that epigenetic processes play a role, at least equally important, in cancer development (Baylin & Jones, 2001). In the 1940s, Conrad Waddington was the first to coin the term 'epigenetics', deriving it from the Greek word 'epigenesis' which means above creation/genetics. The first definition described epigenetics as "*a branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being*", including all molecular events which had the potential of modulating the genotype of phenotype (Lopez et al., 2009). Over time, with advances in research technologies, the definition for epigenetics narrowed to: "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence" (Morris, 2001). Generally, all epigenetic modifications consist of chemical compounds covalently bonded to DNA or histone proteins, which do not disrupt the structure of the DNA, but alter gene expression. Epigenetic modifications such as DNA methylation, hydroxymethylation, demethylation, chromatin remodelling, histone modification, regulation by non-coding RNA and gene imprinting are inheritable and affect genetic information without any mitotic or meiotic interference (Lopez et al., 2009). Among the epigenetic modifications, the role of gene DNA methylation in cervical cancer has been investigated the most.

1.3.1 DNA methylation

DNA methylation is the epigenetic process by which a methyl group is added to DNA, most commonly on the 5' position of a cytosine base (Lorincz et al., 2017). DNA methylation is carried out by a group of proteins called DNA methyltransferases (DNMTs), which catalyse the addition of a methyl group to the 5-position of a cytosine (C) residue, using S-adenosyl methionine (SAM) as a methyl donor, resulting in 5-methylcytosine (Goll & Bestor, 2005). DNMTs can be either *de-novo* (establishing DNA methylation patters during germ cell development; DNMT3A, DNMT3B) or maintenance enzymes (DNMT1) (Lorincz, 2011).

In humans, DNA methylation usually occurs on CpG dinucleotide sequence, where a cytosine nucleotide is followed by a guanine nucleotide in the 5'-3' direction, Around 70–80% of CpGs are methylated by default in somatic cells, however CpG rich sequences (CpG islands) are interspersed in the genome and often un-methylated in germ and somatic cells (Bird et al., 1985; Deaton & Bird, 2011). CpG islands are characterised as regions with an methylated/unmethylated CpG ratio > 0.6 (Strichman-Almashanu et al., 2002). Approximately 70% of gene promoters in the human genome are associated with at least one CpG island

including all housekeeping genes, constitutively active genes typically required for the maintenance of the basic cellular functions (Saxonov, et al., 2006). DNA methylation typically occurs in a tissue and cell type specific pattern. Correct patterns of DNA methylation are crucial for early development and cell differentiation (Mikeska & Craig, 2014).

The alteration of canonical DNA methylation patterns is a hallmark of human cancers and it is typically associated with loss of global genomic DNA methylation (hypomethylation) accompanied by site-specific hypermethylation (Baylin & Jones, 2001; Esteller et al., 2001; Jones et al, 2002) (Figure 1.4). Global DNA hypomethylation may disrupt chromosome condensation and stability (Lorincz et al., 2017), while site-specific DNA hypermethylation occurs on CpG islands present on transcriptional starting sites. CpG islands located on transcriptional starting sites, specifically on gene promoter sites are commonly unmethylated and their hypermethylation is linked to long-term gene silencing (Rincon-Orozco et al., 2009). Aberrant DNA hypermethylation is often associated with the downregulation of tumour suppressor genes (Henken et al., 2007) and cancer development. DNA methylation profiles change with age and so does the risk of cancer. As seen in patients already diagnosed with cancer, aging brings about global DNA hypermethylation, however hypermethylation on specific gene promoters have been associated with cancer development (Esteller et al., 2001).

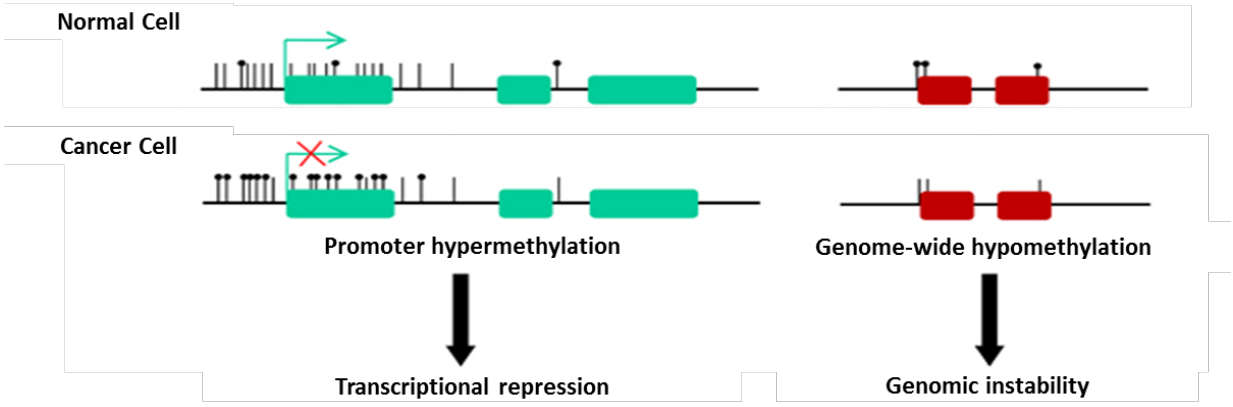

Figure 1.4 - DNA methylation patterns in cancer. Cancer cells show genome-wide hypomethylation accompanied by hypermethylation of specific gene promoters compared to their normal counterparts. Genome-wide hypomethylation is associated with genomic instability, while promoter hypermethylation is associated with transcriptional repression. Each black lollipop represents a methylated cytosine, while the black lines represent unmethylated cytosines. Adapted from Lopez-Serra P *et all*.

1.3.2 HPV-induced host DNA hypermethylation

In this thesis, the focus will be placed on DNA hypermethylation in cervical cancer. DNA hypermethylation occurs when multiple methyl groups are transferred to one cytosine that should not be methylated, thereby resulting in gene silencing. Previous genome-wide methylation profiling studies have found a number of genes with consistently elevated methylation levels in cervical pre-cancers and cancers (Clarke et al., 2017). A recent meta-analysis of 12 TCGA projects identified 1,026 epigenetically silenced genes in cervical cancer cases compared to controls (Cancer Genome Atlas Network, 2017) , indicating that this modification is a genome-wide event. Addition of the methyl group can alter the chromatin conformation and the overall DNA topology inducing transcription factor displacement associated with alterations in gene expression (Plass, et al., 2013).

Following persistent infection, HPV-induced methylation of the host DNA was established to play an important role in viral immune evasion (Bordignon et al., 2017). The HPV oncoproteins E6 and E7 were shown to alter the DNA methylation in host cells by modulating DNMTs activity. HPV-E7 indirectly increases *DNMT1* expression by forming the E7-pRb complex which allows the released transcription factor E2F to activate the promoter of *DNMT1* (Burgers et al., 2012). E7 was also demonstrated to directly bind to DNMT1*,* inducing conformational changes which enhance its methyltransferase activity (Burgers et al., 2012), promoting host DNA hypermethylation.

Cicchini *et al.* identified that the promoter of *CXCL14* is hypermethylated as a consequence of E7 expression in mammalian cell lines (Cicchini et al., 2016). This epigenetic event down-regulates the expression of the chemokine receptor, limiting its interactions with natural killer (NK) cells, T cells and dendritic (DC) cells. This allows HPV-infected cells to evade host immune surveillance and undergo transformation. Evidence also shows E7-induced hypermethylation of the E-cadherin promoter, impacting adhesion efficiency with Langerhans cells (LC) (Shurin et al., 2019). This limits LC movement through affected epithelia, allowing for LC-targeted immune escape of HPV (Schaerli et al., 2005). The E6 viral oncoprotein also plays a role in hijacking the host DNA methylation activity. Knockdown of E6 was associated with a decrease of DNMT1 expression in cervical cancer cell lines (SiHa and CaSki) (Au Yeung et al., 2010; Leonard et al., 2012). Additionally, Rincon-Orozco *et al.* demonstrated the contribution of E6 to the *de novo* methylation of the transcriptional site of interferon (IFN)-k which led to a decrease in gene expression and subsequently, to a decrease in inflammatory responses, which primes carcinogenesis (Rincon-Orozco et al., 2009).

The advances in next generation sequencing (NGS) technologies allowed identification of differentially methylated host genes in patients with cervical disease versus healthy controls. DNA methylation has thus been shown to increase with disease progression (Mikeska & Craig, 2014). The most frequently methylated gene promoters in both SCC and ADC include the cell adhesion molecule (*CADM1),* erythrocyte membrane protein band 4.1 like 3 *(EPB41L3),* Family with sequence similarity 19 (Chemokine (C-C motif-like) member A4 *(FAM19A4),* myelin and lymphocyte *(MAL),* paired box I *(PAX1),* SRY-box I *(SOX1)*, telomerase reverse transcriptase (*TERT*), PR domain containing 14 (*PRDM14*), Junctional Adhesion Molecule 3 (*JAM3*), chromosome 13 open reading frame 18 (*C13ORF18*) and *miR-124* (Cuschieri et al., 2018). These genes were considered promising biomarkers to identify hrHPV-positive women and/or women with abnormal cytology at risk for cervical cancer.

The functional relevance of these markers have been shown through *in vitro* studies, which revealed that the onset of methylation might differ between genes (Henken et al., 2007; Kooter et al., 2015). Additionally, Verlaat *et al.* argued that DNA methylation occurs at the pre-cancer stage and reaches the highest levels after hrHPV-induced carcinogenesis (Verlaat et al., 2018). This might be due to the different susceptibility of specific promoters to DNA methylation by the E6/E7 activated DNMTs (Carvalhopt et al., 2013; Hoque et al., 2011). When analysed in cervical exfoliated cells covering an array of histology types (no disease, CIN2, CIN3 and cancer), all genes analysed except *C13ORF18*, showed a gradual increase in methylation with disease progression. However, in all cases there was a constant increase in methylation levels with disease progression and age (Verlaat et al., 2018).

1.3.3 Viral DNA hypermethylation

A growing number of studies support the role of HPV DNA hypermethylation in cervical cancer development. Compared to the hypermethylation of host gene promoters, which is usually associated with repression of gene transcription, viral DNA hypermethylation may regulate viral gene transcription on multiple levels. DNA hypermethylation of the viral upstream regulatory region (URR) was shown to inhibit E2 binding, therefore inducing E6 and E7 overexpression, the driving force of a transforming HPV infection (Mirabello et al., 2017). The viral late genes L1 and L2 display hypermethylation patterns in persistent infections especially with the high-risk types HPV16, HPV18, HPV31, HPV33 and HPV45 (Clarke et al., 2012). Although there is no conclusive evidence as of why methylation of the viral late gene occurs, targeted methylation within HPV genes might promote a switch between the productive and transformative phases or may serve as a host defence mechanism for silencing viral replication (Vinokurova & Doeberitz, 2011).

Nonetheless, abundant evidence exists to accurately distinguish CIN2+ from transient HPV infections through the viral late gene methylation status (Kalantari et al. 2010; Mirabello et al., 2012; Wentzensen et al., 2012).

1.4 THE CERVICOVAGINAL MICROBIOTA

It is estimated that 15-20% cancer cases are associated with infectious agents including viruses, bacteria and parasites (IARC, 2007). Human oncogenic viruses include hepatitis B and hepatitis C viruses (associated with hepatocellular carcinoma), Epstein-Barr virus (associated with B-cell lymphomas, nasopharyngeal and gastric carcinomas), human T cell lymphotropic virus I (associated with T cell lymphomas), human herpes virus type 8 (associated with Kaposi's sarcoma and primary effusion lymphomas) and HPVs (associated with cervical carcinomas, ano-genital cancers and a subset of head and neck cancers) (Cantalupo et al., 2018; Reshmi et al., 2011). Recent metagenomics sequencing studies detected significant differences in the composition of microbial communities in numerous human cancer cases compared to controls (Bhatt et al., 2018). However, compared to oncogenic viruses which express proto-oncogenes which directly disrupt host signalling pathways (Aadra & Damania, 2013), tumourgenesis associated with microbial dysbiosis is less well understood (Curty et al., 2020)

The human body is an ecosystem with complex interaction network between host cells and the microbiota (Kyrgiou et al., 2018). The collective genome of the microbiota is known as the human microbiome and it is ~100 fold more diverse than the human genome (Savage et al., 1977). Despite the difficulties in culturing many bacteria species (Curty et al., 2020), the development of high-throughput metagenomic sequencing has allowed a deep understanding of the microbiome composition (Virtanen et al., 2017). Metagenomic sequencing is focused only on one gene, the 16S ribosomal RNA gene (16S rRNA) (Bukin et al., 2019), coding for the 30S small component of the prokaryotic ribosome. The 16S rRNA gene is ~1.5kb long and contains nine hypervariable regions (V1 to V9) intercalated by conserved regions, allowing PCR amplification of a variety of bacterial species using universal primers (Bukin et al., 2019; Pinna et al., 2019; Virtanen et al., 2017). The Human Microbiome Project started in 2008 and was the first effort to characterize the large diversity of the microbiota, contributing to the understanding of the microbiome composition in different parts of the body, including the cervicovaginal microbiome and its role in disease development (Curty et al., 2020). Notably, the microbiota has been shown to influence cancer susceptibility, and a large number of cancers have been associated with an altered composition of the microbiota (dysbiosis) (Curty et al.,

2020). This may be partly due to the intricate metabolic capacity exerted by bacteria and the influences on the host immune functions (Srinivasan et al., 2015).

Whether changes in the microbiota promotes the release of carcinogenic agents leading to neoplasia or if the tumour microenvironment modulates its surrounding microbial community still remains unclear. Certain bacteria produce carcinogenic agents that are able to modulate the immune system or modify the physiology of host cells (Srinivasan et al., 2015). For instance, secretion of γ-glutamyl transpeptidase by *H. pylori* is known to abrogate T lymphocyte proliferation and induce cell cycle arrest (Arabski et al., 2005; Schmees et al., 2007). This event is strongly linked to chronic inflammation and gastric adenocarcinoma development. Additionally, colibactin secreted by specific *E. coli* strains was shown to cause DNA double-strand breaks and induce a surge in the production of growth factors, stimulating cell proliferation in cultured intestinal epithelial cell lines (Bossuet-Greif et al., 2018; Secher et al., 2013).

1.4.1 The cervicovaginal microbiota: HPV infection acquisition and persistence

Ravel *et al.* were the first to classify the cervicovaginal microbiota in five community state types (CST I-V), depending on the dominant bacteria species present (Ravel et al., 2010). Generally, the majority of healthy women have stable cervicovaginal microbiota, with relatively low diversity in comparison to other mucosal sites and it is those with a highly diverse microbiota in whom transitions from one CST to another are observed more often (Curty et al., 2020). CSTs I, II, III and V show low diversity in microbiota, with the dominant species being part of the *Lactobacillus* genus: *L. crispatus, L. gasseri, L. iners* and *L. jensenii*, respectively (Ravel et al., 2010; Seta et al., 2019). This genus is characterised by gram-positive, microaerophilic bacteria which produce bacteriocins, L-lactic acid and hydrogen peroxide (H_2O_2) known to prevent colonization of exogenous pathogens (Curty et al., 2020; Oh et al., 2015). However, *L. iners* is considered a transitional species to the dysbiosis state as it produces the D- isoform of lactic acid, associated with susceptibility of viral infections (Amabebe & Anumba, 2018). Conversely, CST IV exhibits higher bacterial diversity with a low abundance of *Lactobacillus* and an increased frequency of anaerobic bacteria from genuses like *Gardnerella, Megashera, Atopobium, Sneathia* and *Prevotella* (Amabebe & Anumba, 2018; Curty et al., 2020). The CST IV microbiota is characterised as dysbiosis and increases susceptibility to HPV infections (Curty et al., 2020) (Figure 1.5).

Evidence supporting a link between HPV persistence and the composition of the cervicovaginal microbiota is emerging. One of the first studies employing 16S rRNA sequencing, a Korean study on twins which included 68 women (23 hrHPV positive, 45 hrHPV negative), found that HPV positive women present higher bacterial diversity with lower proportions of *Lactobacillus spp.* compared to HPV negative women (Lee et al., 2013). Additionally, *Sneathia spp.* was proposed as microbial marker for HPV infection (Lee et al., 2013). Gao *et al.* confirmed the Korean findings and also identified G*. vaginallis* as significantly more abundant in HPV infected women (Gao, et al., 2013). However, both studies are cross-sectional and limited in defining whether the HPV infection alters the cervicovaginal microbiota or that the microbiota influences the persistence of the virus (Kyrgiou et al., 2018). To address this issue, Brotman *et al*. looked at the temporal relationship over 16 weeks between the cervicovaginal microbiota and HPV infection in 32 women who self-sampled twice a week (Brotman et al., 2014). The *L. gasseri*-dominant CST II was associated with the fastest rate of HPV regression while CST IV was associated with the lowest (OR 4.43, 95%CI 1.22-17.70 and OR 0.33, 95%CI 0.12-1.19, respectively) (Brotman et al., 2014). Arokiyaraj *et al*. supports these observations through a longitudinal study following 41 women over 7 years: HPV positive women were observed to have a higher microbiota diversity than HPV negative women, with CST I being the most predominant state in HPV negative women (OR: 8.25, 95%CI 2.13-32.0) (Arokiyaraj et al., 2018). Further, strong associations were observed between *L. johnsonii* and HPV persistence (OR: 16.4, 95%CI 1.77-152.2) (Arokiyaraj et al., 2018).

1.4.2 The cervicovaginal microbiota: cervical cancer progression

Although the majority of HPV infections are transient, viral persistence occurs in almost 10% of infected women, increasing their risk of developing cervical lesions (Gravitt & Winer, 2017; Snijders et al., 2006; Steenbergen et al., 2014). There is still a paucity in our understanding of the exact combinations of factors responsible for persistence as well as those that promote and initiate the carcinogenesis process (Kyrgiou et al., 2018; Steenbergen et al., 2014). However, factors such as immunodeficiency, age, smoking and the use of oral contraceptives have been associated with higher HPV persistence rates (Chen et al., 2014; Mikeska & Craig, 2014). As the cervicovaginal microbiota is the first line of contact against HPV infections, recent studies have highlighted its role in viral acquisition, persistence and the risk of cervical precancer and cancer development (Kyrgiou et al., 2018).

A recent study by Mitra *et al.* was the first to describe the cervicovaginal microbiota in women with cervical disease. The study included 169 women (20 healthy, 52 LSIL, 92 HSIL and 5 invasive cervical cancer) (Mitra et al., 2015a). The rates of high bacterial diversity combined with a low abundance of *Lactobacillus spp*. (CST IV) were two to four times more prevalent in women diagnosed with cervical pre-cancer or cancer, compared to healthy women ($p<0.01$). A negative correlation was observed between the *L. crispatus*-dominant CST I prevalence and disease severity (Mitra et al., 2015a) indicating a possible protective effect against the development of cervical precancerous and cancerous lesions. Additionally, *S. sanguinegens*, *A. tetradius*, *P. anaerobius* and lower levels *L. jensenii* were identified as unique differences between healthy and CIN2+ microbiotas however no bacterial signatures were identified in women with cancer progression (Mitra et al., 2015a). The limitations of the study were its cross-sectional identity and the lack of adjusted risk factors.

Subsequently, more evidence supporting a link between cervicovaginal microbiota composition and cervical disease development has emerged: Oh *et al.* argued that the risk of developing CIN2+ was six-fold higher when a pattern of high abundancies of *A. vaginae, G. vaginalis*, *L. iners* and low *L. crispatus* were present (OR: 5.80, 95%CI 1.73–19.4) (Oh et al., 2015). Additionally, Piyathilake *et al.* argued that risk of developing cervical disease was higher in community types dominated by *L. iners.* compared to those with other diverse *Lactobacillus spp.*, *Bifidobacteriaceae, Clostridiales* and *Allobaculum* (OR: 3.48, 95%CI 1.27-9.55) (Piyathilake et al., 2016). This observation is interesting as *L. iners* dominant microbiota is more likely to transition to CST IV which Mitra *et al.* associated with CIN2+ (Mitra et al., 2015a). Moreover, *L. iners* was associated to the depletion of *L. crispatus* and *L. jensenii* which were both correlated with reduced glutathione levels (Srinivasan et al., 2015). Glutathione is an antioxidant with a protective role against the damage caused by reactive oxygen species, therefore the reduced levels observed suggest that *L. iners* colonisation may promote high oxidative stress. Oxidative stress has been linked to carcinogenesis in microbial-induced gastric adenocarcinoma (Arabski et al., 2005). In the context of cervical cancer development, generation of reactive oxygen species facilitate HPV integration into the host genome(The et al., 2020) and neoplastic transformation, mediated by the HPV E6 oncoprotein (Williams et al., 2014). Viral integration leads to the loss of E1 and E2 genes which control E6 and E7 transcription (Moody & Laimins, 2010). In consequence, oncoprotein transcription is enhanced after viral integration leading to uncontrolled cellular proliferation and decreased apoptosis.

Although there is a possible association between cervical pre-cancer, persistent HPV infection and the composition of the cervicovaginal microbiota, studies so far do not prove causality (Brotman et al., 2010; Gillet et al., 2011). The published studies are limited in interpretation since samples investigated originate mostly from vagina and not from the cervix only(Ling et al., 2011). Although the cervical microbiota is similar to one of the vagina, lower bacterial loads are observed ($p<0.05$) (Ling et al., 2011; The et al., 2020). Additionally, the microbiota collected from swabs is different than that from biopsy samples as swabs might not get to biofilms which are strongly adherent to cells (Patterson et al., 2010). Lastly, most data collected so far comes from cross-sectional studies (Kyrgiou et al., 2018) and do not explore the full longitudinal relationships between microbiota and other factors.

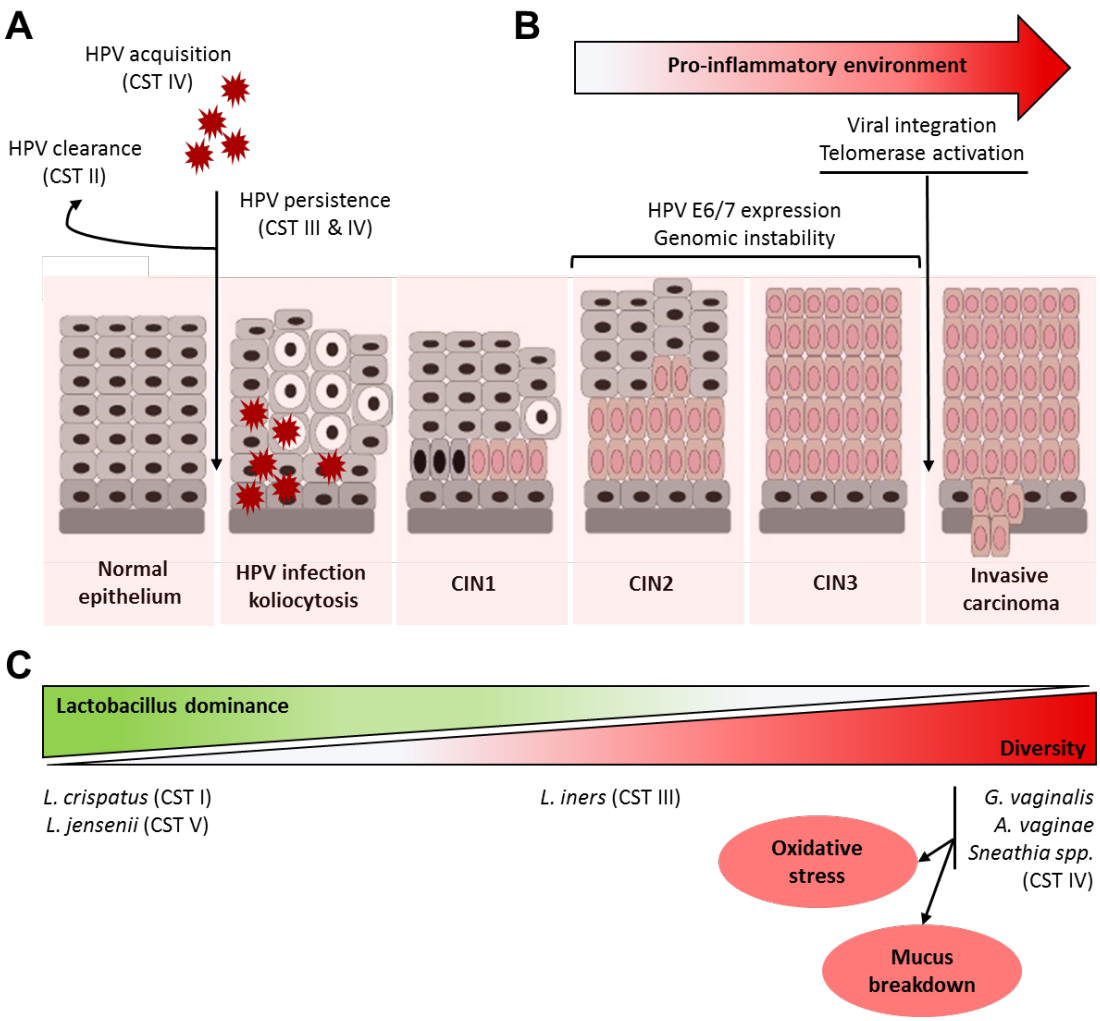

Figure 1.5 – Summary of cervicovaginal microbiota changes related to HPV persistence and cervical cancer progression. A: cervicovaginal microbiota associations with HPV acquisition (CST IV), HPV persistence (CST III and IV) and HPV clearance (CST II). **B:** dysbiosis can be a result of an increased pro-inflammatory environment, which can sustain a number of steps required for viral transformation: increased E6 and E7 expression, genomic instability, viral integration and telomerase activation, which contribute to cervical carcinogenesis. **C:** a lower abundance of *Lactobacillus spp*. combined with an overall higher diversity of other species was associated with disease severity and high grade CIN. The cervicovaginal microbiota of healthy women was characterised by a high abundance in *Lactobacillus spp.*. Particularly, dominance of *L. crispartus* (CST I) and *L. jensenii* (CST V) was associated with a reduced risk of oxidative damage and disease progression. *L. iners* (CST III) does not share many of the protective mechanisms displayed by other *Lactobacillus spp*. and was associated with a transitional phase to dysbiosis. With higher grades in CIN, a dysbiosis state was associated with a cervicovaginal microbiota dominated by a higher diversity in bacteria species, including *G. vaginalis, A. vaginae* or *Sneathia spp*.. Chemicals produced by the species mentioned promote higher levels of oxidative stress and mucus breakdown which have been liked to cervical cancer progression. Abbreviations: HPV: human papillomavirus; CIN 1/2/3: cervical intraepithelial neoplasia grade 1/2/3; CST I: community state type 1; CST II: community state type 2; CST III: community state type 3; CST IV: community state type 4; CST V: community state type 5 (adapted from Mitra et al., 2015b).

1.5 DISEASE CONTROL

1.5.1 Implementation of HPV vaccination

Understanding the aetiological role of HPV has improved the landscape of cervical cancer prevention in multiple ways. Although it is known that multiple genetic, epigenetic and microbial factors may be additionally required to drive carcinogenesis, almost all cervical cancers are still attributable to HPV infection and this allowed the development of three vaccines for cervical cancer primary prevention. Although, HPV immunisation has been introduced from 2008 in most countries, more than 80% of high-income countries (i.e. United Kingdom) have implemented national HPV vaccination programmes compared to less than 30% LMICs (Brisson et al., 2020). Up to date, there are three commercially available HPV vaccines. The bivalent (Ceravix) and quadrivalent (Gardasil) vaccines were developed to protect against initial high-risk HPV16/18 infections, strains responsible for ~70% cervical cancer cases (Drolet et al., 2015). Gardasil also covers the low risk types HPV6 and HPV11. The nonavalent vaccine (Gardasil-9) protects against five other hrHPVs (31, 33, 45, 52, and 58). It is estimated that HPV vaccines may prevent up to 82.5% (bivalent and tetravalent) and up to 95.3% (nonavalent) of HPV-positive cervical cancers (Pirog et al., 2014).

While most vaccination programmes worldwide still use either the bi- or quadrivalent vaccines, the nonavalent vaccine was implemented in national vaccination programmes in the US in 2017 and Australia in 2018. This may contribute to a significant reduction in cervical cancer rates, and allowed Australia to be on the track of becoming the first country to eliminate cervical cancer in the near future (Drolet et al., 2019; Lei et al., 2020). The benefit of introducing nonavalent vaccine as the main vaccine in the United Kingdom vaccination programmes would contribute to a reduction of 36.0% in cervical cancer rates by 2036-40 (Castanon et al., 2018). Additionally, the nonavalent vaccine was licensed in LMICs starting in 2018 and it is not yet routinely used in vaccination programmes due to the higher costs associated (Sankaranarayanan et al., 2019). Further, most HPV vaccination schemes initially targeted girls aged 12 to 13. Developed countries subsequently extended their policies to also include boys of the same ages. In the United Kingdom vaccination for boys started in September 2019 (Cuschieri et al., 2019).

The vaccination strategy does not have any direct effect on women born before 1991, as a significant proportion of the population had not been immunised against HPV prior to becoming sexually active. This means that these women will still require cervical screening until the age of 64 years. With the youngest women in their late 20s, this means cervical screening will remain

a dominant part of cervical cancer prevention for at least 30 years, with the burden of cervical cancer moving towards older age groups. Additionally, population projections in the United Kingdom estimate a substantial increase in the number of women older than 60 years: a 29.0% increase (from 9.4 million in 2012 to 12.1 million in 2037) in the number of women aged 60-74 and a 90.0% increase in those older than 75 years (from 5 million to 9.5 million). Over the next 30 years, the cancer rates in women aged 25-34 are estimated to decrease by 50.0% (Castanon et al., 2018). A recent modelling study predicted that the peak age of cancer diagnosis may shift from 25-29 years in 2011-2015 to two peak structure with first peak at 55-59 years and a second at 75-59 year in 2036-40 ("Office for National Statistics National population projections: 2018-based statistical bulletin," 2019).

Even though the vaccines are potent against initial viral infection, they do not prevent the progression into cervical cancer of already infected individuals. Until HPV vaccination is implemented worldwide, we are facing a need for better screening programmes and diagnosis tools, mostly related to the requirement for effective triage of HPV infected individuals (Liu et al., 2018).

1.5.2 Cervical cancer screening: implementation of hrHPV as the primary screening test

WHO has recently published a report stating that a focus on secondary prevention in women of 30 years or above is the most efficient and cost-effective way to prevent and eventually eliminate cervical cancer (Bray et al., 2018). Secondary prevention focuses on screening and treatment of pre-cancerous lesions. Due to socio-economic factors, access and quality of cervical cancer screening varies greatly among countries (Castanon et al., 2018). Learning from developed countries and adjusting strategies for LMIC might offer a strategy to successfully contain the cervical cancer burden.

There are five main metrics by which the accuracy of a screening test is measured. The area under the curve (AUC) is the measure of how well a test or combination of tests can distinguish between two diagnostic groups (diseased/normal); the higher the AUC, the higher the diagnostic potential of the test. The sensitivity or the rate of true positives measures the proportion of actual positives that are correctly identified as such by the assay. The specificity or the rate of true negatives measures the proportion of actual negatives that are correctly identified as such by the assay. The Positive Predictive value (PPV) measures the probability that individuals with a positive screening test truly have the disease; PPV can be described as the diagnostic accuracy. Lastly, the Negative Predictive value (NPV) measures the probability that individuals with a

negative screening test do not have the disease. The ideal screening algorithm is aiming for AUC>0.9, sensitivity of 85% and specificity of 98% (Castle et al., 2018; Sasieni et al., 2019). Although further research is required to achieve such metrics, Castle *et al.* argues that these can be achieved only through combination of screening tests, with HPV testing as the primary screening (Castle et al., 2018).

The decrease in HPV prevalence associated with HPV vaccination programs favors the worldwide switch to HPV testing as the primary screening method with cytology triage for hrHPV positive women. Turkey was the first to introduce this strategy in 2014, followed by the Netherlands in 2016 and Australia in 2017 (Gultekin et al., 2017). In 2019 the United Kingdom Cervical Screening Programme shifted to HPV16/18 genotyping in conjunction with cytology triage (Cuschieri et al., 2018) (Figure 1.7). As far as LMICs are concerned, Mezei *et al.* argues that HPV testing in combination with cytology triaging might not be as effective due to lack of complex infrastructure and finances. HPV testing combined with acetic acid triage was proposed as a better alternative for screening in LMICs (Mezei et al., 2017).

In the current United Kingdom screening algorithm women are invited for HPV testing every three years if aged 25-49 or five years if aged >50, until 64. This involves HPV testing with cytology triage. For hrHPV negative results, women are invited to screen as usual. For hrHPV positive results, women are invited for cytology triage. If the cytology results are abnormal, they are referred to colposcopy. Women with normal cytology are rescreened in 12 months. In the case of persistent hrHPV positivity over three consecutive screens and normal cytology results, women are referred to colposcopy. Any patient with colposcopy results of CIN2+ are referred for loop excision treatment.

HPV testing is characterized by high clinical sensitivity, a high NPV, and low training requirements, allowing this test to be sufficiently accurate and more cost-effective than cytology-based screening (Castanon et al., 2018; Catarino et al., 2015; Tidy et al., 2016). The United Kingdom Cervical Screening Programme, has been responsible for a decrease of 44% in cervical cancer incidence from 1988 to 2002, but since 2002 this has plateaued (Pesola et al., 2019). Castanon *et al.* have estimated that along with vaccination, introduction of HPV testing as the primary method of screening could tackle this plateau problem and reduce age-standardised rates of cervical cancer at ages 25–64 years by 19% by 2028-32 (Castanon et al., 2018). These findings are supported by a recent HPV-screening pilot in the UK. In a non-randomised cohort of over half of million women, Rebolj *et al.* have noticed that routine HPV screening with follow-ups at 12/24 months increases the detection of CIN2+ and cervical cancer

by approximately 40% and 30%, respectively, compared to cytology-based screening alone (Rebolj et al., 2019).

HPV-testing has been proven to be more sensitive but less specific than cytology-based testing alone (Lorincz, 2016). The increased sensitivity of HPV testing translated into two important benefits: earlier detection of CIN lesions that if treated, results in a reduced incidence of cervical cancer within 4 to 5 years and greater reassurance against after a negative result for many years which allows screening to be extended to a 5-10 years interval (Panse et al., 2009; Ronco et al., 2014; Sankaranarayanan et al., 2009). Even though HPV detection tests used for cervical cancer screening may be positive in presence of any of the 13 or 14 high risk HPV types, they cannot distinguish transient from persistent infections (Smith et al., 2007). The majority of HPV infections become undetectable within a few years and such transient infections are not associated with progression. Gravitt *et al.* argues that due to the non-linear life cycle of the HPV infection, some women might show HPV-latency (Gravitt et al., 2011). This means that the infection is not cleared by the immune system. Although present, such latent infection is not continuously detected via HPV-testing. Kostrikis *et al.* supports the fact that HPV-testing leads to a large number of clinically insignificant positives, resulting in more false referrals for colposcopy and biopsy, decreased correlation with the histological presence of disease and unnecessary treatment of healthy women (Kostrikis et al., 2018).

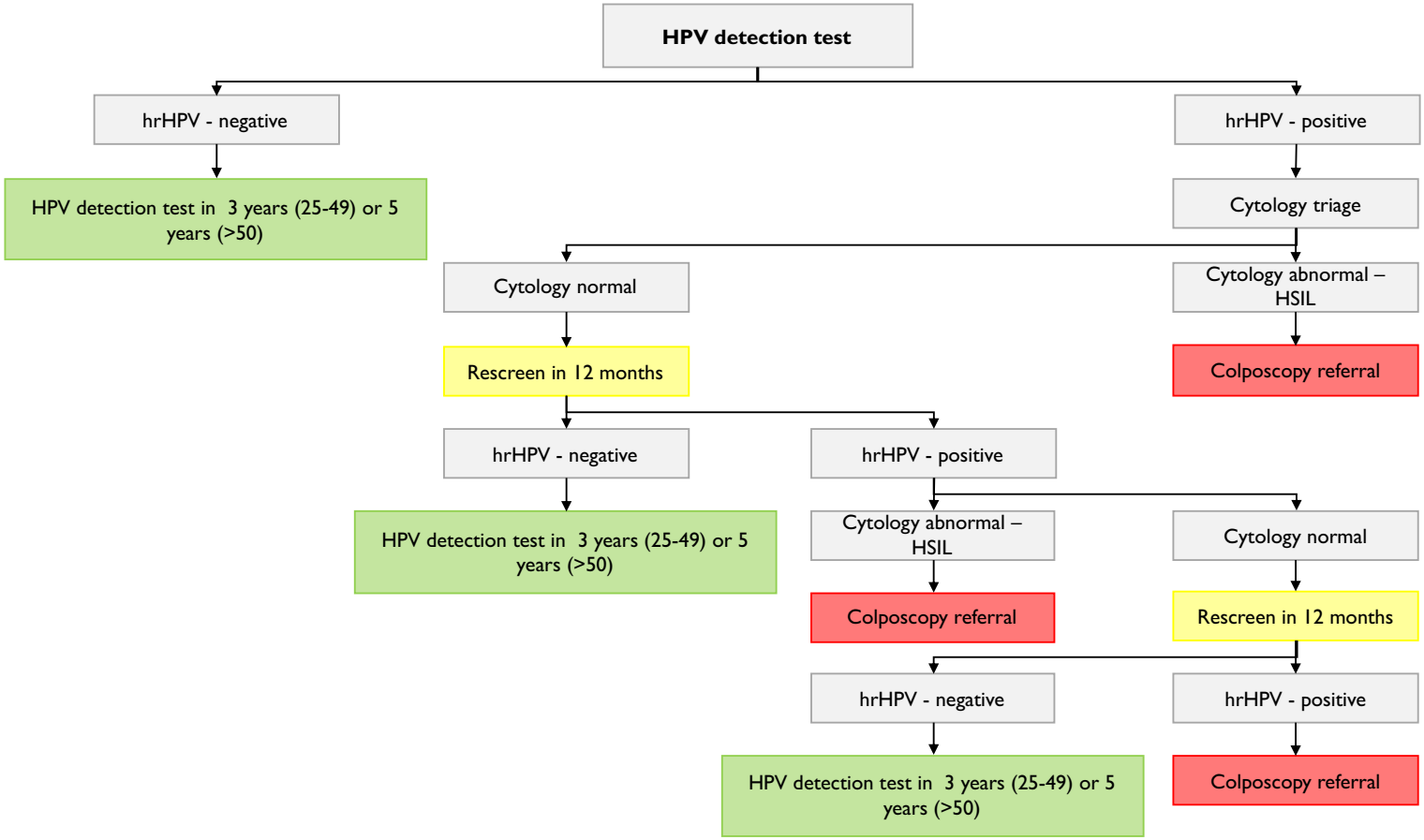

Figure 1.7 – HPV test-based screening flowchart. Since 2019 cervical screening in the United Kingdom shifted towards HPV detection test in conjunction with cytology triaging. Women are invited for screening every 3 years if aged 25-49 or 5 years if aged >50. For hrHPV negative results, women are invited to screen as usual. For hrHPV positive results, women are invited for cytology screening. If the cytology results are abnormal (HSIL), they are referred to colposcopy. Women with normal cytology are rescreened in 12 months for. In the case of persistent hrHPV positivity over 3 consecutive screens and normal cytology results, women are referred to colposcopy. Any patient with colposcopy results of CIN2+ are referred for surgical treatment. Abbreviations: hrHPV: high risk human papillomavirus; CIN2+: cervical intraepithelial neoplasia grade 2 or above.

1.5.3 Triaging hrHPV positive women

Cytology testing (Papanicolau test) is historically the most widely used method for cervical cancer screening and currently used as the triaging method for women testing positive after primary HPV screening. Sample collection for cytology screening involves the use of a small plastic brush to collect cervical cells which are stored in a liquid transport medium. This collection strategy is known as liquid-based cytology (LBC) sampling and allows the preparation glass slides for microscopical investigation (Zielinski et al., 2001).

Women who are hrHPV-positive are generally triaged by cytology in developed countries. The abnormalities caused by an acute hrHPV infection, are classified as LSIL, HSIL or borderline results such as atypical squamous cells of undetermined significance (ASCUS) (Clifford et al., 2005; Smith et al., 2007). Notably, all hrHPV-positive women with abnormal cytology results are referred for histology testing (colposcopy) to receive the appropriate CIN diagnostics (Gravitt et al., 2011). However, women can be misdiagnosed at this stage if the histology sample does not include the diseased tissue (Gravitt et al., 2011). Moreover, over-diagnosis can be a consequence interpretational mistakes due to hrHPV-positivity bias (Lorincz et al., 2016).

All CIN2+ women are treated by removal of the abnormal tissue. CIN2 lesions might also regress, but this is difficult to know in advance, so surgical treatment is the default option. Loop Electrosurgical Excision Procedure (LEEP) is the most commonly used treatment for CIN2+ and involves the use of a small electrical wire loop to remove abnormal cells from the cervix (Wiik et al., 2019). Cytology-based triaging has a significant false positive rate (15-50%) for CIN2+ and cannot be automated (Sahasrabuddhe et al., 2013; Yim & Park, 2006). Consequently, some women are likely to be over-treated, young women, especially being at great risk of having their fertility impaired due to LEEP (Vink et al., 2015).This process requires skilled workforce and constant financial investment to assure systematic screening, being mostly available to countries with the relevant infrastructure (Cuschieri et al., 2018). Despite efforts to improve the performance of cytology testing, low sensitivity was a problem.

Among others, Catarino *et al.* argues that although this systematic screening process has potential of keeping the disease under control it involves a great deal of financial investment in workforce training and health-care infrastructure and it might not be the best solution for LMIC where cervical cancer incidence is actually the highest (Catarino, 2015).

While cytology testing was historically implemented in systemic screening programmes in developed countries like the United Kingdom, LMICs lack complex infrastructure for systematic

screening, have limited quality assurance as well as a paucity in education resources for practitioners. For instance estimated CIN2+ sensitivity for cytology in LMICs showed an average of 67.3% with large variations from 33.0% to 100.0% due to differences in infrastructure and interpretation of results by practitioners (Mezei et al., 2017; Castanon et al., 2017). In consequence, visual inspection with acetic acid (VIA) screening is currently recommended by the WHO as a cervical cancer screening strategy when more accurate approaches are not available(WHO, 2013). VIA involves a vaginal speculum exam during which a health care practitioner applies dilute (3- 5%) acetic acid to the cervix (Catarino et al., 2015). Abnormal tissue temporarily appears white when exposed to the acetic acid. VIA results are instant but also depend on the visual interpretation of the practitioners performing the test. Although VIA shows a lower CIN2+ sensitivity compared to cytology (60.4% vs 67.3%), the cost per sample for VIA triage is more than 2 times lower than the one for cytology, making it more cost-effective for screening in LMICs. However, VIA shows moderate specificity, leading to higher numbers of colposcopy referrals (Muwonge et al., 2013). These increases the work-load and costs to public health services, which might come as an issue especially in LMIC.

1.5.4 Alternative triaging opportunities

Although current screening tools have contributed to a sizeable reduction in the rates of CIN2+ diagnosed, there are still improvements to be made. It is still important to avoid miss-diagnosis of hrHPV-positive women and to establish whether women who cleared a past viral infection have any risk of acquiring new HPV infection and progressing to CIN. Although there is evidence supporting that HPV testing is more sensitive for the detection of CIN2+, the specificity is low. Triage to colposcopy generally relies on cytology, which might not have the required sensitivity to rule out HPV positive women without evidence of disease (Cuschieri et al., 2018). High numbers of colposcopy referrals increase the work-load and costs to public health services, which might come as an issue especially in LMICs, where most cervical cancer deaths occur (Allen-Leigh et al., 2017). The p16/Ki67 immunostaining test is more sensitive than cytology and identifies women with high risk of CIN2+, but interpretation is subjective and the test is expensive and difficult to perform without adequate training (Kisser & Zechmeister-Koss, 2015; Luttmer et al., 2016). The challenges of triaging to avoid unnecessary colposcopy referrals without high costs can be overcome via molecular testing.

To date, molecular biomarkers, have been investigated for molecular triage purposes, they offer an accurate and cost-effective alternative by identifying the cases with the highest risk of progressing into invasive cancer (Lorincz et al., 2011). Compared to the current morphologically-

based triage methodologies, molecular testing offers the advantage of potential automatization and is less prone to training and interpretational mistakes (Lorincz et al., 2016). Molecular testing can be performed on the same clinician-collected or self-collected specimens used for HPV-testing, without having women go to additional clinician visits (as in the case of cytology or VIA triaging) (Boers et al., 2014; Snoek et al., 2018). Although current technologies may not yet be affordable in LMIC, rapid technological advances may allow molecular screening for cervical cancer to be implemented in those setting in the not too distant future.

Specifically, methylation biomarkers, have been investigated for molecular triage purposes. They offer an accurate and cost-effective alternative by identifying the cases with the highest risk of progressing into invasive cancer (Lorincz et al., 2011). A recent meta-analysis described the performance of DNA methylation assays in early detection of CIN2+ and CIN3+ (Kelly et al., 2019). This included data from 43 studies and covered DNA methylation of the host *CADM1, MAL, mir-124-2*, *FAM19A4, POU4F3, EPB41L3, PAX1, SOX1* and the viral HPV16 L1/L2. In this study, all the genes were pooled and a set specificity of 70.0% was required for improving the hrHPV positive women triage to colposcopy, the reported sensitivity was 68.6% for CIN2+ and 71.1% for CIN3+ detection (Kelly et al., 2019). Although the sensitivity was lower than cytology, specificity was higher (54.8% for CIN2+ and 51.8% for CIN3+ detection) (Kelly et al., 2019). A high specificity is of particular advantage in low-, medium-resource areas where there is high prevalence of hrHPV (i.e. sub-Saharan Africa or China).

So far, only a limited number of host cell methylation biomarkers have been tested for their triage potential (Cuschieri et al., 2018). Unlike HPV genotyping, methylation biomarkers have the potential to distinguish between persistent and transient HPV infections (Clarke et al., 2012). Although most biomarkers show good sensitivities in detecting CIN2 lesions, the specificity is still moderate (Table 1.1) For instance, the dual-biomarker panel composed of the host genes *MAL* and *CADM1* has been validated in a population-based screening study. The multiplexing test can differentiate CIN2 women within a HPV-positive cohort of women with a sensitivity of 84%, specificity of 52% (Bierkens et al., 2013; Snellenberg et al., 2012). Additionally, the AUC was 0.72. In the same cohort analysed, the sensitivity and specificity values for cytology testing were 66% and 79% (Cuschieri et al., 2018).

Genes	Sensitivity %	Specificity%	PPV%	AUC
CADM1, MAL (Overmeer et al., 2011)	82	54	25	0.72
***CADM1, MAL*, miR-124** (Snellenberg et al., 2012)	89	50	52	0.8
***EPB41L3* & HPV16/18/31/33** (Lorincz et al., 2016)	90	49	51	0.84
FAM19A4 (Luttmer et al., 2016)	70	66	55	NA
JAM3, TERT, EPB41L3 (Boers et al., 2014)	55	90	NA	NA
***ASTN1, DLX1, ITGA4, RXFP3, SOX17,* and *ZNF6* (GynTect®)** (Clarke et al., 2017)	59	98	74	NA

Table 1.1. – Methylation Biomarker tests: Component genes and performance characteristics for CIN2 endpoint identification. Abbreviations: AUC: area under the curve; NA: not available; PPV: positive predictive value.

1.5.5 The S5 DNA methylation classifier

Aberrant DNA methylation, has been reported to increase with cervical disease progression (Lorincz et al., 2016), allowing this epigenetic event to be used as a temporal biomarker, with the potential to predict whether HPV infection will lead to CIN2+ lesion or if the infection will resolve (Kelly et al., 2019). Combining the knowledge of methylation on host and viral genes, Lorincz *et al.* developed the S5 DNA methylation classifier. In brief, the S5 classifier involves testing the levels of DNA methylation on CpGs from the host *EPB41L3* and viral genes: HPV16-L1, HPV16-L2, HPV18-L2, HPV31-L1 and HPV33-L2. Table 1.1 shows a history of the S5 DNA methylation classifier development.

Historically, it was the association between hypermethylation of viral HPV genes and CIN or cervical cancer that primed the development of methylation biomarkers for diagnostic and triage purposes (Lorincz et al., 2013; Mirabello et al., 2012). Although HPV16/18 infections account for more than 70% CIN and cancer cases, HPV distribution worldwide is not uniform (Clifford et al., 2005). This might be due to the complex geographical and biological interplay between HPV types and host immunogenic variations (McKay et al., 2017). Apart from HPV16/18, HPV types 31, 33 and 45 infections have been found to be common infections in women from Europe or North America (Smith et al., 2007). HPV types 35, 52, 56, and 58 are more common in HPV-positive women in sub-Saharan Africa than in Europe (all $p<0.001$). Additionally, in South America, HPV types 51 and 58 are fairly common in HPV-positive women. However, HPV-type distribution in Asia is heterogeneous, potentially due to the broad geographical and cultural range in Asian populations (Smith et al., 2007). For instance, infections

with HPV58 are common in Taiwan, but not India (Sir, 1997). Due to its development in the United Kingdom, the S5 DNA methylation classifier was built to test for the HPV types most common in Europe. Initially, HPV types 16, 18, and 31 were used for cervical lesion classification. Subsequently, HPV33 was added to the test for improved specificity (Clarke et al., 2012; Lorincz et al., 2013).

Although the viral components performed well in CIN prediction, host genes were investigated in the classifier's composition with the aim of improving specificity. Vasilievic *et al.* showed that the promoter of *EPB41L3 was* differentially methylated in a cross-sectional case-control study, thus being selected as the host component of S5 (Vasilievic et al., 2014; Brentnall et al., 2014.). The *EPB41L3* gene codes for the membrane Band 4.1-like protein 3 which acts as a tumour suppressor inhibiting cell proliferation while promoting apoptosis (Moody & Laimins, 2010; Schmees et al., 2007; Zeng et al., 2018). The product of *EPB41L3* was found to modulate protein arginine N-methyltransferases (PRMTs), involved in the methylation of arginine residues on histone tails, thus playing an indirect role in epigenetic regulation of gene transcription (Jiang et al., 2005; Singh et al., 2004). Hypermethylation of CpG islands on the *EPB41L3* promoter leads to a decrease in gene expression, which was associated with the progression of multiple cancers including cervical (Clarke et al., 2017; Lorincz et al., 2016), lung (Tran et al., 1999) and esophageal cancer (Zeng et al., 2018).

The S5 DNA methylation classifier was initially validated in a British cohort of randomly selected 710 women attending routine screening. In the cohort, 38 women were diagnosed with CIN2+ within a year after triage to colposcopy based on cytology and 341 were hrHPV positive (Lorincz et al., 2016). The results showed an accurate separation of women with CIN2+ and cancer from those with CIN1 or negative cytology (Lorincz et al., 2016). The threshold methylation score i.e. cut-off value for CIN2+ detection depends on HPV load and screening availability of a country. Within this validation cohort and considering the systematic screening in the United Kingdom, the predefined threshold for the S5 classifier was set at 0.80. Any woman scoring above 0.80 has a high risk of developing cervical disease.

The S5 DNA methylation classifier showed a good diagnostics potential with an area under the curve (AUC) of 0.84, 74% sensitivity, 65% specificity and 51% positive predictive value for CIN2+ diagnosis (Lorincz et al., 2016). Additionally, the S5 classifier showed a significantly improved performance, compared to HPV16/18 genotyping on the same samples ($p<0.0001$) (Lorincz et al., 2016; Lorincz et al., 2016). A meta-analysis on the performance of DNA methylation assays indicated that S5 had a higher sensitivity for CIN2+ detection than considering *EPB41L3*

methylation alone, without compromising specificity (Kelly et al., 2019). This indicates that the combination of host cell and viral gene targets improves the accuracy for CIN2+ detection.

Additionally, the S5 has demonstrated improved triage performance compared to hrHPV genotyping or cytology alone or combined in women outside the United Kingdom (Cuschieri et al., 2018). It has recently been validated in a HPV-positive cohort of women as part of the Canadian FOCAL clinical trial, in the FRIDA screening trial in Mexico and the Colombian ASC-US-COL trial (Cook et al., 2018; Hernández-López et al., 2019; Ramirez et al., 2021). Further, the S5 classifier was also proven as a potential prognostic test, being able to identify women with progressive CIN2 (Louvanto et al., 2019). Together, these data may mark the prospect of using the S5 classifier as a tool for objectively identifying clinically significant cervical abnormalities and predicting their outcomes.

Year	Milestone in S5 DNA methylation development
2012	Elevated HPV16 DNA methylation is associated with high grade CIN development (Mirabello et al., 2013).
2013	Methylation of HPV18, HPV31, HPV45 viral DNA indicates transforming infections (Wentzensen et al., 2012).
2013	HPV16 L1 and L2 DNA methylation predicts high-grade CIN in women with mildly abnormal cervical cytology (Lorincz et al., 2013).
2014	Aberrant methylation of several human genes (*EPB41L3, EDNRB, LMX1, DPYS, MAL* and *CADM1*) are associated with CIN2/3 development. *EPB41L3* shows the most promising results (Vasilievic et al., 2014).
2014	The S4 DNA methylation classifier based on host EPB41L3 and viral HPV16, HPV18 and HPV31 genes can predict CIN lesions(Brentnall et al., 2014).
2014	Methylation of host *EPB41L3, LMX1* and viral of HPV16 L1 can predict cervical lesions associated with HPV16 infection (Louvanto et al., 2015).
2015	**The S5 DNA methylation classifier:** HPV33 is added to the S4 structure (Brentnall et al., 2015).
2016	S5 is validated Predictors 3 study based on a United Kingdom population and the 0.8 threshold is defined (Lorincz et al., 2016).
2018	S5 is validated in the FOCAL HPV randomized cervical cancer screening trial based on Canadian women aged 25-65. Baseline methylation testing using the S5 provides triage performance similar to an algorithm relying on cytology and HPV genotyping (Cook et al., 2018).
2019	S5 is validated in the FRIDA screening trial based in Mexico. S5 methylation testing on hrHPV-positive women significantly increases diagnostic accuracy compared to triage by HPV16/18 plus cytology. S5 might have clinical utility as an additional test to the screening algorithm with the aim to decrease colposcopy referrals (Hernández-López et al., 2019).
2020	S5 shows high potential as a prognostic biomarker to identify progressive CIN2 (Louvanto et al., 2019).
2020	The S5 classifier was used with a cut-off of 2.85 to triage Chinese women with hrHPV positive and/or cytology abnormal results. Triage with S5 could reduce the numerous unnecessary colposcopy referrals and avoid overtreatment(Gu et al., 2020).
2021	S5 testing shows better sensitivity than HPV16/18 or cytology and comparable specificity for CIN2+ and CIN3+ detection in hrHPV-positive Colombian women. Triage with S5 shows comparable sensitivity and significantly fewer false positives than cytology and HPV16/18 combination (Ramirez et al., 2021).

Table 1.2 – History of the S5 classifier development

1.5.6 Microbial markers for cervical cancer progression

In addition to DNA methylation, microbial biomarkers may provide detailed information about the health status of women reproductive system. For instance, high diversity microbiota was frequently shown to correlate with HPV status and cervical disease progression (Kyrgiou et al., 2018). A large number of studies reported that HPV positive women exhibited a higher diversity in the cervicovaginal microbiota than their HPV negative counterparts (Arokiyaraj et al., 2018; Paola et al., 2017; Virtanen et al., 2017). Bacteria from the genus of *Sneathia* was highly associated with hrHPV positivity and high risk of cervical disease development (OR: 8.25, 95%CI 2.13-32.0) (Audirac-Chalifour et al., 2016; Mitra et al., 2015b; Paola et al., 2017). Additionally, the presence of high levels of *A. vaginae*, *G. vaginalis* and *L. iners* combined with low levels of *L. crispatus* was suggested as the most hazardous combination for CIN development (OR: 5.80, 95%CI 1.73–19.4) (Oh et al., 2015). Further, a cervicovaginal microbiota dominated by *L. iners* was significantly associated with CIN2+ progression in HPV positive women (Piyathilake et al., 2016). However, to validate the above-mentioned findings, further investigations should focus on the causal link between cervicovaginal microbiota and clinical outcome in longitudinal samples from screening trials.

Although the 16S rRNA microbiome sequencing is the most efficient technology to detect the complexity of bacteria in a sample, it is too expensive to be implemented routinely. Microbial markers can be included in real-time PCR assays to test for specific bacteria in the cervicovaginal microbiota from the routinely collected smear tests (Nugent et al., 1991). For instance, commercial tests designed to detect *G. vaginalis* shows a high sensitivity (90-94%) and specificity (81-97%) in detecting bacterial vaginosis (Coleman & Gaydos, 2018; Nugent et al., 1991), characterised by a dysbiotic state also associated with CIN development (Curty et al., 2020; Kyrgiou et al., 2018). Multiplex real-time PCR assays were also designed for *A. vaginae*, *Megasphaera spp.*, *L. crispatus* and *L. jensenii*, also showing high sensitivities (90-94%) and variable specificities (70.2-95%) (Hilbert et al., 2016). In combination with other biomarkers for cervical cancer progression (i.e. methylation markers), the microbial markers might have the potential to improve the specificity for triaging women after HPV testing.

1.6 PROJECT AIMS

The S5-classifier has so far demonstrated improved triage performance compared to hrHPV genotyping, cytology or the combination thereof and has been validated in HPV-positive cohorts of women as part of the Canadian FOCAL, Mexican FRIDA and Colombian ASC-US COL clinical trials (Cook et al., 2018; Hernández-López et al., 2019; Ramirez et al., 2021). There is no validation of the S5-classifier in a large number of cancer samples from both high-income countries and LMICs. This is required to demonstrate that the methylation test can consistently detect the vast majority of cervical cancers worldwide including non-squamous types. Extensive validation of the S5 classifier will support implementation of the test in global screening programmes and routine patient management clinics. The first aim of this thesis is to analyse the performance and consistency of S5 in detecting high grade lesions and cervical cancers from diverse geographical settings. This includes quantifying the degree of epigenetic separation between women who are cytologically negative versus those with high grade CIN and cervical cancer, including various histologies and HPV-negative cancers. The first objective is to show that the S5 scores are consistent regardless of sample type. The second objective is to identify a suitable S5 cut-off for LMICs. The third objective here is to analyse the individual S5-classifier components and their contribution to the performance of the test.

To date, molecular biomarkers have been investigated for triaging purposes. Specifically, the focus has been placed on epigenetic DNA methylation assays as they may offer a more accurate and cost-effective alternative to cytology. For instance, the S5 methylation scores have been shown to increase with disease severity allowing this test to be a strong contented for triaging. As epigenetic DNA methylation is tissue specific, the local microbial environment may influence the rate of DNA methylation (Barros & Offenbacher, 2009). Therefore, correlations between the S5 methylation score and specific microbial signatures may indicate an interplay between epigenetic changes and the disease environment. The second aim of this thesis is to characterise cervicovaginal microbiota changes over time in young hrHPV-positive women who developed CIN3 and recovered from LEEP treatment. We also investigate potential microbial markers priming CIN3 development by comparing the cervicovaginal microbiota composition in women with CIN3 to that of women with persistent hrHPV infections and normal cytology results. Finally, the third aim is to show that specific microbial signatures are associated with epigenetic changes (S5) and disease development.

CHAPTER 2 – MATERIALS AND METHODS

2.1 STUDY DESIGN AND CLINICAL SPECIMENS

2.1.1 Bisulfite conversion kit determination

A total of 336 women were initially included in the study. Specimens were collected as following: exfoliated cells from liquid-based cytology (LBC) specimens only from 216 women, formalin-fixed paraffin-embedded (FFPE) specimens only from 48 women, paired LBC and FFPE specimens from 72 women. A total of 5 women were excluded from the study due to inadequate DNA concentration (n=3) or DNA quality (n=2) as described in Figure 2.1. The 48 FFPE specimens were obtained from archived material from the Predictors studies collected from 2005 to 2009 (Ganti et al., 2015; Szarewski et al., 2008). Additional specimens included in the study were obtained from i) The International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) for Bhutan FFPE material and all LBC samples from Colombia, Georgia, India and Philippines, South Africa and Spain; ii) The Scottish HPV Archive, MRC Centre for Reproductive Health, The University of Edinburgh for FFPE and LBC samples from the UK; iii) Department of Microbial, Cellular and Molecular Biology, Addis Ababa University for all samples from Ethiopia; iv) Department of Pathology, University of New Mexico for FFPE and LBC samples from USA – New Mexico. All specimens in i, ii, iii and iv are part of a larger study involving the DNA methylation analysis with the S5-classifier as described in section 2.2. Final cytological reports and histopathological diagnoses were based on reviews by at least two pathologists a diagnosed as: histology negative cervix (NEG), CIN1, CIN2, CIN3 or squamous cell carcinoma (SCC). The highest grade of abnormality seen in the biopsy was used. Final details regarding the country of origin of the patients are described below (**Table 2.1).**

Country of Origin	**Specimen type**	
	LBC, n = 288 (%)	*FFPE, n = 120 (%)*
Colombia	24 (8.33)	-
Bhutan	24 (8.33)	24 (20.86)
Ethiopia	24 (8.33)	-
Georgia	24 (8.33)	-
India	24 (8.33)	-
Philippines	24 (8.33)	-
South Africa	24 (8.33)	-
Spain	24 (8.33)	-
United Kingdom	48 (16.66)	72 (58.26)
USA (New Mexico)	48 (16.66)	24 (20.86)

Table 2.1 – LBC vs FFPE sample characteristics

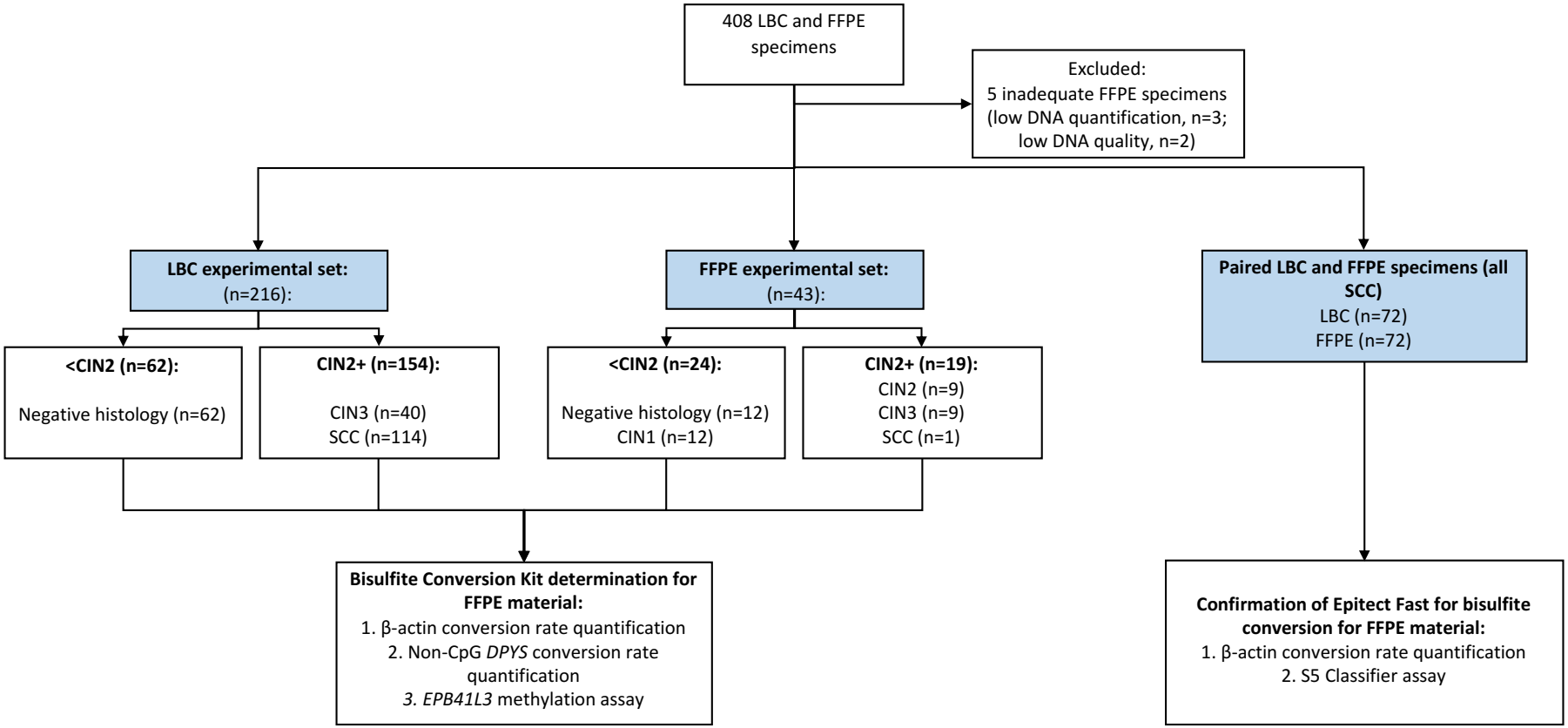

Figure 2.1 – **LBC vs FFPE study design flowchart.** Abbreviations: LBC: liquid-based cytology; FFPE: formalin-fixed paraffin-embedded; CIN1/2/3: cervical intraepithelial neoplasia grade 1/2/3; SCC: squamous cell carcinoma

2.1.2 Worldwide quantification of DNA methylation using the S5 classifier

Here, we present a cross-sectional retrospective study including women from ten countries to evaluate the S5 methylation classifier performance to detect CIN3 and cervical cancer. The institutions involved in sample acquisition were: The International Agency for Research on Cancer (IARC) of the World Health Organization (WHO); The Scottish HPV Archive, MRC Centre for Reproductive Health, The University of Edinburgh; Department of Microbial, Cellular and Molecular Biology, Addis Ababa University; University of New Mexico Human Research Review Committee; Wolfson Institute of Preventive Medicine, Queen Mary University of London. Samples were collected from 1991 to 2014 and sored accordingly in archives. Sample size calculations for the present study has been done based on previous published studies of S5 (Cook et al., 2018; Hernández-López et al., 2019; Ramirez et al., 2021). The ratio Cyt(-):Cancer =1:3.The study design is presented in Figure 2.2.

The samples used for this study originate from archives and have received ethical and research approvals from: i) IARC - WHO for all samples from Bhutan, Colombia, Georgia, India and Philippines, South Africa and Spain; ii) The Scottish HPV Archive, MRC Centre for Reproductive Health, The University of Edinburgh for all samples from the United Kingdom; iii) Department of Microbial, Cellular and Molecular Biology, Addis Ababa University for all samples from Ethiopia; iv) University of New Mexico Human Research Review Committee for all samples from USA – New Mexico. The study protocol has been approved by Queen Mary University of London's Ethical Committee (MTA-2018-ICE-0426, MTA-2018-ICE-1353, 11/06/2018).

Cervical cell swabs and biopsies were collected from a total of 973 patients aged 21-64 as described in the referenced papers (Alibegashvili et al., 2011; Bosch et al., 1992.; Bosch et al., 1995; Clifford et al., 2005; Tshomo et al., 2014; Vuyst et al., 2012). Details about specimens in each study are detailed below in table 2.1.2. Cytology and histopathology results for all samples are detailed in Figure 2.2. We excluded CIN1 and CIN2 from our study because the latter is a highly misclassified diagnostic category consisting mainly of either CIN1 or CIN3 diagnoses that cannot be distinguished accurately by morphological examination. Our study focussed mainly on cervical cancer stages I and II in order to have a sharper view of the epigenetic contrast between CIN3 versus early cancer.

Study	Country	Overview specimens	Genotyping
Alibegashvili et al., 2011	Georgia	Cervical cell swabs from 1309 women aged 18–59 years and biopsies from 91 women histologically confirmed with invasive cervical cancers (squamous-cell carcinoma)	GP5+/6+-based PCR assay
Bosch et al., 1992	Spain	Biopsies from 436 women histologically confirmed with squamous-cell carcinoma and from 387 women diagnosed with CIN3.	GP5+/6+-based PCR assay
Bosch et al., 1995	22 countries including Colombia, India Philippines	Biopsies from 1000 women histologically confirmed with squamous cell-carcinoma and CIN3 were collected and stored frozen at 32 hospitals in 22 countries.	GP5+/6+-based PCR assay
Clifford et al., 2005	13 countries including India, Spain, Colombia	Cervical cell swabs from 15 613 women aged 15-74 years without cytological abnormalities were included in a pooled analysis.	GP5+/6+-based PCR assay
Tshomo et al., 2014	Bhutan	Biopsies from 211 women diagnosed with CIN3 and 112 women histologically confirmed with invasive cervical cancer cases. Also, the study included cervical cell swabs from 2 505 sexually active women aged 18-69 years and underwent cytology and HPV screening.	GP5+/6+-based PCR assay
Vuyst et al., 2012	Kenya and South Africa	Biopsies from 106 HIV-positive women (mean age 40.8 years) and 129 HIV-negative women (mean age 45.7) histologically confirmed with squamous cell carcinoma were included.	GP5+/6+-based PCR assay

Table 2.1.2 Characteristics of specimens included in selected studies

Details regarding patient characteristics are described in Table 4.1.

With respect to cervical liquid-based cytology samples, the cellular material from cytology specimens were collected in PreservCyt® (Hologic Corporation, Marlborough MA, USA) medium for storage until DNA extraction. A subset of specimens from Bhutan (n=10), Ethiopia (n=49), India (n=10), Spain (n=20), United Kingdom (n=51), Colombia (n=20), USA (n=50) were selected for negative cytology results. CIN3 samples originated from Spain (n=50), United Kingdom (n=54), Colombia (n=50) and USA (n=50). Cancer samples originated from Ethiopia (n=81), South Africa (n=49), Bhutan (n=50), India (n=50), Philippines (n=50), Georgia (n=42), Spain (n=50), United Kingdom (n=51), Colombia (n=46), USA (n=86). All CIN3 and cancer samples were collected from patients showing abnormal cytology and histology results though colposcopy referral and diagnosis according to specific country recommendations. Biopsy samples were collected and stored at -70˚C (IARC and Scottish HPV Archive, Addis Ababa University); formalin-fixed paraffin-embedded (FFPE) samples were stored at room temperature (Bhutan-IARC and University of New Mexico) until DNA was extracted.

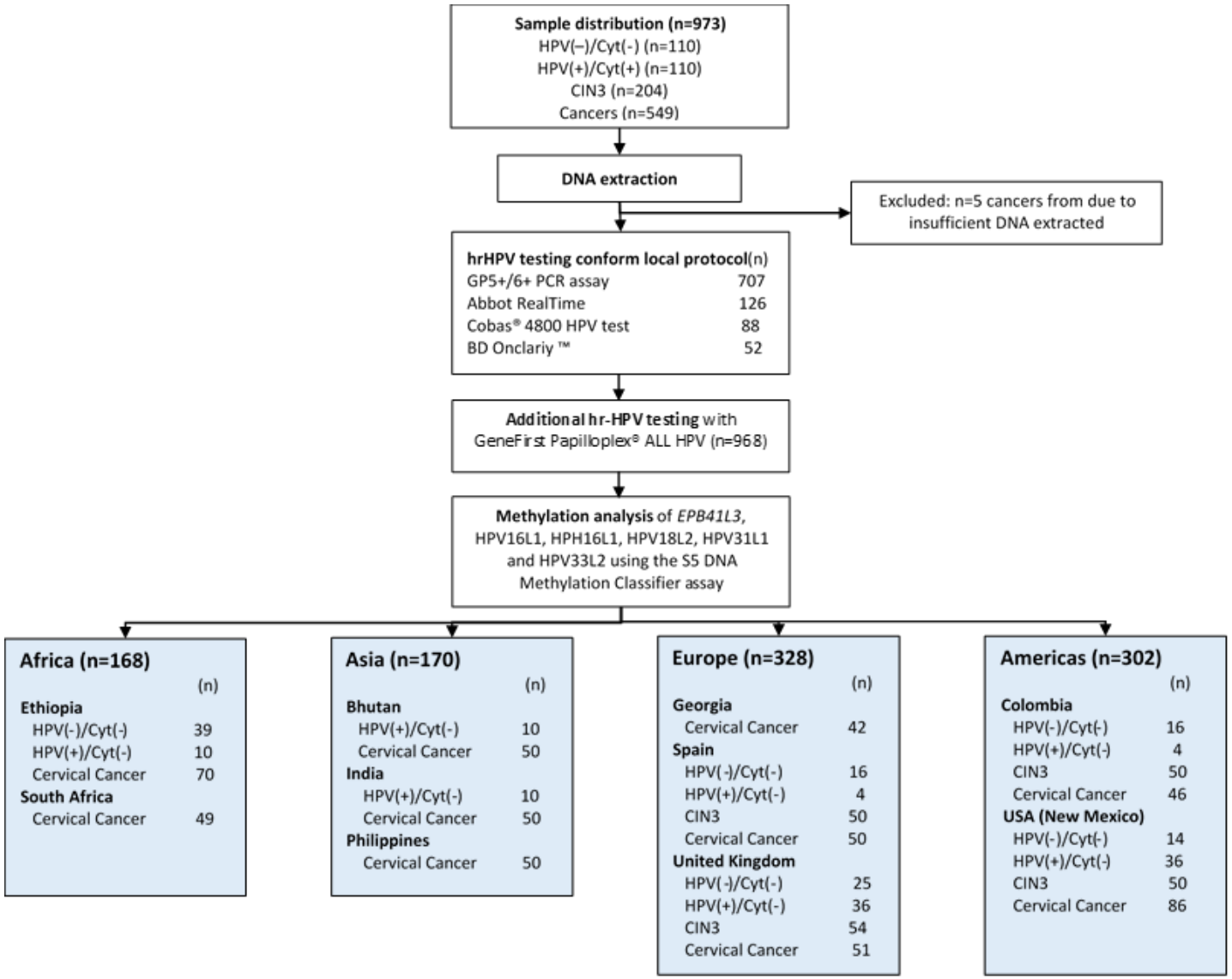

Figure 2.2 – S5 Classifier world-wide validation study design flowchart. Abbreviations: HPV, human papillomavirus, CIN3, cervical intraepithelial neoplasia grade 3

2.1.3 Investigation of the cervicovaginal microbiota

We investigated the microbiota composition of cervical swabs (LBC) originated from women who participated in the ARTISTIC screening trial. Every woman in this present longitudinal case-control study was samples three times over a course of six years. Three distinct groups were compared: i) hrHPV-negative women (n=14); ii) persistently hrHPV positive women (n=15) and iii) hrHPV positive women who developed CIN3 and were treated (n=30). Metagenomic sequencing was performed on Ion S5™ Metagenomics Sequencing and sequencing results were compared within and between the groups.

The samples included in this study were part of the ARTISTC trial which was designed to collect cervical samples from women aged 20-39 years attending for cervical screening (NHS programme) in Greater Manchester UK, between July 2001 and October 2003. In total, 24,510 participants had a cervical sample collected for liquid based cytology and HPV testing. Participating women were followed beyond the original protocol, until March 2008 and participated in three main screening rounds, each approximatively 3 years apart. Women found to develop CIN3 were referred to colposcopy, surgically treated through LEEP and screened three-year post-treatment. The use of the ARTISTIC samples in this study have been approved by the NRES Committee South East Coast - Brighton and Sussex (Study title: Long-term follow-up of ARTISTIC cervical screening trial cohort) with the REC reference: 14/LO/0627 IRAS project ID: 153311

Initially, 65 women were selected for metagenomics sequencing of the cervical swabs collected part of the ARTISTIC trial. All women participated in all three screening rounds, approximately 3 years apart. Due to set sequencing reads criteria, 6 women were excluded and the remaining 59 were divided in three arms as described in Figure 2.3. The three arm of the study are: i) group A consisting cervical swabs collected from healthy women who were hrHPV negative (n=14); ii) group B consisting of cervical swabs collected from healthy women who were hrHPV positive throughout the trial (n=15); and iii) group C including cervical swabs from initially hrHPV positive healthy women (C1) who developed CIN3 at the first follow-up (C2) and were surgically treated for the disease, testing HPV-negative (C3)(n=30).

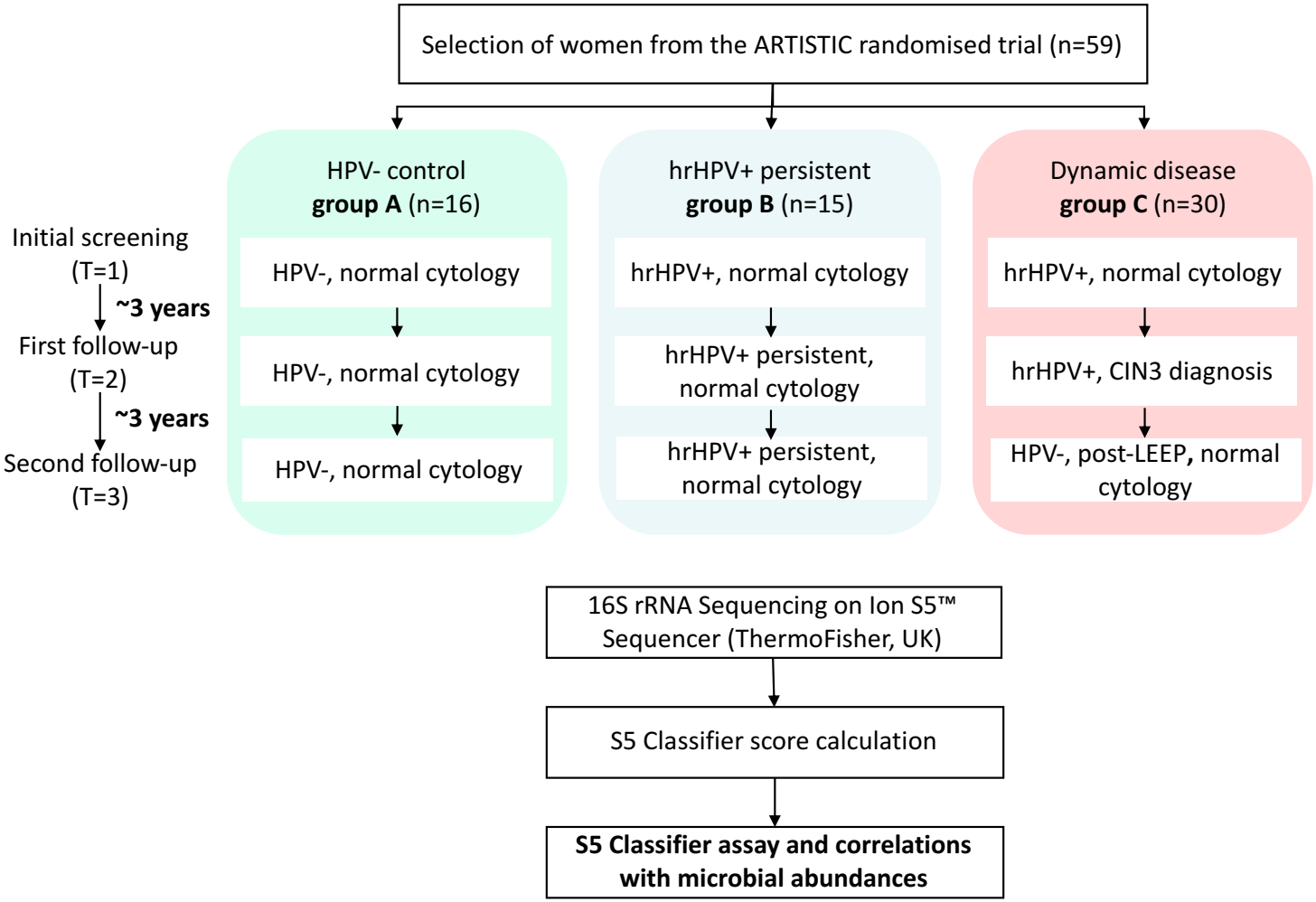

Figure 2.3 – **Longitudinal microbiome study design flowchart**. Abbreviations: hrHPV, high-risk human papillomavirus, CIN3, cervical intraepithelial neoplasia grade

2.2 DNA EXTRACTION AND QUALITY ASSESSMENT

2.2.1 LBC DNA extraction

Specimens were collected in sterile PreservCyt® (Hologic Corporation, Marlborough MA, USA) medium and stored immediately -80°C to stop any bacterial growth. DNA was extracted directly from PreservCyt® using: i) QIAsymphony DSP DNA Mini Kit (Qiagen, Hilden, Germany) for all samples in the IARC biobank; ii) Abbott M2000 system (Abbott Laboratories, USA) for all samples in the Scottish HPV Archive biobank; iii) DNA from all other specimens was manually extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), as per manufacturer's notes. Isolated DNA was stored at -70˚C until required. The DNA concentration was measured using a Qubit Fluorometer with the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA).

2.2.2 FFPE DNA extraction

FFPE DNA was extracted from 115 FFPE tissue samples. Ten 5 µm FFPE sections were cut using a new blade for each block and directly put in a tube and stored at -70°C until DNA extraction. DNA was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions with an initial incubation step at 56°C for 16-18 hours with proteinase K and a 1-hour incubation at 90°C. The DNA concentration was measured using a Qubit Fluorometer with the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA).

2.3 DNA BISULFITE CONVERSION

2.3.1 Bisulfite conversion optimisation

One hundred nanograms of all LBC DNA were used in the bisulfite conversion reaction where un-methylated cytosines were converted to uracils with the EZ DNA methylation kit (Zymo Research, CA, USA). 43/115 FFPE samples: 24 <CIN2 (12 NEG and 12 CIN1) and 19 CIN2+ samples (9 CIN2, 9 CIN3 and 1 SCC) were bisulfite converted using four different kits: the EZ DNA methylation kit (Zymo Research, Irvine, USA, abbreviated as Zymo Standard), the EZ DNA Methylation-Lightning kit (Zymo Research, CA, USA, abbreviated as Zymo Lightning), the Epitect Bisulfite kit (Qiagen, Hilden, Germany, abbreviated as Epitect Standard) and the Epitect Fast Bisulfite kit (Qiagen, Hilden, Germany, abbreviated as Epitect Fast). The reminder 72/115 FFPE samples (72 SCC) were bisulfite converted only with Zymo Standard and EpiTect Fast kits. All kits were used following the manufacturer's instructions with an input of 40ul of DNA for Zymo Standard, Epitect Standard and Epitect Fast and 20 µl (maximum recommended volume) for Zymo Lightning. Differences between kits are highlighted in Table 2.2.

	Zymo Standard	Epitect Standard	Zymo Lightning	Epitect Fast
DNA protect buffer		Tetrahydrofurfuryl alcohol		Tetrahydrofurfuryl alcohol
Conversion reagent	Sodium metabisulphite & sodium hydroxide	Sodium metabisulphite	Ammonium bisulfite	Ammonium hydrogensulphite
Denaturation	15 min 37°C in sodium hydroxide	5 min 95°C	8 min 98°C	5 min 95°C
Incubation	16 hours 50°C	25 min 60°C	60 min 54°C	10 min 60°C
Denaturation		5 min 95°C		5 min 95°C
Incubation		85 min 60°C		10 min 60°C
Denaturation		5 min 95°C		
Incubation		175 min 60°C		
Binding to column	Guanidinium chloride	Guanidine thiocyanate + carrier RNA	Guanidinium chloride	Guanidine thiocyanate + carrier RNA
Desulphonation agent	Sodium hydroxide, ethanol, propanol	Sodium hydroxide	Sodium hydroxide, ethanol, propanol	Sodium hydroxide

Table 2.2 - Composition and procedures of the four bisulfite conversion kits investigated.

2.3.2 Quality assessment

LabChip® GX Genomic DNA Quantification technology (Caliper Life Sciences, USA) was used to analyse the quality and concentration of the bisulfite converted DNA originating from LBC and FFPE material. The assay was run on 48 samples at time. All reagents were equilibrated at room temperature for 20 min prior to the assay. To prime the LabChip® (Caliper Life Sciences, USA), 3.44 µl DNA Dye Concentrate (Caliper Life Sciences, USA) were added to 275 µl Genomic DNA Matrix (Caliper Life Sciences, USA). The mixture was vortexed and transferred into 2 spin filteres to be centrifuged at 9200 rcf for 7.5 min. The mixture was added on the LabChip® along with 120 µl Genomic DNA Marker (Caliper Life Sciences, USA), according to the manufacturer's notes. On a 384-well plate, 2 µl of DNA and 18 µl of Nuclease-Free water were added to each well. All samples were ran in duplicates. The primed LabChip® and the 384-well plate were placed in the LabChip® GX Touch Nucleic Acid Analyser. Scores ranging from 0 to 5 (0-low quality, 5-high quality) were attributed for each sample according to DNA quality.

2.3.3 Real-time PCR primer design

Real-time PCR primers were designed against an in silico reference sequence of the gene of interest and validated in NCBI Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and Pyromark Assay Design software (Qiagen, Hilden, Germany).

2.3.4 Bisulfite conversion efficiency quantification

Bisulfite conversion efficiency measures the percentage of cytosines in CpG sites converted into uracils. Any conversion at non-CpG sites indicates an incomplete conversion reaction and allows us to estimate the bisulfite conversion efficiency. The efficiency of bisulfite conversion was measured by an in-house real-time PCR using two pairs of primers ("converted" and "unconverted") targeting a region of the β-Actin gene containing non-CpG sites. After bisulfite conversion, the cytosines should all be fully converted and anneal to the 'converted' primers. If some cytosines were left unconverted, the DNA strand will anneal to the 'unconverted' primers. Both F and R converted/unconverted primers were targeting the same fragment of DNA. The "converted" primer pair anneals to the fully converted DNA, the "unconverted" primer pair to unconverted DNA. Primer are described in Table 2.3. The samples were run in duplicate on a Quantstudio 5 (ThermoFisher Scientific, UK) using a 384-well block. For each 10 µl PCR, 1.25 µl of bisulfite-converted DNA (diluted 1:5), 6.25µL 2x KAPA SYBR FAST qPCR Master Mix, 0.625 µL µl each forward and reverse primer (final concentrations of 0.5 µM), and 44.375 µL RNAse Free Water (Qiagen, Hilden, Germany) were used. Real-time PCR conditions were 95°C for 10 min followed by 40 cycles of 95°C for 15 s, 60°C for 1 min with data acquisition after each cycle. At the end, a dissociation curve was added ranging from 60°C to 95°C. Threshold of cycles (CT) values were calculated in duplicate. CTs represents the amplification cycle when amplification starts. The mean of the duplicate CT values was calculated for the converted primer and unconverted primer values. The delta CT was calculated as the average of "converted" CT minus the average of "unconverted" CT value and the percentage of unconversion was calculated as 100%/ (1 + (2^ (delta CT). The converstion efficiency was calculated as 100% minus percentage of unconversion.

Primer	Primer name	Primer sequence 5' to 3'
Converted β-Actin Control	ACT (+) F	TGGTGATGGAGGAGGTTTAGTAAGT
	ACT (+) R	AACCAATAAAACCTACTCCTCCCTTAA
Unconverted β-Actin Control	ACT (-) F	TGGTGATGGAGGAGGCTCAGCAAGT
	ACT (-) R	AGCCAATGGGACCTGCTCCTCCCTTGA

Table 2.3 – Conversion rate real-time PCR primers

2.4 TARGETED METHYLATION SEQUENCING

2.4.1 PCR amplification

TA total of 2µL converted DNA were used per PCR reaction. The converted DNA was amplified using one methylation independent biotinylated primer and one non-biotinylated primer (Table 2.4). For each 23 µL PCR reaction master-mix, the following were added: 2.5 µL CoralLoad 10x (Qiagen, Hilden, Germany), 5.3 µL RNAse Free Water (Qiagen, Hilden, Germany), 12.5 µL Pyromark PCR mix (Qiagen, Hilden, Germany), 2.5 µLM MgCl2 (Qiagen, Hilden, Germany) and 0.2 µM biotinylated primer mix (Sigma-Aldrich, Germany). PCR cycling conditions were 15 minutes at 94°C, followed by 45 cycles of 94°C, 54°C, 72°C each for 30 seconds and a final extension at 72°C for 10 minutes.

Gene	Primer name	Primer sequence 5' to 3'	Annealing T (°C)	No Cycles	[Primer] (mM)
EPB41L3	EPB41L3 F	GGGGGATTTGTGTAAATTGG	54	45	0.2
	EPB41L3 R (btn)	(btn) ACCTAAAAACCTCCCTAAAATC			
HPV16L1	HPV16 L1.3 F (btn)	[btn]-AGTGAAGTTTTATTGGATATTTGTAT	51	40	0.3
	HPV16 L1.3 R	CAACAACACCAACCCTATTAAAT			
HPV16L2	HPV16 L2 F	GTATGTTTTATAAAGTTGGGTAG	51	50	0.3
	HPV16 L2 R (btn)	[btn]-TTAATAAACTATTATCACTTAACAATAC			
HPV18L2	HPV18L2 6CpG F	GTATAGGTTGTTTTATATAGTGTATTGT	54	50	0.2
	HPV18L2 6CpG R (btn)	[btn]-TCCACCTTAAAAACAACATCAAATAA			
HPV31L1	HPV31 L1 6284 F (btn)	[btn]-ATTTGTGTATTTGAAGTAATTATGGAG	49	45	0.2
	HPV31 L1 6284 R	TCCAAATTATCTTAAAATAATTACTAAACC			
HPV33L2	HPV33 L2 F	AGGTAGGTATATTGTGGTTTTATTAGGT	53	45	0.2
	HPV33 L2 R (btn)	[btn]-CACATCTAACCCATTTATTCCTATTTC			
DPYS	DPYS F	GGTTTGGGGTGTTTTTTTGTAAGG	56	45	0.2
	DPYS R (btn)	[btn]-AAACTCCAACCCAACCTTCC			

Table 2.4 – S5 classifier PCR primers and PCR conditions. Abbreviations: F: forward primer; R: reverse primer; Btn: biotinylated primer

2.4.2 Pyrosequencing

Pyrosequencing is a targeted DNA sequencing method. It relies on the detection of light produced in a luciferase-catalysed reaction by a charge coupled device (CCD) camera. Each light signal is proportional to the number of nucleotides incorporated and it is seen as a peak in a pyrogram™. Pyrosequencing involves the hybridisation of the biotinylated PCR amplicons with sequencing primers via incubation with DNA polymerase, ATP sulfurylase, luciferase, adenosine 5' phosphosulfate (APS) and apyrase (Figure 2.4).

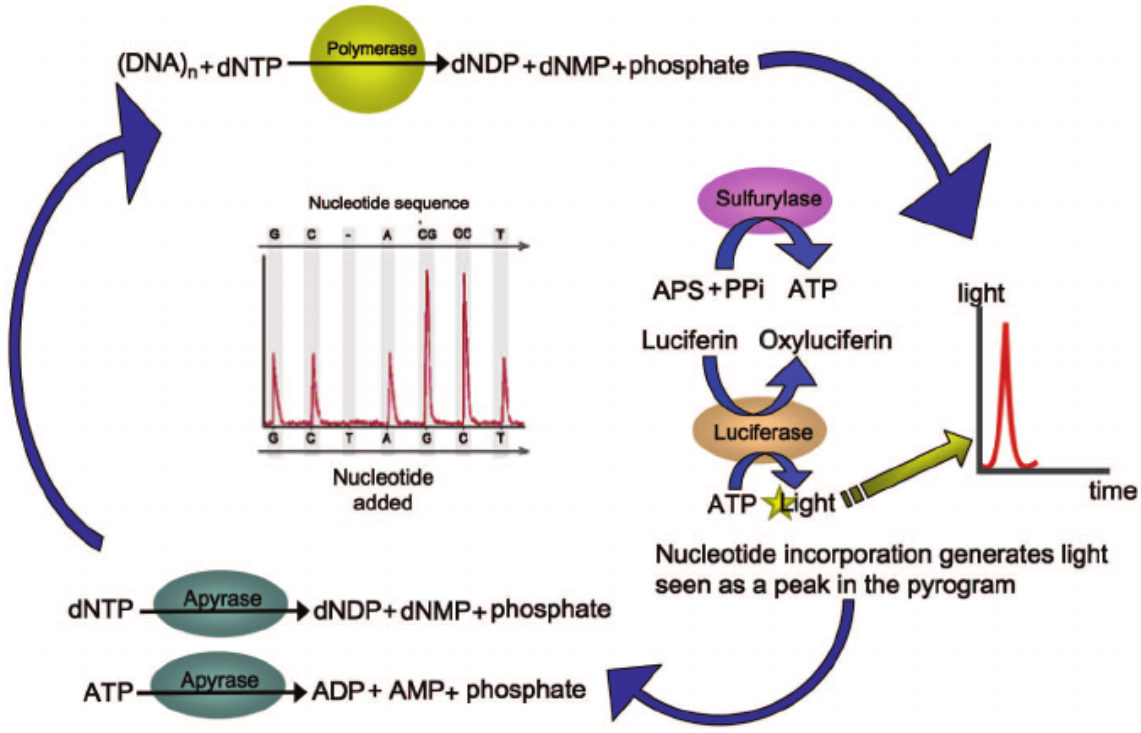

Figure 2.4 – Pyrosequencing reaction flowchart. Upon adding the deoxyribonucleotide triphosphate (dNTP) mix to the reaction, the DNA polymerase catalyses the addition of dNTPs to the sequencing primer, if complementary to the base in the template strand. Incorporation of one dNTP results in the release of one molecule of pyrophosphate (PPi). Further, ATP sulfurylase converts PPi to ATP in the presence of APS. The resulted ATP drives the luciferase-catalysed conversion of luciferin into oxyluciferin which leads to the generation of visible light. The light produced is then detected by CCD camera. Each light signal is proportional to the number of nucleotides incorporated and it is seen as a peak in a pyrogram™. Moreover, apyrase is a nucleotide degrading enzyme which continuously degrades unincorporated nucleotides and ATP to prepare the space for the next incorporation event. Addition of dNTPs is done sequentially to allow the possible incorporation of any. dNTP at every incorporation site (Adapted from Petrosino et al., 2010).

The biotinylated PCR amplicons were pyrosequenced using a PyroMark™Q96 ID (Qiagen, Hilden, Germany) instrument as previously described (Vasilievic et al., 2011). A total of 3µl streptavidin beads (GE Healthcare, Buckinghamshire, UK), 37 µl PyroMark binding buffer (Qiagen), 25 µl PCR product and 25 µl nuclease-free water were mixed and incubated for 10 min on a shaking table at 1300 rpm. Using the Biotage Q96 Vaccum Workstation, amplicons were separated, denatured, washed and subsequently mixed with 45 µl annealing buffer containing 0.33 µM of pyrosequencing primer. DNA methylation analysis on CpG islands of *EPB41L3* (CpG 438, 427, 425) and viral late genes (L1 and L2) coding for proteins involved in viral capsule formation: HPV16L1 (CpG 6367, 6389), HPV16L2 (CpG 4275, 4268, 4259, 4247, 4238), HPV18L2 (CpG 4256, 4261, 4265, 4269, 4275, 4282), HPV31L1 (CpG 6352, 6367) and HPV33L2 (CpG 5557, 5560, 5566, 5572) as previously described(Lorincz et al., 2016). Percentage methylation was taken as the mean for CpG sites involved in each case. Sequencing primers are described in Table 2.5**.** DNA methylation analysis on the CpG positions of *DPYS* (CpG 6, 13, 16, non-CpG, 24) using 5'-GGTTTGGGGTGTTTTTTTGTAAGG-3'.

Percentage methylation was taken as the mean for CpG sites involved in each case. Primer annealing was performed by incubating the samples at 80◦C for 2 min and allowed to cool to room temperature for 5 minutes prior to pyrosequencing. PyroGold reagents were used for the pyrosequencing reaction and the signal was analysed using the PyroMark™Q96 ID (Qiagen, Hilden, Germany). Target CpGs were evaluated by instrument software (PSQ96MA 11.4) which converts the pyrograms to numerical values for peak heights and calculates proportion of methylation at each base as a C/T ratio. All pyrosequencing runs included positive controls of known methylation level (0%, 50% and 100%) (Millipore, Billerica, MA, USA) to allow standardised direct comparisons between different primer sets and a negative control.

Gene	Primer name	Primer sequence 5' to 3'	CpG positions
EPB41L3	EPB41L3 s	GGGATTTGTGTAAATTGG	438, 427, 425
HPV16L1	HPV16L1.3 6367 s	CTAACAAACATTTATTCCCTTC	6389, 6367
HPV16L2	HPV16L2.5CpG 5s	TTTTATAAAGTTGGGTAGT	4275, 4268, 4259, 4247, 4238
HPV18L2	HPV18 L2s	TGTATTTTTGTAATAAAAGTATGGTA	4256, 4261, 4265, 4269, 4275, 4281
HPV31L1	HPV31 6284 sB	CATTTTTTTAATAAATCAAACAC	6352, 6364
HPV33L2	HPV3333 L2 5572s	GGATATTTGTAAAAAAATATGG	5557, 5560, 5566, 5572

Table 2.5 - Pyrosequencing primers and target CpG positions investigated.

2.5 THE S5 DNA METHYLATION CLASSIFIER

The S5 DNA methylation analysis involves the bisulfite conversion of unmethylated cytosines to thymidines followed by PCR amplification and pyrosequencing (Bock et al., 2016). During sodium bisulfite treatment, unmethylated cytosines are converted to uracils through a deamination reaction, while 5-methylcytosine remains intact. This results in the conversion of the usually undetectable epigenetic information into detectable sequence information at base pair resolution. The S5 classifier was developed and validated on cells collected in LBC medium (Lorincz et al., 2016; Lorincz et al., 2013). S5 was calculated here for DNA extracted from LBC and FFPE material. For this test, the threshold methylation score i.e. cut-off value for CIN2+ detection depends on HPV load and screening availability. Due to systematic screening in the United Kingdom, the predefined threshold for the S5 classifier has been set at 0.80 (Lorincz et al., 2016). S5 is defined as the sum of six methylation components:

S5 = 30.9(EPB41L3) + 13.7(HPV16L1) + 4.3(HPV16L2) + 8.4(HPV18L2) + 22.4(HPV31L1) + 20.3 (HPV33L2).

All components, except HPV16L2, are calculated as the percentage of average methylation of the CpG investigated (3 for EPB41L3, 2 for HPV16L1, 6 for HPV18L2, 2 for HPV31L1 and 3 for HPV33L2). For HPV16L2, the proportion of 3 CpGs (sites: 4238, 4259, 4275) with methylation values >0 is used (Lorincz et al., 2013).

2.6 HPV GENOTYPING

The clinical samples used in the study have been initially genotyped as being part of previous research studies or stored in biobanks. The following technologies were used: GP5+/6+ PCR assay (n= 641) (Bosch et al., 1992; Bosch et al., 1995, 1998; Vuyst et al., 2012), Abbot RealTime assay (n=126), Cobas® 4800 HPV test (n= 88), BD Onclariy ™ assay (n=52). Additionally, the Papilloplex High Risk HPV test (GeneFirst, UK), a multiplexed PCR-based system that simultaneously detects 14 hrHPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68) was used to identify HPV infection. The assay was based on multi-probe amplification technology which allowed differentiation of up to six targets per each of the following florescence channels: FAM, JOE/HEX, ROX and Cy5. The technology is using a combination of PCR primers and probes (dual labelled fluorescent probe and partially complimentary oligo hybrid) to detect each specific target. Each HPV-specific probe had a unique melting profile, consisting of a unique melting temperature and shape. With reference to the baseline, any change in the characteristic melting profile showed the sample to be positive for the respective

HPV type(s). Genotyping was run on QuantStudio 5 Real-Time PCR System (TheomoFicher Scientific, UK) and targeted the L1 region of all the genotypes described. Papilloplex High Risk HPV internal negative and positive controls were used as baselines and analysis was performed on the GeneFirst software.

All samples included in the microbiota study presented in Chapter 5 have been hrHPV genotyped as part of the ARTISTIC screening trial guidelines with either one of the following methodologies: Abbot RealTime assay (n=86), Cobas® 4800 HPV test (n= 51), BD Onclariy ™ assay (n=40). Sample characteristics are described in Table 5.1.

We investigated the following categories: 1) HPV16 positive, 2) HPV18 positive, 3) HPV31 positive, 4) HPV33 positive and 5) other hrHPV positive (non-HPV16/18/31/33). In cases with multiple hrHPV-infections the more prevalent genotype was attributed (e.g. an HPV16-positive sample was placed in category 1).

2.7 16S rRNA METAGENOMIC SEQUENCING

2.7.1 16S rRNA library preparation

The 16S rRNA gene is highly conserved among all bacteria species. The 16S gene also contains nine hypervariable regions (V1-V9), ranging from 30 to 100 base pairs long. The degree of conservation varies widely between the hypervariable regions. Sequencing of the 16S hypervariable regions allows differentiation of bacteria at species level. The 16S rRNA hypervariable regions were amplified using the Ion 16S™ Metagenomics Kit (ThermoFisher, UK) as per manufacturer's notes. Two individual PCR reactions per, including primers covering almost all 16S hypervariable regions: 2, 4, 8 (V2-4-8; PSI) and 3, 6, 7, 9 (V3-6, 7-9; PSII) were run per sample on 96 well plates (primer structure referenced) (Barb et al., 2016). For each 30 µl reaction, the following were added: 15 µL 2X Environmental Master Mix, 3 µL of either PSI or PSII primer sets, 10 µL nuclease free water and 2 µL genomic DNA. Both positive and negative controls including *E.coli* DNA (ThermoFisher, UK) or nuclease free water instead of genomic DNA were added per run. PCR cycling conditions were 10 minutes at 95°C, followed by 25 cycles at 95°C (30 seconds), 58°C (30 seconds) and 72°C (20 seconds) and a final extension at 72°C for 7 minutes. Corresponding PSI and PSII PCR products were mixed together and purified using Agencourt® AMPure XP® Magnetic Beads (ThermoFisher, UK) as per manufacturer's notes. The DNA concentration was measured using a Qubit Fluorometer with the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) and amplicons (50 ng) were pooled.

Library preparation was performed using the Ion Plus Fragment Kit (Cat. No. 4471252, ThermoFisher, UK) in 1.5mL Eppendorf® LoBind® tubes. Amplicons ends were enzymatically processed to allow unique barcode binding. Reactions of 100 µl which included 20 µl 5X End Repair Buffer, 1 µl End Repair Enzyme and 79 µl pooled amplicons were incubated at room temperature for 20 minutes. The product was purified using Agencourt® AMPure XP® Magnetic Beads as per manufacturer's notes.

Barcoded libraries were prepared in 0.2mL PCR tubes. For each 100 µl reaction, the following were added: 25 µl DNA product, 2 µl 10X Ligase Buffer, 2 µl P1 Adapter (ThermoFisher, UK), 2 µl Ion Xpress™ Barcode X (X being the unique barcode chosen per sample) (ThermoFisher, UK), 2 µl dNTP Mix, 49 µl Nuclease-free Water, 2 µl DNA Ligase and 8 µl Nick Repair polymerase. PCR cycling conditions were 15 minutes at 25°C, followed by 5 minutes at 72°C and hold at 4°C. The resulted barcoded libraries were purified using Agencourt® AMPure XP® Magnetic Beads (ThermoFisher, UK) as per manufacturer's notes.

The sequencing input library concentration was quantified by qPCR on Quantstudio 5 (ThermoFisher, UK), using the Ion Universal Quantification Kit (Cat. No. A26217; ThermoFisher, UK). Four sequential 10-fold dilutions using *E. coli* DH10B Ion Control Library: 6.8 pM, 0.68 pM, 0.068 pM and 0.0068 pM for standard curve determination. The libraries were diluted 1:10, 1:100, 1:1000 and 1:10000. Standard 20 µl qPCR reactions included 10 µl TaqMan® Fast Universal PCR Master Mix (2X), 1 µl Ion Library TaqMan® Quantification Assay (2X), 4 µl nuclease-free water and 5 µl diluted library per sample per dilution. The thermocycling conditions were 20 seconds at 95°C, followed by 40 cycles at 95°C (3 seconds) and 60°C (30 seconds). A selection of 50 pM, 20 pM, 10 pM, 5 pM, 2.5 pM and 1 pM input for template preparation were calculated for every barcoded library. The 2.5 pM library input was found to be the most optimal.

2.7.2 Ion S5™ metagenomics sequencing

Template preparation was performed on Ion Chef™ Instrument (ThermoFisher, UK) using the Ion 510 & Ion 520 &Ion 530 Chef Kit (ThermoFisher, UK) as per manufacturer's notes (Pub, No. MAN0016854). The 2.5 pM library inputs were templated on Ion 520™ Chips (ThermoFisher, UK), covering for library lengths of <300bp. Metagenomics sequencing of the templated Ion 520™ Chips was performed on Ion S5™ Sequencer as per manufacture's notes (Pub. No. MAN0016855).

2.7.3 Bioinformatics analysis of the 16S rRNA metagenomics sequencing data

Sequencing reads were obtained in the form of FASTQ files and quality of reads was assessed using FastQC (Babraham Bioinformatics). Sequencing reads were aligned to MicroSEQ® ID 16S rRNA reference database (>15,000 organisms) along with the curated GreenGenes database (>400,000 records) for bacterial genome references using the Ion Reporter™ Software 4.2 (ThermoFisher, UK). Sequence reads were quality checked and normalised to the lowest number of reads. The resulting SAM file was processed into a sorted and indexed BAM file using Samtools v1.3. BAM files were imported onto the Galaxy Bioinformatics platform and the free LEfSe tool was used to perform machine-learning clustering of the bacterial species. OTUs were defined using an adjusted cut-off parameter of 98% in Ion Reporter™ Software 4.2 and resulted data was analysed using the Vegan package within R statistical package for assessment of microbial composition and diversity (R Development Core Team 2008). OTU taxonomies were generated from Phylium to Genus and Alpha and Beta indices were calculated from the two databases in the 'vegan' package in R v 3.4.1.

2.7.4 Markov chain analysis on transition probabilities and simulations

The Markov analysis defined the probability of a future action given the current state of a system. Probabilities for transition of each bacterial community state type (CST) were calculated in 2 different scenarios i) transitions from C1 to C2 and ii) transitions from C2 to C3. To investigate the likelihood of reaching C2 and C3, specifically, the Markov Model for i) and ii) was calculated by multiplying the transition probabilities of all CSTs with the probability of reaching a specific state at the selected timepoints: C2 and C3, respectively. Simulations of 1000 iterations were run based on the Markov models in scenarios i) and ii). Analysis was performed in Matlab Online R2019b.

2.8 STATISTICAL ANALYSIS

2.8.1 Bisulfite conversion kit determination

Comparisons of methylation levels between groups were made using Kruskall-Wallis and Dunn's multiple comparison tests. The ability of methylation scores to separate <CIN2 from CIN2+ samples was assessed with receiver operating characteristic (ROC) curves. The area under the curve (AUC) is calculated through ROC analysis. The calculated AUC are given with 95% confidence intervals. The differences between the AUC of the ROC curves from exfoliated cells and FFPE biopsies was tested using the DeLong's test. A Wilcoxon matched-pairs signed rank test was used to assess the difference in S5 methylation values between the paired FFPE biopsies

and LBC samples. Reproducibility of the assays were evaluated using Spearman's correlation test. All p-values were two sided with $\alpha \leq 0.05$ considered significant. Statistical analyses were conducted using GraphPad Prism v8.0 and R v 3.4.4.

2.8.2 Worldwide quantification of DNA methylation using the S5 classifier

The main groups being compared were: HPV-negative cytology negative: HPV (-)/Cyt (-), HPV-positive cytology negative: HPV (+)/Cyt (-), CIN3 and cervical cancer including stage I (CSI), stage II (CSII), stage III (CSIII) and stage IV (CSIV) cancers as per FIGO stage classification. We hypothesised that uniformity in proportions and / or levels of methylation status among the different lesion categories may imply prognostic value as a clinical marker.

We validated the performance sensitivity of the S5 classifier CIN3 and cervical cancer samples from the global population samples in this study. We used the mean of methylation scores for selected *EPB41L3* CpG sites (438, 427, 425), HPV16L1 (6367, 6389), HPV18L2 (4256, 4261, 4265, 4269, 4275, 4282), HPV31L1 (6352, 6367) and HPV33L2 (5557, 5560, 5566) (Lorincz, 2014; Lorincz et al., 2013, 2016, Mirabello et al., 2012, 2013). Proportion of methylation on HPV16L2 was calculated using the following three CpG sites: 4275, 4259 and 4238 as they were the most reproducible sites in this region. The S5-methylation score was calculated by using the following weighted average defined in 2016 by Lorincz A *et al.* (Lorincz et al., 2016).

We compared differences in methylation levels between countries and between groups using Mann Whitney and Dunn's multiple comparison tests and the Cuzick test for trend to assess any methylation trend with disease progression.

There are three main metrics by which the accuracy of a methylation biomarker is assessed. Firstly, the AUC is the measure of how well a test or combination of tests can distinguish between two diagnostic groups (diseased/normal); the higher the AUC, the higher the diagnostic potential of the test. The sensitivity or the rate of true positives measures the proportion of actual positives that are correctly identified as such by the assay. The specificity or the rate of true negatives measures the proportion of actual negatives that are correctly identified as such by the assay.

The performance of the methylation test in different backgrounds was measured by receiver operating characteristic (ROC) analysis, by comparing the AUC. Specificity and sensitivity were also calculated for CIN3 and cancer endpoints. Diagnostic category methylation differences were studied with respect to the standard S5 cut-off of 0.8. In addition, exploratory new methylation

test cut-off points were calculated using Youden-J index and as described by Hernandes-Lopez R *et al*, 2019 (Hernández-lópez et al., 2019). McNemar's test with continuity correction was used for differences in sensitivity and specificity.

We also used unconditional logistic regression to study the relationship between methylation in the invasive cervical cancer group and the covariates - stage of cancer, type of cancer, age, demographics and HPV status. Additionally, we calculated the odds ratios (ORs) for the associations between HPV16/18 positivity, S5 classifier sensitivity at different cut-offs and CIN3 or cervical cancer (CSI-IV, FIGO stage unknown included) diagnosis in a selected HPV positive group of women with negative biopsies by histology (presumed healthy women who were hrHPV positive). All p-values were two sided with $\alpha \leq 0.05$ considered significant. Analysis was performed with GraphPad Prism v8.0 as well as R v 3.4.1 for Cuzick tests, ORs and for unconditional logistic regression analyses.

2.8.3 Investigation of the cervicovaginal microbiota

Subjects were analysed in 3 groups according their hrHPV status and disease severity: i) group A: hrHPV-negative controls (n=14); ii) group B: hrHPV persistent controls (n=15) and iii) group C: hrHPV positive women who developed CIN3 and were treated (n=30). We included women with hrHPV changes (from CIN3: C2 to post-LEEP: C3). We compared hrHPV positive versus hrHPV negative women, irrespective of disease status. We compared histology negative hrHPV positive or negative with CIN3 and post-treatment time points.

Taxonomic α-diversity was estimated as the number of observed OTUs, non-parametric Shannon and Inverse Simspon indeces. Differences in the observed OTUs, non-parametric Shannon, and Inverse Simspon indeces (α-diversity metrics) according to hrHPV and CIN status were tested by the Wilcoxon rank-sum test. Data was subject to multivariate analysis. Principal coordinates analysis (PCoA) of Bray-Curtis indices (β-diversity metrics) were used to visualise differences in microbial community structure according to CST distribution using Bray-Curtis dissimilarity as the distance measure. Beray-Curtis dissimilarity is 0 for identical samples and 1 for communities that do not share any species. Hierarchical clustering (HCA) using Euclidian centroid clustering of the most abundant 37 species was performed by the nearest neighbour linkage with a clustering density threshold of 0.75. P-values and Q-values calculated using linear regression. P-value was calculated using Fishers exact test, Q value was calculated using Benjamini-Hochberg false discovery rate (FDR) method.

Although differences in measures of diversity provide evidence that microbial communities are different, it is possible for measures of diversity to be similar while the composition of the microbial communities is different. Linear discriminate analysis effect size (LEfSe) analysis was conducted to find significant differences the relative abundance taxa within and between the three groups investigated, using the default values of α=0.05 and LDA=2.0 (Segata et al., 2011). Hierarchical clustering of the CST per sample was performed using the 'vegan' package in R. The S5 Score was calculated as described previously in section 2.2 and correlations with bacterial abundancies were made in R. All p-values were two sided with $\alpha \leq 0.05$ considered significant and adjusted for false discovery rate (Benjamin & Hochberg). Analysis was performed with GraphPad Prism v8.3 and with the 'vegan' package in R v 3.4.4.

2.8.4 Associations between microbial abundances and the S5 classifier score

The null hypothesis for this experiment was: 'there is no monotonic associations between microbial abundances and the S5 classifier score'. To identify any relevant associations, a linear regression analysis was performed. Correlation analysis explores the association between two or more variables. We chose to perform Spearman's correlation analysis due to the ordinal nature of the data in the study i.e. there is no known linear relationship between the variables analysed (microbial abundance and the S5 classifier score). Spearman's correlation is a nonparametric measurement of the strength and direction of monotonic association between two variables. This examined the relationship between the microbial abundances in the form of operational taxonomic units (OTUs) and the S5 classifier score distribution in the sample group. The OTU data from both bacterial genuses and species was considered for the analysis. Bacterial genuses and species were subsequently filtered for the highest OTUs in ≥25 samples in the study. Only correlation coefficients of ρ (rho) = +/- 0.5 threshold were considered significant. These indicate the threshold for strong correlations. Both positive (i.e. any rho > 0.5) and negative correlation (i.e. rho < -0.5) were taken into account. BAM files with bacterial OTU matrices were uploaded and analysed in Python v3.8.

CHAPTER 3 - BISULFITE CONVERSION KIT DETERMINATION FOR FORMALIN-FIXED PARAFFIN-EMBEDDED MATERIAL

3.1 INTRODUCTION

Epigenetic DNA methylation is an essential process for gene expression, chromosome stability, cell differentiation and embryonic development (Snijders et al., 2006; Steenbergen et al., 2014). DNA methylation generally occurs on the C_5 position of cytosines within CpG regions in the genome. Aberrant DNA methylation has been linked with loss of DNA homeostasis and genomic instability, leading to the development of various diseases including cancer (Feinberg et al., 2016; Plass et al., 2013). Around 60–80% of CpGs are constitutively methylated in somatic cells, however CpG rich sequences (CpG islands) are often un-methylated (Moody & Laimins, 2010). The biological importance of variations in DNA methylation creates an urgent demand for effective methods with high sensitivity and reliability to explore innovative diagnostic and therapeutic strategies.

One of the most accurate methods to measure DNA methylation is bisulfite conversion of unmethylated cytosines to thymidines followed by PCR amplification and pyrosequencing (Bock et al., 2016). During sodium bisulfite treatment, unmethylated cytosines are converted to uracils through a deamination reaction, while 5-methylcytosine (5-mC) are protected and remain unchanged. This enables downstream PCR amplification to recognize uracils as thymines and 5-mC as cytosines thereby discriminating between the methylated and unmethylated bases(Leontiou et al., 2015). This results in the conversion of the usually undetectable epigenetic information into detectable altered sequence information at base pair resolution (Figure 3.1).

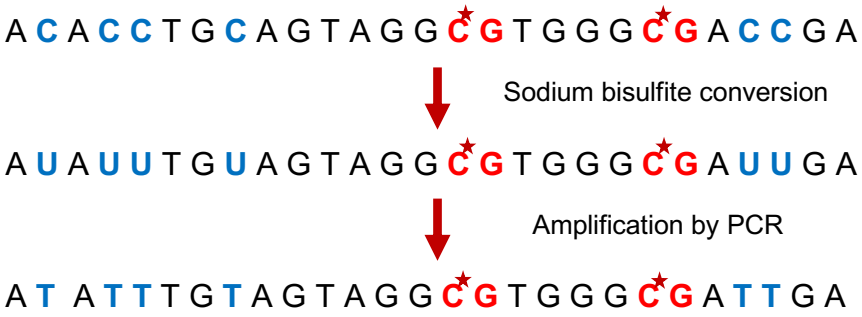

Figure 3.1 – DNA bisulfite conversion reaction outline. The un-methylated cytosines are represented in blue. The red stars represent DNA methylation events on cytosines within CpG islands (red). Abbreviations: A – adenine; T – thymine, G – guanine, C – cytosine.

One of the caveats of bisulfite conversion is the aggressive chemical environment required for the deamination reaction, which can lead to a high extent of DNA degradation. Conversely, less aggressive treatments may lead to a lower conversion rate of the un-menthylated cytosines. Elevated levels of unconverted 5-mC are associated with artificially increased methylation levels on the pyrosequencing assay (Leontiou et al., 2015). Therefore, a highly efficient bisulfite conversion reaction is required to ensure an accurate estimation of DNA methylation events in the pyrosequencing assay by keeping background noise to a minimum (Bock et al., 2016).

In cervical cancer, methylation biomarkers are promising diagnostic and prognostic tools that could be used for disease prevention in clinical settings (Clarke et al., 2012; Lorincz et al., 2016; Mirabello et al., 2012). Genome-wide methylation profiling studies have found about 10 genes with consistently elevated methylation levels in cervical pre-cancers and cancers (Clarke et al., 2017), including erythrocyte membrane protein band 4.1 like 3 *(EPB41L3)* (Cuschieri et al., 2018). Hyper-methylation on the *EPB41L3* promoter was associated with cervical disease progression and is demonstrated to distinguish between low and high grade cervical intraepithelial neoplasia (<CIN2 and CIN2+ respectively) (Eijsink et al., 2011; Vasilievic et al., 2014). Methylation analysis of selected CpG sites on *EPB41L3* as well as on the viral late genes of the high-risk human papilloma virus (HPV) strains HPV16, HPV18, HPV31 and HPV33 are part of the S5 DNA methylation classifier. The S5 classifier is an assay validated for detection of cancer and pre-cancerous lesions (CIN2+) from fresh or frozen liquid based-cytology (LBC) samples (Lorincz et al., 2013, 2016; Nedjai et al., 2018). Briefly, the process consists of bisulfite conversion, PCR amplification and pyrosequencing to yield the S5 score. The bisulfite conversion step in the S5 test is performed with the Zymo EZ DNA Methylation kit as standard and shows median unmethylated cytosine to uracil conversion rates of approximately 95% which is generally regarded as acceptable for valid DNA methylation measurement (Lorincz et al., 2016; Nedjai et al., 2018; Torres-Ibarra et al., 2019).

Good biomarkers should perform well in different sample types such as LBC, fresh, frozen or formalin-fixed paraffin-embedded (FFPE) biopsies but this is not always the case. Best results are usually obtained with fresh DNA that is not treated with a fixative. In contrast, DNA extracted from FFPE material is frequently of lower quality and concentration than from fresh or frozen material (Holmes et al., 2014). Tissue fixation in buffered formalin can cause DNA fragmentation, reducing the amount of amplifiable templates for PCR amplification (Holmes et al., 2014). Further sequence artefacts such as inter-strand DNA or protein-DNA crosslinks, deamination of cytosines and conformational change of the DNA can also occur in DNA extracted from FFPE

samples decreasing the bisulfite conversion efficiency and possibly leading to less accurate results (Tournier et al., 2012). Although DNA methylation assays performed on FFPE samples usually have a lower success rate than on fresh or frozen material (Holmes et al., 2014), they can still provide adequate information for epidemiological studies and to inform clinical decisions.

Here, we aim to identify the bisulfite conversion kit with the highest reaction efficiency for the treatment of FFPE material for use in the S5 classifier methylation assays. We extracted DNA from both FFPE and LBC samples to measure the conversion efficiency of four commercially available bisulfite conversion kits. The DNA extraction protocol for both FFPE and LBC samples is described in Chapter 2. An in-house real-time PCR methylation-specific assay targeted a region of β-actin gene containing no cytosines within the context of CpG sites (no 3' adjacent guanine) and we also measured the conversion efficiency on additional non-CpG cytosines in a control gene *DPYS* (Vasilievic et al., 2014). Less than 100% conversion at non-CpG cytosines indicates an incomplete conversion reaction and allows us to estimate the bisulfite conversion efficiency.

We compared Zymo EZ DNA Methylation kit (Zymo Standard), the Zymo EZ DNA Methylation-Lightning kit (Zymo Lightning), the Qiagen Epitect Bisulfite kit (Epitect Standard) and the Qiagen Epitect Fast Bisulfite kit (Epitect Fast). The Zymo Standard and Epitect Standard kits were chosen due to their less aggressive chemical environment and long incubation times at high temperatures required for full cytosine denaturation, while Epitect Fast and Zymo Lightning were chosen for their more aggressive chemical treatment and short incubation times which may lead to a reduced level of DNA degradation. Further, we tested the kits' ability to differentiate <CIN2 from CIN2+ using *EPB41L3* methylation assay. We discuss differences between the two sample types and the four bisulfite conversion kits in terms of conversion efficiency with the β-actin and *DPYS* assays and compare to methylation levels of CpG sites on *EPB41L3.* Finally, we validated a bisulfite conversion kit for FFPE treatment by assessing the performance of the S5 classifier in a set of paired LBC and FFPE material collected from the same women.

3.2 INVESTIGATION OF BISULFITE CONVERSION RATES IN LIQUID-BASED CYTOLOGY AND FORMALIN-FIXED PARAFFIN-EMBEDDED SAMPLES

The bisulfite conversion rate was investigated in 216 randomly selected LBC samples. The study included samples from women diagnosed with negative histology (n=62), cervical intraepithelial neoplasia grade 3 (CIN3, n=40) and squamous cell carcinoma (SCC, n=114). The LBC material originated equally (11.11% each) from Colombia, India, Ethiopia, Georgia, Philippines, South Africa, Spain, the United Kingdom and USA-New Mexico. The LBC samples were bisulfite converted with the validated Zymo EZ DNA Methylation kit, a protocol with which our laboratory have extensive experience of good results. The bisulfite conversion rate was measured using an in-house real-time PCR methylation-specific assay targeting a region of β-actin gene containing no CpG sites. Median conversion rates were above 95% for LBC samples from all countries, ranging from 96.42% (95%CI 96.15 – 97.20) to 98.11% ((5%CI 97.49 – 98.60). The overall median non-CpG cytosine conversion rate was 97.69% (95%CI 97.54 – 97.82).

To discover the best conversion rate of FFPE samples, four bisulfite conversion kits were tested on DNA extracted from 43 FFPE biopsies. These included material from women diagnosed with histology negative (n=12), CIN1 (n=12), CIN2 (n=9), CIN3 (n=9) and SCC (n=1). All FFPE material originated from the UK. The bisulfite conversion kits investigated were: Zymo Standard, Zymo Lightning, Epitect Standard and Epitect Fast. The conversion rate of all kits was measured by the in-house PCR assay, as previously described in Chapter 2 of this thesis. A median bisulfite conversion rate of 95.93% (95%CI 91.74 – 97.68) was observed for Epitect Standard, 94.30% (95%CI 86.24 – 97.81) for Epitect Fast, 72.85% (95%CI 67.15 – 77.72) for Zymo Standard and 92.94% (95%CI 87.70 – 98.16) for Zymo Lightning. The conversion rate of FFPE material using Zymo Standard was significantly lower than using any other kit tested (Mann Whitney U test, all $p<0.0001$) and was also much lower for FFPE compared to the overall conversion rate of LBC samples using the Zymo Standard kit (Mann Whitney U test, $p<0.0001$), therefore it appears that formalin-fixation of DNA substantially reduces apparent cytosine conversion efficiency with the Zymo Standard kit as measured by the in-house PCR assay. No other significant difference in conversion rates was observed (LBC material treated with Zymo Standard vs FFPE-EpiTect Fast, FFPE-EpiTect Fast, FFPE-Zymo Lightning) (Figure 3.2 A).

Figure 3.2 B shows the genomic quality score of the bisulfte converted DNA from LBC and FFPE material, assessed on the LabChip® GX scale of 0 to 5, with 0 representing extremely low-quality DNA and 5 indicating high quality DNA. The mean quality score of the bisulfite converted LBC-DNA (4.18, 95%CI 4.05–4.31) was higher than the bisulfite converted FFPE-DNA across all kits investigated (2.87, 95%CI2.44-3.30 for Zymo Standard; 3.31, 95%CI3.04-3.59 for Epitect Standard; 3.47, 95%CI3.19-3.76 for Epitect Fast; 3.00 95%CI2.55-3.47 for Zymo Lightning; all p<0.001). No significant differences were observed in terms of quality scores among the kits used for FFPE-DNA bisulfite conversion.

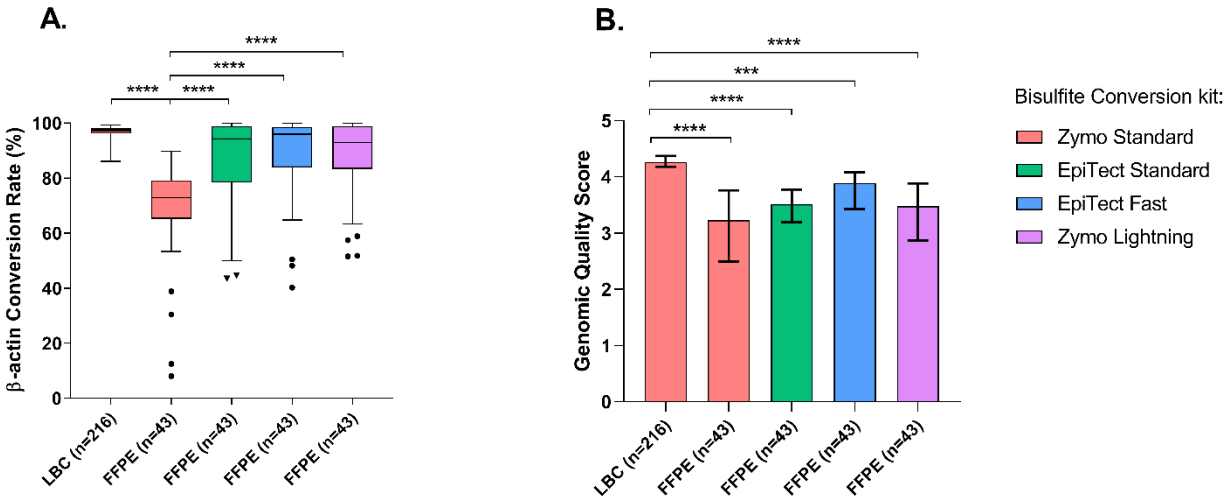

Figure 3.2 – **Bisulfite conversion kit performance and genomic quality score in LBC and FFPE samples**. (**A**) Shows a comparison between conversion efficiency rates. The Epitect Standard (green), Epitect Fast (blue) and Zymo Lightning (pink) kits performed well and their rates of conversion were not significantly different from LBC material. The Zymo Standard kit (light red) showed a lower conversion rate than the LBC samples and the other kits investigated. The outliers are plotted as individual points for each sample type. Statistical difference between samples was analysed by Mann Whitney U test.; (**B**) shows the comparison of genomic quality score of LBC and FFPE samples after bisulfite conversion. LBC material treated with Zymo Standard showed higher DNA quality than FFPE materials treated with any of the kits. The error bars represent 95% confidence intervals of the median genomic scores. Statistical difference between samples was analysed by Kruskall Wallis and Dunn's multiple comparisons tests. Only significant comparisons were shown as following: ***P<0.001, ****P<0.0001.

3.3 Alternative measurement of bisulfite conversion rate using non-CpG DPYS pyrosequencing assay

The conversion of cytosines at non-CpG sites on *DPYS* was analysed to assess conversion efficiency rate. The assay involved bisulfite conversion, followed by PCR amplification and pyrosequencing. The assay contains a cytosine not followed by guanine (non-CpG site) that should be fully converted during the bisulfite reaction. Failure to convert at non-CpG sites indicates an incomplete conversion reaction. The non-CpG *DPYS* methylation assay was performed twice and showed good reproducibility rates (Spearman r ranged from 0.574 to 0.778, $p<0.05$).

Mean conversion failure rate for at the non-CpG site in *DPYS* for LBC <CIN2 was 1.54% (95%CI 1.23 – 1.81) and for LBC CIN2+ were 1.51% (95%CI 1.31 – 1.98). No significant difference in non-CpG site in *DPYS* methylation was observed between either LBC <CIN2 or CIN2+ (Mann Whitney U test, $p>0.05$) (Figure 3.3A).

The non-CpG *DPYS* conversion rate was calculated by subtracting the conversion failure rated from 100. Overall, the conversion rate from the non-CpG *DPYS* assay was 98.67% (95%CI 98.53-99.00) for all LBC material bisulfite converted with the Zymo Standard kit**.** For FFPE material, the conversion rates were 98.19% (95%CI 97.68 – 99.05) for treatment with EpiTect Standard kit, 98.74% (95%CI 98.53-99.68) with EpiTect Fast kit, 97.23% (95%CI 96.25-98.42) with Zymo Standard kit and 98.57% (95%CI 98.21-99.09) with Zymo Lightning kit. Similar to what was observed for the in-house PCR conversion studies, the non-CpG *DPYS* conversion rate for FFPE material treated with the Zymo Standard kit was significantly lower than compared to treatment with the EpiTect Fast kit (Mann Whitney U test, $p=0.002$) or compared to data from the LBC samples converted with the Zymo Standard (Mann Whitney U test, $p=0.033$) kit (Figure 3.2). In conclusion, the conversion of LBC samples with Zymo Standard was comparable to the conversion of FFPE samples with Epitect Fast.

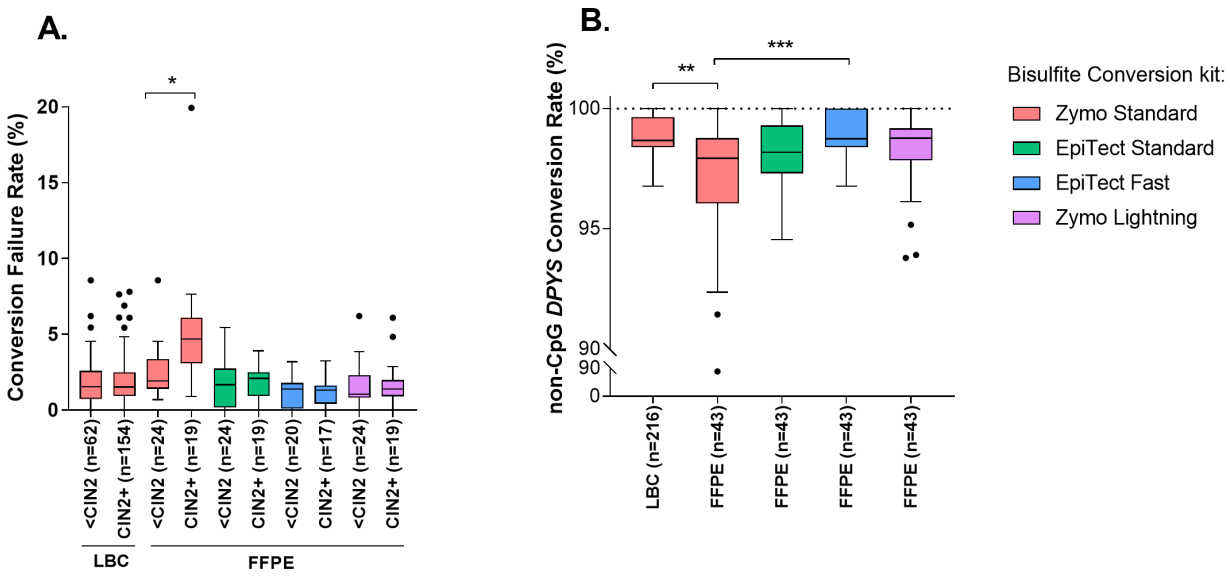

Figure 3.3 – non-CpG *DPYS* methylation assay performance; (**A**) Comparison of mean non-CpG methylation values in LBC and FFPE material as per bisulfite conversion kit; (**B**) Comparison of mean non-CpG conversion rate values in LBC and FFPE material as per bisulfite conversion kit. Comparisons were made between LBC treated with Zymo Standard and FFPE treated with all four bisulfite conversion kits investigated. The outliers are plotted as individual points for each sample type. Statistical difference between samples was analysed by Mann Whitney U test. Only significant comparisons were shown as following: *P<0.05 **P<0.01, and ***P<0.001.

3.4 *EPB41L3* METHYLATION ASSAY: DIFFERENTIATION BETWEEN <CIN2 AND CIN2+

To assess the performance of differentiating <CIN2 from CIN2+, *EPB41L3* methylation assays were tested using bisulfite converted LBC and FFPE material. The samples set was randomly selected to represent <CIN2 and CIN2+ from both LBC and FFPE sample types. The LBC set consisted of 216 LBC samples organised in: 62 <CIN2 (62 negative histology women) and 154 CIN2+ (40 CIN3 and 114 SCC) bisulfite converted with Zymo Standard. The FFPE set consisted of 43 FFPE biopsies organised in: 24 <CIN2 (12 NEG and 12 CIN1) and 19 CIN2+ samples (9 CIN2, 9 CIN3 and 1 SCC) bisulfite converted with the four bisulfite conversion kits. The *EPB41L3* methylation assay was performed twice and showed high reproducibility rates for all bisulfite conversion kits used (Spearman r ranged from 0.619 to 0.918, p<0.0001).

The *EPB41L3* methylation assay could differentiate well between LBC <CIN2 and CIN2+ samples converted using the Zymo Standard (Mann Whitney U test, p<0.0001). In contrast, differentiation between FFPE <CIN2 and CIN2+ was not statistically significant when the FFPE material was converted with Zymo Standard (13.17%, 95%CI 8.10-17.90 for <CIN2 and 15.18%, 95%CI 13.20-30.41 for CIN2+; p=0.0582) or with Zymo Lightning (12.30%, 95%CI 6.62-17.10 for <CIN2 and 14.30%, 95%CI 8.11-26.96 for CIN2+, p=0.2216). *EPB41L3* methylation showed significant differentiation of FFPE CIN2+, when bisulfite conversion was performed with either EpiTect Standard (6.65%, 95%CI 4.50-9.69 for <CIN2 and 12.68%, 95%CI 5.21-26.56 for CIN2+; p=0.0313) or EpiTect Fast (6.65%, 95%CI 4.10-9.50 for <CIN2 and 8.30%, 95%CI 5.80-26.20 for CIN2+; p=0.0427) (Figure 3.4 A).

The different receiver operating characteristic (ROC) curves in Figure 3.4 B highlight the performance in differentiating between <CIN2 and CIN2+ based on *EPB41L3* methylation. A high area under the curve (AUC) indicated a high differentiating power between two disease categories. As expected, the highest AUC observed was for the LBC-Zymo Standard (AUC: 0.80, 95%CI 0.74-0.85, p<0.0001). The AUC results for FFPE-Zymo Standard and FFPE-Zymo Lightning were not statistically significant (p=0.053 and p=0.093, respectively), indicating no significant differentiation between <CIN2 and CIN2+. The AUCs for FFPE DNA converted with EpiTect Standard (0.71, 95%CI 0.56-0.87, p=0.014) was significantly lower than the AUC for LBC-Zymo Standard (DeLong's test, p=0.04). FFPE-EpiTect Fast showed similar performance in differentiating between the two disease groups as LBC-Zymo Standard. The AUC for FFPE samples converted with the EpiTect Fast kit was the highest out of the four kits investigated: 0.76 (95%CI 0.61-0.91, p=0.005) and showed no difference compared to LBC samples converted with the Zymo Standard kit (DeLong's tests, p=0.126).

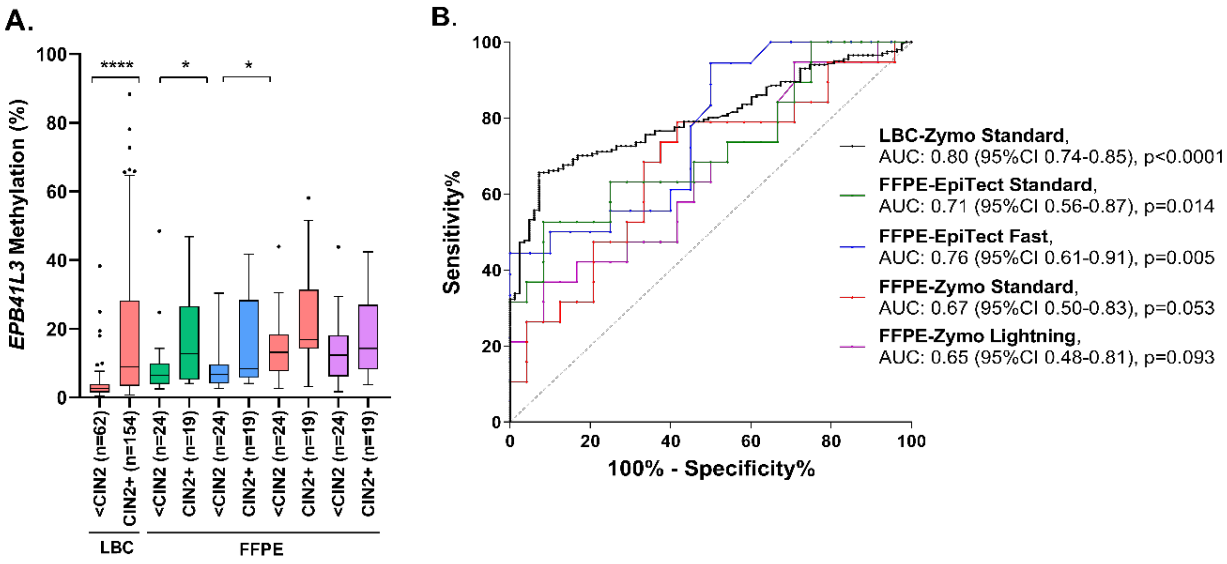

Figure 3.4 – *EPB41L3* methylation assay performance; (**A**) Comparison of mean EPB41L3 methylation values in LBC and FFPE material as per bisulfite conversion kit. Good <CIN2 from CIN2+ differentiation was observed in the LBC-Zymo Standard (light red), FFPE-EpiTect Standard (green) and FFPE-EpiTect Fast (blue) groups. The outliers are plotted as individual points for each sample type. Statistical difference between samples was analysed by Mann Whitney U test. Only significant comparisons were shown as following: *P<0.05 and ****P<0.0001; (**B**) shows the receiver operating characteristic (ROC) curves on <CIN2 vs CIN2+ differentiation based on *EPB41L3* methylation. The AUC of FFPE-Zymo Standard and FFPE-Zymo Lightning were not statistically significant (p>0.05) showing lack of differentiation between <CIN2 and CIN2+. The AUC of FFPE-EpiTect Fast was not significantly different from the AUC on the reference LBC-Zymo Standard (DeLong's tests). AUC: area under the ROC curve, CI: confidence interval.

3.5 COMPARISON OF S5 METHYLATION SCORES IN FFPE MATERIAL TREATED WITH EPITECT FAST AND ZYMO STANDARD BISULFITE CONVERSION KITS

As the Epitect Fast data has shown consistently the best conversion results it was decided to go ahead and compare it with the S5 validated Zymo Standard kit. To investigate the performance of S5 classifier on DNA extracted from LBC versus FFPE, we tested a set of paired LBC (n=72) and one of FFPE (n=72) material collected from the same women from UK, Bhutan and USA (i.e. one LBC and one FFPE sample per woman). Samples were acquired at only one timepoint. All women were previously diagnosed with SCC. We used the S5-validated Zymo Standard kit and the Epitect Fast kit to perform bisulfite conversion on LBC and FFPE DNA, for comparison.

The median bisulfite conversion rate for LBC DNA is consistent, regardless of the bisulfite conversion kit used (Zymo Standard: 97.30%, 95%CI 97.14 – 98.01; and Epitect Fast: 97.22, 95%CI 97.10 – 97.92). However, the same consistency could not be replicated in the DNA derived and converted from the FFPE samples. The median conversion rate for FFPE material using the Zymo Standard kit (98.07%, 95%CI 97.87 – 98.11) was significantly lower than by using Epitect Fast (Zymo Standard: 76.16%, 95%CI 70.25-81.20; and Epitect Fast: 98.07%, 95%CI 97.87 – 98.11; Wilcoxon matched-pairs signed rank test, $p<0.0001$). Bisulfite conversion rates of LBC-Zymo Standard were consistent with the bisulfite conversion rates of FFPE-Epitect Fast (Wilcoxon matched-pairs signed rank test, $p=0.241$) as shown in Figure 3.5 A**.**

Figure 3.5 B shows the S5 performance in LBC and FFPE material. The median S5 methylation score from FFPE material treated with Zymo Standard was 20.80 (95%CI 18.23 – 25.05) and was observed to be higher than the S5 scores of the paired LBC samples treated with the same bisulfite conversion kit (16.16, 95%CI 10.74 – 21.58; Wilcoxon matched-pairs signed rank test, $p<0.0001$). However, there was a great deal of similarity between FFPE material treated with EpiTect Fast and the paired LBC material treated with the Zymo Standard (Wilcoxon matched-pairs signed rank test, $p=0.426$).

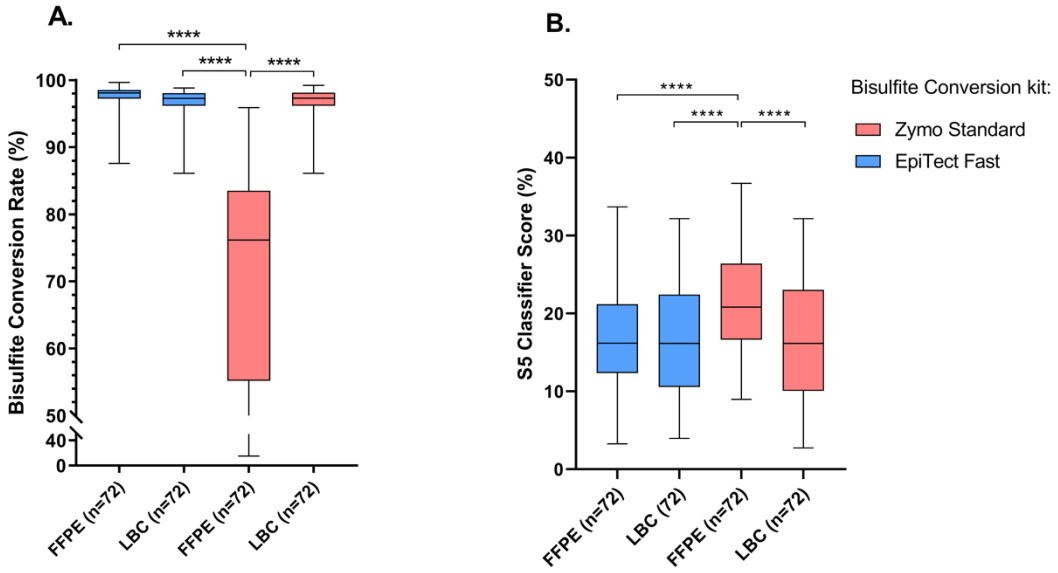

Figure 3.5 - **Performance of Zymo Standard and Epitect Fast kits in LBC and FFPE samples**. Assay has been performed in a paired set of LBC and FFPE material collected from the same women; (**A**) Shows the conversion efficiency rate of LBC and FFPE material treated with both Zymo Standard (orange) and EpiTect Fast (blue). No significant difference was observed between LBC-Zymo Standard and FFPE-EpiTect Fast; (**B**) shows the S5 Classifier scores for LBC and FFPE material treated with both Zymo Standard and EpiTect Fast. No significant difference was observed between LBC-Zymo Standard and FFPE-EpiTect Fast. Statistical differences between samples were analysed by Wilcoxon matched-pairs signed rank test. Only significant comparisons were shown as following: ****P<0.0001.

3.6 DISCUSSION

Aggressive bisulfite treatment protocols (long incubation, high temperatures and harsh conversion reagents) assure a maximum conversion of cytosines to uracils, however, at the cost of the DNA integrity. Conversely, less aggressive treatments bear the risk of incomplete conversion and can lead to overestimation of the resulting methylation levels (Holmes et al., 2014). One of the most important parameters of bisulfite conversions is the success rate or efficiency of the reaction. It is crucial that the yield and integrity of recovered DNA post treatment is sufficiently high for downstream analyses (Leontiou et al., 2015). The incomplete conversion during bisulfite treatment is another important factor and should be considered carefully as artificial results could arise. Here, we investigated four different bisulfite conversion kits on the FFPE material to select the most appropriate one for this type of sample. The kits chosen were: the Zymo EZ DNA Methylation kit (Zymo Standard) which was the reference kit since it was used to develop the S5 classifier, the Zymo EZ DNA Methylation-Lightning (Zymo Lightning), the Qiagen Epitect Bisulfite Kit (Epitect Standard) and the Qiagen Epitect Fast Bisulfite Kit (Epitect Fast).

Although the Zymo Standard was the bisulfite conversion kit used in the development of the S5 classifier in LBC samples, it showed poor efficiency on FFPE material bisulfite conversion (97% on LBC vs 73% on FFPE). The other kits (Epitect Standard, Epitect Fast and Zymo Lightning) had a conversion rate of FFPE similar to the one in LBC samples treated with Zymo Standard (95%, 94% and 93% respectively).

DNA methylation occurs generally only on CpG sites. The DPYS assay was used as another mean to estimate the bisulfite conversion efficiency. The assay contains a cytosine not followed by guanine (non-CpG site) that should be fully converted during the bisulfite reaction. Failure to convert at non-CpG sites indicates an incomplete reaction and allows for an estimation of the efficiency of the kit: 98.19% EpiTect Standard kit, 98.74% for EpiTect Fast, 97.23% Zymo Standard and 98.57% for Zymo Lightning kit. These estimates correlate with previous studies which investigated non-CpG conversion rates found a conversion efficiency of 98.6% for Epitect Standard[5] and 99.2% for EpiTect Fast kit (Holmes et al., 2014).

Conversion failure rates assessed with the non-CpG site in *DPYS* assay were higher for FFPE bisulfite conversion with Zymo Standard kit than with Epitect Fast ($p=0.0017$). No significant difference was observed in the assay. Artificially high methylation levels are expected in FFPE

material due to inter-strand, protein-DNA and histone-DNA crosslinks caused by the formaldehyde fixation (Do & Dobrovic, 2015). This causes incomplete DNA denaturation during bisulfite conversion that could vary by genomic location. Although the alternative measurement of conversion rate using the non-CpG DPYS assay involves pyrosequencing of the bisulfite converted material, compared to the β-actin assay, the non-CpG DPYS assay is limited by not including more than one non-CpG sites for deeper accurate analysis.

Hyper-methylation of *EPB41L3* has previously been shown to be a biomarker for cervical disease progression. We showed that *EPB41L3* methylation analysis performed well at discriminating normal and low-grade lesions (<CIN2) from high grade CIN (CIN2+) on FFPE material bisulfite converted with only with Epitect Standard and Epitect Fast kits. Methylation was higher in CIN2+ than in <CIN2 samples ($p<0.05$). No significant differences in methylation between the two groups was observed using either Zymo Standard of Zymo Lightning for FFPE bisulfite conversion. The reproducibility of the *EPB41L3* methylation assay on FFPE material converted with Epitect Standard and Epitect Fast was good as indicated by a very high Spearman r values (0.867 for Epitect Standard and 0.918 for Epitect Fast, $p<0.0001$). This is important since it indicates that the assay is technically robust for both kits used.

Performance of the ROC curve shows the ability of a biomarker to separate two groups. The only significant AUCs for *EPB41L3* methylation differentiation of <CIN2 and CIN2+ on the FFPE biopsies were 0.71 with the Epitect Standard and 0.76 with the Epitect Fast kit. The AUC obtained from the LBC samples converted with the Zymo Standard kit was slightly higher: 0.81. The data showed that there was similar performance in the methylation assay when LBC samples were treated with Zymo Standard kit while FFPE material was treated with either Epitect Standard or Epitect Fast kits (both not significant, DeLong's test).

In conclusion, although the Zymo Standard was the reference bisulfite conversion kit for the development and validation of the S5 classifier on LBC material, bisulfite conversion efficiency of FFPE material using this kit was poor. Consistently high rates of the bisulfite conversion reaction and uniform performance in the S5 methylation assay can be observed using the Zymo Standard kit for bisulfite conversion of LBC samples, whilst the Epitect Fast was shown to be more consistent when using DNA derived from FFPE material. The Epitect Fast kit has been selected to bisulfite convert FFPE material for use in other experiments described in Chapter 4 of this thesis.

CHAPTER 4 –CLINICAL PERFORMANCE METHYLATION AS A BIOMARKER FOR CERVICAL CARCINOMA-IN-SITU AND CANCER DIAGNOSIS

4.1 INTRODUCTION

Persistent infection with high risk human papilloma virus (hrHPV) is a necessary, but not sufficient factor for development of cervical cancer and its precursor; cervical intra-epithelial neoplasia (CIN). However, most hrHPV infections are transient and persistence beyond 2 years occurs in <10% of women (Doorbar et al., 2012). The implementation of cervical cancer screening programme by systematic cytology screening has contributed to a reduction in cervical cancer-associated deaths (Catarino et al, 2015). Yet, cervical cancer is currently the fourth most common cancer in women worldwide with 604,000 cases in 2020, accounting for 7.5% of all female cancer deaths (Sung et al., 2021). Cervical cancer incidence, ranges from 5 to 50 per 100,000 women depending on setting and while hrHPV testing is sensitive for the detection of disease, specificity is less optimal given the benign trajectory of most infections. To allow a further reduction in the incidence of cervical cancer, screening is shifted towards high risk human papilloma virus (hr-HPV) testing with triage of HPV positive women. Triage generally relies on cytology as an option for a secondary triage test (Cuschieri et al., 2018). Secondary triage tests are urgently needed to identify the minority of hrHPV positive women with histologically relevant high-grade disease (Cuschieri et al., 2018). Furthermore, triage tests that rely on molecular, rather than morphological signatures (such as cytology) remove the requirement for subjective assessment.

Methylation biomarkers can offer an accurate and cost-effective alternative for screening and identifying women who have the highest risk of progressing into invasive cervical cancer (Lorincz, 2011; Nedjai et al., 2018). Aberrant DNA methylation, has been reported to increase with cervical cancer disease progression (Lorincz, 2016), allowing this epigenetic event to be used as a temporal biomarker, with a potential to accurately predict whether hrHPV infection will lead to cervical intraepithelial neoplasia grade 2 or above (CIN2+) or disappear (Lorincz, 2011; Louvanto et al., 2019).

To date, a number of studies have shown that DNA methylation levels in cervical cancer samples increase with disease severity (Vink et al., 2015). In such manner, methylation biomarker tests including the S5 methylation classifier which tests for methylation on the host *EPB41L3* and viral late genes (L1 and L2) of HPV16, HPV18, HPV31 and HPV33, can accurately separate women with CIN2/3 and cancer from those with CIN1 or less (cytology negative) (Lorincz et al., 2016). The S5-classifier accurately measures DNA methylation by using bisulfite conversion of unmethylated cytosines followed by PCR amplification and pyrosequencing (Mirabello et al., 2013). For this test, the threshold methylation score i.e. cut-off value differentiating CIN2 from healthy patients depends on HPV load and screening availability. Due to systematic screening in the UK, the predefined threshold for S5 has been set at 0.80 in this country. Other countries might benefit from higher thresholds of the classifier. This is because in unscreened populations HPV prevalence is generally higher. A higher methylation cut-off would give a lower false positive diagnosis rate.

The S5-classifier has demonstrated improved triage performance compared to hrHPV genotyping, cytology or the combination thereof and has been validated in a HPV-positive cohort of women as part of the Canadian FOCAL, Mexican FRIDA and Colombian ASC-US COL clinical trials (Cook et al., 2018; Hernández-López et al., 2019; Ramirez et al., 2021). Additionally, the S5-classifier demonstrated a potential prognostic utility, in its ability to identify women with progressive CIN2 (Louvanto et al., 2019). Together, these data support the prospect of using the S5-classifier as a molecular tool to identify clinically significant cervical abnormalities and predicting their clinical course.

Validation of the S5-classifier in a large number of cancer samples from both high and low-middle income countries (LMICs) is required to demonstrate that this methylation test can consistently detect the vast majority of cervical cancers worldwide including non-squamous types. Extensive validation of the S5 classifier will support implementation of the test in global screening programmes and routine patient management clinics. The main aim of the present study is to analyse the performance and consistency of S5 in detecting high grade lesions and cervical cancers from diverse settings that reflect demographics in Asia, Europe, Africa and the Americas. This includes quantifying the degree of epigenetic separation between women who are cytologically negative versus those with high grade pre-cancer (CIN3) and cervical cancer, including histologies such as adenocarcinoma, adenosquamous carcinoma and other rare types of cervical cancer such as HPV-negative cancers and neuroendocrine small cell carcinoma. The second aim of the study is investigate how an increase in the classifier's cut-off affects the S5

specificity for CIN3 and cancer detection. The third aim is to analyse the individual S5-classifier components and their contribution to the performance of the test.

4.2 CHARACTERISTICS AND SELECTION CRITERIA

Here, we propose a cross-sectional retrospective study including 973 women from ten countries to evaluate the S5 methylation classifier performance to detect CIN3 and cervical cancer. We selected 220 HPV negative or positive women, all with negative cytology (HPV(-/+)/Cyt(-)), 204 women diagnosed with CIN3 and 544 women with invasive cervical cancer as described in Figure 1. Baseline characteristics of the women are presented in Table 1. All women diagnosed with CIN3 and invasive cervical cancer were part of previous studies. The invasive cervical cancer group wass further divided into the various histopathologies (hystotypes): squamous cell carcinomas (SCC, n = 510), adenocarcinomas (ADC, n = 29), adenosquamous cell carcinomas (ADS, n = 1) and neuroendocrine small cell carcinoma (SNEC, n = 4). The median age for women with negative cytology was 38 years (IQR, 30 – 47), for CIN3 was 31 years (IQR, 28 – 39) and for women with invasive cervical cancer was 45 years (IQR, 38 – 55).

4.3 HRHPV PREVALENCE IN THE SAMPLED GROUPS

The Papilloplex High Risk HPV genotyping data was in 98.23% (95%CI 96.38-99.99%) agreement with previous genotyping methodologies used. The grouped prevalence of 13 types of hrHPV plus HPV66 (now regarded as a low risk HPV) was 50.00% in the histology negative women, 95.09% in the CIN3 group and 95.21% in the cancer group (Table 4.1). Each type of hrHPV in women infected by multiple HPV types was counted as described in section 2.2.7. Figure 4.1 shows the hrHPV types prevalence stratified by histology diagnosis and country of origin of the sample set. A total of 71.67% of hrHPV infections in the study group were attributed to HPV16, HPV18, HPV31 and HPV33, the viral components of the S5 classifier. The most frequent hrHPV type infection in women was HPV16. Overall, the five most prevalent hrHPV types in the sample set were HPV16 (51.98%) followed by HPV18 (9.34%), HPV31 (6.19%), HPV45 (4.71%) and HPV33 (4.16%). Ethiopia and the Philippines showed a high prevalence of HPV45 and HPV59 and a low prevalence of HPV33. In the USA (New Mexico), the second most prevalent infection was with HPV39, while in the Spain sample set HPV52 infection was the most prevalent after HPV16.

A high level of diversity of hrHPV types was observed in HPV positive women with normal cytology HPV(+)/Cyt(-). Although HPV16 was the major infectious agent (31.81%), the majority of infections were not associated with HPV16. Interestingly, HPV51 had a high prevalence within the HPV(+)/Cyt(-) group (9.18%) but was less prevalent in CIN3 (3.09%) or cancer (1.25%) groups.

In the cancer groups, 374 / 544 women were positive for HPV16 (68.91%), 37 women for HPV18 (6.91%), 21 women for HPV31 (3.90%), 16 women for HPV33 (3.01%) and 68 women for other hrHPV types (12.41%). A total of 25 (4.78%) cervical cancers tested negative for any hrHPV type

The adenocarcinoma (ADC) subgroup of cancers showed a slightly different hrHPV profile than the majority of cancers analysed. Although HPV16 was generally the most prevalent viral infection in the cancer group, the ADC subgroup showed higher HPV18 infection rates (38.29%) than HPV16 (34.04%) as described in Figure 4.1. Additionally, higher prevalence of HPV52 (8.51%) and HPV59 (4.25%) were observed, compared to the general hrHPV prevalence distribution.

Characteristics		HPV(-)Cyt(-) n = 110 (%)	HPV(+)Cyt(-) n = 110 (%)	CIN3 n = 204 (%)	Cervical Cancer – FIGO stage			
					Stage I n = 245 (%)	Stage II n = 249 (%)	Stage III n = 28 (%)	Stage IV n = 22 (%)
Histotype of cervical cancer	Squamous cell carcinoma				230 (93.9)	236 (94.8)	26 (92.8)	19 (86.5)
	Adenocarcinoma				14 (5.7)	10 (4.0)	2 (7.2)	2 (9.0)
	Adenosquamous cell carcinoma				-	1 (0.4)	-	-
	Neuroendocrine carcinoma				1 (0.4)	2 (0.8)	-	1 (4.5)
HrHPV status*	Consistently Negative	110 (100)	-	10 (4.90)	14 (5.7)	10 (4.0)	2 (7.1)	-
	HPV16+	-	35 (31.8)	120 (58.8)	160 (65.3)	179 (72.1)	20 (71.4)	19 (86.4)
	HPV18+	-	14 (12.7)	7 (3.4)	19 (7.7)	15 (6.0)	1(3.5)	-
	HPV31+	-	6 (5.5)	30 (14.7)	8 (3.2)	9 (3.6)	2 (7.1)	1 (4.5)
	HPV33+	-	6 (5.5)	7 (3.4)	9 (3.6)	8 (3.2)	-	-
	Other hrHPV+	-	49 (44.5)	30 (14.7)	35 (14.3)	28 (10.8)	3 (10.7)	2 (9.1)
Sample type	Cervical Scrape	110 (100)	110 (100)	204 (100)	142 (58.0)	192 (77.0)	28 (100)	22 (100)
	FFPE Tissue	-	-	-	103 (42.0)	57 (22.9)	-	-
Age	<25	20 (18.18)	12 (10.9)	14 (6.9)	5 (2.0)	2 (0.8)	-	-
	25-29	27 (24.5)	25 (22.7)	59 (28.9)	27 (11.0)	10 (4.0)	-	1 (4.5)
	30-39	28 (25.4)	26 (23.6)	81 (39.7)	83 (33.8)	43 (17.3)	4 (14.2)	3 (13.6)
	40-49	21 (19.0)	23 (20.9)	30 (14.7)	56 (22.8)	84 (33.8)	8 (28.5)	6 (27.4)
	50-59	8 (7.2)	12 (10.9)	14 (6.9)	46 (18.7)	63 (25.4)	9 (32.1)	11 (50.0)
	>60	6 (5.5)	12 (10.9)	6 (2.9)	28 (11.4)	47 (18.5)	7 (25.0)	1 (4.5)
Country of Origin	Bhutan	-	10 (9.0)	-	28 (11.4)	22 (8.8)	-	-
	Colombia	16 (14.5)	4 (3.6)	50 (24.2)	21 (8.5)	25 (10.0)	-	-
	Ethiopia	39 (35.5)	10 (9.0)	-	2 (0.8)	18 (7.2)	28 (100)	22 (100)
	Georgia	-	-	-	40 (16.3)	2 (0.8)	-	-
	India	-	10 (9.0)	-	12 (4.9)	38 (15.3)	-	-
	Philippines	-	-	-	15 (6.1)	35 (14.1)	-	-
	South Africa	-	-	-	2 (0.8)	47 (18.9)	-	-
	Spain	16 (14.5)	4 (3.6)	50 (24.2)	28 (11.4)	22 (8.8)	-	-
	United Kingdom	25 (22.7)	36 (32.7)	54 (27.1)	48 (19.5)	3 (1.2)	-	-
	USA (New Mexico)	14 (12.7)	36 (32.7)	50 (24.2)	50 (20.4)	36 (14.5)	-	-

Table 4.1 - Baseline characteristics of the of the S5 validation study participants. All women were diagnosed as histology negative (healthy women), cervical intraepithelial neoplasia grade 3 (CIN3) and cervical cancer stages I-IV (highlighted).

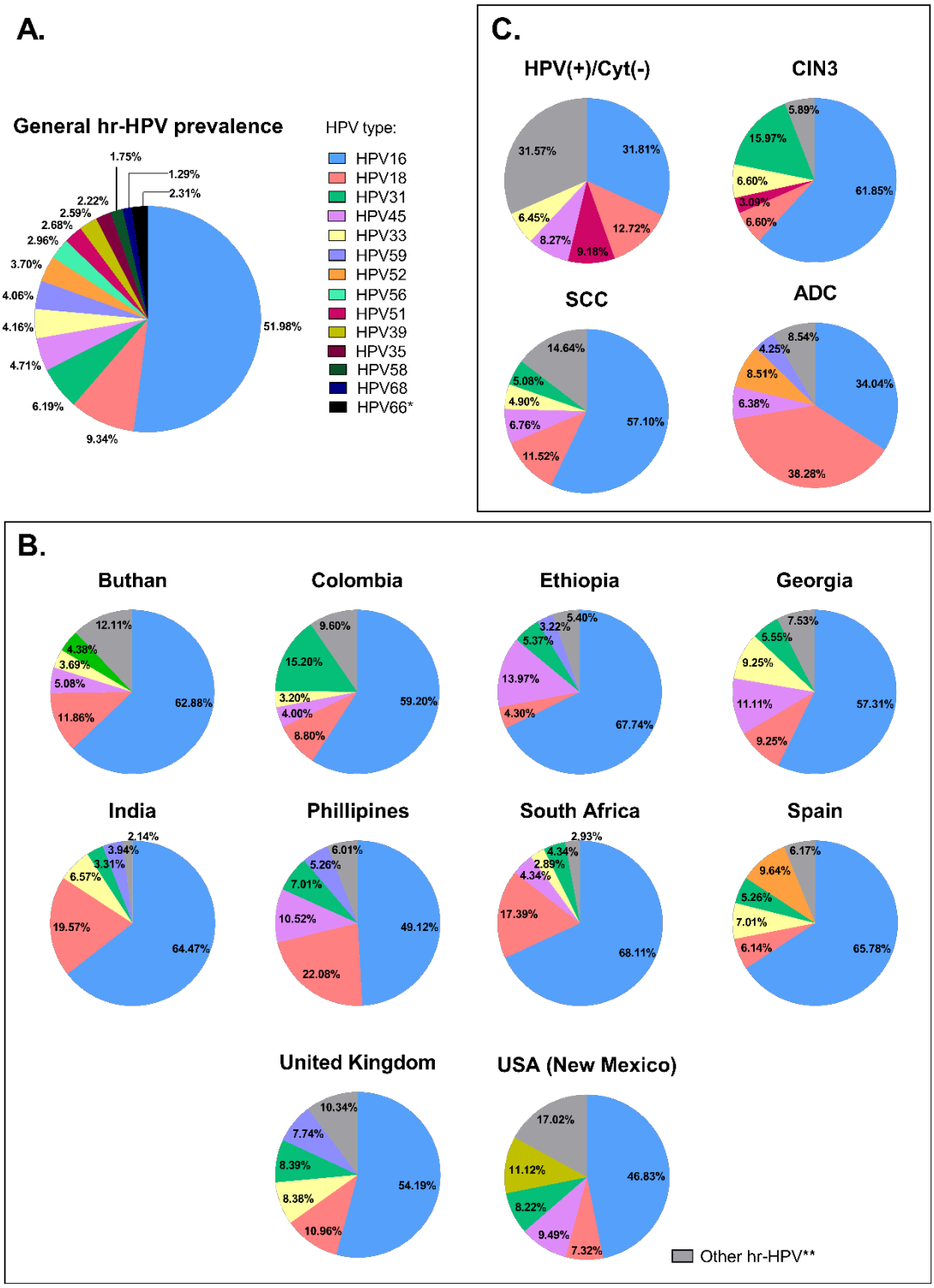

Figure 4.1 – Global and country-specific hrHPV prevalence. Each analysed group and subgroup investigates HPV positive (all 13 types of hrHPV and HPV66* now regarded as low-risk) only women. B and C show the percentages of the 5 most prevalent hrHPV type in women per subgroup, any other hrHPV types different than the top 5 were marked as Other hrHPV (**). In C, only the most abundant cervical cancer subgroups were considered: squamous cell carcinoma (SCC, n=510) and adenocarcinoma (ADC, n=29). The adenosquamous cell carcinomas (n = 1) and neuroendocrine small cell carcinoma (n = 4) were deliberately not included. Abbreviations: HPV(+)/Cyt(-), HPV positive with cytology negative results; CIN, cervical intraepithelial neoplasia (of grade 3); SCC, squamous cell carcinoma; ADC, adenocarcinoma.

4.4 INCREASING TREND IN THE S5 METHYLATION SCORES WITH DISEASE SEVERITY

The S5 methylation scores were clustered according to severity of cervical abnormality. An outline of the methylation scores per country is provided in Figure 4.2. Median methylation score was 0.66 (95%CI 0.60 – 0.78) in HPV(-)/Cyt(-), 0.91 (95%CI 0.86 – 0.94) in HPV(+)/Cyt (-), 5.64 (95%CI 3.99 – 6.88) in CIN3, 17.58 (95%CI 15.99 – 20.33) in cervical cancer stage I (CSI), 23.71 (95%CI 21.67 – 25.24) in cervical cancer stage II (CSII), 24.44 (95%CI 22.10 – 32.14) in cervical cancer stage III (CSIII) and 24.95 (95%CI 22.34 – 32.85) in cervical cancer stage IV (CSIV). The distribution of the S5 scores based on the disease diagnosis and hrHPV status is shown in Figure 4.3. Kruskal-Wallis and Dunn's multiple comparisons tests show significant separation in the following paired comparisons: HPV(-)/Cyt(-) vs CIN3 ($p<0.0001$), HPV(-)/Cyt(-) vs CSI-IV (all $p<0.0001$), HPV(+)/Cyt(-) vs CIN3 ($p<0.0001$), HPV(+)/Cyt(-) vs CSI-IV (all $p<0.0001$), CIN3 vs CSI-IV (all $p<0.0001$). No other significant differences were identified. The S5-classifier scores increased significantly with the severity of lesions (Cuzick test for trend: $z = 9.23$, $p<0.0001$).

HPV genotyping was 95.38% (95%CI 91.38-98.33%) in agreement with methylation data revealing HPV16/18/31/33 infection.

4.5 S5-CLASSIFIER SENSITIVITY IN CERVICAL CANCERS AT THE 0.80 PREDEFINED CUT-OFF

The S5-classifier methylation score was successfully measured in all 544 initial women histopatologically diagnosed with cervical cancer included in the study. A total of 543 out of 544 cancer patients tested positive for S5 at 0.80, yielding a sensitivity of 99.81% (95%CI 98.34-99.96). Table 4.2 shows the S5 sensitivity stratified per cancer histotype, FIGO stage, hrHPV status, hrHPV type, sample type, age and country of origin. At the 0·80 cut-off, cervical cancers which were consistently hrHPV-negative when tested with multiple hrHPV genotyping assays were 96.15% (95%CI 94.38-98.25) identified by the S5 classifier. The performance of the S5 classifier was uniform among all stratified groups. There were no significant differences in S5 sensitivity among histotype, FIGO stage, hrHPV type, sample type, age and country of origin (Fishers' test, all $p>0.05$).

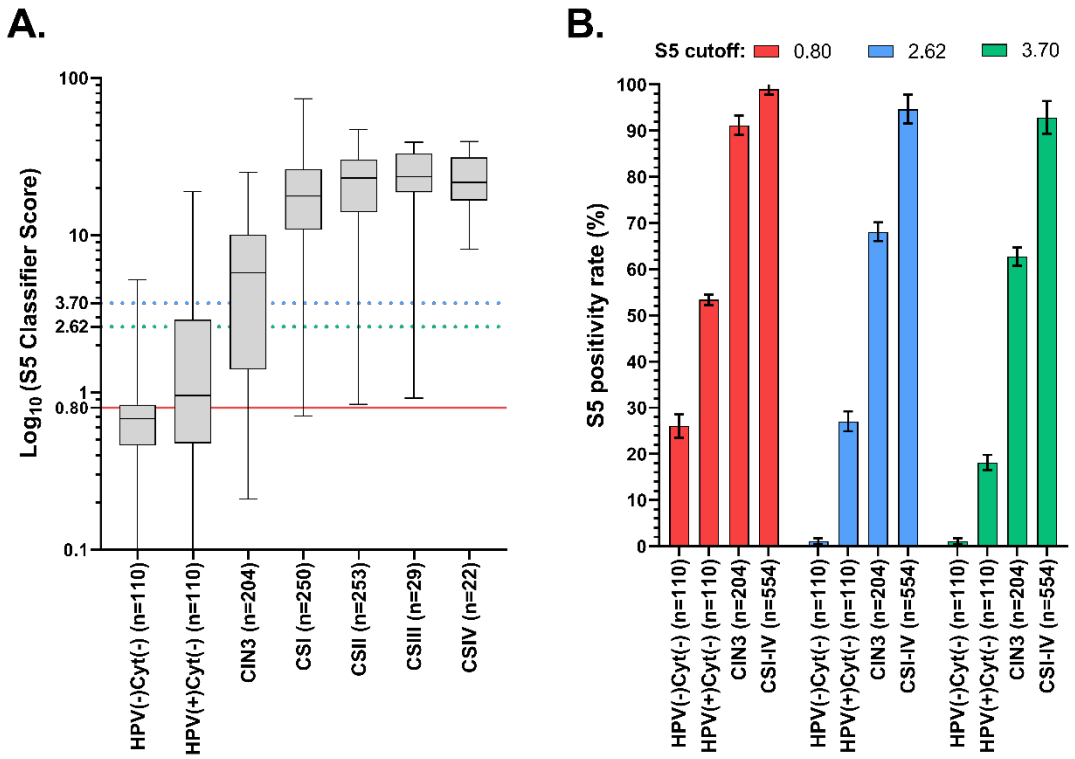

Figure 4.2 – **S5 classifier performance in the study group. A.** Distribution of the S5 scores based on the histopathological diagnostic of the patient. Data was plotted as $\log_{10}$ of the S5 score according to lesion grade: HPV(-)/Cyt(-), HPV(+)/Cyt(-), CIN3 and all stages of cervical cancer (CSI-IV). The S5 classifier was significantly different between the following group comparisons: HPV(-)/Cyt(-) vs CIN3 ($p < 0{\cdot}0001$), HPV(-)/Cyt(-) vs CSI-IV (all, $p < 0{\cdot}0001$), HPV(+)/Cyt(-)vs CIN3 ($p < 0{\cdot}0001$), HPV(+)/Cyt(-)vs CSI-IV (all, $p < 0{\cdot}0001$) and CIN3 vs CSI-IV (all, $p < 0{\cdot}0001$). Other comparisons were not significant (HPV(-)/Cyt(-)vs HPV(+)/Cyt(-), among CSI-IV). The proposed cut-offs for analysis are: 0·80 (red), 2·62 (blue) and 3·70 (green) The top of box represents the upper quartile (p75), bottom the lower quartile (p25), and the line the median (p50). The upper whisker extends to the largest point of the inter-quartile range from the upper quartile. The lower whisker extends to the smallest point of the inter-quartile range from the lower quartile. The Cuzick test for trend was highly significant ($p < 0.0001$). **B.** S5 sensitivity per lesion grade at 3 cut-offs: 0·80 (red), 2·62 (blue) and 3·70 (green). Median values are shown by each bar and error bars show 95% CI of the median. Abbreviations: HPV(-)/Cyt(-),HPV negative and cytology negative; HPV(+)/Cyt(-), HPV positive and cytology negative; CIN, cervical intraepithelial neoplasia (of grade 3); CSI-IV, cervical cancer stages I-IV.

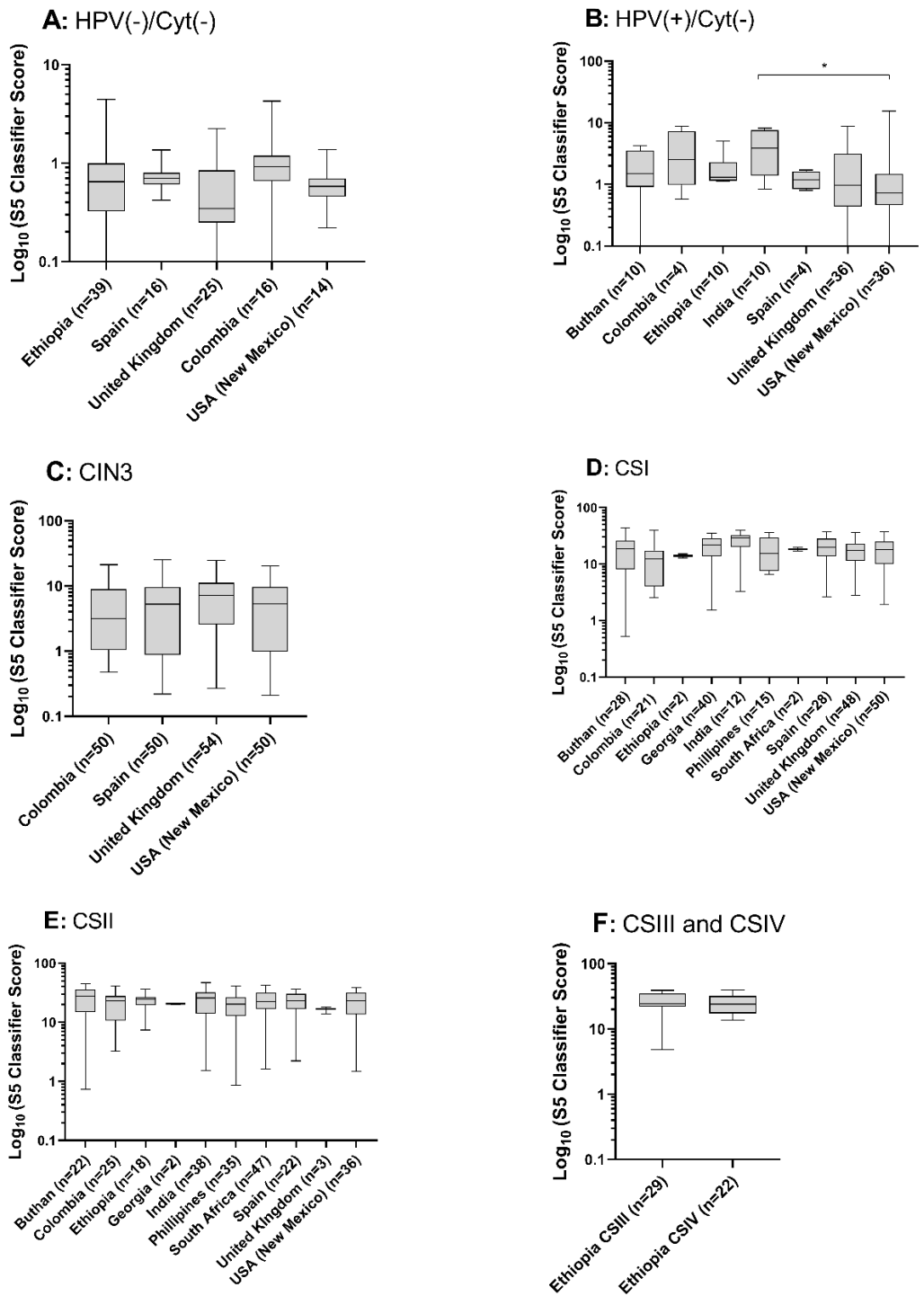

Figure 4.3 – S5 classifier score distribution per country and disease endpoint. Data is shown in $\log_{10}$ scale per country, measured at different endpoints (HPV(-)/Cyt(-), HPV(+)/Cyt(-), CIN3, CSI, CSII, CSIII, CSIV). Outlier values are represented by black dots. Significant differences between countries are marked on top of the graph with '*' for $p < 0.01$ and '**' for $p < 0.001$. Abbreviations: HPV(-/+)Cyt(-), HPV positive/negative and cytology negative; CIN, cervical intraepithelial neoplasia (of grade 3); CSI-IV, cervical cancer stages I-IV

	S5 sensitivity at cut-off 0.80			
	n / N*	**%**	**95% CI**	**P-value****
Histotype of cervical carcinoma				
Squamous cell carcinoma	509 / 510	99.80	(99.10-99.95)	
Adenocarcinoma	29 / 29	100.0	(97.34-100.0)	0.837
Adenosquamous cell carcinoma	1 / 1	100.0	(12.8-100.0)	
Neuroendocrine carcinoma	4 / 4	100.0	(26.86-100.0)	
FIGO Stage				
Stage I	244 / 245	99.60	(99.25-99.83)	
Stage II	249 / 249	100.0	(99.43-100.0)	0.687
Stage III	28 / 28	100.0	(97.34-100.0)	
Stage IV	22 / 22	100.0	(96.45-100.0)	
HPV status***				
HPV-positive	518 / 518	100.0	(99.46-100.0)	
HPV16	379 / 379	100.0	(99.36-100.0)	0.465[a]
HPV18	36 / 36	100.0	(98.82-100.0)	0.587[b]
HPV31	20 / 20	100.0	(96.45-100.0)	
HPV33	17 / 17	100.0	(93.12-100.0)	
Other hrHPV	66 / 66	100.0	(98.93-100.0)	
HPV-negative	25 / 26	96.15	(94.38-98.25)	
Sample type				
LBC sample	383 / 384	99.73	(98.34 –99.96)	0.917
FFPE tissue	160 / 160	100.0	(98.76-100.0)	
Age				
<25	7 / 7	100.0	(64.87-100.0)	
25-29	38 / 38	100.0	(98.62-100.0)	
30-39	133 / 133	100.0	(99.32-100.0)	0.989
40-49	153 / 154	99.39	(98.74-99.86)	
50-59	129 / 129	100.0	(99.22-100.0)	
>60	83 / 83	100.0	(99.02-100.0)	
Country of Origin				
Bhutan	50 / 50	100.0	(98.89-100.0)	
Colombia	46 / 46	100.0	(98.87-100.0)	
Ethiopia	70 / 70	100.0	(98.98-100.0)	
Georgia	42 / 42	100.0	(98.84-100.0)	0.892
India	50 / 50	100.0	(98.89-100.0)	
Philippines	50 / 50	100.0	(98.89-100.0)	
South Africa	49 / 49	100.0	(98.88-100.0)	
Spain	50 / 50	100.0	(98.89-100.0)	
United Kingdom	50 / 51	98.03	(95.99-99.05)	
USA (New Mexico)	86 / 86	100.0	(99.12-100.0)	
Total	**543/544**	**99.81**	**(98.34-99.96)**	

Table 4.2 - S5 Classifier sensitivity at the 0.80 cut-off in a cervical cancer referral group. Data was stratified per histotype of cervical carcinoma, FIGO stage, hrHPV status, sample type, age or country of origin.* n = number of positive samples in a specified group; N = group total;** hrHPV genotype grouping performed by hierarchical genotype attribution, as detailed in materials and methods;*** Determined by performing Fishers' exact test of independence; [a] P-value among all subgroups in the HPV-positive group; [b] P-value between HPV-positive and HPV-negative subgroups.

4.6 S5 CLASSIFIER CUT-OFF ADJUSTED PER COUNTRY TO OPTIMISE TRIAGE CAPACITY

Cervical cancer incidence is directly linked with the presence of a screening programme in the country. Hence the importance to introduce difference modality for the implementation of a molecular triage reflecting the country clinical setting. We investigated the false positive rates in women with HPV(-)/Cyt(-), HPV(+)/Cyt(-), CIN3 and CSI-IV at the UK-predefined cut-off of 0.80, the Youden-J index cut-off based on the S5 methylation scores of cervical cancers: 2.62 and the previously explored LMIC cut-off of 3.70 (Hernández-López et al., 2019). For all groups analysed, the false positive rate, decreased with the increase in cut-off. The most important decrease was observed in the HPV(-)/Cyt(-) and HPV(+)/Cyt(-) groups. At a cut-off of 0.80, the false positive rate in HPV(-)/Cyt(-) women was 26.32% (95%CI 23.90 – 29.94) which would suit a country with a strong health care system and decreased to 0.92% (95%CI 0.36 - 1.82) at both 2.62 and 3.70 cut-offs (McNemar test χ^2=27.1, p<0.0001) which would suit better countries with minimal screening capacity. However, the false positive rate of S5 in the HPV(+)/Cyt(-) was 52.74% (95%CI 49.71 – 55.63) at 0.80 and showed a significant decrease trend to 27.22% (95%CI 24.94 – 29.53) at 2.62 and 18.26% (95%CI 16.62 – 20.24) at 3.70 (Cuzick test for trend, p<0.0001). These possible adjustments will allow to customise S5 cut-off according to the country of clinical implementation.

Although we observed a significant decrease in false positive rate with the increase of the cut-off, a similar but less pronounced decrease trend was observed in S5 sensitivity for CIN3 and cancer detection. CIN3 sensitivity decreased from 87.26% (95%CI 84.42 – 89.93) at 0.80 to 62.74% (95%CI 60.13 – 65.25) at 3.70 (Cuzick test for trend, p<0.0001). Further, the S5 sensitivity for cancer decreased from 99.81% (95%CI 98.34-99.96) at 0.80 to 95.77% (95%CI 92.39- 97.40) at 3.70 (Cuzick test for trend, p=0.005).

4.7 DIAGNOSTIC POTENTIAL OF S5 CLASSIFIER IN LMIC WITH LIMITED RESOURCES

Table 4.3 presents the associations between HPV16/18 and S5 classifier sensitivity at different cut-offs for the identification of CIN3+, compared to the HPV(+)/Cyt(-). HPV16/18 positivity was strongly associated with CIN3+ development. The univariate odds ratios (OR) of HPV16/18 positivity for CIN3 was 2.86 (95%CI 1.77 – 4.62), while for cancer the OR was approximatively two times higher: 4.80 (95%CI 3.13 – 7.96). The univariate ORs for all S5 cut-offs were higher than the univariate ORs for HPV16/18 (p<0.0001). Increased ORs were observed for the bi-variable associations of HPV16/18 and the S5 test regardless of the cut-off or the geographic location (all, p<0.0001). This indicates stronger associations between the combination of

HPV16/18 positivity and S5 sensitivity and CIN3+ development. Although the ORs for the bi-variable analysis of HPV16/18 positivity and S5 0.80 cut-off were significantly higher than the univariate HPV16/18 ORs (p<0.0001), the highest associations for CIN3+ development were observed for the combination of HPV16/18 positivity with the S5 sensitivity at the 3.70 cut-off (OR: 5.01, 95%CI 2.82 - 8.90 for CIN3; and OR: 14.90, 95%CI 8.69 – 25.56 for cervical cancer).

	Variables	OR	95% CI	Z value	P value[a]
CIN3	**HPV 16/18**	2.86	1.77 - 4.62	4.30	Reference
	S5 0.80	4.50	2.71 – 7.46	5.83	<0.0001
	S5 2.62	5.63	3.26 – 9.73	6.19	<0.0001
	S5 3.70	6.42	3.67 – 11.24	6.52	<0.0001
	***HPV 16/18 and S5 0.80**	3.26	2.01 - 5.30	4.79	<0.0001
	***HPV 16/18 and S5 2.62**	3.56	2.12 - 5.98	4.80	<0.0001
	***HPV 16/18 and S5 3.70**	5.01	2.82 - 8.90	5.49	<0.0001
Cervical Cancer	**HPV 16/18**	4.80	3.13 - 7.36	7.19	Reference
	S5 0.80	20.94	7.89 - 51.71	7.22	<0.0001
	S5 2.62	36.21	20.9 - 62.73	12.80	<0.0001
	S5 3.70	45.55	24.67 – 73.38	13.49	<0.0001
	***HPV 16/18 and S5 0.80**	6.32	4.08 - 9.80	8.25	<0.0001
	***HPV 16/18 and S5 2.62**	9.87	6.11 - 15.96	9.35	<0.0001
	***HPV 16/18 and S5 3.70**	14.90	8.69 - 25.56	9.81	<0.0001

Table 4.4 – Univariate and bivariate odds ratios (OR with 95%CI confidence intervals) for the associations between the different clinical outcomes and HPV16/18 and/or S5 cut-offs in a HPV(+)/Cyt(-) baseline. * Bivariate OR for the associations between HPV16/18 and S5 cut-offs (0·80, 2·62 and 3·70) and CIN3 and cervical cancer; [a] P value indicating an increased ORs compared to the reference HPV16/18. P value determined using Fishers' test of independence

4.8 S5-CLASSIFIER PERFORMANCE IN DETECTING CERVICAL CANCERS AT THE 3.70 CUT-OFF

Due to a lack of organised screening systems, LMIC might benefit from the S5 classifier, with an increased cut-off and an associated decreased rate of false positives. Increasing the cut-off to 3.70, the false positive rate was 9.54% (95%CI 8.49 – 10.76) and approximatively 4 fold lower than at the 0.80 cut-off (39.54%, 95%CI 37.20 – 41.86). The decrease in the false positive rate correlates to an increase in specificity of the S5 classifier (data not shown). A total of 520 out of 543 women with any cancer type tested positive for S5 at the 3.70 cut-off, yielding a sensitivity rate of 95.77% (95%CI 92.39- 97.40). A lower proportion of the hrHPV-negative cancer group tested S5 positive at a cut-off 3.70 compared to the hrHPV-positive cancer group: 73.07% (95%CI 56.85 – 86.82) versus 98.45% (95%CI 92.72 – 99.46) (Fishers' test, p<0.0001) (Table 4.4). There were no other significant differences in S5 sensitivity among cancer histotype, FIGO stage, hrHPV type, sample type, age and country of origin. These results underscored the rather uniform

performance of the S5 classifier in our referral population of cancer patients from different continents.

	S5 sensitivity at cut-off 3.70			
	n / N*	**%**	**95% CI**	**P-value****
Histotype of cervical carcinoma				
Squamous cell carcinoma	491 / 510	96.22	(91.32 – 98.35)	
Adenocarcinoma	28 / 29	96.55	(91.71 – 99.23)	
Adenosquamous cell carcinoma	1 / 1	100.0	(91.81 – 100.0)	0.837
Neuroendocrine carcinoma	4 / 4	100.0	(94.24 – 100.0)	
FIGO Stage				
Stage I	230 / 245	93.87	(89.42 – 96.91)	
Stage II	242 / 249	97.18	(93.62 – 99.00)	
Stage III	27 / 28	96.42	(90.32 – 99.42)	0.687
Stage IV	22 / 22	100.0	(97.76 – 100.0)	
HPV status***				
HPV-positive	510 / 518	98.45	(92.72 – 99.46)	
HPV16	372 / 379	98.15	(92.52 – 99.32)	
HPV18	34 / 36	94.44	(91.60 – 98.72)	0.465[a]
HPV31	19 / 20	95.00	(92.31 – 96.42)	0.587[b]
HPV33	16 / 17	94.11	(90.22 – 95.41)	
Other hrHPV	61 / 66	92.42	(81.35– 94.15)	
HPV-negative	19 / 26	73.07	(56.85 – 86.82)	
Sample type				
LBC sample	371 / 384	96.61	(91.70 – 98.62)	
FFPE tissue	150 / 160	93.75	(90.82 – 96.62)	0.917
Age				
<25	6 / 7	85.71	(65.55 – 90.22)	
25-29	36 / 38	94.73	(90.45 – 98.65)	
30-39	128 / 133	96.24	(91.24 – 98.75)	
40-49	148 / 154	96.10	(92.54 – 98.12)	0.989
50-59	122 / 129	94.57	(89.79 – 96.12)	
>60	81 / 83	97.59	(94.05 – 98.92)	
Country of Origin				
Bhutan	47 / 50	94.00	(92.32 – 96.82)	
Colombia	40 / 46	86.95	(78.35 – 92.92)	
Ethiopia	68 / 70	97.14	(94.92 – 98.96)	
Georgia	40 / 42	95.23	(92.98 – 97.35)	
India	48 / 50	96.00	(94.12 – 97.59)	0.892
Philippines	48 / 50	96.00	(94.12 – 97.59)	
South Africa	49 / 49	100.0	(98.88 - 100.0)	
Spain	48 / 50	96.00	(94.12 – 97.59)	
United Kingdom	48 / 51	94.11	(93.72 – 96.59)	
USA (New Mexico)	84 / 86	97.67	(94.72 – 98.68)	
Total	**521 / 544**	**95.77**	**(92.39- 97.40)**	

Table 4.4 - S5 Classifier sensitivity at the 3·70 cut-off in a cervical cancer referral group. Data was stratified per histotype of cervical cancer, FIGO stage, hrHPV status, sample type, age or country of origin. * n = number of positive samples in a specified group; N = group total;** hrHPV genotype grouping performed by hierarchical genotype attribution, as detailed in materials and methods;*** Determined by performing Fishers' exact test of independence; [a] P-value among all subgroups in the HPV-positive group; [b] P-value between HPV-positive and HPV-negative subgroups.

4.9 S5 CLASSIFIER PERFORMANCE: SENSITIVITY AND SPECIFICITY

The performance of the S5-classifier was assessed through receiver operating characteristic (ROC) curves for detecting CIN3 and invasive cervical cancer (all histological types, grades and stages combined) from either HPV(-)/Cyt(-) women (approximating a highly vaccinated population) or HPV(+)/Cyt(-) women (approximating a currently relevant triage population). Sensitivity and specificity were estimated at all three S5 cut-offs described earlier in this chapter.

The ROC curves for detecting CIN3 and cancers in a HPV(-)/Cyt(-) baseline population gave areas under the curve (AUC) of 0.90 (95%CI 0.87-0.93) and 0.99 (95%CI 0.98-0.99), respectively for all S5 cut-off points. Although the sensitivity of detecting CIN3 and cervical cancer was high at cut-off 0.80 (91.18%; 95%CI 86.48 – 94.16 for CIN3 and 99.81%; 95%CI 98.56 – 99.99 for cervical cancer), the specificity was lower: 65.12% (95%CI, 54.59 – 74.35). The exploratory cut-off of 2.62 dramatically improved specificity at a cost on S5 sensitivity. Although, sensitivity for CIN3 was significantly lower at 2.62 than sensitivity at 0.80 (68.13%; 95%CI 61.46 – 74.15), McNemar test: χ^2=15.27, p<0.0001), no significant differences were observed in sensitivity for cancer (95.21%; 95%CI 93.12 – 96.68; χ^2=3.52, p=0.076). The 3.70 cut-off was observed to reduce the rate of positives even more: 62.74% (95%CI, 55.93 – 69.09) for CIN3 and 93.26% (95%CI, 90.89– 95.05) for cancer. On the other hand, specificity at cut-offs 2.62 and 3.70, were 100% (95.19 – 100.0) for both disease endpoints (Table 4.5).

In a HPV(+)/Cyt(-) baseline population, we observed lower AUCs for the detection of both disease endpoints than when using HPV(-)Cyt(-) women as baseline: 0.80 (95%CI 0.76-0.84) for CIN3 (DeLong's test, p=0.005) and an AUC of 0.97 (95%CI 0.96-0.98) for cervical cancer detection (DeLong's test; not significant, p>0.05) at the S5 cut-offs analysed. Sensitivity for all cut-offs were the same as described in the HPV(-) population. For CIN3 detection, specificity was higher with the increase in cut-off: at 0.80 was 48.81% (95%CI, 41.36 – 56.31), 78.57% (95%CI 71.76 – 84.10) at 2.62 and 83.33% (95%CI 76.97 – 88.21) at 3.70. For cancer detection, the same trend was observed: specificity increased from 50.60% (95%CI, 43.11 – 58.06) at 0.80 to 83.33% (95%CI 76.97 – 88.21) at 3.70.

Increasing specificity of the S5-classifier was significant by adjusting the cut-off from 0.80 to 2.62 (McNemar test: χ^2=15.17, p<0.0001) and 3.70 (χ^2=16.48, p<0.0001). However, this came at the cost of a decrease in sensitivity and a larger number of women missed though testing with S5 (Table 4.5). However, considering the referral nature of the sample, the sensitivity and specificity identified are just an estimation and cannot be translated to a real screening scenario.

		AUC	95% CI	Cut-off	Sensitivity (%)	95% CI	McNemar p-value*	Specificity (%)	95% CI	McNemar p-value*	Women missed N (% total)
HPV(-)/Cyt(-) vs	CIN3 (n=204)	**0.90**	0.87 – 0.93	0.80	**91.18**	86.48 - 94.16		**65.12**	54.59 – 74.35	<0.0001	16 (8.22)
				2.62	**68.13**	61.46 – 74.15	<0.0001	**100**	95.19 - 100		65 (31.87)
				3.70	**62.74**	55.93 – 69.09	<0.0001	**100**	95.72 - 100	<0.0001	76 (37.26)
	Cancer (n=544)	**0.99**	0.98 – 0.99	0.80	**99.81**	98.56 – 99.99		**65.12**	54.59 – 74.35	0.0025	2 (0.36)
				2.62	**95.21**	93.12 – 96.68	0.036	**100**	95.19 – 100		27 (4.88)
				3.70	**93.26**	90.89– 95.05	0.002	**100**	95.72 - 100	<0.0001	38 (6.74)
HPV(+)/Cyt(-) vs	CIN3 (n=204)	**0.80**	0.76 – 0.84	0.80	**91.18**	86.48 - 94.16		**48.81**	41.36 - 56.31	<0.0001	16 (8.22)
				2.62	**68.13**	61.46 – 74.15	<0.0001	**78.57**	71.76 – 84.10		65 (31.87)
				3.70	**62.74**	55.93 – 69.09	<0.0001	**83.33**	76.97 – 88.21	<0.0001	76 (37.26)
	Cancer (n=544)	**0.97**	0.96 – 0.98	0.80	**99.81**	98.56 – 99.99		**50.60**	43.11 – 58.06	0.0051	2 (0.36)
				2.62	**95.21**	93.12 – 96.68	0.036	**78.57**	71.76 – 84.10		27 (4.79)
				3.70	**93.26**	90.89– 95.05	0.002	**83.33**	76.97 – 88.21	<0.0001	38 (6.74)

Table 4.5 – Sensitivity and specificity of the S5 methylation classifier in identifying CIN3 and invasive cancer based on different classifier cut-offs. 0.80 is the pre-defined cut-off developed in the United Kingdom and validated for developed countries, 2.62 is the Youden-J index cut-offs based on the methylation scores of invasive cervical cancers and 3.70 as cut-off explored in Mexico and proposed for underdeveloped countries.

4.10 S5 CLASSIFIER COMPONENT BREAKDOWN

On a component basis, individual *EPB41L3* methylation was observed to increase in a sigmoidal manner with disease progression. The Cuzick test for trend was significant for HPV(-)/Cyt(-), CIN3, CSI and CSII ($p < 0.0001$). The second most important component of the S5 classifier, HPV16 methylation, showed an increasing trend with disease progression (Cuzick test for trend, both, $p < 0.001$) as well as a steep increase from CIN3 to CSI-IV. Additionally, both HPV18 and HPV33 methylation showed a linear increase with disease progression (Cuzick test for trend, both, $p < 0.01$) as described in Figure 4.4.

The weight of each component of the S5-classifier was plotted for HPV(+)Cyt(-), CIN3 and the cancer CSI-IV groups in Figure 4.5. The Cuzick test for trend showed an increasing trend of *EPB41L3* weight with disease lesion, $z = 8.21$ ($p<0.0001$). *EPB41L3* weight plateaus at CSII. Unconditional logistic regression models showed the strength of the association between *EPB41L3* methylation, severity of lesion, and age. The relationship between *EPB41L3* methylation was stronger for severity of lesion: $F = 367.5$, $p<0.0001$ than age ($F = 81.0$, $p<0.0001$). This indicates that host *EPB41L3* methylation might have a good potential to predict disease progression, independent of increasing natural epigenetic methylation levels occurring with age.

Interestingly, the relative weights of the HPV components of the S5-classifier decreased slightly with severity of lesion (Cuzick test for trend, $z = -6.52$, $p<0.0001$), informing on the possible changes in the dynamics of the viral infection. HPV16 had the highest weight out of all viral components, however this was 1.8 times lower than the weight of *EPB41L3* in CSII+ specimens. The weights of the HPV18, HPV31 and HPV33 did not show a consistent trend.

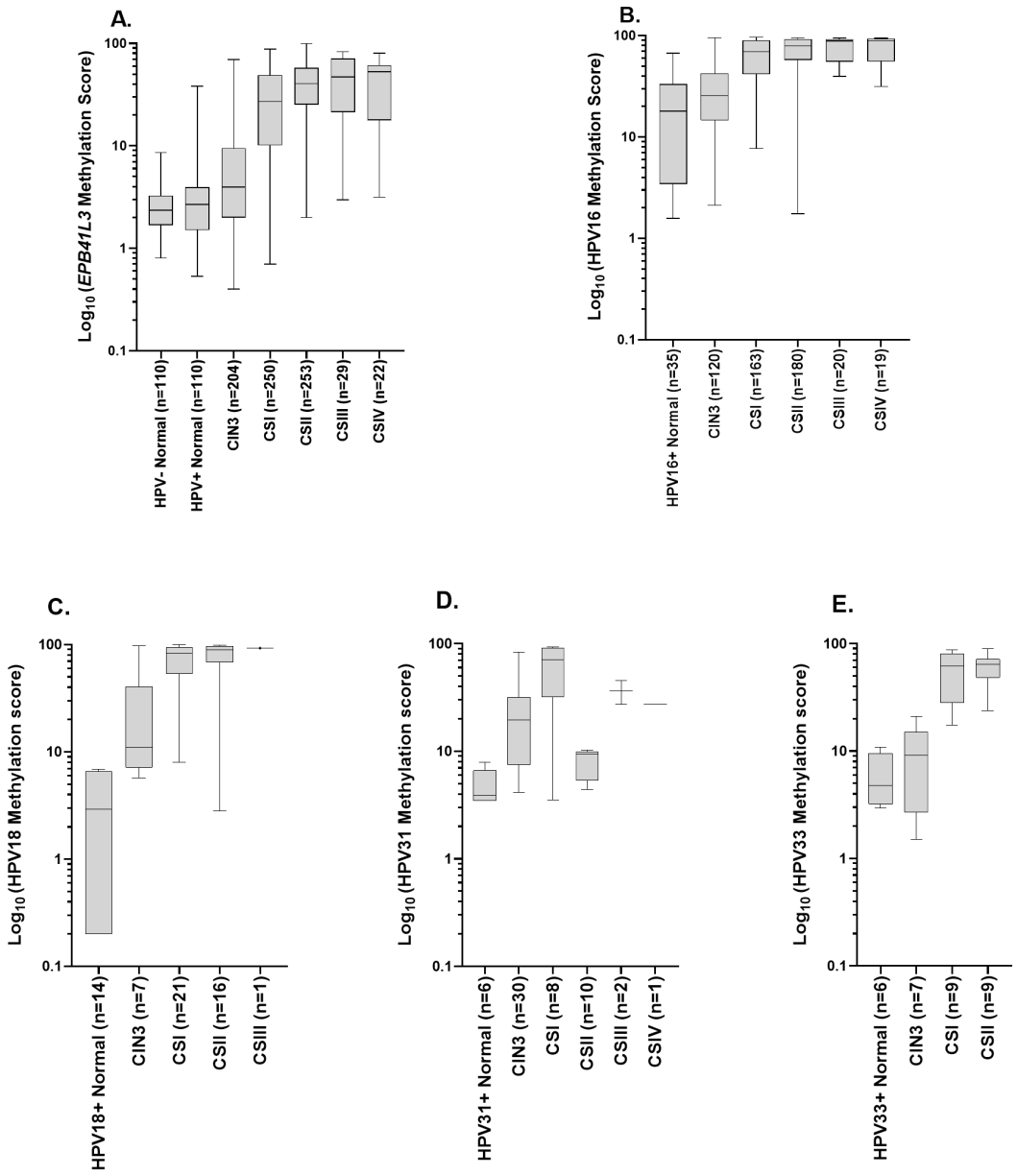

Figure 4.4 - Distribution of *EPB41L3*, HPV16, HPV18, HPV31L1 and HPV33L2 methylation scores based on the histopathological diagnostic of the patient. Absolute methylation was calculated as an average of the selected CpGs on: *EPB41L3* **(A)**, HPV16 L1 and L2 late genes **(B)** HPV18 L2 **(C),** HPV31 L1 **(D)** and HPV33 L2 **(E)** and plotted as log10 of methylation score. Significant differences were found between the following group comparisons in **A:** HPV(-)/Cyt(-) vs CIN3 ($p = 0.049$), HPV(-)/Cyt(-) vs CSI-IV (all, $p < 0.0001$), HPV(+)/Cyt(-) vs CSI-IV (all, $p < 0.0001$), CIN3 vs CSI-IV (all, $p < 0.0001$) and CSI vs CSII ($p < 0.0001$); in **B:** HPV16(+)/Cyt(-) vs CSI-IV (all, $p < 0.0001$) and CIN3 vs CSI-IV (all, $p < 0.0001$). **C**: HPV18(+)/Cyt(-) vs CSI ($p = 0.013$), HPV18(+)/Cyt(-) vs CIN3 ($p = 0.023$), HPV18(+)/Cyt(-) vs CSII ($p = 0.003$), CIN3 vs CSII ($p = 0.010$); in **D**: HPV31(+)/Cyt(-) vs CSI ($p < 0.001$); in **E**: HPV33(+)/Cyt(-) vs CSI ($p = 0.025$), HPV33(+)/Cyt(-) vs CSII ($p = 0.007$), CIN3 vs CSI ($p = 0.007$), CIN3 vs CSII ($p < 0.001$). Other comparisons were not significant. Abbreviations: HPV(-/+)Cyt(-), HPV positive/negative and cytology negative; CIN, cervical intraepithelial neoplasia (of grade 3); CSI-IV, cervical cancer stages I-IV. The top of box represents the upper quartile (p75), bottom the lower quartile (p25), and the line the median (p50). The upper whisker extends to the largest point of the inter-quartile range from the upper quartile. The lower whisker extends to the smallest point of the inter-quartile range from the lower quartile. The outliers are plotted as individual points for each lesion grade.

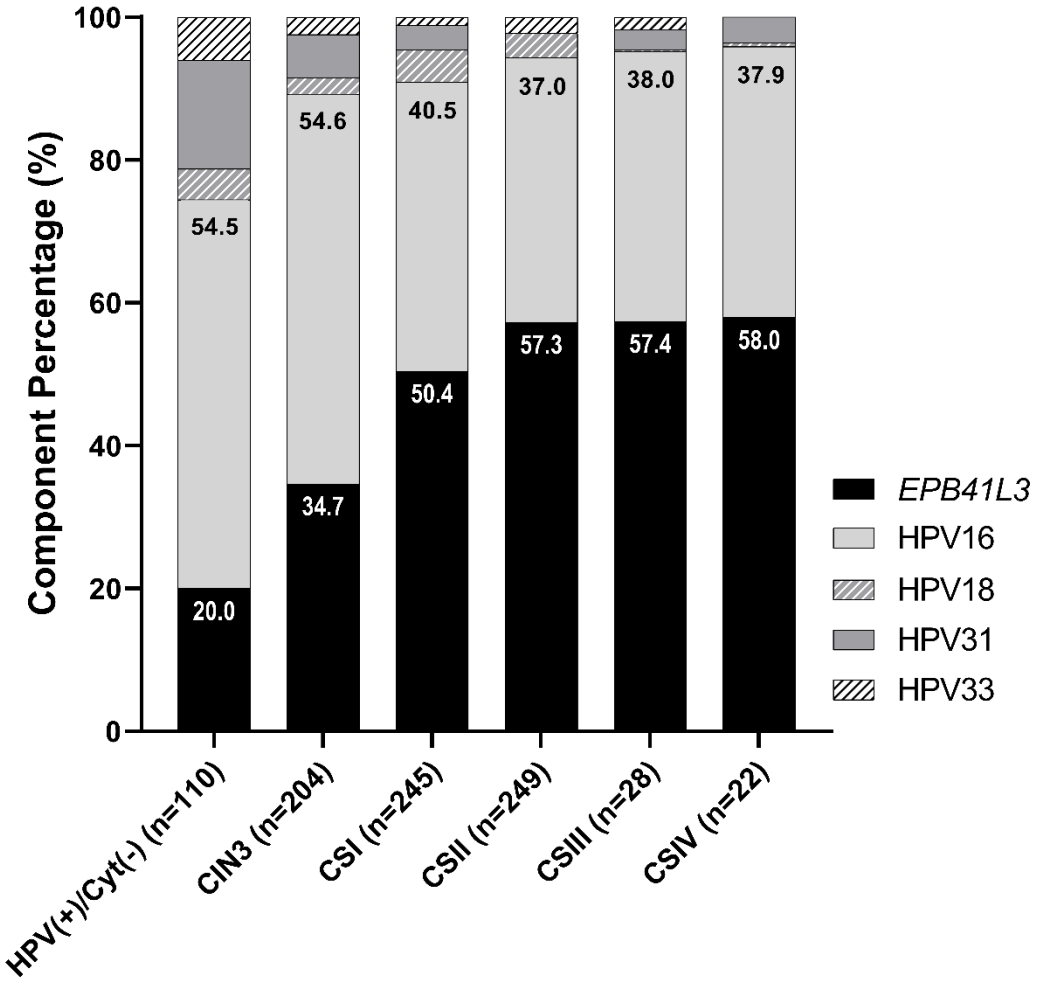

Figure 4.5 – S5 classifier component breakdown in HPV(+)Cyt(-), CIN3 and CSI-IV samples. The weight of the methylation on the S5 components: *EPB41L3*, HPV16, HPV18, HPV31 and HPV33 has been calculated for each group. Percentages of EPB41L3 and HPV16 weights in the classifier are noted at the top of the corresponding bars. HPV methylation becomes less important versus *EPB41L3* as the lesions become more advanced, however *EPB41L3* weight in the classifier does not change after stage II cancer

4.11 DISCUSSION

To our knowledge, our study is the largest and most comprehensive appraisal of viral and host cell DNA methylation in invasive cervical cancer to date (Vink et al., 2020). We show that there is a strongly increasing trend of S5 DNA methylation score with cervical disease severity and that this is evident in a global collection of samples. Our results also show a very high S5 sensitivity for CIN3+ at the United Kingdom predefined cut-off of 0.80, while there was high to moderate sensitivity at the Youden's and the LMIC suggested exploratory cut-offs of 2.62 and 3.70 respectively.

In our geographically diverse samples we found a DNA methylation test sensitivity of 91.18% (186/204) for CIN3 and 99.81% (543/544) for cervical cancer detection at the predefined cut-off of 0.80 (Lorincz et al., 2016) which was developed and validated for developed countries. On note, at this cut-off S5 sensitivity for cervical cancer was higher than the sensitivity of HPV DNA testing: 95.21% (518/544) (McNemar χ^2=5.08, p=0.032). Additionally, at the S5 cut-off of 3.70, previously explored for use in LMIC (Hernández-López et al., 2019), we found a sensitivity of 62.74% (128/204) for CIN3 and 95.77% (521/544) for cervical cancer.

Through the examination of cervical specimens from 220 HPV(-/+)/Cyt(-) women, 204 CIN3 and 544 cervical cancer cases, our data shows that increased methylation of the host *EPB41L3* and viral late genes of HPV16, 18, 31, 33 appear to compose an apparently universal biomarker panel for high-grade precancerous and malignant cervical cancer worldwide. This study complements our previous results in Mexico, Canada and Colombia by including data on cancers from additional countries, describing a larger study for cervical cancer identification (Cook et al., 2018; Hernández-López et al., 2019; Ramirez et al., 2021). A recent Dutch study of 519 cervical cancer samples, FAM19A4/miR124-2 methylation analysis yielded a sensitivity of 98.30% (510/519) (Vink et al., 2020). Additionally, the four-gene methylation marker panel comprising of the host genes *JAM3*, *EPB41L3*, *TERT* and *C13ORF18* identified 94.20% (65/69) of cervical cancers (Eijsink et al., 2012). S5 demonstrated a slightly higher sensitivity of 99.81% (543/544) compared to the above-mentioned tests (p=0.047 and p=0.029, respectively). More importantly, we identified 25 out of 26 hrHPV-negative cancers which were not explored in other studies.

We explored the performance of the S5 classifier at three cut-offs, 0.80 and two additional cut-offs. We calculated a cut-off based on the Youden-J index from the methylation scores of the HPV(-)Cyt(-) women versus cervical cancer (2.62) and another cut-off previously explored in Mexico (3.70). At cut-off 0.80, 26.32% of HPV(-)Cyt(-) women tested positive for the S5 test, which is quite high and indicates either a potential specificity issue of the tool in our selected group of HPV(-)Cyt(-) samples or a higher than expected prevalence of occult disease in these women. The lowest false positive rate was observed at the 3.70 cut-off (0.92%), a cut-off at witch 95.77% (521/544) of cancer cases were still identified. Our data on S5 sensitivity combined with our earlier results from several studies in the United Kingdom and Canada suggest that the prevalence of HPV infection as well as the difference in screening capacity and performance of populations can affect disease prevalence, thereby arguably the optimal cut-off of S5 could be made 'setting specific'.

Selection of a threshold for the test will depend on whether the test is recommended in screening or triaging setting. In a screening setting every programme (any setting) requires a highly specific test (with moderate to high sensitivity) to minimize harms and save resources. However, a highly sensitive screening test is particularly important in the LMICs as screening women too frequently is logistically challenging. Less frequent screening with a highly sensitive test also saves resources for the LMICs. In a triaging setting a moderate sensitivity is acceptable (any setting) as the triage negative women are requested to return after one year. In a triaging context (any setting) a very high sensitivity (as is seen with a lower cut off of the S5) signifies that the triage-negative women may be asked to return after two years rather than one year.

This study shows that an increased cut-off with a lower number of false positives rate would maximise the detection of cancer, given the lack of resources in LMICs. Indeed, in less developed regions and in countries with weak screening programmes one can increase the cut-off (Cook et al., 2018; Cuschieri et al., 2019; Hernández-López et al., 2019). However, this study shows that depending on the screening policy chosen any of these cut-offs could be implemented to reflect the desired clinical outcome.

The major strengths of our study are its size, the incorporation of sample sets that reflect diverse settings from ten different countries spanning five continents. Our study highlights that the general trend of increasing DNA methylation with disease progression is independent of the population.

Although HPV infection is an important co-factor in cervical cancer development, a small proportion of cervical carcinoma samples in the study were tested hrHPV negative as confirmed by HPV testing with multiple assays. Though rare, these cases represent a challenge for detection in the current primary hrHPV screening programme. Our data shows that these cancers (25/26) where nearly all identified by the S5 classifier at a cut-off of 0.80. Regardless of the cut-off examined, performance of the S5 classifier was uniform among the stratified groups: histology, FIGO stage, hrHPV status, hrHPV type, sample type, age and country of origin.

A limitation to this study is that all CIN3 and cervical cancer cases come from referral populations and do not accurately represent those that may be apparent in the screening population or those that do not present to clinics. The test performance observed with DNA extracted from tissue samples may not be replicable when the test is performed on cervical swab samples or self-collected vaginal samples. The number of rare cervical cancer histology in our study was also small, so this element would benefit from further investigation. Moreover, much more emphasis was placed on cancers FIGO stage I and II as previously published data indicates that aberrant methylation is an early event in cervical carcinogenesis (Verlaat et al., 2018). An intentional limitation of our study is that we excluded CIN1 and CIN2 which would be present in a real-world setting. Addition of these samples to our study in realistic proportions would likely lower the sensitivity and specificity of the S5 test. There is a further limitation in our selection of the cytology negative women who were presumed to have no disease on the basis of cytological testing. Although we divided these women into HPV+ and HPV- groups these women may not be representative of the routine screening populations in many geographic locations including in Europe and the USA. This said, we have demonstrated that the S5 test had a low sensitivity in women with no apparent abnormal pathology and that its performance as a triage was superior to either cytology or HPV16/18 or combination for the detection of CIN, which is completely consistent with all other published studies to date on S5 from different countries. Therefore, our focus for the present study was to assess a larger panel of CIN3+ samples to confirm the sensitivity and robustness of the assay for the detection of significant disease.

The findings of our study highlight the major weight of the host *EPB41L3* methylation in the S5-classifier score. We showed that the relationship between *EPB41L3* methylation was approximatively 4.5 times stronger for severity of lesion than age ($p<0.0001$). This indicates that host *EPB41L3* methylation might have a strong potential to predict disease progression, independent of increasing natural epigenetic methylation levels occurring with age. Indeed, methylation level of *EPB41L3* might be a prognostic indicator of progressive disease. This would

be in line with previously published data on the S5-classifier where it was more likely to identify women with CIN2 that was more likely to progress (Louvanto et al., 2019). Additionally, the weight of *EPB41L3* methylation shows an increasing trend ($p<0.0001$, Cuzick test for trend), up to cervical cancer FIGO stage II, where it plateaus. However, the strength of this observation is limited by the decreased number of cervical cancer samples of FIGO stage III and IV, included in the study.

In conclusion, our study shows that the S5 classifier at a cut-off of 0.80 identifies more than 90% CIN3 cases and almost 100% of cervical cancers, independent of histology, FIGO stage hrHPV status, hrHPV genotype, sample type and geographical origin. Adjustment of the cut-off leads to an increase in specificity with only a small decrease in sensitivity. The 3.70 cut-off could allow for a better triage modality for LIMC where screening is not performed as systematically as in higher income countries. Additionally, high methylation levels on the host gene component of the S5 classifier, *EPB41L3* is associated with higher severity of the disease, indicating a possible prognostic marker potential. Thus, considering the growing acceptability of self-sampling, our results support the utility of the S5 classifier as a credible tool for enhanced risk stratification of women in cancer screening programmes.

CHAPTER 5 – MICROBIAL BIOMARKERS IN WOMEN WHO DEVELOPED CIN3 – A PILOT STUDY

5.1 INTRODUCTION

Persistent infection with hrHPV is necessary, but not sufficient for cervical cancer development. Although evidence suggests that the cervicovaginal microbiota plays a functional role in the persistence of HPV infections, this has yet to be investigated in longitudinal studies, to understand the exact microbial factors associated with cervical intra-epithelial neoplasia (CIN) (Gravitt & Winer, 2017; Snijders et al., 2006; Steenbergen et al., 2014).

The cervicovaginal microbiota has an important role in female reproductive health. Advances in 16S-rRNA next generation sequencing (NGS) have facilitated the characterization of the healthy cervicovaginal microbiota and reported that hrHPV positive women exhibited a higher diversity in the cervicovaginal microbiota than their HPV negative counterparts (Arokiyaraj et al., 2018; Paola et al., 2017; Virtanen et al., 2017). The cervicovaginal microbiota is divided into five community state types (CSTs) (Curty et al., 2020; Ma & Li, 2017; Seta et al., 2019). CST I, CST II, CST III and CST V show low diversity in microbiota, with the dominant species being: *L. crispatus, L. gasseri, L. iners* and *L. jensenii*, respectively (Ravel et al., 2010; Seta et al., 2019). Conversely, CST IV shows higher bacterial diversity with a low abundance of *Lactobacillus* and an increased frequency of anaerobic bacteria from genuses like *Gardnerella, Megasphera, Atopobium, Sneathia* or *Prevotella* (Amabebe & Anumba, 2018; Curty et al., 2020). An abundance of one or more *Lactobacillus* species is thought to defend against pathogens, through the maintenance of an acidic pH, secretions of specific chemical metabolites, through adherence to the vaginal mucus and disruption of biofilms (Curty et al., 2020; Kyrgiou et al., 2018).

Specific microbial signatures were associated with hrHPV positivity and disease progression. For instance, *Sneathia* species were highly associated with hrHPV positivity and high risk of cervical disease development (Audirac-Chalifour et al., 2016; Mitra et al., 2015b; Paola et al., 2017). A cervicovaginal microbiota dominated by *L. Gasseri* was related to faster hrHPV clearance while microbiotas with lower *Lactobacillus* levels and high *Atopobium*, *Gardnerella* and *Prevotella* were associated with slower rates of clearance (Brotman et al., 2014). Specifically, high levels of *A. vaginae*, *G. vaginalis* and *L. iners* combined with low levels of *L. crispatus* were suggested as the most risky combination for CIN development (Oh et al., 2015). Further, a cervicovaginal microbiota dominated by *L. iners* was significantly associated with CIN2+ development in hrHPV

positive women(Piyathilake Chandrika, Ollberding Nicholas, Ranjit Kumar, Maurizio Macaluso, Roland Alvarez, 2016).

The current cervical screening program in the United Kingdom relies on the combination of HPV testing and cytology triaging for the identification of CINs and cancer (Castanon et al., 2018). Due to the high incidence of HPV infections (~80% in the general population), HPV testing leads to a large number of clinically insignificant positives (Cuschieri et al., 2019; Lorincz et al., 2014). Cytology triage depends on human judgement, many women may be considered morphologically abnormal when they are known to be positive for hrHPV, resulting in more false referrals for colposcopy and biopsy (Lorincz et al., 2014).

Generally, women diagnosed with CIN2+ undergo surgical treatment. Loop Electrosurgical Excision Procedure (LEEP) is the most commonly used treatment for CIN2+ and involves the use of a small electrical wire loop to remove abnormal cells from the cervix. A large proportion of women treated are young and have not started or completed their reproductive career (Wiik et al., 2019). Women who underwent LEEP treatment were associated with an increased risk of preterm delivery in subsequent pregnancies (Amabebe & Anumba, 2018). The exact mechanisms of how excisional treatment increases the risk of preterm delivery are not clear, however, the lack of mechanical support of the cervix may be part of the explanation. Ascending bacterial infections from the lower genital tract to the uterine cavity may be a possible pathological mechanism for preterm delivery (Bobdiwala et al., 2020). LEEP treatment of CIN2+ may alter the cervicovaginal microbiota and hence increase the risk of bacterial infections when a woman becomes pregnant. So far, only two studies have investigated how LEEP influences the cervicovaginal microbiota. Both Zhang *et al.* and Wiik *et al.* have found a decrease in cervicovaginal microbiota with an increase in *Lactobacillus spp.* three and twelve months, respectively after LEEP (Wiik et al., 2019; Zhang et al., 2018).

To date, molecular biomarkers have been investigated for triaging purposes. Specifically, the focus has been placed on epigenetic DNA methylation assays as they may offer a more accurate and cost-effective alternative to cytology. For instance, the S5 methylation classifier has been validated for the detection of CIN2+ and involves the methylation analysis on the host *EPB41L3* gene and viral genes of HPV types 16, 18, 31 and 33. The S5 methylation scores increase with disease severity. As epigenetic DNA methylation is tissue specific, the local microbial environment may influence the rate of DNA methylation (Barros & Offenbacher, 2009). Therefore, correlations between the S5 methylation score and specific microbial signatures may indicate an interplay between epigenetic changes and the disease environment.

The present study aims to investigate how the cervicovaginal microbiota changes over time in young hrHPV-positive women who developed CIN3 and recovered from LEEP treatment. Additionally, we investigate potential microbial markers priming CIN3 development by comparing the cervicovaginal microbiota composition in women with CIN3 to that of women with persistent hrHPV infections and normal cytology results. Finally, we show that specific microbial signatures are associated with epigenetic changes and disease development.

5.2 THE CERVICOVAGINAL MICROBIOTA COMPOSITION ACCORDING TO DISEASE AND HPV STATUS

We investigated the cervicovaginal microbiota of 59 women in a longitudinal study run over six years. All 59 women had cervicovaginal swabs taken as part of the ARTISTIC screening trial at three time points: initial screening, first follow-up and second follow up which were all approximately three years apart and in accordance with cytology screening regulations in the United Kingdom. The women included in the study were classified into 3 groups: A, B and C, based on the HPV typing and cytology results and each group was divided in 3 subgroups according to the collection time point. Group A included 14 women who tested negative for HPV typing and showed normal cytology results at all three time points (A1, A2 and A3). Group B included 15 women who consistently tested positive for hrHPV and showed normal cytology results throughout the length of the study (B1, B2 and B3). Lastly, group C included 30 women who tested hrHPV positive and had normal cytology results at the initial screening (C1); they were diagnosed with CIN3 at the first follow-up (C2) and were screened approximately three years after LEEP treatment (C3). Table 5.1 summarises the characteristics of each group. No difference was determined in the mean age of the women included in the longitudinal study. (Fishers' exact test, p=0.97).

Characteristics	Group A (n=14)	Group B (n=15)	Group C (n=30)
Ethnicity	Caucasian	Caucasian	Caucasian
Age range at initial screening (mean ± se)	20-37 (28 ± 6)	20-37 (27 ± 7)	20-39 (27 ± 5)
Pre-menopausal, n (%)	100 (100%)	100 (100%)	100 (100%)
Available HPV genotyping, n (%)	0 (0%)	15 (100%)	30 (100%)
Colposcopy report, n (%)	0 (0%)	0 (0%)	CIN3; 30 (100%)

Table 5.1 – Characteristics of the women included the microbiota characterisation longitudinal study. For age range, mean and standard error are indicated. With n (%), number of women and percentage are reported. Group A: HPV negative women with normal cytology results; Group B: persistently hrHPV

positive women with normal cytology results; Group C: women who were hrHPV positive and cytologically normal at the initial screen, developed CIN3 at the first follow-up screen and underwent local excision treatment prior to the second follow-up screen.

The investigation of cervicovaginal microbiota was performed via metagenomics sequencing of the V2, V3, V4, V6, V7, V8 and V9 hypervariable regions of the 16S rRNA gene. In total 15 846 102 reads were generated from 177 samples with an average number of reads per sample of 89 529 and mean and median lengths of 232bp and 259bp respectively. To avoid sequencing bias, operational taxonomic units (OTUs) were assigned for each sample. A total of 236 OTUs were identified, with a mean of 40.4 OTUs per sample.

Initial assessment of the cervicovaginal microbiota structure in the context of HPV status and disease grade was performed by using principal component analysis (PCA) based on Bray Curtis dissimilarities (Figure 5.1). The five major clusters identified correlated with bacterial communities dominated by *L. crispatus, L. gaseri, L. iners, L. jensenii* and a depletion of *Lactobacillus spp*. (<60% of *Lactobacillus* relative abundance per sample) combined with a higher bacterial diversity of anaerobic mixed species.

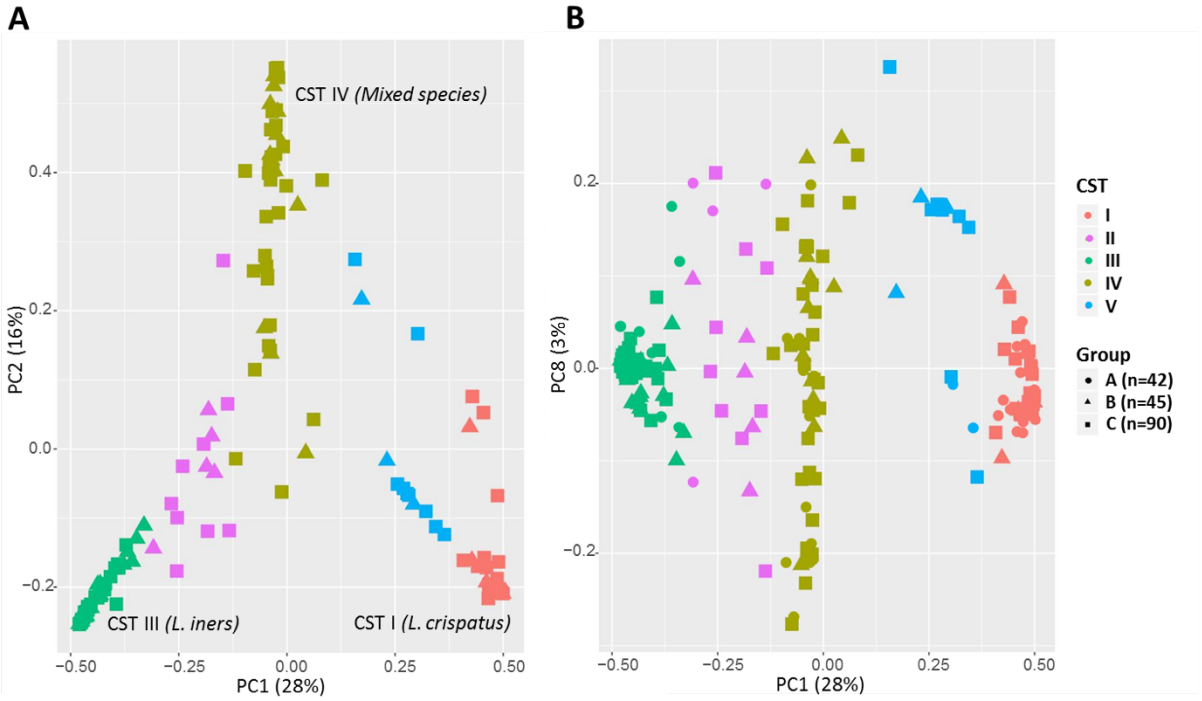

Figure 5.1 – Bacterial species beta diversity in the longitudinal study. Principal component analysis (PCA) based on Bray Curtis dissimilarities identified 5 major clusters corresponding to samples dominated by the following bacteria*: L. iners* (CST III), *L. crispatus* (CST I), *L. jensenii* (CST V), mixed species (CST IV) and *L. gasseri* (CST II). Samples belonging to different CSTs are indicated with different colour dots. **A**) Shows PCA analysis comparing principal component (PC) 1 to PC2, while **B**) shows PC1 vs PC8 for a more comprehensive view of the sample distribution. Abbreviations: CST: Community state type.

Hierarchical clustering analysis (HCA) using Euclidian centroid clustering of the most abundant 37 species also identified five major bacterial clusters. These showed a bacterial community structure in accordance with the community state types (CST); CST I: *L. crispatus*-dominant; CST II: *L gasseri*-dominant; CST III: *L. iners*-dominant; CST IV: *Lactobacillus spp.* depleted combined with a higher diversity of mixed species; CST V: *L. jensenii*-dominant (Figure 5.2).

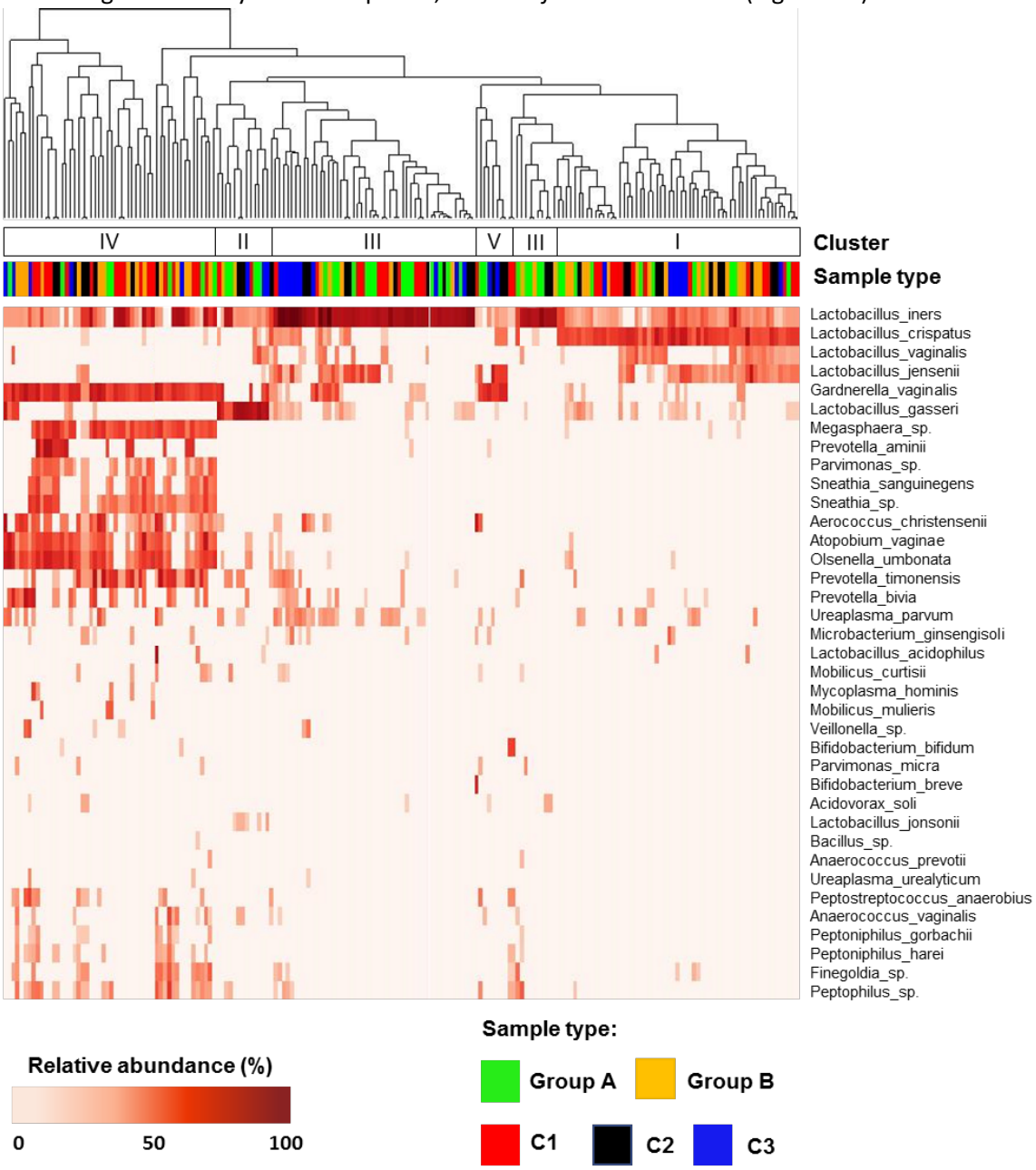

Figure 5.2 – Cervicovaginal microbiota composition based on relative bacterial abundance cording to disease and HPV status. Hierarchical clustering analysis (HCA) using Euclidian centroid clustering of the most abundant 37 species showed the community state type (CST) clustering for HPV and cytology negative women (group A - green), persistently HPV positive and cytology negative women (group B - orange) and for HPV positive, cytology negative women (C1 - red) who developed CIN3 (C2 - black) and who subsequently underwent LEEP and tested HPV negative (C3 - blue). Abbreviations: Group A: HPV negative women with normal cytology results; Group B: persistently hrHPV positive women with normal cytology results; C1: subgroup of C, including hrHPV women; C2: subgroup of C, including CIN3 diagnosed women; C3: subgroup of C, including women recovering post-LEEP treatment;

The rates and distribution of the different CSTs in groups A, B, and C are presented in Table 5.2. CST III (54/177, 30.5% of all samples) was the most frequent CST in the sample set, followed by CST I (53/177, 29.9%), CST IV (42/177, 23.6%), CST II (20/177, 11.2%) and CST V (7/177, 3.8%). In group A, the most frequent microbial state was CST III (21/72, 29.2%), while CST I was the most frequent in group B (19/75, 25.3%). Although, CST III was overall the most frequent microbial state in group C (26/90, 28.8%), results from the C1 subgroup correlated with group B, showing CST I as the most frequent microbial state (10/30, 33.3%). Conversely, CST IV was the most frequent in C2 (10/30, 33.3%), while CST III was the most frequent in C3 (10/30, 33.3%). However, results did not reach significance, likely due to the sample size of the subgroups.

HPV status was available for all samples included in the study. Table 5.2 shows the CST distribution within the groups and subgroups of samples analysed. CST III was most frequently observed in HPV negative compared to hrHPV positive samples (32/72, 44.4% versus 23/105, 21.9%). Conversely, CST I and CST IV were most frequently observed among hrHPV positive samples (35/105, 33.3% and 32/105, 30.4% respectively). The rate of CST IV was 3 times higher for samples positive for HPV16 (22/58, 37.9%) when compared to HPV18 (2/16, 12.5%) and 1.5 times higher compared to other high-risk oncogenic types (8/31, 25.8%).

	CST I *L. crispatus* n/N (%)	**CST II** *L. gasseri* n/N (%)	**CST III** *L. iners* n/N (%)	**CST IV** *Mixed* n/N (%)	**CST V** *L. jensenii* n/N (%)	**Total** n/N
Group A						
A1	3/14 (21.4)	2/14 (14.2)	7/14 (50.0)	2/14 (14.2)	0/14 (0.0)	14/42 (33.3)
A2	3/14 (21.4)	0/14 (0.0)	7/14 (50.0)	3/14 (21.4)	1/ 14 (7.1)	14/42 (33.3)
A3	4/14 (28.5)	0/14 (0.0)	7/14 (50.0)	3/14 (21.4)	0/14 (0.0)	14/42 (33.3)
Total group	10/42(23.8)	2/42 (4.8)	21/42(50.0)	8/42 (19.0)	1/42 (2.4)	42/42 (100)
P value	1.00	0.48	1.00	1.00	1.00	-
Q value	1.00	1.00	1.00	1.00	1.00	-
Group B						
B1	6/15 (40.0)	0/15 (0.0)	2/15 (13.3)	4/15 (26.6)	3/15 (20.0)	15/45 (33.3)
B2	6/15 (40.0)	0/15 (0.0)	4/15 (26.6)	5/15 (33.3)	0/15 (0.0)	15/45 (33.3)
B3	7/15 (46.6)	2/15(13.3)	1/15 (6.6)	4/15 (26.6)	1/15 (6.6)	15/45 (33.3)
Total group	19/45(42.2)	2/45 (4.5)	7/45 (15.5)	13/45 (28.9)	4/45 (8.9)	45/45 (100)
P value	1.00	0.48	0.34	1.00	0.32	-
Q value	1.00	1.00	1.00	1.00	0.87	-
Group C						
C1	10/30(33.4)	3/30 (10.0)	8/30 (26.6)	8/30 (26.6)	1/30 (3.4)	30/90 (33.3)
C2	6/30 (20.0)	6/30 (20.0)	8/30 (26.6)	10/30 (33.4)	0/30 (0.0)	30/90 (33.3)
C3	8/30 (26.6)	7/30 (23.3)	10/30 (33.4)	3/30 (10.0)	2/30 (6.7)	30/90 (33.3)
Total group	24/90 (26.6)	16/90 (17.8)	26/90 (28.9)	21/90 (23.3)	2/90 (2.2)	90/90 (100)
P value	0.76	0.31	0.79	0.20	0.49	-
Q value	0.87	0.74	1.00	0.65	1.00	-
HPV status						
Negative	18/72 (25.0)	10/72 (13.8)	31/72 (43.0)	11/72 (15.3)	2/72 (2.7)	72/177 (40.6)
Positive	35/105(33.3)	11/105(10.5)	23/105(21.9)	32/105(30.4)	4/105 (3.8)	105/177 (59.4)
Total group	53/177(29.9)	21/177(11.9)	54/177(30.5)	43/177(24.3)	6/177 (3.4)	177/177 (100)
P value	0.42	0.64	0.04	0.08	1.00	-
Q value	1.00	1.00	1.00	0.68	1.00	-
HPV genotype						
HPV 16	19/58 (32.7)	4/58 (6.9)	13/58 (22.4)	22/58 (37.9)	0/58 (0.0)	58/105 (55.2)
HPV 18	10/16 (62.5)	2/16 (12.5)	2/16 (12.5)	2/16 (12.5)	0/16 (0.0)	16/105 (15.2)
Other hr-HPV	6/31 (19.4)	5/31 (16.1)	8/31 (25.8)	8/31 (25.8)	4/31 (12.9)	31/105 (29.6)
Total group	35/105(33.3)	11/105(10.5)	23/105(21.9)	32/105(30.4)	4/105 (3.8)	105/105 (100)
P value	0.62	0.27	1.00	0.48	0.01	-
Q value	0.67	1.00	0.24	0.19	0.24	-

Table 5.2 – Rates of CST in accordance to disease severity and HPV status/genotype at a species level. CST: Community state type; CST I = Lactobacillus crispatus dominant, CST II = Lactobacillus gasseri-dominant; CST III = Lactobacillus iners-dominant; CST IV = high-diversity and Lactobacillus spp-depleted; CST V = Lactobacillus jensenii-dominant; HPV: human papilloma virus; hrHPV: high-risk human papilloma virus; n: number of samples in a specified category; N: total number of samples in a specified group. P-value and Q-value calculated using linear regression. P-value was calculated to assess subgroup similarities using Fishers exact test, Q value was calculated using Benjamini-Hochberg false discovery rate (FDR) method.

5.3 THE CERVICOVAGINAL MICROBIAL DIVERSITY ASSOCIATED WITH DISEASE AND HPV STATUS

We measured the alpha diversity, based on the number of observed species (OTUs), non-parametric Shannon and Inverse Simpson indices in the CSTs identified. A significantly higher number of species were observed in samples classified as CST IV compared to CST I (P<0.0001), CST II (P<0.0001), CST III (P<0.0001) and CST V (P<0.001). Diversity was also significantly higher in CST IV classified samples as assessed by the non-parametric Shannon and Inverse Simpson indices compared to CST I (P<0.0001), CST II (P<0.0001) and CST III (P<0.0001) (Figure 5.3). Microbial diversity was observed to be higher in women showing persistent hrHPV positivity (B2 and B3) and in CIN3-diagnosed women (C2). However, there were no statistical differences between any of the data collection points, due to the high variability observed among samples and the modest number of samples included in each group.

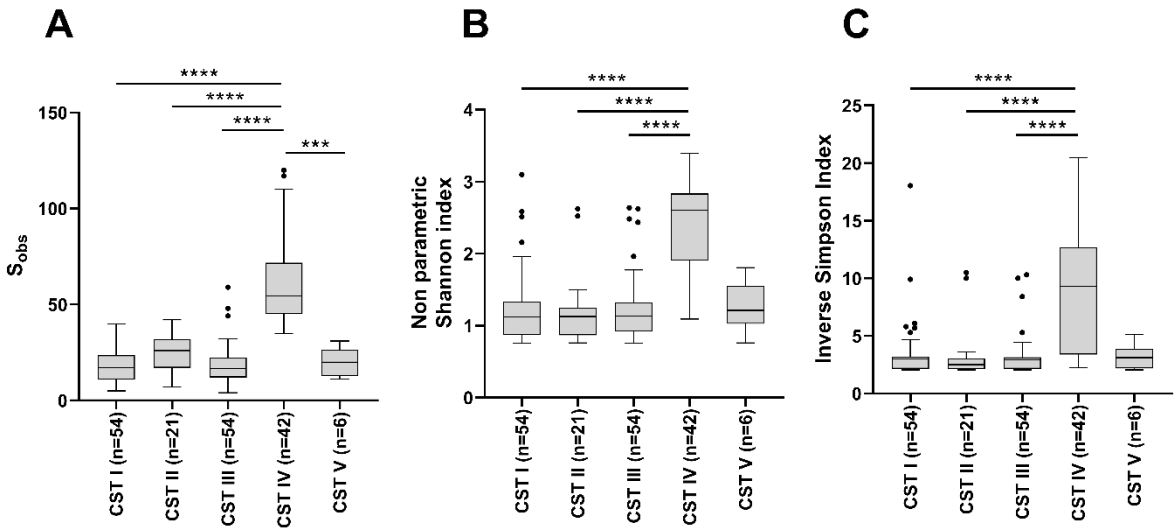

Figure 5.3 – Cervicovaginal microbiota richness and diversity with attributed community state types from samples in the longitudinal study. A significantly higher number of species observed in samples classified as CST IV compared to CST I (P<0.0001), CST II (P<0.0001), CST III (P<0.0001) and CST V (P<0.001) (**A**). Diversity was also significantly higher in CST IV classified samples as assessed by the non-parametric Shannon (**B**) and Inverse Simpson (**C**) indices compared to CST I (P<0.0001), CST II (P<0.0001), CST III (P<0.0001). Analysis was performed using Kruskall-Wallis and Dun's multiple comparisons tests. Only significant differences were shown: ***=P<0.001, ****=P<0.0001. Abbreviations: CST: Community state type, S_{obs}: Species observed.

5.4 THE CERVICOVAGINAL MICROBIOTA DYNAMICS LEADING TOWARDS CIN3 DEVELOPMENT AND RECOVERY AFTER LEEP TREATMENT

To understand the dominant CST transitions leading to CIN3 development and recovery of the cervix after LEEP treatment, the transition probabilities between CSTs were investigated using Markov chain analysis. A Markov analysis is a stochastic model describing a sequence of possible transitions between one or more states, in which the probability of each transition depends only on the state attained in the previous event. Additionally, to further investigate the general stationary distribution of CSTs, iterations of 1000 simulations based on the Markov models were performed, assuming that the transition probabilities between states was constant**.**

Considering the heterogenous nature of the patients' characteristics in group C, the Markov analysis was split in two subgroups. First, to understand the cervicovaginal microbiota dynamics towards CIN3 development, the CST transitions between C1 and C2 were investigated (Figure 5.4 A). The main transition flux observed was among CST I-IV, with no transitions or from CST V. Although the results of the analysis did not attain significance given the modest sample size, there appears to be a correlation of CST IV and increasing disease severity. CST IV was observed to be the most prevalent cervicovaginal microbiota state, encouraging CIN3 development. CST IV showed the highest probability for transitioning back into itself (76.0%). Additionally, the Markov simulations which reached CST IV after 75.0% of iterations.

Secondly, to investigate the cervicovaginal microbiota states associated with recovery after LEEP treatment and hrHPV clearance, the CST transitions between C2 and C3 were investigated (Figure 5.4 B). The main transition flux was observed to be among CST I-V with CST II (*L. gasseri* dominant) showing a high probability for transitioning back into itself (49.0%). Additionally, all other states (CST I, III, IV and V) showed transitions towards CST II. While both CST I and CST III showed high stability (probability for transitioning back into CST I: 83.0%; and CST III: 76.0%), CST IV showed a trend of transitioning towards other CSTs than transitioning back to itself. Moreover, the Markov simulations highlighted CST I and CST II as the dominant states for recovery after LEEP treatment and hrHPV clearance (CST I resulted after 46.0% iterations while CST II resulted after 35.0% iterations).

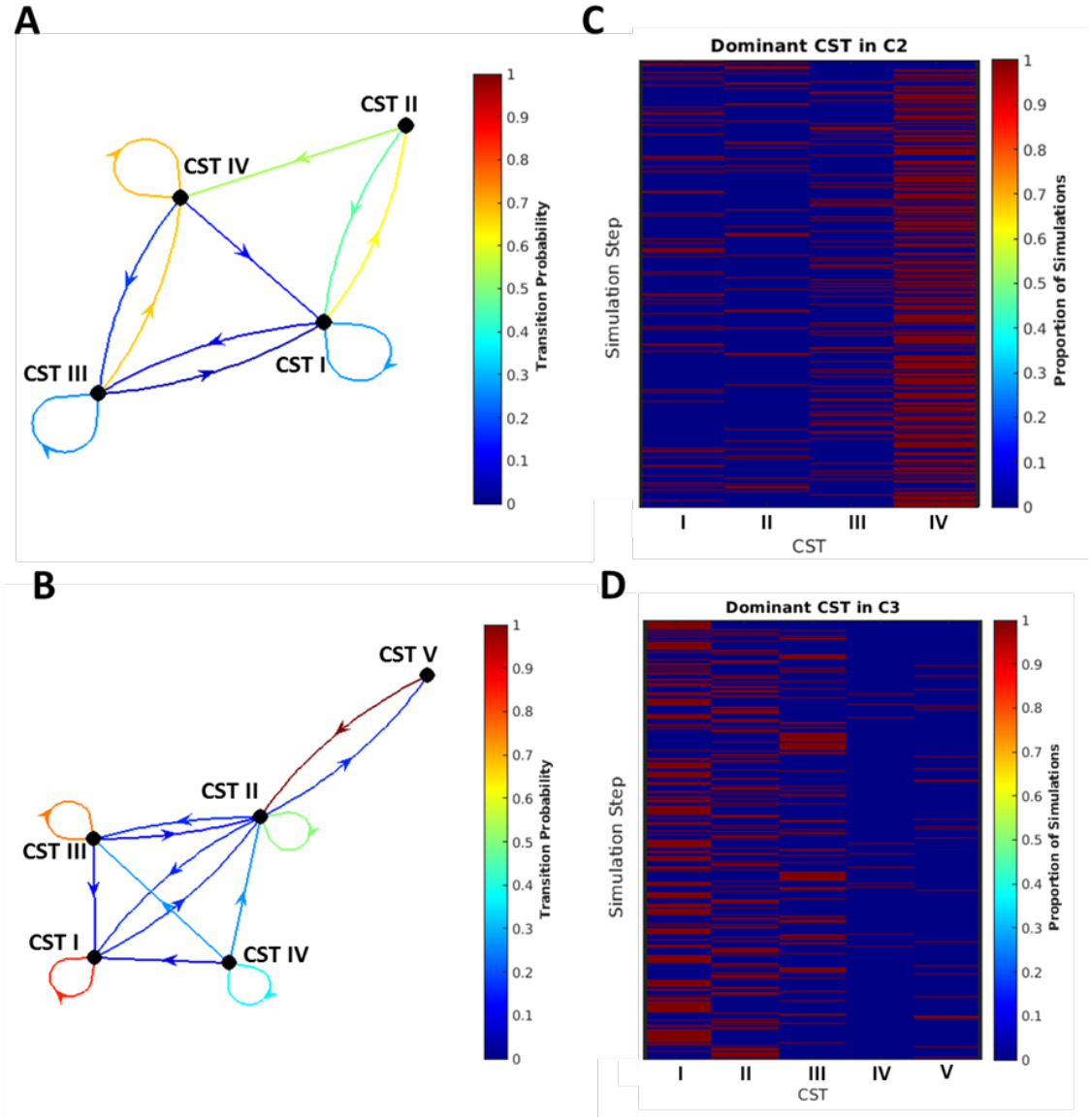

Figure 5.4 – Transition probabilities and stationary distribution of bacterial community state types in group C. The transition probabilities were calculated using Markov chain analysis for paired HPV-positive with normal cytology and CIN3 diagnosed women i.e. C1 vs C2 **(A);** and between paired CIN3 diagnosed and HPV-negative post-LEEP women i.e. C2 vs C3 **(B).** The two diagrams describe a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. Depending on the bacterial composition, four and five CSTs, respectively, have been identified to transition from one to another. The transitions between states are represented by arrows. The probability of one state to transition into another is colour coded according to the legend on the right-side of the Markov chain models. Assuming that the transition probabilities between CSTs remains constant, **C** and **D** show the CST reached after simulating of the corresponding Markov chain models for C2 and C3, respectively. Simulations were performed 1000 times. In the long run, the stationary distribution of CSTs showed CST IV as dominant in C2 (**C**). CST I was found dominant in C3, followed by CST II and CST III (**D**). For the simulations, transition probabilities from one CST to another were multiplied by the percentages of the corresponding CST outcomes in C2 and C3, respectively. Analysis was performed in Matlab Online R2019b. Abbreviations: C1: subgroup of C, including hrHPV women; C2: subgroup of C, including CIN3 diagnosed women; C3: subgroup of C, including women recovering post-LEEP treatment; CST: Community state type.

5.5 IDENTIFICATION OF CERVICOVAGINAL MICROBIOTA MARKERS FOR CIN3 DEVELOPMENT

To identify the unique differences in the cervicovaginal microbiota composition over time where CIN3 had developed, we used linear discriminant analysis (LDA) effect size (LEfSe) modelling. We compared the sequencing results in the three groups with an emphasis on group C. We compared women pre-CIN3 (C1) to the paired data point for CIN3 diagnosis (C2) and post-LEEP treatment (C3). Additionally, we compared C2 with cytology normal women showing persistent hrHPV positivity over six years (B3) and the reference cytology normal women showing hrHPV negativity (group A).

Within group C, we identified unique species associated with CIN3 development and recovery after LEEP-treatment. Compared to the other subgroups, women in C2 showed a significant overrepresentation of the following microbial species: *S. amnii* ($p<0.01$), *M. genomosp.* ($p<0.01$), *P. anaerobius* ($p<0.05$), *A. spanius* ($p<0.05$). Conversely, women in C3 were found to have a significant overrepresentation of *Lactobacillus* species, especially *Lactobacillus gasseri* ($p<0.01$), which was also associated with hrHPV clearance (Figure 5.5).

To investigate the unique differences in the cervicovaginal microbiotas of women showing persistent hrHPV positivity and normal cytology results versus those who developed CIN3, we compared B3 with C2. Consistent with the previous results, *S. amnii* was identified as a microbial biomarker for CIN3 development ($p<0.01$). Conversely, *Lactobacillus* spp. were significantly overrepresented in persistently hrHPV positive women: *L. helveticus* ($p<0.01$), *L. suntoryeus* ($p<0.01$) and *L. vaginalis* ($p<0.01$) (Figure 5.6).

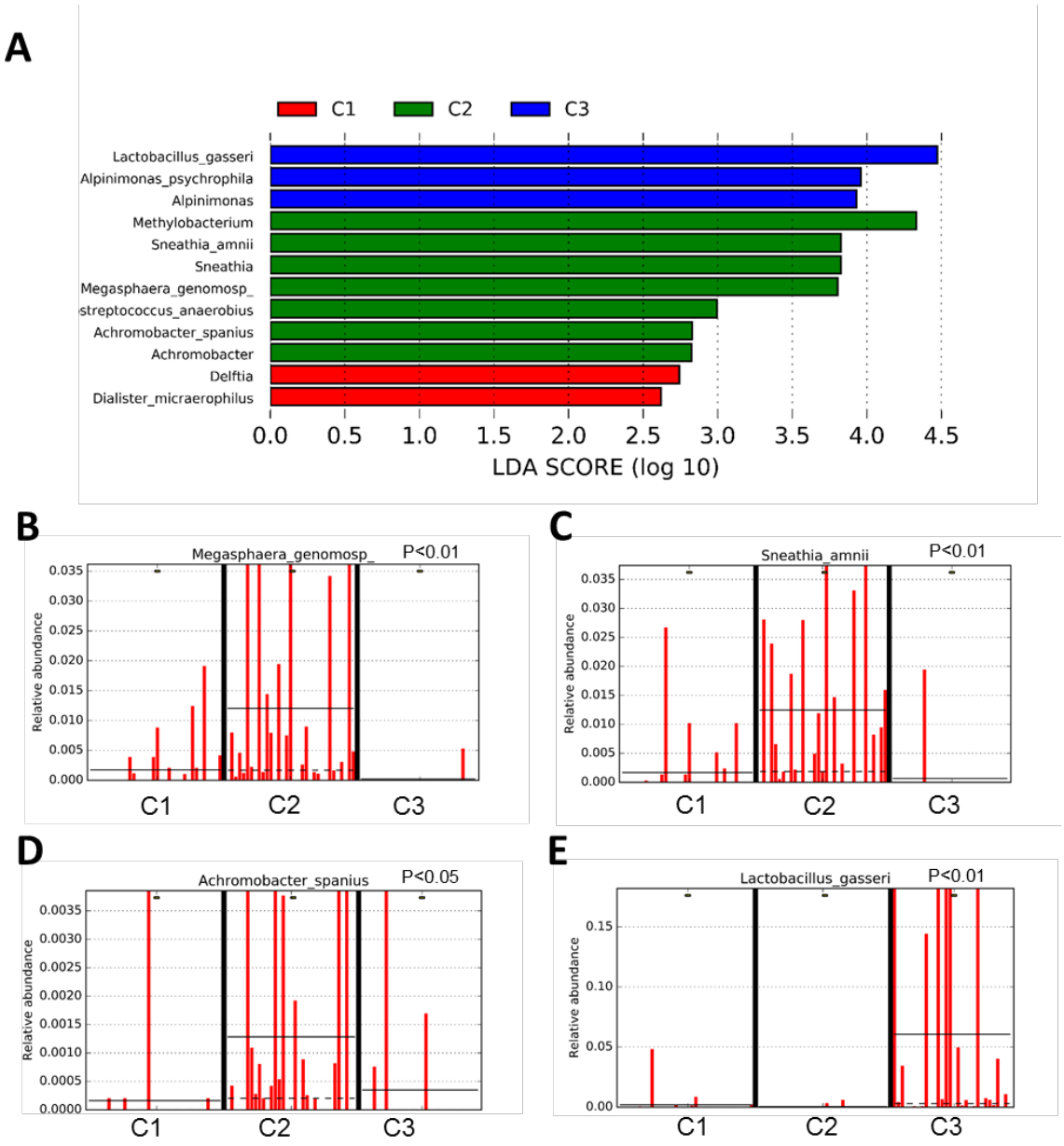

Figure 5.5 – CIN3 progression microbial biomarkers by LEfSe analysis. (A) Histogram of the LDA scores computed for features differentially abundant in each subgroup (C1, C2, C3). Relative abundance counts of *Megasphaera genomosp. (***B***)*, *Sneathia amnii (***C***)* and *Achromobacter spanius (***D***)* which were found to be significantly overexpressed in C2, whereas *Lactobacillus gasseri* **(E)** was enriched in C3 (Welch's t-test). The threshold for the logarithmic LDA score was 2.0. Abbreviations: C1: subgroup of C, including hrHPV women; C2: subgroup of C, including CIN3 diagnosed women; C3: subgroup of C, including women recovering post-LEEP treatment; LDA score: Linear discriminant analysis score; LEfSe: Linear discriminant analysis effect size.

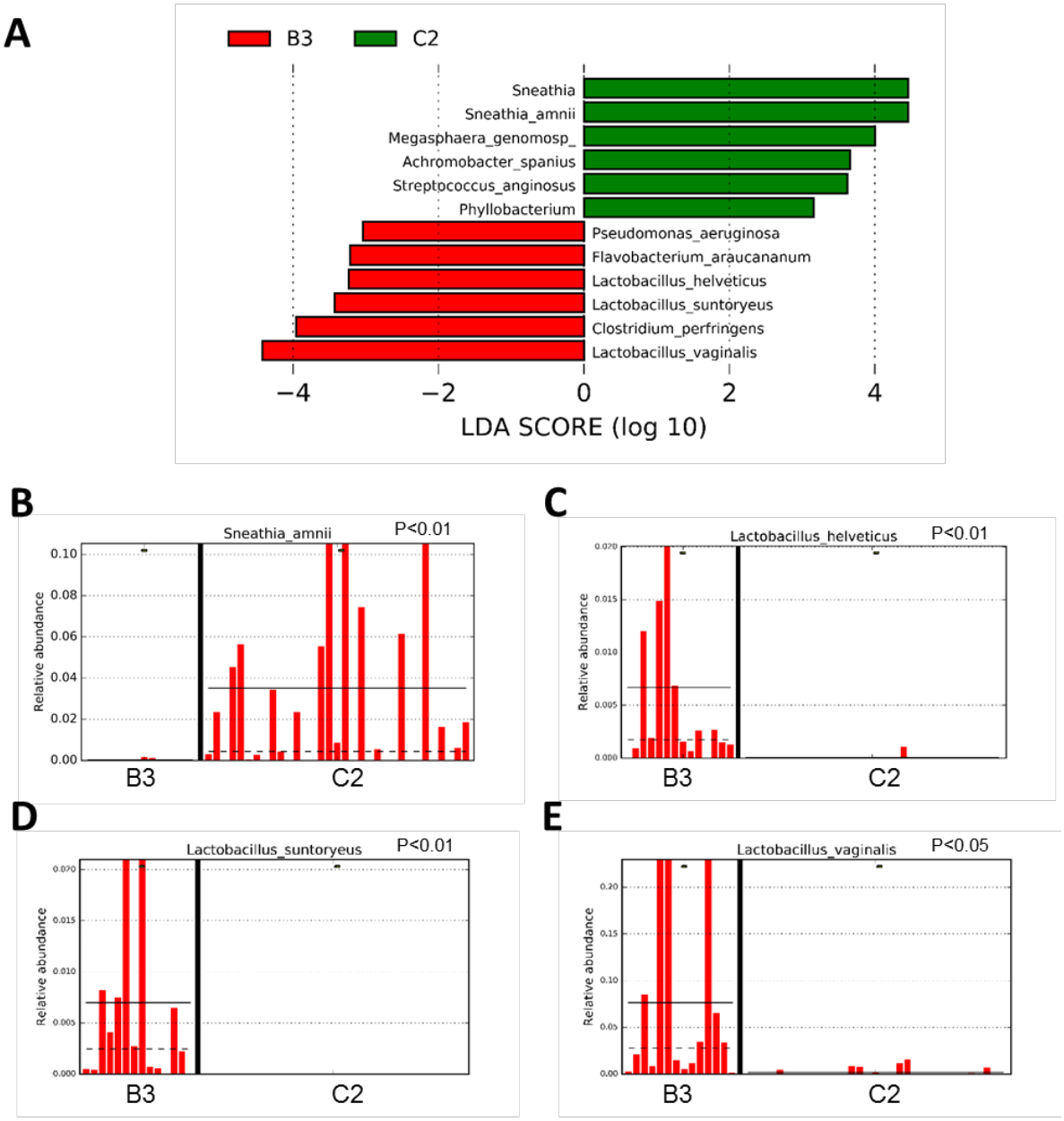

Figure 5.6 – HPV persistence microbial biomarkers by LEfSe analysis. The discovery analysis was performed by comparing persistently HPV positive women with normal cytology (B3) vs CIN3 diagnosed women (C2). **(A)** Histogram of the LDA scores computed for features differentially abundant in each subgroup (C2 vs B3). Relative abundance counts of *Sneathia amnii* **(B)** which was found to be significantly overexpressed in C2, whereas *Lactobacillus helveticus* **(C),** *Lactobacillus suntoryeus* **(D)** and *Lactobacillus vaginalis* **(E)** were enriched in B3 (Welch's t-test). The threshold for the logarithmic LDA score was 2.0. Abbreviations: B3: subgroup of B, including persistently hrHPV positive (over 6 years) women with normal cytology results; C2: subgroup of C, including CIN3 diagnosed women; LDA score: Linear discriminant analysis score. LEfSe: Linear discriminant analysis effect size.

5.6 CORRELATIONS BETWEEN THE S5 CLASSIFIER SCORES AND BACTERIAL ABUNDANCES

The S5-classifier methylation score was successfully measured in all 59 women included in the study. The distribution of the S5 scores in each group included in the study (Figure 5.7). S5 score was low as expected for group A and increased with disease progression, a significant decrease was seen after treatment i.e. between group C2 and C3. Kruskall-Wallis and Dunn's multiple comparisons tests show significant separation in the following paired comparisons: A1 vs B2 ($p<0.01$), A1 vs B3 (all $p<0.001$), A1 vs C1 ($p<0.0001$), A1 vs C2 ($p<0.0001$), A1 vs C3 ($p<0.01$), A2 vs C2 ($p<0.001$), A2 vs C3 ($p<0.0001$), A3 vs C1 ($p<0.01$), A1 vs C2 ($p<0.0001$), B1 vs C1 ($p<0.05$), B1 vs C2 ($p<0.01$), C1 vs C3 ($p<0.01$) and C2 vs C3 ($p<0.001$). No other significant differences were identified. The sensitivity for identifying CIN3 was 93.3% at the 0.80 predefined cut-off.

To investigate the relationship between individual bacterial abundancies and the S5 classifier scores we calculated the Spearman correlation coefficients (Figure 5.8). Thresholds of significance were set at the Spearman coefficient ρ (rho) at +/-0.5 as these indicate the threshold for strong correlations. Both positive (i.e. any rho > 0.5) and negative correlation (i.e. rho < -0.5) were taken into account.

Species from the genus *Sneathia s*howed positive relationships with the increase in the S5 score ($rho=0.55$, $p<0.05$). Specifically, the previously identified marker for CIN3 progression, *Sneathia amnii* showed the strongest correlation with the S5 classifier scores ($rho=0.57$, $p<0.05$). Conversely, bacteria from the genuses *Sphingonomas* and *Finegoldia* showed negative relationships with the S5 classifier scores ($rho= -0.50$, $p>0.05$ and $rho= -0.53$, $p<0.05$ respectively). At a species level, *Finegoldia magna* was correlated with the S5 classifier scores ($rho= -0.52$, $p<0.05$).

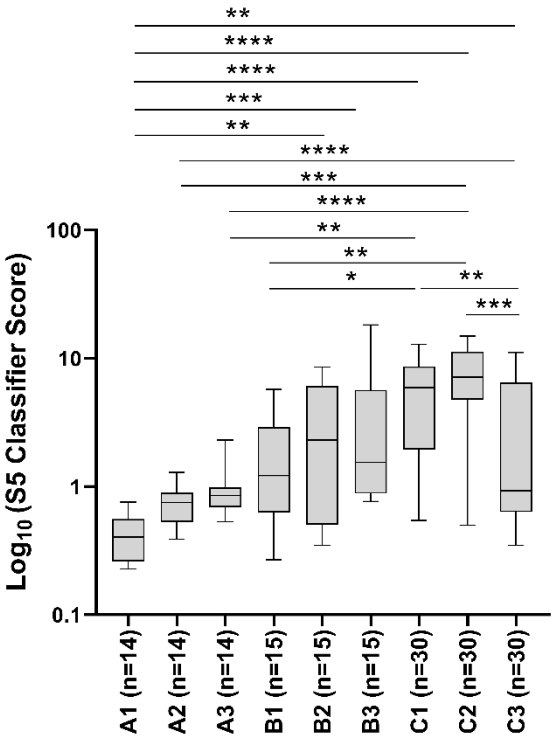

Figure 5.7 – S5 classifier performance in the longitudinal study. Data was plotted as $\log_{10}$ of the S5 score according to group attribution: The top of box represents the upper quartile (p75), bottom the lower quartile (p25), and the line the median (p50). The upper whisker extends to the largest point of the inter-quartile range from the upper quartile. The lower whisker extends to the smallest point of the inter-quartile range from the lower quartile. Analysis was performed using Kruskall-Wallis and Dun's multiple comparisons tests. Only significant differences were shown: *=P<0.05, **=P<0.01, ***=P<0.001, ****=P<0.0001. Abbreviations: A1/2/3: subgroups of A, including HPV negative women with negative cytology results; B1/2/3: subgroups of B, including hrHPV positive women with normal cytology results. C1: subgroup of C, including hrHPV women; C2: subgroup of C, including CIN3 diagnosed women; C3: subgroup of C, including women recovering post-LEEP treatment.

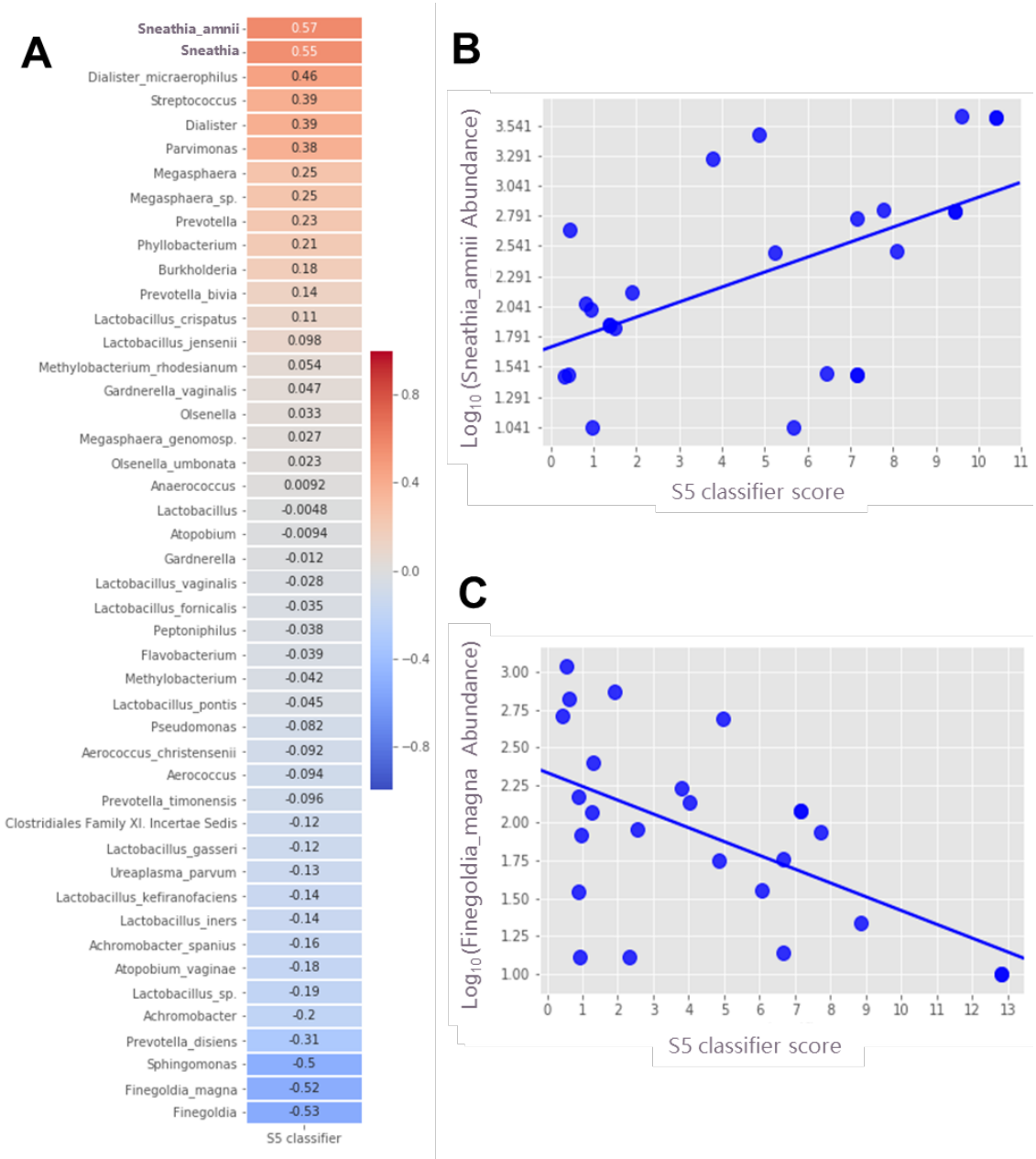

Figure 5.8 – Associations between the S5 classifier score and microbial abundances. Spearman correlations coefficients (rho) were used to assess the relationship between the S5 classifier and $\log_{10}$ of operational taxonomic units (OTUs) or bacterial abundance. This relationship was represented as a heatmap showing both positive (orange) and negative (blue) relationships between the two variables analysed (**A**). Only r^2>0.5 were considered significant. Only two bacterial species have shown significant correlations with the S5 classifier scores. *Sheathia_amnii* abundance increased directly proportional with S5 classifier scores (**B**), while *Finegoldia_magna* abundance decreased directly proportional with the S5 classifier scores (**C**).

5.7 DISCUSSION

The 16S-rRNA metagenomic sequencing was used to assess the cervicovaginal microbiota composition in regard to HPV and cervical disease status. So far, evidence shows a lower abundance of *Lactobacillus spp.*, and higher diversity of anaerobic species in hrHPV positive women compared to HPV negative counterparts. Similar observations were made when investigating disease severity (Freitas et al., 2017; Mitra et al., 2015b; Oh et al., 2015). Similar observations were made when investigating disease severity. To our knowledge, the present study is the first longitudinal investigation following healthy hrHPV positive women who developed CIN3 and underwent LEEP treatment. Further, the study included women who were either hrHPV positive or negative and cytologically normal for comparisons. Data collection was carried over six years and included the analysis of three data points, each approximatively three years apart. All women included in the study were reproductively active, with an overall age at initial screening ranging from 20 to 39 years.

Cervicovaginal microbiota composition was assessed in three separate ways: through principal component analysis (PCA), hierarchical clustering analysis (HCA) and through alpha diversity indices. Both the PCA and HCA identified five clusters, depending on the bacterial composition of each. This finding was in line with the five bacterial community state types (CSTs) described in the literature (Curty et al., 2020; Kyrgiou et al., 2018; Piyathilake et al., 2016). The most frequent CST observed in HPV negative women was the *L. iners*-dominant CST III, while the *L. crispatus*-dominant CST I was most frequently observed in hrHPV positive women with normal cytology results. As described by Kyrgiou *et al.,* the mixed species CST IV was found to be more frequent in women diagnosed with CIN3, highlighting possible associations between a dysbiotic state and CIN3 development (Kyrgiou et al., 2018). The alpha diversity analysis showed a higher overall diversity of CST IV compared to the other CSTs identified. Additionally, no statistical differences were observed based on HPV or disease status, suggesting that there is no specific CST strictly associated with HPV or disease status.

We found that *S. amnii*, *M. genomosp.*, *P. anaerobius* and *A. spanius* were strongly associated to CIN3 development (all $p \leq 0.05$). However, only *S. amnii* was a consistent biomarker in all group comparisons performed. This finding complements data from Mitra *et al.* who identified another species of the genus *Sneathia*, namely, *S. sanguinegens* as a biomarker for high grade squamous intraepithelial lesions (HSIL) (Mitra et al., 2015b). Further, comparing women who developed CIN3 to those with persistent hrHPV positivity over the length of the study but with negative cytology results, higher proportions of *L. helveticus*, *L. suntoryeus* and *L. vaginalis (all $p \leq 0.05$)*

were found as unique biomarkers. These findings highlight a potential protective role of these *Lactobacillus* species against CIN3 development.

One strength of the study is the novel correlations between cervicovaginal microbial species and the S5 methylation classifier. *S. amnii* and *F. magna* were the species showing the highest correlations with the S5 classifier scores (rho= 0.57 and rho= -0.52, respectively). The increase in *S. amnii* abundance was directly proportional to the increase of S5 classifier scores. As the S5 classifier score increases with cervical disease severity, *S. amnii* might be a marker of disease progression, although further investigations are required. *S. amnii* thrives under anaerobic conditions and has been previously identified as an opportunistic pathogen of the female genito-urinary tract (Petrosino et al., 2010). Conversely, the increase in *F. magna* abundance was indirectly proportional to the increase of S5 classifier scores indicating the potential health benefits associated with *F. magna*. Although, frequently found in the cervicovaginal microbiota, Neumann A *et al*. have recently shown its pathogenic nature (Neumann et al., 2020). When abundant, proteins secreted by *F. magna* promote inflammation and activate neutrophils to release reactive oxygen species as a host-defence mechanism against the pathogen (Neumann et al., 2020).

Although the development of high grade CIN alone was linked to an increased risk of preterm delivery, undergoing LEEP treatment increases this even further. The severity of preterm delivery has been correlated to the cone depth of the surgical excision (Berretta et al., 2013). Alterations in the cervicovaginal microbiota could potentially be one of the mechanisms increasing the risk for preterm delivery. Previous studies have shown that women who suffer spontaneous preterm delivery show altered cervicovaginal microbiotas with an increase of *M. hominis, U. pavrum, G. vagi*nalis and Bacteriodes (Bobdiwala et al., 2020). However, the data does not show any significant differences in the abundances of the named bacteria between cervicovaginal microbiotas of women diagnosed with CIN3 vs post-LEEP treatment. In the present longitudinal study, LEEP did not increase dysbiosis, instead there was a tendency in the reduction of bacterial species present with an increase in *Lactobacillus* spp. three years post-LEEP as described by the Markov analysis. These results possibly reflect the impact of disease removal. These results support recent post-LEEP observations at three-month follow-up screens by Zang *et al*, and at six and twelve months by Wiik *et al.* and Zang *et al.* In particular (Seta et al., 2019; Wiik et al., 2019). Also, the LEfSe analysis reveals *L. gasseri* as a microbial biomarker for LEEP recovery Brotman *et al.* showed that the *L. gasseri*-dominant CST II was associated with the fastest rate of HPV clearance (Brotman et al., 2014). In this study, women who underwent

LEEP also were HPV negative, indicating that the higher abundance of *L. gasseri* might as an indicator of HPV clearance. These observations support the hypothesis of mechanical factors rather than altered cervicovaginal microbiotas to act as possible risk factor for preterm delivery after LEEP.

The limitations of the current study include the lack of ethical diversity and of epidemiological data collected for each woman. The history of recent sexual activity, the use of contraceptives, the menopausal status and the use of postmenopausal hormone replacement therapy was unknown, and these factors may influence the results. Second, the highest diversity and instability of the cervicovaginal microbiota is usually observed at the time of menstruation when oestrogen and progesterone are at their lowest levels within the menstrual cycle. Although the exact time of the last menstruation was unknown, none of the samples were collected during menstruation. Third, although longitudinal by design, the study included a modest sample sizes for all groups analysed. This was a pilot study but highlights the need for more data to be collected for accurate, statistically significant results. Another limitation was not including cervical cancer samples for the longitudinal analysis. Among women monitored in the longitudinal study, the ones who developed CIN3 were immediately recommended LEEP, so monitoring the progression to cancer was not possible.

In conclusion, our pilot longitudinal identified *S. amnii* as a consistent microbial biomarker for CIN3 development. The increase in *S. amnii* abundance was directly proportional to the increase of S5 classifier scores. As the S5 methylation score increases with cervical disease severity, *S. amnii* might play a role in altering the epigenetic landscape of the cervicovaginal space. Further investigations are required to establish a link between the identified microbial signature and its influence on epigenetic mechanisms. Conversely, higher proportions of L. *helveticus*, *L. suntoryeus* and *L. vaginalis* showed a potential protective role against CIN3 development in women with persistent hrHPV infections. Our study also found *L. gasseri* as a possible indicator of LEEP recovery and HPV clearance. However, whether distinct microbial signatures promote disease development/regression or the disease microenvironment modulates its surrounding microbial community still remains unclear. Future studies must include larger numbers of women of different races, ethnicities and socio-economic backgrounds, as considerable differences have been reported in the cervicovaginal microbiota composition among women of different ethnic origins(Łaniewski et al., 2020).

CHAPTER 6 – CONCLUSIONS AND FUTURE DIRECTIONS

6.1 DISCUSSION & FUTURE DIRECTIONS

6.1.1 Strategies for cervical cancer prevention worldwide

Early 2020 the World Health Organisation (WHO) set targets to eliminate cervical cancer by 2030. These include having: 90% of girls vaccinated against hrHPV by the age of 15; 70% women screened at least twice between 35 to 45 years of age and 90% women diagnosed with either CIN or cancer to be given the appropriate healthcare (Sung et al., 2021). Currently, cervical cancer incidence rates are disproportionally high in LMICs compared to developed countries (18.8 vs 11.3 cases per 100,000) (Sung et al., 2021). It is believed that by achieving these three targets we will reach the threshold of 4 cases per 100,000/year, which is the definition established by WHO for cervical cancer elimination (Bray et al., 2018; Sung et al., 2021). However, these goals might be delayed due to the global COVID-19 pandemic. Management of COVID-19 has led to re-distribution of funding and other logistical challenges which limited access to proper cervical cancer screening (Maringe et al., 2020). Also, it is estimated that the global cancer burden will be around 28.4 million cases in 2040, a 47% rise from 2020 (Sung et al., 2021). To cope with the new potential cases of CIN or cancer that may arise, efforts should be placed towards creating reliable screening systems for both patients and the healthcare professionals.

The average cervical screening coverage in the United Kingdom is approximatively 73.6% (Castanon et al., 2018). This is approximatively 44% in LMICs (Lemp et al., 2021). It is essential to increase screening coverage in women aged 30-34 as only a small proportion of this age group would have been vaccinated before becoming sexually active (Castanon et al., 2018). As far as the United Kingdom is concerned, Castanon *et al.* estimated that along with vaccination, introduction of HPV testing (with HPV16/18 genotyping) as the primary screening would reduce age-standardised rates of cervical cancer at ages 25–64 years by 19.0% by 2032 (Castanon et al., 2018). To optimize cost and screening coverage, self-sampling strategies for HPV testing have also been investigated. Self-sampling strategies are based on the dissemination of postal testing kits, which contain swabs for the self-collection of exfoliated vaginal samples or urine as presented by our group (Cadman et al., 2021). Firstly, self-sampling eliminates the costs of visiting a healthcare practitioner. The medical cost for self-collected HPV testing was estimated to be approximatively 1.5 times lower than the one collected by a practitioner (Mezei et al., 2017; Basu et al., 2019). Secondly, Yeh *et al*. showed that self-sampling is well-accepted by

women from a wide variety of countries and ethnic groups, having potential to increase screening coverage (Yeh et al., 2019). Self-sampling strategies might prove a good alternative to reduce the load of face-to-face meetings of healthcare professionals, thus limiting potential air-borne pathogen exchange.

Using HPV testing as the primary screening tool presents several advantages over cytology-based screening, including a higher sensitivity, reduced costs of training and the prospect of lengthening screening intervals for women with negative hrHPV results. The currently proposed triaging method for hrHPV positive women in developed countries is cytology. hrHPV positive women with abnormal cytology results are referred to colposcopy. Cytology triaging depends on the ability of healthcare practitioners to collect good samples and accurately distinguish morphological abnormalities in cells under the microscope. These require patient – healthcare professional contact, extensive training and continuous quality assurance.

A major drawback of cytology is its subjective nature: examining cells under the microscope relies on human judgement to interpret what is seen (Lorincz et al., 2014). Because of this, many more women may be considered morphologically abnormal when they are known to be positive for hrHPV (Lorincz et al., 2014). Evidence for this comes from a Finish trial conducted in 2009, where women under 35 were referred to colposcopy based on cytology triage results (Leinonen et al., 2009). Although showing a higher specificity in identifying CIN2+ than HPV testing, cytology-triaging might not have the capacity to rule out hrHPV positive women without evidence of disease leading to a higher number colposcopy referrals (Mezei et al., 2017; Castanon et al., 2017). Cytology triage can be automation of reading of slides with LBC samples makes cytology analysis less human resource intense. However, this is more likely to occur in high income countries (Kamineni et al., 2019).

Cytology performance in LMICs varies from 33% to 100%, due to inadequate facilities and limited quality assurance along with poor training opportunities for healthcare practitioners (Mezei et al., 2017; Castanon et al., 2017). HPV testing combined with visual inspection with VIA triage was proposed as a cost-effective alternative for screening algorithms in LMICs (Mezei et al., 2017). shows moderate specificity, leading to higher numbers of colposcopy referrals (Mezei et al., 2017). These increases the work-load and costs to public health services, which might come as an issue especially in LMIC, where 85% of the worldwide registered cervical cancer deaths occur (Allen-Leigh et al., 2017).

The current screening algorithms worldwide have been focusing on delaying colposcopy referrals until the optimal moment, involving a minimal number of screens performed in a woman's lifetime. In LMICs, especially, the ability to screen target risk populations to detect high-grade CIN or cancer at early stages could have a big impact on cervical cancer mortality while keeping colposcopy referrals and LEEP referrals minimal. To date, cervical screening targeted the detection of CIN2+. Molecular markers have the potential to advance this beyond the current status quo and provide data on the true carcinogenic potential of hrHPV infections, irrespective of histological disease grade. This allows the prediction of women who are likely to have a progressive infection and/or disease. As local excisional treatment impairs reproductive abilities, avoiding unnecessary treatments has major benefits on women's health.

Compared to the current morphologically-based triage methodologies or VIA, molecular testing offers an objective interpretation of the results. Additionally, the advantage of automatization leads to lower training requirements and interpretational mistakes (Lorincz et al., 2016). Molecular testing can be performed on the same clinician-collected or self-collected specimens used for HPV-testing. The specimens can be transported to the research laboratory as a dry swab eliminating the need for expensive and inconvenient media required for the preservation or morphology as needed in cytology testing (Boers et al., 2014; Snoek, et al., 2018). Although current technologies may not yet be affordable in LMIC, rapid technological advances may allow molecular screening for cervical cancer to be implemented in these settings in the near future.

So far, methylation biomarkers have been investigated for molecular triage purposes. However, for methylation testing to be introduced in the current screening algorithms, their performance has to be better than cytology (for high resource settings) and VIA (for low resource settings). A recent meta-analysis focused on most relevant 43 studies describing the performance of DNA methylation assays in early detection of CIN2+ and cancer (Kelly et al., 2019). In this study, all the genes (host *CADM1, MAL, mir-124-2*, *FAM19A4, POU4F3, EPB41L3, PAX1, SOX1* and the viral HPV16 L1/L2) were pooled at a set specificity of 70.0% which is required for improving hrHPV positive women triage to colposcopy. A triage test, DNA methylation had higher specificity than cytology. The reported sensitivity was 68.6% for CIN2+ and 71.1% for CIN3+ detection (Kelly et al., 2019). Although the sensitivity was lower than cytology (relative methylation sensitivity for CIN2+: 0.81 and relative methylation sensitivity for CIN3+: 0.87), the set specificity was higher (relative cytology specificity for CIN2+: 1.25 and relative cytology specificity for CIN3+: 1.37). Additionally, the pooled CIN2+ sensitivity achieved by DNA methylation analysis was higher than VIA testing in LMICs (68.6% vs 60.3%) (Kelly et al., 2019; Mezei et al., 2017).

6.1.2 The S5 classifier – universal biomarker for CIN3 and cervical cancer detection

In this thesis we validate the performance of the S5 classifier in detecting high-grade CIN and invasive cancers worldwide. To our knowledge, the current study is the most comprehensive appraisal of viral and host cell DNA methylation in invasive cervical cancer to date (Vink et al., 2020). The findings of this thesis highlight that the S5 classifier methylation score increases with disease severity worldwide. The results show a high S5 sensitivity for CIN3 and cervical cancer detection at the United Kingdom predefined and the low and LMICs suggested exploratory cut-offs.

With the adjustment of the bisulfite conversion kit, the S5 classifier shows a good performance, irrespective of sample type used. We found a test sensitivity of 91.2% (186/204) for CIN3 and 99.8% (543/544) for cervical cancer detection at the UK-predefined cut-off of 0.80 which was designed and validated for developed countries such as the United Kingdom (Lorincz et al., 2016). At this cut-off S5 sensitivity for cervical cancer was higher than the sensitivity of HPV genotyping: 95.2% (518/544) (p=0.032). Although HPV infection is an important co-factor in cervical cancer development, a small proportion of cervical cancer samples in the study tested hrHPV negative as confirmed by HPV testing with multiple assays. Though rare, these cases represent a challenge for detection in the current primary hrHPV screening programme. Our data shows that these cancers can be identified up to 96.1% (25/26) at a S5 classifier at a cut-off of 0.8. Regardless of the cut-off examined, performance of the S5 classifier was uniform among the stratified groups: histotype, FIGO stage, hrHPV status, hrHPV type, sample type, age and country of origin.

In a recent study investigating the S5 classifier performance in a Mexican population part of the FRIDA clinical trial, Hernándes-Lopez *et al.* noticed higher levels of methylation in already triaged women. DNA methylation rates were influenced by region-specific environmental factors. So far, several studies have described consistent differences in DNA methylation patterns observed between developed and LMICs (Bock et al., 2016; Dongen et al., 2018; Elliott et al., 2016; Nilsson & Ling, 2017). These observations led to the proposal of an increased S5 cut-off being more beneficial for LMICs. At the S5 cut-off of 3.7, we found a sensitivity of 62.7% (128/204) for CIN3 and 95.7% (521/544) for cervical cancer. These results complement the Mexican - FRIDA trial, where, despite increasing the cut-off to 3.7, the S5 retained a better performance in detecting cancer (n=3) than cytology, HPV16/18 genotyping (Hernández-López et al., 2019). Our data on S5 sensitivity combined with earlier results from several screening studies in the United Kingdom and Canada suggests that the prevalence of HPV infection as well as the screening intervals of

populations can affect disease prevalence due to many factors including treatments and dietary or other lifestyle interventions. Thus, this affects the optimal cut-off of S5, indicating that in LMICs with limited resources, it is more important to compromise sensitivity for an increased specificity of diagnosis to prioritise disease detection.

Increasing the S5 cut-off from 0.8 to 3.7 leads to a 28 fold decrease in the false positive rate, highlighting a possible positive impact on the specificity of the tool. In triaging hrHPV positive women, the estimated specificity for CIN3 detection increased by ~70% for CIN3 detection with the S5 cut-off adjustment (48.8% to 83.3%). A similar uphill trend was observed with cancer detection (50.6% to 83.3%) (Table 4.5). Considering the referral nature of the sample population specificity cannot be translated to a real screening scenario. However, these estimated results fall in line with previously published modelling data showing that S5 methylation testing with a cut-off of 3.7 could reduce colposcopy referrals by up to 50% with no loss of sensitivity for CIN2+ and cancer compared to HPV16/18 genotyping plus cytology (Hernández-López et al., 2019). In the FRIDA study, women testing HPV16/18 positive but showing normal cytology results were called for colposcopy. It was suggested that the use of the S5 classifier as a triage test would have the potential to reduce colposcopy referrals by 50% for a CIN2+ endpoint ($p<0.00001$) or 43% for a CIN3+ endpoint ($p<0.00001$). An increased specificity could lead to a decrease in the number of colposcopy referrals, which would be beneficial for LMICs due to lack of resources available for healthcare infrastructures (Cook et al., 2018; Cuschieri et al., 2019; Hernández-López et al., 2019).

The strategy of adjusting the S5 classifier cut-off depending on the country/region screening needs has been recently investigated for triage purposes in a Chinese cohort. The S5 classifier was used with a cut-off of 2.85 to triage Chinese women with hrHPV positive and/or cytology abnormal results. The sensitivity of S5 CIN3+ was 89.1% and the specificity was 76.6%. The study concluded that S5 could reduce unnecessary colposcopy referrals by 74% with no loss of sensitivity for CIN3+ and avoid overtreatment (Gu et al., 2020).

Several studies have highlighted the value of methylation in predicting HPV persistence and development of high-grade CIN (Bowden et al., 2019; Lorincz et al., 2016; Mirabello et al., 2013). One of the largest longitudinal cohorts including pre-diagnostic samples from Costa Rican women, demonstrated that high methylation of HPV16 L1/L2 was associated with increased risk of future CIN2+, displaying an OR of up to 9.3 (95%CI 2.3–45.1) (Mirabello et al., 2013). Additionally, the advances in NGS technologies allowed the identification of more than 10 host genes showing constant differential methylation in patients with cervical disease versus healthy

controls (Cuschieri et al., 2018). Over the past 10 years, combinations of these biomarkers have been tested for clinical use. However, only the S5 classifier showed evidence of being a potential prognostic test, being able to identify women with progressive CIN2 (Louvanto et al., 2019).

Here we highlight the major weight of the host *EPB41L3* methylation in the S5-classifier score. We showed that the relationship between *EPB41L3* methylation was approximatively 4.5 times stronger for severity of lesion than age ($p<0.0001$). This indicates that host *EPB41L3* methylation might have a strong potential to predict disease progression, independent of increasing natural epigenetic methylation levels occurring with age. This data is in line with previously published findings by Louvanto *et al.* on the S5-classifier which has been proven as a potential prognostic test, being able to identify women with progressive CIN2 (Louvanto et al., 2019). Together, these data may support the prospect of using the S5-classifier as a tool for objectively identifying clinically significant cervical abnormalities and predicting their outcomes.

To date, there are two commercially available methylation assays based on human genes. The GynTect® assay investigates the methylation of *ASTN1, DLX1, ITGA4, RXFP3, SOX17* and *ZNF671* while the QIAsure Methylation Test Kit looks at the promoter hypermethylation of *FAM19A4* and *mir124-2*. In comparison, the S5 classifier utilizes a combination of host and viral methylation targets. The *EPB41L3* gene, was found to have the best performance, compared to a selection of other genes in predictors 1 and 2 discovery studies (Lorincz et al., 2013; Vasilievic et al., 2014). In a study including only 29 cervical cancer patients, the GynTect® assay showed a sensitivity for cervical cancer detection of 100.0% (29/29) (Schmitz et al., 2018). Using liquid based cytology samples from Africa, Asia, Americas, Vink *et al.* reported that the QIAsure sensitivity for cancer detection was 98.3% (510/519) (Vink et al., 2020). The S5 classifier demonstrated a comparable performance to both GynTect® and QIAsure for both cut-offs investigated. Moreover, the S5 sensitivity at the 0.8 cut-off was significantly better compared to QIAsure ($p=0.047$).

As hypermethylation of the HPV L1 and L2 have been found to be associated with CIN and cancer development, customising the HPV components to the S5 classifier, might increase specificity of the tool. HPV distribution worldwide is not uniform. This might be due to the complex geographical and biological interplay between HPV types and host immunogenetic variations. Apart from HPV16/18, HPV31, HPV33 and HPV45 infections are common in Europe and North America (Smith et al., 2007). HPV35, HPV45, HPV52, HPV56, and HPV58 are all more common in HPV-positive women in sub-Saharan Africa than in Europe (all $p<0.001$). In South America, HPV51 and HPV58 are fairly common in HPV-positive women. However, HPV-type distribution

in Asia is heterogeneous, potentially due to the broad geographical and cultural range in Asian populations. For instance, infections with HPV58 are common in Taiwan, but not India (Clifford et al., 2005).

The COVID-19 pandemic points towards a shift to self-sampling for hrHPV primary screening to reduce the burdens on the healthcare professionals and access women who do not respond to screening invitations. Having the possibility to triage hrHPV positive women from the same self-collected specimen would bring many advantages including a reduction in logistical associated to systematic screening as well as eliminating objectivity as self-samples may display distinct signatures compared with physician-collected cervical specimens (Snoek et al., 2019; Stanczuk et al., 2016; Verlaat et al., 2018). As such, the accuracy of DNA methylation tests, including the S5 classifier have been tested in cervical self-samples. A study led by De Strooper *et al.* showed that the QiAsure methylation test can be performed in the same self-collected samples used for primary hrHPV testing. The performance of QiAsure to detect CIN3+ was measured. The sensitivity of the test was 69.4% and the specificity was 76.4%. In combination with HPV16/18 genotyping on the same samples, the sensitivity increased to 84.7% while specificity decreased to 54.9% (Strooper et al., 2016). Further, the ability of S5 classifier to differentiate between <CIN2 and CIN2+ in self-samples and urine was measured in a study including 1105 women attending a colposcopy clinic in London. S5 showed a statistically significant separation between <CIN2 and CIN2+ samples for both urine and cervical self-samples ($p \leq 0.0001$). At the pre-defined cut-off of 0.8, the sensitivity for cervical self-samples was 71% and specificity 68% and for urine samples was 66% and specificity 72% (Cadman et al., 2021).

This present study showed a good performance of the S5 classifier in high-grade precancer and cancer samples. However, the sample excluded CIN1/CIN2 diagnoses which are quite frequent in the general population. Considering this limitation and the current trend emerging, further validation in large population cohorts should be performed to determine acceptability and 'real world' performance of self-sampling S5 testing in diagnosing cervical diseases.

6.1.3 Microbial biomarkers predicting CIN3 progression – a pilot longitudinal study

A recent survey on KOLs' views on future directions of HPV research included the investigation of the cervicovaginal microbiota associated to HPV infection and cervical disease high priority. This thesis looks into potential microbial biomarkers for cervical disease progression in a cross-sectional longitudinal cohort of 59 women. This included 30 healthy hrHPV positive women who developed CIN3 and underwent LEEP treatment. To date, the investigations of microbial

biomarkers have linked anaerobes from the genuses of *Sneathia, Megasphaera*, *Gardnerella* and *Prevotella* to hrHPV positivity and high risk of developing cervical disease (Audirac-Chalifour et al., 2016; Mitra et al., 2015b; Paola et al., 2017). Using LEfSe and machine learning algorithms, this thesis demonstrates that *S. amnii*, *M. genomosp.*, *P. anaerobius* and *A. spanius* are associated with CIN3 development ($p \leq 0.05$). However, only *S. amnii* was a consistent biomarker in all group comparisons performed. These results complement the findings of Mitra *et al.* who proposed another species of the genus *Sneathia*, namely, *S. sanguinegens* as a biomarker for CIN2+ development (Mitra et al., 2015a).

Subsequent analysis of correlations between the cervicovaginal microbiotas of women and their associated S5 classifier scores was in line with the microbial biomarker discovery results. As such, *S. amnii* was one of the species showing the highest correlations with the S5 classifier scores (rho=0.57). The increase in *S. amnii* abundance was directly proportional to the increase of S5 classifier scores. As the S5 classifier score increases with cervical disease severity, *S. amnii* abundance might be a marker of disease progression, although further investigations in larger cohorts are required to validate this result.

S. amnii thrives under anaerobic conditions and has been previously identified as an opportunistic pathogen of the female genito-urinary tract (Wiik et al., 2019). Clinically, *S. amnii* was identified among the most prevalent species in the proinflammatory cervicovaginal microbiota community types and was strongly correlated with proinflammatory cytokines (Anahtar et al., 2015). *S. amnii* produced significant upregulation of IL-1α, IL-β, TNF-α, and IL-8 in human vaginal epithelial cells *in vitro* (Anahtar et al., 2015). Specifically, upregulation of TNF-α might directly damage the columnar epithelial barrier of the endocervix by disrupting tight junctions, which enables both bacterial and viral translocation, leading to susceptibility of viral acquisition (Masson et al., 2015). Additionally, cytokine-induced chronic inflammation was reported to induce hypermethylation of CpG islands on the DNA (Suzuki et al., 2009). Direct links between *S. amnii*-induced inflammation and aberrant methylation of the S5 components has not been proven yet.

Research so far indicated that a *Lactobacillus*-dominated cervicovaginal microbiota is independently associated with a lower risk of disease progression (Mitra et al., 2015b). Conversely, a reduced relative abundance of *Lactobacillus* spp. and an increased frequency of anaerobes is associated with worse outcomes (Oh et al., 2015). *Lactobacillus* species support an acidic cervicovaginal environment. Population-based studies indicated that a low cervicovaginal pH is associated with lower hrHPV positivity rates (Amabebe & Anumba, 2018; Norenhag & Du,

2019; Oh et al., 2015). Motevaseli *et al.* showed that *Lactobacillus spp.* have a cytotoxic effect on HPV18 infected HeLa cells independently of lactate and pH, this effect was not observed in noninfected epithelial cell lines (Motevaseli et al., 2013). This indicates that other factors in addition to pH may contribute to the protective characteristics of *Lactobacillus species.* Such observations have been reinforced by more recent work which demonstrated the inhibitory effect of *Lactobacillus*-derived metabolites on HPV16 infected CaSki cells (Wang, Xu et al., 2018).

Using machine learning modelling to compare women who developed CIN3 versus those with persistent hrHPV infection and negative cytology results, we identified *L. helveticus, L. suntoryeus* and *L. vaginalis* as protective biomarkers. McLean *et al.* observed that the hydrogen peroxide-producing *Lactobacillus* strains retain bactericidal activity (McLean et al., 1996). Specifically, *L. helveticus* KS300, a hydrogen peroxide-producing strain isolated from the human *vagina*, was shown to inhibit the growth and reduce the viability of dysbiosis-associated bacteria *G. vaginalis* and *P. bivia* (Atassi et al., 2006). *L. helveticus KS300 also* adheres to epithelial cervical and vaginal cells, displacing *G. vaginalis,* therefore limiting the pathogenic effects of the latter (Boris et al., 1998). These findings highlight a potential protective role of this *Lactobacillus* species against CIN3 development.

The results of the present longitudinal study are in alignment with previously published cross-sectional studies i.e. *Sneathia* genus was found to be pathogenic while *L. Gasseri* is associated with HPV clearance. However, extensive validation is required for any microbes to be confirmed as biomarkers for disease progression. To achieve this, samples that are properly stored could be collected from the existing or past screening trials for cervicovagianal microbiota to be investigated.

The cervicovaginal microbiome is a dynamic network of microorganisms able to modulate the host's immune responses and induce aberrant DNA methylation patterns to alter gene transcription, rendering an individual more susceptible to HPV infection. However, there is a much more complex relationship between the cervicovaginal microbiota, susceptibility to HPV infection and cervical disease development. The evaluation of the complex interactions can be challenging as the cervicovaginal microbiota shows intra-individual variability. Although microbial biomarker discovery may be key for accurate diagnostic design, it is not enough. Understanding the mechanisms by which the microbial biomarkers contribute to carcinogenesis is important to define causality. It remains essential for future research to include functional assays of pathogenesis such as proteomics, metabolomics and peptidomes. The impact of the cervicovaginal microbiota on host cell functions is likely to arise due to changes in the production

of numerous compounds secreted by the microbiota, which can result in alterations of tissue specific epigenetics, cell signaling and metabolic pathways. Additionally, future research in the field must consider the impact that the viriome may play in this complex relationship. Although, HPV infection leads to dysplastic lesions of the cervix, whole genome shotgun sequencing has been used to demonstrate that there are many other types of virus in the cervicovaginal area (Wylie et al., 2014). It is possible that these viruses could also play a role in the dynamics of cervicovaginal environment and should be investigated in future longitudinal studies.

6.1.4 Integrating DNA methylation and microbial biomarkers

The S5 DNA methylation score is currently derived from bisulfite pyrosequencing data. Bock *et al.* coined this targeted sequencing method as the best for analysing clinical biomarkers due its highly reproducible results when tested on short DNA sequences (Bock et al., 2016). This method requires bisulfite treatment of the DNA to convert unmethylated cytosines to uracil, increasing the complexity of library preparation and the potential of artefacts and biases from incomplete chemical conversion. Additionally, bisulfite pyrosequencing requires a high DNA input for methylation quantification as the chemical treatment leads to a high extent of DNA degradation.

Recent development in sequencing technologies allows for bisulfite-free DNA methylation sequencing. The Oxford Nanopore is an example of such technology. This 'third-generation' sequencing technology records sequence-dependent changes in ionic current over time as single-stranded DNA passes through a protein nanopore inserted in a lipid membrane (Simpson et al., 2017). Just as reproducible patterns in current can be interpreted as specific short DNA sequences, methylated bases can generate reproducible and interpretable signatures, allowing for DNA methylation to be recorded without any chemical treatment of the DNA input.

To date, Nanopore sequencing was reported to achieve a 95% correlation with bisulfite sequencing and it can predict methylation states of 5% more CpGs than bisulfite sequencing (Rand et al., 2017; Simpson et al., 2017). Additionally, recent technological updates describe the potential of Nanopore sequencing to simultaneously provide information on HPV status and cervicovaginal microbiota composition. The approach combines multiplex PCR of HPV16 E6/E7 and full-length 16S rRNA with Nanopore sequencing. Although initial Nanopore studies were performed on limited numbers of samples, metagenomic sequencing results were in 98.4% agreement with the conventionally used 16S rRNA Illumina V4 sequencing method (Helmersen & Aamot, 2020). As Oxford Nanopore technologies are expanding in metagenomic sequencing, HPV genotyping, DNA methylation and microbial biomarkers could be analysed from the same

samples (i.e. self- or clinician collected), on the MinION platform. This thesis has shown that methylation of the S5 components had a similar diagnostic accuracy to that of hrHPV for detecting CIN3+. DNA methylation may be an alternative screening method. Additionally, there are potential associations between vaginal microbial species and CIN3 development. Combining these markers into a multi-marker classifier might lead to an improvement of sensitivity and specificity of triaging or predicting cervical disease. However, this hypothesis needs to be validated in screening trials.

The Oxford Nanopore technologies could be the best platform to integrate a multi-marker classifier. Additionally, the sequencing technology is in continuous development and the advantage of it being portable could make it useful for cervical cancer screening in areas without access to research laboratories. Costs associated with such a complex multi-marker is currently around $60-80 per self-sample (Helmersen & Aamot, 2020). A cost-effectiveness analysis needs to be performed to demonstrate potential applicability and cost-utility across different health systems.

6.2 CONCLUDING REMARKS

Here we confirm that the S5 DNA-methylation classifier can accurately detect CIN3 and cervical cancers. A fundamental aspect of our test is its ability to prioritise women for treatment. The S5-score is directly proportional to disease severity Therefore a high score would mean 'treat now'. This would help decision making by providing healthcare professionals with tools to prioritise for disease risk. The S5 could be adjusted according to any clinical setting. In a LMIC setting, an increased threshold with a lower number of false-positives would maximise detection of cancer and save resources. The 3.70 cut-off might be best suited for LMICs as this triage method increases specificity and reduces the number of colposcopy referrals which is a desirable outcome in low resource settings.

Although technological advancements may have the potential to revolutionise molecular testing for cervical cancer prevention, based on the preliminary results on cervicovaginal microbiome correlations with the S5 classifier, it is not yet certain whether including microbial markers into the S5 classifier score composition could lead to an increase in the specificity of the biomarker. This thesis presents microbial biomarkers derived from the first longitudinal study following healthy hrHPV positive women who developed CIN3 and underwent LEEP treatment. However, validation of these biomarkers in a larger longitudinal epidemiological study is required to prove a causality. The validation studies must include women of different races, ethnicities and socio-economic backgrounds, as considerable differences have been reported in the cervicovaginal microbiota composition among women of different ethnic origins (Łaniewski et al. 2020).

This thesis demonstrates the superiority of S5 classifier to consistently detect CIN3 and cancer throughout several existing multi-ethnic cohorts. The S5 classifier is a serious contender of cytology and also HPV typing for triage, which will potentially transform current screening programmes. My work provided the team with strong evidences-based data to support S5 utility as triage biomarker for hrHPV positive women and justified its validation in self-collected samples which is a highly transformative step towards cervical cancer eradication.

BIBLIOGRAPHY

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477

Alibegashvili, T., Clifford, G. M., Vaccarella, S., Baidoshvili, A., Gogiashvili, L., Tsagareli, Z., Kureli, I., Snijders, P. J., Heideman, D. A., van Kemenade, F. J., Meijer, C. J., Kordzaia, D., & Franceschi, S. (2011). Human papillomavirus infection in women with and without cervical cancer in Tbilisi, Georgia. *Cancer epidemiology*, *35*(5), 465–470. https://doi.org/10.1016/j.canep.2010.12.006

Allen-Leigh, B., Uribe-Zúñiga, P., León-Maldonado, L., Brown, B. J., Lörincz, A., Salmeron, J., & Lazcano-Ponce, E. (2017). Barriers to HPV self-sampling and cytology among low-income indigenous women in rural areas of a middle-income setting: a qualitative study. *BMC cancer*, *17*(1), 734. https://doi.org/10.1186/s12885-017-3723-5

Alizon, S., Murall, C. L., & Bravo, I. G. (2017). Why Human Papillomavirus Acute Infections Matter. *Viruses*, *9*(10), 293. https://doi.org/10.3390/v9100293

Amabebe, E., & Anumba, D. (2018). The Vaginal Microenvironment: The Physiologic Role of *Lactobacilli*. *Frontiers in medicine*, *5*, 181. https://doi.org/10.3389/fmed.2018.00181

Anahtar, M. N., Byrne, E. H., Fichorova, R. N., Kwon, D. S., Anahtar, M. N., Byrne, E. H., … Yamamoto, H. S. (2015). Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Article Cervicovaginal Bacteria Are a Major Modulator of Host Inflammatory Responses in the Female Genital Tract. *Immunity*, *42*(5), 965–976. https://doi.org/10.1016/j.immuni.2015.04.019

Hodgson, A., & Park, K. J. (2019). Cervical Adenocarcinomas: A Heterogeneous Group of Tumors With Variable Etiologies and Clinical Outcomes. *Archives of pathology & laboratory medicine*, *143*(1), 34–46. https://doi.org/10.5858/arpa.2018-0259-RA

Anthony F Jerant. (1997). Adenocarcinoma of the Uterine Cervix, *10*(1), 36–42.

Arabski, M., Klupinska, G., Chojnacki, J., Kazmierczak, P., Wisniewska-jarosinska, M., Drzewoski, J., & Blasiak, J. (2005). DNA damage and repair in Helicobacter pylori -infected gastric mucosa cells, *570*, 129–135. https://doi.org/10.1016/j.mrfmmm.2004.10.006

Arokiyaraj, S., Seo, S. S., Kwon, M., Lee, J. K., & Kim, M. K. (2018). Association of cervical microbial community with persistence , clearance and negativity of Human Papillomavirus in Korean women : a longitudinal study. *Scientific Reports*, (October), 1–9. https://doi.org/10.1038/s41598-018-33750-y

Atassi, F., Brassart, D., Grob, P., Graf, F., & Servin, A. L. (2006). Lactobacillus strains isolated from the vaginal microbiota of healthy women inhibit Prevotella bivia and Gardnerella vaginalis in coculture and cell culture. *FEMS immunology and medical microbiology*, *48*(3), 424–432. https://doi.org/10.1111/j.1574-695X.2006.00162.x

Au Yeung, C. L., Tsang, W. P., Tsang, T. Y., Co, N. N., Yau, P. L., & Kwok, T. T. (2010). HPV-16 E6 upregulation of DNMT1 through repression of tumor suppressor p53. *Oncology reports*, *24*(6), 1599–1604. https://doi.org/10.3892/or_00001023

Audirac-Chalifour, A., Torres-Poveda, K., Bahena-Román, M., Téllez-Sosa, J., Martínez-Barnetche, J., Cortina-Ceballos, B., López-Estrada, G., Delgado-Romero, K., Burguete-García, A. I., Cantú, D., García-Carrancá, A., & Madrid-Marina, V. (2016). Cervical

Microbiome and Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. *PloS one*, *11*(4), e0153274. https://doi.org/10.1371/journal.pone.0153274

Banister, C. E., Liu, C., Pirisi, L., Creek, K. E., & Buckhaults, P. J. (2017). Identification and characterization of HPV-independent cervical cancers. *Oncotarget*, *8*(8), 13375–13386. https://doi.org/10.18632/oncotarget.14533

Barb, J. J., Oler, A. J., Kim, H. S., Chalmers, N., Wallen, G. R., Cashion, A., Munson, P. J., & Ames, N. J. (2016). Development of an Analysis Pipeline Characterizing Multiple Hypervariable Regions of 16S rRNA Using Mock Samples. *PloS one*, *11*(2), e0148047. https://doi.org/10.1371/journal.pone.0148047

Barros, S. P., & Offenbacher, S. (2009). Epigenetics: connecting environment and genotype to phenotype and disease. *Journal of dental research*, *88*(5), 400–408. https://doi.org/10.1177/0022034509335868

Basu, P., Mahajan, M., Patira, N., Prasad, S., Mogri, S., Muwonge, R., Lucas, E., Sankaranarayanan, R., Iyer, S., Naik, N., & Jain, K. (2019). A pilot study to evaluate home-based screening for the common non-communicable diseases by a dedicated cadre of community health workers in a rural setting in India. *BMC public health*, *19*(1), 14. https://doi.org/10.1186/s12889-018-6350-4

Baylin, S. B., & Jones, P. A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nature reviews. Cancer*, *11*(10), 726–734. https://doi.org/10.1038/nrc3130

Berretta, R., Gizzo, S., Dall'Asta, A., Mazzone, E., Monica, M., Franchi, L., Peri, F., Patrelli, T. S., & Bacchi Modena, A. (2013). Risk of preterm delivery associated with prior treatment of cervical precancerous lesion according to the depth of the cone. *Disease markers*, *35*(6), 721–726. https://doi.org/10.1155/2013/686027

Beskow, A. H., Moberg, M., & Gyllensten, U. B. (2005). HLA class II allele control of HPV load in carcinoma in situ of the cervix uteri. *International journal of cancer*, *117*(3), 510–514. https://doi.org/10.1002/ijc.21204

Beskow, A. H., & Gyllensten, U. B. (2002). Host genetic control of HPV 16 titer in carcinoma in situ of the cervix uteri. *International journal of cancer*, *101*(6), 526–531. https://doi.org/10.1002/ijc.90010

Bhatla, N., Berek, J. S., Cuello Fredes, M., Denny, L. A., Grenman, S., Karunaratne, K., Kehoe, S. T., Konishi, I., Olawaiye, A. B., Prat, J., Sankaranarayanan, R., Brierley, J., Mutch, D., Querleu, D., Cibula, D., Quinn, M., Botha, H., Sigurd, L., Rice, L., Ryu, H. S., … Natarajan, J. (2019). Revised FIGO staging for carcinoma of the cervix uteri. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, *145*(1), 129–135. https://doi.org/10.1002/ijgo.12749

Bhatt, A. P., & Damania, B. (2013). AKTivation of PI3K/AKT/mTOR signaling pathway by KSHV. *Frontiers in immunology*, *3*, 401. https://doi.org/10.3389/fimmu.2012.00401

Bhatt, A. P., Redinbo, M. R., & Bultman, S. J. (2017). The role of the microbiome in cancer development and therapy. *CA: a cancer journal for clinicians*, *67*(4), 326–344. https://doi.org/10.3322/caac.21398

Bierkens, M., Hesselink, A. T., Meijer, C. J., Heideman, D. A., Wisman, G. B., van der Zee, A. G., Snijders, P. J., & Steenbergen, R. D. (2013). CADM1 and MAL promoter methylation levels in hrHPV-positive cervical scrapes increase proportional to degree and duration of

underlying cervical disease. *International journal of cancer*, *133*(6), 1293–1299. https://doi.org/10.1002/ijc.28138

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Bobdiwala, S., Al-Memar, M., Fourie, H., Mannino, R., Lee, Y. S., Smith, A., Marchesi, J. R., Timmerman, D., Bourne, T., Bennett, P. R., & MacIntyre, D. A. (2020). The association between vaginal bacterial composition and miscarriage: a nested case-control study. *BJOG : an international journal of obstetrics and gynaecology*, *127*(2), 264–274. https://doi.org/10.1111/1471-0528.15972

Bock, C., Halbritter, F., Carmona, F. J., Tierling, S., Datlinger, P., Assenov, Y., … Zhang, K. (2016). Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nature Biotechnology*, *34*(7), 726–737. https://doi.org/10.1038/nbt.3605

Boers, A., Bosgraaf, R. P., Leeuwen, R. W. Van, Schuuring, E., Heideman, D. A. M., & Massuger, L. F. A. G. (2014). DNA methylation analysis in self-sampled brush material as a triage test in hrHPV-positive women. *British Journal of Cancer*, *111*(6), 1095–1101. https://doi.org/10.1038/bjc.2014.392

Bordignon, V., Di Domenico, E. G., Trento, E., D'Agosto, G., Cavallo, I., Pontone, M., Pimpinelli, F., Mariani, L., & Ensoli, F. (2017). How Human Papillomavirus Replication and Immune Evasion Strategies Take Advantage of the Host DNA Damage Repair Machinery. *Viruses*, *9*(12), 390. https://doi.org/10.3390/v9120390

Bosch, F. X., Muñoz, N., de Sanjosé, S., Izarzugaza, I., Gili, M., Viladiu, P., Tormo, M. J., Moreo, P., Ascunce, N., & Gonzalez, L. C. (1992). Risk factors for cervical cancer in Colombia and Spain. *International journal of cancer*, *52*(5), 750–758. https://doi.org/10.1002/ijc.2910520514

Bosch, F. X., Manos, M. M., Muñoz, N., Sherman, M., Jansen, A. M., Peto, J., Schiffman, M. H., Moreno, V., Kurman, R., & Shah, K. V. (1995). Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *Journal of the National Cancer Institute*, *87*(11), 796–802. https://doi.org/10.1093/jnci/87.11.796

Bossuet-Greif, N., Vignard, J., Taieb, F., Mirey, G., Dubois, D., Petit, C., Oswald, E., & Nougayrède, J. P. (2018). The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. *mBio*, *9*(2), e02393-17. https://doi.org/10.1128/mBio.02393-17

Boris, S., Suárez, J. E., Vázquez, F., & Barbés, C. (1998). Adherence of human vaginal lactobacilli to vaginal epithelial cells and interaction with uropathogens. *Infection and immunity*, *66*(5), 1985–1989. https://doi.org/10.1128/IAI.66.5.1985-1989.1998

Bowden, S. J., Kalliala, I., Veroniki, A. A., Arbyn, M., Mitra, A., Lathouras, K., Mirabello, L., Chadeau-Hyam, M., Paraskevaidis, E., Flanagan, J. M., & Kyrgiou, M. (2019). The use of human papillomavirus DNA methylation in cervical intraepithelial neoplasia: A systematic review and meta-analysis. *EBioMedicine*, *50*, 246–259. https://doi.org/10.1016/j.ebiom.2019.10.053

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *68*(6), 394–424.

https://doi.org/10.3322/caac.21492

Brentnall, A. R., Vasiljević, N., Scibior-Bentkowska, D., Cadman, L., Austin, J., Szarewski, A., Cuzick, J., & Lorincz, A. T. (2014). A DNA methylation classifier of cervical precancer based on human papillomavirus and human genes. *International journal of cancer*, *135*(6), 1425–1432. https://doi.org/10.1002/ijc.28790

Brentnall, A. R., Vasiljevic, N., Scibior-Bentkowska, D., Cadman, L., Austin, J., Cuzick, J., & Lorincz, A. T. (2015). HPV33 DNA methylation measurement improves cervical pre-cancer risk estimation of an HPV16, HPV18, HPV31 and EPB41L3 methylation classifier. *Cancer biomarkers : section A of Disease markers*, *15*(5), 669–675. https://doi.org/10.3233/CBM-150507

Brisson, M., Kim, J. J., Canfell, K., Drolet, M., Gingras, G., Burger, E. A., Martin, D., Simms, K. T., Bénard, É., Boily, M. C., Sy, S., Regan, C., Keane, A., Caruana, M., Nguyen, D., Smith, M. A., Laprise, J. F., Jit, M., Alary, M., Bray, F., … Hutubessy, R. (2020). Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet (London, England)*, *395*(10224), 575–590. https://doi.org/10.1016/S0140-6736(20)30068-4

Brotman, R. M., Klebanoff, M. A., Nansel, T. R., Yu, K. F., Andrews, W. W., Zhang, J., & Schwebke, J. R. (2010). Bacterial vaginosis assessed by gram stain and diminished colonization resistance to incident gonococcal, chlamydial, and trichomonal genital infection. *The Journal of infectious diseases*, *202*(12), 1907–1915. https://doi.org/10.1086/657320

Brotman, R. M., Shardell, M. D., Gajer, P., Tracy, J. K., Zenilman, J. M., Ravel, J., & Gravitt, P. E. (2014). Interplay between the temporal dynamics of the vaginal microbiota and human papillomavirus detection. *The Journal of infectious diseases*, *210*(11), 1723–1733. https://doi.org/10.1093/infdis/jiu330

Brown, D. R., Kitchin, D., Qadadri, B., Neptune, N., Batteiger, T., & Ermel, A. (2006). The human papillomavirus type 11 E1--E4 protein is a transglutaminase 3 substrate and induces abnormalities of the cornified cell envelope. *Virology*, *345*(1), 290–298. https://doi.org/10.1016/j.virol.2005.09.048

Bryan, J. T., & Brown, D. R. (2000). Association of the human papillomavirus type 11 E1()E4 protein with cornified cell envelopes derived from infected genital epithelium. *Virology*, *277*(2), 262–269. https://doi.org/10.1006/viro.2000.0599

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V, Bukin, S. V, & Zakharenko, A. S. (2019). Data Descriptor : The effect of 16 S rRNA region choice on bacterial community metabarcoding results. *Nature Publishing Group*, *6*, 1–14. https://doi.org/10.1038/sdata.2019.7

Burgers, W. A., Blanchon, L., Pradhan, S., de Launoit, Y., Kouzarides, T., & Fuks, F. (2007). Viral oncoproteins target the DNA methyltransferases. *Oncogene*, *26*(11), 1650–1655. https://doi.org/10.1038/sj.onc.1209950

Cadman, L., Reuter, C., Jitlal, M., Kleeman, M., Austin, J., Hollingworth, T., Parberry, A. L., Ashdown-Barr, L., Patel, D., Nedjai, B., Lorincz, A. T., & Cuzick, J. (2021). A randomised comparison of different vaginal self-sampling devices and urine for human papillomavirus testing - Predictors 5.1. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, cebp.1226.2020. Advance online publication. https://doi.org/10.1158/1055-9965.EPI-20-1226

Cancer, T., & Atlas, G. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, *513*(7517), 202–209. https://doi.org/10.1038/nature13480

Cantalupo, P. G., Katz, J. P., & Pipas, J. M. (2018). Viral sequences in human cancer. *Virology*, *513*, 208–216. https://doi.org/10.1016/j.virol.2017.10.017

Carrington, M., Wang, S., Martin, M. P., Gao, X., Schiffman, M., Cheng, J., Herrero, R., Rodriguez, A. C., Kurman, R., Mortel, R., Schwartz, P., Glass, A., & Hildesheim, A. (2005). Hierarchy of resistance to cervical neoplasia mediated by combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci. *The Journal of experimental medicine*, *201*(7), 1069–1075. https://doi.org/10.1084/jem.20042158

Carvalhopt, D. D. De. (2013). DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell*, *21*(5), 655–667. https://doi.org/10.1016/j.ccr.2012.03.045.DNA

Castanon, A., Landy, R., Pesola, F., Windridge, P., & Sasieni, P. (2018). Prediction of cervical cancer incidence in England, UK, up to 2040, under four scenarios: a modelling study. *The Lancet Public Health*, *3*(1), e16–e23. https://doi.org/10.1016/S2468-2667(17)30222-0

Castanon, A., Landy, R., Michalopoulos, D., Bhudia, R., Leaver, H., Qiao, Y. L., Zhao, F., & Sasieni, P. (2017). Systematic Review and Meta-Analysis of Individual Patient Data to Assess the Sensitivity of Cervical Cytology for Diagnosis of Cervical Cancer in Low- and Middle-Income Countries. *Journal of global oncology*, *3*(5), 524–538. https://doi.org/10.1200/JGO.2016.008011 Torres-ibarra, L., Cuzick, J., Lorincz, A. T., Spiegelman, D., Lazcano-ponce, E., & Franco, E. L. (2019). Comparison of HPV-16 and HPV-18 Genotyping and Cytological Testing as Triage Testing Within Human Papillomavirus – Based Screening in Mexico, 1–15. https://doi.org/10.1001/jamanetworkopen.2019.15781

Castle, P., Feldman, S., & Perkins, R. B. (2018). The Next Generation of Cervical Cancer Screening: Should Guidelines Focus on Best Practices for the Future or Current Screening Capacity?. *Journal of lower genital tract disease*, *22*(2), 91–96. https://doi.org/10.1097/LGT.0000000000000378

Catarino, R., Petignat, P., Dongui, G., & Vassilakos, P. (2015). Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices. *World journal of clinical oncology*, *6*(6), 281–290. https://doi.org/10.5306/wjco.v6.i6.281

Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., … Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, *45*(10), 1113–1120. https://doi.org/10.1038/ng.2764

Chen, D., Enroth, S., Ivansson, E., & Gyllensten, U. (2014). Pathway analysis of cervical cancer genome-wide association study highlights the MHC region and pathways involved in response to infection. *Human molecular genetics*, *23*(22), 6047–6060. https://doi.org/10.1093/hmg/ddu304

Chen, D., Juko-Pecirep, I., Hammer, J., Ivansson, E., Enroth, S., Gustavsson, I., … Gyllensten, U. (2013). Genome-wide association study of susceptibility loci for cervical cancer. *Journal of the National Cancer Institute*, *105*(9), 624–633. https://doi.org/10.1093/jnci/djt051

Chen, D., Juko-Pecirep, I., Hammer, J., Ivansson, E., Enroth, S., Gustavsson, I., Feuk, L., Magnusson, P. K., McKay, J. D., Wilander, E., & Gyllensten, U. (2013). Genome-wide association study of susceptibility loci for cervical cancer. *Journal of the National Cancer Institute*, *105*(9), 624–633. https://doi.org/10.1093/jnci/djt051

Chen, Y., Williams, V., Filippova, M., Filippov, V., & Duerksen-Hughes, P. (2014). Viral carcinogenesis: factors inducing DNA damage and virus integration. *Cancers*, *6*(4), 2155–2186. https://doi.org/10.3390/cancers6042155

Cheung, J. L., Lo, K. W., Cheung, T. H., Tang, J. W., & Chan, P. K. (2006). Viral load, E2 gene disruption status, and lineage of human papillomavirus type 16 infection in cervical neoplasia. *The Journal of infectious diseases*, *194*(12), 1706–1712. https://doi.org/10.1086/509622

Cicchini, L., Westrich, J. A., Xu, T., Vermeer, D. W., Berger, J. N., Clambey, E. T., Lee, D., Song, J. I., Lambert, P. F., Greer, R. O., Lee, J. H., & Pyeon, D. (2016). Suppression of Antitumor Immune Responses by Human Papillomavirus through Epigenetic Downregulation of CXCL14. *mBio*, *7*(3), e00270-16. https://doi.org/10.1128/mBio.00270-16

Clarke, M. A., Luhn, P., Gage, J. C., Bodelon, C., Dunn, S. T., Walker, J., … Wentzensen, N. (2017). Discovery and validation of candidate host DNA methylation markers for detection of cervical precancer and cancer. *International Journal of Cancer*, *141*(4L), 701–710. https://doi.org/10.1002/ijc.30781

Clarke, M. A., Wentzensen, N., Mirabello, L., Ghosh, A., Wacholder, S., Harari, A., … Burk, R. D. (2012). Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *21*(12), 2125–2137. https://doi.org/10.1158/1055-9965.EPI-12-0905

Clifford, G. M., Rana, R. K., Franceschi, S., Smith, J. S., Gough, G., & Pimenta, J. M. (2005). Human papillomavirus genotype distribution in low-grade cervical lesions: comparison by geographic region and with cervical cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *14*(5), 1157–1164. https://doi.org/10.1158/1055-9965.EPI-04-0812

Clifford, G. M., Gallus, S., Herrero, R., Muñoz, N., Snijders, P. J., Vaccarella, S., Anh, P. T., Ferreccio, C., Hieu, N. T., Matos, E., Molano, M., Rajkumar, R., Ronco, G., de Sanjosé, S., Shin, H. R., Sukvirach, S., Thomas, J. O., Tunsakul, S., Meijer, C. J., Franceschi, S., … IARC HPV Prevalence Surveys Study Group (2005). Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis. *Lancet (London, England)*, *366*(9490), 991–998. https://doi.org/10.1016/S0140-6736(05)67069-9

Coleman, J. S., & Gaydos, C. A. (2018). Molecular Diagnosis of Bacterial Vaginosis: an Update. *Journal of clinical microbiology*, *56*(9), e00342-18. https://doi.org/10.1128/JCM.00342-18

Cook, D. A., Krajden, M., Brentnall, A. R., Gondara, L., Chan, T., Law, J. H., Smith, L. W., van Niekerk, D. J., Ogilvie, G. S., Coldman, A. J., Warman, R., Reuter, C., Cuzick, J., & Lorincz, A. T. (2019). Evaluation of a validated methylation triage signature for human papillomavirus positive women in the HPV FOCAL cervical cancer screening trial. *International journal of cancer*, *144*(10), 2587–2595. https://doi.org/10.1002/ijc.31976

Curty, G., de Carvalho, P. S., & Soares, M. A. (2019). The Role of the Cervicovaginal Microbiome on the Genesis and as a Biomarker of Premalignant Cervical Intraepithelial Neoplasia and Invasive Cervical Cancer. *International journal of molecular sciences*, *21*(1), 222. https://doi.org/10.3390/ijms21010222

Cuschieri, K., Lorincz, A. T., & Nedjai, B. (2019). Human Papillomavirus Research: Where Should We Place Our Bets?. *Acta cytologica*, *63*(2), 85–96. https://doi.org/10.1159/000493800

Cuschieri, K., Ronco, G., Lorincz, A., Smith, L., Ogilvie, G., Mirabello, L., … Franceschi, S. (2018). Eurogin roadmap 2017: Triage strategies for the management of HPV-positive women in cervical screening programs. *International Journal of Cancer*. https://doi.org/10.1002/ijc.31261

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & development*, *25*(10), 1010–1022. https://doi.org/10.1101/gad.2037511

Do, H., & Dobrovic, A. (2015). Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clinical chemistry*, *61*(1), 64–71. https://doi.org/10.1373/clinchem.2014.223040

Dongen, J., Bonder, M. J., Dekkers, K. F., Nivard, M. G., van Iterson, M., Willemsen, G., Beekman, M., van der Spek, A., van Meurs, J., Franke, L., Heijmans, B. T., van Duijn, C. M., Slagboom, P. E., Boomsma, D. I., & BIOS consortium (2018). DNA methylation signatures of educational attainment. *NPJ science of learning*, *3*, 7. https://doi.org/10.1038/s41539-018-0020-2

Doorbar J. (2005). The papillomavirus life cycle. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, *32 Suppl 1*, S7–S15. https://doi.org/10.1016/j.jcv.2004.12.006

Doorbar, J., Quint, W., Banks, L., Bravo, I. G., Stoler, M., Broker, T. R., & Stanley, M. A. (2012). The Biology and Life-Cycle of Human Papillomaviruses. *Vaccine*, *30*, F55–F70. https://doi.org/10.1016/j.vaccine.2012.06.083

Drolet, M., Bénard, É., Pérez, N., Brisson, M., & HPV Vaccination Impact Study Group (2019). Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. *Lancet (London, England)*, *394*(10197), 497–509. https://doi.org/10.1016/S0140-6736(19)30298-3

Eijsink, J. J. H., Yang, N., Lendvai, A., Klip, H. G., Volders, H. H., Buikema, H. J., … Van Der Zee, A. G. J. (2011). Detection of cervical neoplasia by DNA methylation analysis in cervico-vaginal lavages, a feasibility study. *Gynecologic Oncology*, *120*(2), 280–283. https://doi.org/10.1016/j.ygyno.2010.10.029

Eijsink, J. J., Lendvai, Á., Deregowski, V., Klip, H. G., Verpooten, G., Dehaspe, L., de Bock, G. H., Hollema, H., van Criekinge, W., Schuuring, E., van der Zee, A. G., & Wisman, G. B. (2012). A four-gene methylation marker panel as triage test in high-risk human papillomavirus positive patients. *International journal of cancer*, *130*(8), 1861–1869. https://doi.org/10.1002/ijc.26326

Elliott, H. R., Walia, G. K., Duggirala, A., Groom, A., Reddy, S. U., Chandak, G. R., Gupta, V., Laakso, M., Dekker, J. M., RISC Consortium, Walker, M., Ebrahim, S., Smith, G. D., & Relton, C. L. (2013). Migration and DNA methylation: a comparison of methylation patterns in type 2 diabetes susceptibility genes between indians and europeans. *Journal of diabetes research & clinical metabolism*, *2*, 6. https://doi.org/10.7243/2050-0866-2-6

Esteller, M., Corn, P. G., Baylin, S. B., & Herman, J. G. (2001). A gene hypermethylation profile of human cancer. *Cancer research*, *61*(8), 3225–3229.

Feinberg, A. P., Koldobskiy, M. A., & Göndör, A. (2016). Epigenetic modulators, modifiers and

mediators in cancer aetiology and progression. *Nature reviews. Genetics*, *17*(5), 284–299. https://doi.org/10.1038/nrg.2016.13

Freitas, A. C., Chaban, B., Bocking, A., Rocco, M., Yang, S., Hill, J. E., Money, D. M., & VOGUE Research Group (2017). The vaginal microbiome of pregnant women is less rich and diverse, with lower prevalence of Mollicutes, compared to non-pregnant women. *Scientific reports*, *7*(1), 9212. https://doi.org/10.1038/s41598-017-07790-9

Ganti, K., Broniarczyk, J., Manoubi, W., Massimi, P., Mittal, S., Pim, D., … Banks, L. (2015). The human papillomavirus E6 PDZ binding motif: From life cycle to malignancy. *Viruses*, *7*(7), 3530–3551. https://doi.org/10.3390/v7072785

Gao, W., Weng, J., Gao, Y., & Chen, X. (2013). Comparison of the vaginal microbiota diversity of women with and without human papillomavirus infection : a cross-sectional study. *BMC Infectious Diseases*, *13*(1), 1. https://doi.org/10.1186/1471-2334-13-271

Gillet, E., Meys, J. F. A., Verstraelen, H., Bosire, C., Sutter, P. De, Temmerman, M., & Broeck, D. Vanden. (2011). Bacterial vaginosis is associated with uterine cervical human papillomavirus infection : a meta-analysis. *BMC Infectious Diseases*, *11*(1), 10. https://doi.org/10.1186/1471-2334-11-10

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual review of biochemistry*, *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Goodwin, E. C., & DiMaio, D. (2000). Repression of human papillomavirus oncogenes in HeLa cervical carcinoma cells causes the orderly reactivation of dormant tumor suppressor pathways. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(23), 12513–12518. https://doi.org/10.1073/pnas.97.23.12513

Gravitt, P. E. (2011). The known unknownsof hpv natural History. *The Journal of Clinical Investigation*, *121*(12), 4599. https://doi.org/10.1172/JCI57149.live

Gravitt, P. E., & Winer, R. L. (2017). Natural History of HPV Infection across the Lifespan: Role of Viral Latency. *Viruses*, *9*(10), 267. https://doi.org/10.3390/v9100267

Gu, W., Li, M., Zhao, W. M., Fang, N. X., Bu, S., Frazer, I. H., & Zhao, K. N. (2004). tRNASer(CGA) differentially regulates expression of wild-type and codon-modified papillomavirus L1 genes. *Nucleic acids research*, *32*(15), 4448–4461. https://doi.org/10.1093/nar/gkh748

Gu, Y. Y., Zhou, G. N., Wang, Q., Ding, J. X., & Hua, K. Q. (2020). Evaluation of a methylation classifier for predicting pre-cancer lesion among women with abnormal results between HPV16/18 and cytology. *Clinical epigenetics*, *12*(1), 57. https://doi.org/10.1186/s13148-020-00849-x

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1016/s0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Harper, D. M., Franco, E. L., Wheeler, C., Ferris, D. G., Jenkins, D., Schuind, A., … Dubin, G. (2004). Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet*, *364*(9447), 1757–1765. https://doi.org/10.1016/S0140-6736(04)17398-4

Heino, P., Zhou, J., & Lambert, P. F. (2000). Interaction of the papillomavirus transcription/replication factor, E2, and the viral capsid protein, L2. *Virology*, *276*(2),

304–314. https://doi.org/10.1006/viro.2000.0342

Helmersen, K., & Aamot, H. V. (2020). DNA extraction of microbial DNA directly from infected tissue: an optimized protocol for use in nanopore sequencing. *Scientific reports*, *10*(1), 2985. https://doi.org/10.1038/s41598-020-59957-6

Henken, F. E., Wilting, S. M., Overmeer, R. M., van Rietschoten, J. G., Nygren, A. O., Errami, A., Schouten, J. P., Meijer, C. J., Snijders, P. J., & Steenbergen, R. D. (2007). Sequential gene promoter methylation during HPV-induced cervical carcinogenesis. *British journal of cancer*, *97*(10), 1457–1464. https://doi.org/10.1038/sj.bjc.6604055

Hernández-López, R., Lorincz, A. T., Torres-Ibarra, L., Reuter, C., Scibior-Bentkowska, D., Warman, R., Nedjai, B., Mendiola-Pastrana, I., León-Maldonado, L., Rivera-Paredez, B., Ramírez-Palacios, P., Lazcano-Ponce, E., Cuzick, J., Salmerón, J., & FRIDA Study Group (2019). Methylation estimates the risk of precancer in HPV-infected women with discrepant results between cytology and HPV16/18 genotyping. *Clinical epigenetics*, *11*(1), 140. https://doi.org/10.1186/s13148-019-0743-9

Hilbert, D. W., Smith, W. L., Chadwick, S. G., Toner, G., Mordechai, E., Adelson, M. E., Aguin, T. J., Sobel, J. D., & Gygax, S. E. (2016). Development and Validation of a Highly Accurate Quantitative Real-Time PCR Assay for Diagnosis of Bacterial Vaginosis. *Journal of clinical microbiology*, *54*(4), 1017–1024. https://doi.org/10.1128/JCM.03104-15

Holl, K., Nowakowski, A. M., Powell, N., McCluggage, W. G., Pirog, E. C., Collas De Souza, S., Tjalma, W. A., Rosenlund, M., Fiander, A., Castro Sánchez, M., Damaskou, V., Joura, E. A., Kirschner, B., Koiss, R., O'Leary, J., Quint, W., Reich, O., Torné, A., Wells, M., Rob, L., … Jenkins, D. (2015). Human papillomavirus prevalence and type-distribution in cervical glandular neoplasias: Results from a European multinational epidemiological study. *International journal of cancer*, *137*(12), 2858–2868. https://doi.org/10.1002/ijc.29651

Holmes, E. E., Jung, M., Meller, S., Leisse, A., Sailer, V., Zech, J., Mengdehl, M., Garbe, L. A., Uhl, B., Kristiansen, G., & Dietrich, D. (2014). Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PloS one*, *9*(4), e93933. https://doi.org/10.1371/journal.pone.0093933

Holmes, K. K., Chen, K. C., Lipinski, C. M., & Eschenbach, D. A. (1985). Vaginal redox potential in bacterial vaginosis (nonspecific vaginitis). *The Journal of infectious diseases*, *152*(2), 379–382. https://doi.org/10.1093/infdis/152.2.379

Hoque, M. O., Kim, M. S., Ostrow, K. L., Liu, J., Wisman, G. B., Park, H. L., Poeta, M. L., Jeronimo, C., Henrique, R., Lendvai, A., Schuuring, E., Begum, S., Rosenbaum, E., Ongenaert, M., Yamashita, K., Califano, J., Westra, W., van der Zee, A. G., Van Criekinge, W., & Sidransky, D. (2008). Genome-wide promoter analysis uncovers portions of the cancer methylome. *Cancer research*, *68*(8), 2661–2670. https://doi.org/10.1158/0008-5472.CAN-07-5913

IARC. (2007). *INTERNATIONAL AGENCY FOR RESEARCH ON CANCER IARC Monographs on the Evaluation of Carcinogenic Risks to Humans VOLUME 90 Human Papillomaviruses*.

Cancer Genome Atlas Research Network, Albert Einstein College of Medicine, Analytical Biological Services, Barretos Cancer Hospital, Baylor College of Medicine, Beckman Research Institute of City of Hope, Buck Institute for Research on Aging, Canada's Michael Smith Genome Sciences Centre, Harvard Medical School, Helen F. Graham Cancer Center &Research Institute at Christiana Care Health Services, HudsonAlpha Institute for

Biotechnology, ILSbio, LLC, Indiana University School of Medicine, Institute of Human Virology, Institute for Systems Biology, International Genomics Consortium, Leidos Biomedical, Massachusetts General Hospital, McDonnell Genome Institute at Washington University, Medical College of Wisconsin, … Washington University in St Louis (2017). Integrated genomic and molecular characterization of cervical cancer. *Nature*, *543*(7645), 378–384. https://doi.org/10.1038/nature21386

Jiang, W., Roemer, M. E., & Newsham, I. F. (2005). The tumor suppressor DAL-1/4.1B modulates protein arginine N-methyltransferase 5 activity in a substrate-specific manner. *Biochemical and biophysical research communications*, *329*(2), 522–530. https://doi.org/10.1016/j.bbrc.2005.01.153

Jones, P. A., & Baylin, S. B. (2002). The fundamental role of epigenetic events in cancer. *Nature reviews. Genetics*, *3*(6), 415–428. https://doi.org/10.1038/nrg816

Kalantari, M., Chase, D. M., Tewari, K. S., & Bernard, H. U. (2010). Recombination of human papillomavirus-16 and host DNA in exfoliated cervical cells: a pilot study of L1 gene methylation and chromosomal integration as biomarkers of carcinogenic progression. *Journal of medical virology*, *82*(2), 311–320. https://doi.org/10.1002/jmv.21676

Karbalaie Niya, M. H., Keyvani, H., Safarnezhad Tameshkel, F., Salehi-Vaziri, M., Teaghinezhad-S, S., Bokharaei Salim, F., Monavari, S., & Javanmard, D. (2018). Human Papillomavirus Type 16 Integration Analysis by Real-time PCR Assay in Associated Cancers. *Translational oncology*, *11*(3), 593–598. https://doi.org/10.1016/j.tranon.2018.02.017

Kelly, H., Benavente, Y., Pavon, M. A., De Sanjose, S., Mayaud, P., & Lorincz, A. T. (2019). Performance of DNA methylation assays for detection of high-grade cervical intraepithelial neoplasia (CIN2+): a systematic review and meta-analysis. *British journal of cancer*, *121*(11), 954–965. https://doi.org/10.1038/s41416-019-0593-4

Kisser, A., & Zechmeister-Koss, I. (2015). A systematic review of p16/Ki-67 immuno-testing for triage of low grade cervical cytology. *BJOG: An International Journal of Obstetrics and Gynaecology*, *122*(1), 64–70. https://doi.org/10.1111/1471-0528.13076

Klingelhutz, A. J., Foster, S. A., & McDougall, J. K. (1996). Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature*, *380*(6569), 79–82. https://doi.org/10.1038/380079a0

Kostrikis, L. G. Chrysostomou, A. C., Stylianou, D. C., Constantinidou, A. (2018). Cervical Cancer Screening Programs in Europe: The Transition Towards HPV Vaccination and Population-Based HPV Testing. *Viruses*, *10*(12), 729. https://doi.org/10.3390/v10120729

Kyrgiou, M., Mitra, A., & Moscicki, A. B. (2017). Does the vaginal microbiota play a role in the development of cervical cancer?. *Translational research : the journal of laboratory and clinical medicine*, *179*, 168–182. https://doi.org/10.1016/j.trsl.2016.07.004

Łaniewski, P., Ilhan, Z. E., & Herbst-Kralovetz, M. M. (2020). The microbiome and gynaecological cancer development, prevention and therapy. *Nature reviews. Urology*, *17*(4), 232–250. https://doi.org/10.1038/s41585-020-0286-z

Lee, J. E., Lee, S., Lee, H., Song, Y. M., Lee, K., Han, M. J., Sung, J., & Ko, G. (2013). Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PloS one*, *8*(5), e63514. https://doi.org/10.1371/journal.pone.0063514

Lei, J., Ploner, A., Lagheden, C., Eklund, C., Nordqvist Kleppe, S., Andrae, B., Elfström, K. M.,

Dillner, J., Sparén, P., & Sundström, K. (2018). High-risk human papillomavirus status and prognosis in invasive cervical cancer: A nationwide cohort study. *PLoS medicine*, *15*(10), e1002666. https://doi.org/10.1371/journal.pmed.1002666.

Lei, J., Ploner, A., Elfström, K. M., Wang, J., Roth, A., Fang, F., Sundström, K., Dillner, J., & Sparén, P. (2020). HPV Vaccination and the Risk of Invasive Cervical Cancer. *The New England journal of medicine*, *383*(14), 1340–1348. https://doi.org/10.1056/NEJMoa1917338

Leinonen, M., Nieminen, P., Kotaniemi-Talonen, L., Malila, N., Tarkkanen, J., Laurila, P., & Anttila, A. (2009). Age-specific evaluation of primary human papillomavirus screening vs conventional cytology in a randomized setting. *Journal of the National Cancer Institute*, *101*(23), 1612–1623. https://doi.org/10.1093/jnci/djp367

Lemp, J. M., De Neve, J. W., Bussmann, H., Chen, S., Manne-Goehler, J., Theilmann, M., Marcus, M. E., Ebert, C., Probst, C., Tsabedze-Sibanyoni, L., Sturua, L., Kibachio, J. M., Moghaddam, S. S., Martins, J. S., Houinato, D., Houehanou, C., Gurung, M. S., Gathecha, G., Farzadfar, F., Dryden-Peterson, S., … Geldsetzer, P. (2020). Lifetime Prevalence of Cervical Cancer Screening in 55 Low- and Middle-Income Countries. *JAMA*, *324*(15), 1532–1542. https://doi.org/10.1001/jama.2020.16244

Leonard, S. M., Wei, W., Collins, S. I., Pereira, M., Diyaf, A., Constandinou-Williams, C., Young, L. S., Roberts, S., & Woodman, C. B. (2012). Oncogenic human papillomavirus imposes an instructive pattern of DNA methylation changes which parallel the natural history of cervical HPV infection in young women. *Carcinogenesis*, *33*(7), 1286–1293. https://doi.org/10.1093/carcin/bgs157

Leontiou, C. A., Hadjidaniel, M. D., Mina, P., Antoniou, P., Ioannides, M., & Patsalis, P. C. (2015). Bisulfite Conversion of DNA: Performance Comparison of Different Kits and Methylation Quantitation of Epigenetic Biomarkers that Have the Potential to Be Used in Non-Invasive Prenatal Testing. *PloS one*, *10*(8), e0135058. https://doi.org/10.1371/journal.pone.0135058

Leventakos, K., Tsiodras, S., Kelesidis, T., Kefala, M., Kottaridi, C., Spathis, A., … Panayiotides, I. (2017). γH2Ax Expression as a Potential Biomarker Differentiating between Low and High Grade Cervical Squamous Intraepithelial Lesions (SIL) and High Risk HPV Related SIL. *PLOS ONE*, *12*(1), e0170626. https://doi.org/10.1371/journal.pone.0170626

Li, X., Tian, R., Gao, H., Yang, Y., Williams, B. R. G., Gantier, M. P., … Gao, Y. (2017). Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. *Scientific Reports*, *7*(1), 16495. https://doi.org/10.1038/s41598-017-16472-5

Ling, Z., Liu, X., Chen, X., Zhu, H., Nelson, K. E., Xia, Y., Li, L., & Xiang, C. (2011). Diversity of cervicovaginal microbiota associated with female lower genital tract infections. *Microbial ecology*, *61*(3), 704–714. https://doi.org/10.1007/s00248-011-9813-z

Litwin, T. R., Clarke, M. A., Dean, M., & Wentzensen, N. (2017). Somatic Host Cell Alterations in HPV Carcinogenesis. *Viruses*, *9*(8), 206. https://doi.org/10.3390/v9080206

Liu, L., Ying, C., Zhao, Z., Sui, L., Zhang, X., Qian, C., … Wu, J. (2018). Identification of reliable biomarkers of human papillomavirus 16 methylation in cervical lesions based on integration status using high-resolution melting analysis. *Clinical Epigenetics*, *10*, 10. https://doi.org/10.1186/s13148-018-0445-8

Liu, Y., Lu, Z., Xu, R., & Ke, Y. (2016). Comprehensive mapping of the human papillomavirus

(HPV) DNA integration sites in cervical carcinomas by HPV capture technology. *Oncotarget*, *7*(5), 5852–5864. https://doi.org/10.18632/oncotarget.6809

Lopez, J., Percharde, M., Coley, H. M., Webb, A., & Crook, T. (2009). The context and potential of epigenetics in oncology. *British journal of cancer*, *100*(4), 571–577. https://doi.org/10.1038/sj.bjc.6604930

Lorincz, A. T. (2011). The Promise and the Problems of Epigenetics Biomarkers in Cancer. *Expert Opinion on Medical Diagnostics*, *5*(5), 375–379. https://doi.org/10.1517/17530059.2011.590129

Lorincz A. T. (2014). Cancer diagnostic classifiers based on quantitative DNA methylation. *Expert review of molecular diagnostics*, *14*(3), 293–305. https://doi.org/10.1586/14737159.2014.897610

Lorincz A. T. (2016). Virtues and Weaknesses of DNA Methylation as a Test for Cervical Cancer Prevention. *Acta cytologica*, *60*(6), 501–512. https://doi.org/10.1159/000450595

Lorincz, A. T., Brentnall, A. R., Scibior-Bentkowska, D., Reuter, C., Banwait, R., Cadman, L., … Vasiljević, N. (2016). Validation of a DNA methylation HPV triage classifier in a screening sample. *International Journal of Cancer*, *138*(11), 2745–2751. https://doi.org/10.1002/ijc.30008

Lorincz, A. T., Brentnall, A. R., Vasiljević, N., Scibior-Bentkowska, D., Castanon, A., Fiander, A., … Sasieni, P. (2013). HPV16 L1 and L2 DNA methylation predicts high-grade cervical intraepithelial neoplasia in women with mildly abnormal cervical cytology. *International Journal of Cancer*, *133*(3), 637–644. https://doi.org/10.1002/ijc.28050

Lorincz, A. T., Nathan, M., Reuter, C., Warman, R., Thaha, M. A., Sheaff, M., … Sasieni, P. (2017). Methylation of HPV and a tumor suppressor gene reveals anal cancer and precursor lesions. *Oncotarget*, *8*(31), 50510–50520. https://doi.org/10.18632/oncotarget.17984

Lorincz, A., Castanon, A., Wey Lim, A. W., & Sasieni, P. (2013). New strategies for human papillomavirus-based cervical screening. *Women's health (London, England)*, *9*(5), 443–452. https://doi.org/10.2217/whe.13.48

Louvanto, K., Franco, E. L., Ramanakumar, A. V., Vasiljević, N., Scibior-Bentkowska, D., Koushik, A., Cuzick, J., Coutlée, F., Lorincz, A. T., & Biomarkers of Cervical Cancer Risk Study Team (2015). Methylation of viral and host genes and severity of cervical lesions associated with human papillomavirus type 16. *International journal of cancer*, *136*(6), E638–E645. https://doi.org/10.1002/ijc.29196

Louvanto, K., Aro, K., Nedjai, B., Bützow, R., Jakobsson, M., Kalliala, I., Dillner, J., Nieminen, P., & Lorincz, A. (2020). Methylation in Predicting Progression of Untreated High-grade Cervical Intraepithelial Neoplasia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *70*(12), 2582–2590. https://doi.org/10.1093/cid/ciz677

Luttmer, R., Dijkstra, M. G., Snijders, P. J. F., Berkhof, J., van Kemenade, F. J., Rozendaal, L., … Meijer, C. J. L. M. (2016). p16/Ki-67 dual-stained cytology for detecting cervical (pre)cancer in a HPV-positive gynecologic outpatient population. *Modern Pathology*, *29*(8), 870–878. https://doi.org/10.1038/modpathol.2016.80

Ma, Z. S., & Li, L. (2017). Quantifying the human vaginal community state types (CSTs) with the species specificity index. *PeerJ*, *5*, e3366. https://doi.org/10.7717/peerj.3366

Maringe, C., Spicer, J., Morris, M., Purushotham, A., Nolte, E., Sullivan, R., … Aggarwal, A. (2020). Articles The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England , UK : a national , population-based , modelling study. *Lancet Oncology*, *21*(8), 1023–1034. https://doi.org/10.1016/S1470-2045(20)30388-0

Masson, L., Passmore, J. A., Liebenberg, L. J., Werner, L., Baxter, C., Arnold, K. B., Williamson, C., Little, F., Mansoor, L. E., Naranbhai, V., Lauffenburger, D. A., Ronacher, K., Walzl, G., Garrett, N. J., Williams, B. L., Couto-Rodriguez, M., Hornig, M., Lipkin, W. I., Grobler, A., Abdool Karim, Q., … Abdool Karim, S. S. (2015). Genital inflammation and the risk of HIV acquisition in women. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *61*(2), 260–269. https://doi.org/10.1093/cid/civ298

McKay, J., Tenet, V., Franceschi, S., Chabrier, A., Gheit, T., Gaborieau, V., … Vaccarella, S. (2017). Immuno-related polymorphisms and cervical cancer risk: The IARC multicentric case-control study. *PLoS ONE*, *12*(5), 1–13. https://doi.org/10.1371/journal.pone.0177775

McLean, N. W., & McGroarty, J. A. (1996). Growth inhibition of metronidazole-susceptible and metronidazole-resistant strains of Gardnerella vaginalis by Lactobacilli in vitro. *Applied and environmental microbiology*, *62*(3), 1089–1092. https://doi.org/10.1128/AEM.62.3.1089-1092.1996

Mezei, A. K., Armstrong, H. L., Pedersen, H. N., Campos, N. G., Mitchell, S. M., Sekikubo, M., Byamugisha, J. K., Kim, J. J., Bryan, S., & Ogilvie, G. S. (2017). Cost-effectiveness of cervical cancer screening methods in low- and middle-income countries: A systematic review. *International journal of cancer*, *141*(3), 437–446. https://doi.org/10.1002/ijc.30695

Mikeska, T., & Craig, J. M. (2014). DNA methylation biomarkers: cancer and beyond. *Genes*, *5*(3), 821–864. https://doi.org/10.3390/genes5030821

Mirabello, L., Schiffman, M., Ghosh, A., Rodriguez, A. C., Vasiljevic, N., Wentzensen, N., Herrero, R., Hildesheim, A., Wacholder, S., Scibior-Bentkowska, D., Burk, R. D., & Lorincz, A. T. (2013). Elevated methylation of HPV16 DNA is associated with the development of high grade cervical intraepithelial neoplasia. *International journal of cancer*, *132*(6), 1412–1422. https://doi.org/10.1002/ijc.27750

Mirabello, L., Sun, C., Ghosh, A., Rodriguez, A. C., Schiffman, M., Wentzensen, N., Hildesheim, A., Herrero, R., Wacholder, S., Lorincz, A., & Burk, R. D. (2012). Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. *Journal of the National Cancer Institute*, *104*(7), 556–565. https://doi.org/10.1093/jnci/djs135

Mirabello, L., Yeager, M., Yu, K., Clifford, G. M., Xiao, Y., Zhu, B., … Schiffman, M. (2017). HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell*, *170*(6), 1164–1174.e6. https://doi.org/10.1016/j.cell.2017.08.001

Mitra, A., Macintyre, D. A., Lee, Y. S., Smith, A., Marchesi, J. R., Lehne, B., & Bhatia, R. (2015a). Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Nature Publishing Group*, (October), 1–11. https://doi.org/10.1038/srep16865

Moody, C. A., & Laimins, L. A. (2010). Human papillomavirus oncoproteins: Pathways to transformation. *Nature Reviews Cancer*, *10*(8), 550–560. https://doi.org/10.1038/nrc2886

Morris, J. R.& Wu Ct, (2001). Genes, genetics, and epigenetics: a correspondence. *Science (New York, N.Y.)*, *293*(5532), 1103–1105. https://doi.org/10.1126/science.293.5532.1103

Motevaseli, E., Shirzad, M., Akrami, S. M., Mousavi, A. S., Mirsalehian, A., & Modarressi, M. H. (2013). Normal and tumour cervical cells respond differently to vaginal lactobacilli, independent of pH and lactate. *Journal of medical microbiology*, *62*(Pt 7), 1065–1072. https://doi.org/10.1099/jmm.0.057521-0

Nedjai, B., Reuter, C., Ahmad, A., Banwait, R., Warman, R., Carton, J., Boer, S., Cuzick, J., & Lorincz, A. T. (2018). Molecular progression to cervical precancer, epigenetic switch or sequential model?. *International journal of cancer*, *143*(7), 1720–1730. https://doi.org/10.1002/ijc.31549

Nedjai, B., Reuter C., Banwait R., Cuzick, J., Lorincz A. T. (2019). Non-invasive methylation test to detect cervical pre-cancer in self-collected vaginal and urine specimens. *National Cancer Research Institute Conference* (p. [Poster]). Glasgow: https://abstracts.ncri.org.uk/year_published/2019/

Neumann, A., Björck, L., & Frick, I. M. (2020). *Finegoldia magna*, an Anaerobic Gram-Positive Bacterium of the Normal Human Microbiota, Induces Inflammation by Activating Neutrophils. *Frontiers in microbiology*, *11*, 65. https://doi.org/10.3389/fmicb.2020.00065

Nilsson, E., & Ling, C. (2017). DNA methylation links genetics, fetal environment, and an unhealthy lifestyle to the development of type 2 diabetes. *Clinical epigenetics*, *9*, 105. https://doi.org/10.1186/s13148-017-0399-2

Norenhag, J., Du, J., Olovsson, M., Verstraelen, H., Engstrand, L., & Brusselaers, N. (2020). The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis. *BJOG : an international journal of obstetrics and gynaecology*, *127*(2), 171–180. https://doi.org/10.1111/1471-0528.15854

Nugent, R. P., Krohn, M. A., & Hillier, S. L. (1991). Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *Journal of clinical microbiology*, *29*(2), 297–301. https://doi.org/10.1128/JCM.29.2.297-301.1991.

Office for National Statistics National population projections: 2018-based statistical bulletin. 2018. (2018), (https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2018based#national-population-projections-data).

Oh, H. Y., Kim, B. S., Seo, S. S., Kong, J. S., Lee, J. K., Park, S. Y., Hong, K. M., Kim, H. K., & Kim, M. K. (2015). The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, *21*(7), 674.e1–674.e6749. https://doi.org/10.1016/j.cmi.2015.02.026

Ojesina, A. I., Lichtenstein, L., Freeman, S. S., Pedamallu, C. S., Imaz-Rosshandler, I., Pugh, T. J., Cherniack, A. D., Ambrogio, L., Cibulskis, K., Bertelsen, B., Romero-Cordoba, S., Treviño, V., Vazquez-Santillan, K., Guadarrama, A. S., Wright, A. A., Rosenberg, M. W., Duke, F., Kaplan, B., Wang, R., Nickerson, E., … Meyerson, M. (2014). Landscape of genomic alterations in cervical carcinomas. *Nature*, *506*(7488), 371–375. https://doi.org/10.1038/nature12881 Manuscript, A., & Dysfunction, E. (2015). NIH Public Access, *25*(8), 713–724. https://doi.org/10.1097/MCA.0000000000000178.Endothelial

Overmeer, R. M., Louwers, J. A., Meijer, C. J. L. M., Van Kemenade, F. J., Hesselink, A. T., Daalmeijer, N. F., … Steenbergen, R. D. M. (2011). Combined CADM1 and MAL promoter methylation analysis to detect (pre-)malignant cervical lesions in high-risk HPV-positive women. *International Journal of Cancer*, *129*(9), 2218–2225. https://doi.org/10.1002/ijc.25890

Paola, M. Di, Sani, C., Clemente, A. M., Iossa, A., Perissi, E., Castronovo, G., … Torcia, M. G. (2017). Characterization of cervico-vaginal microbiota in women developing persistent high-risk Human Papillomavirus infection. *Scientific Reports*, (July), 1–12. https://doi.org/10.1038/s41598-017-09842-6

Patterson, J. L., Stull-Lane, A., Girerd, P. H., & Jefferson, K. K. (2010). Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of Gardnerella vaginalis relative to other bacterial-vaginosis-associated anaerobes. *Microbiology (Reading, England)*, *156*(Pt 2), 392–399. https://doi.org/10.1099/mic.0.034280-0

Pesola, F., & Sasieni, P. (2019). Impact of screening on cervical cancer incidence in England: a time trend analysis. *BMJ open*, *9*(1), e026292. https://doi.org/10.1136/bmjopen-2018-026292

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*, *55*(5), 856–866. https://doi.org/10.1373/clinchem.2008.107565

Pinna, N. K., Dutta, A., Monzoorul Haque, M., & Mande, S. S. (2019). Can Targeting Non-Contiguous V-Regions With Paired-End Sequencing Improve 16S rRNA-Based Taxonomic Resolution of Microbiomes?: An *In Silico* Evaluation. *Frontiers in genetics*, *10*, 653. https://doi.org/10.3389/fgene.2019.00653

Pirog, E. C., Lloveras, B., Molijn, A., Tous, S., Guimerà, N., Alejo, M., Clavero, O., Klaustermeier, J., Jenkins, D., Quint, W. G., Xavier Bosch, F., Alemany, L., de Sanjosé, S., & RIS HPV TT study group (2014). HPV prevalence and genotypes in different histological subtypes of cervical adenocarcinoma, a worldwide analysis of 760 cases. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, *27*(12), 1559–1567. https://doi.org/10.1038/modpathol.2014.55 5

Piyathilake, C. J., Ollberding, N. J., Kumar, R., Macaluso, M., Alvarez, R. D., & Morrow, C. D. (2016). Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses. *Cancer prevention research (Philadelphia, Pa.)*, *9*(5), 357–366. https://doi.org/10.1158/1940-6207.CAPR-15-0350

Plass, C., Pfister, S. M., Lindroth, A. M., & Bogatyrova, O. (2013). Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nature Publishing Group*, (October). https://doi.org/10.1038/nrg3554

Ramírez, A. T., Sánchez, G. I., Nedjai, B., Agudelo, M. C., Brentnall, A. R., Cuschieri, K., Castañeda, K. M., Cuzick, J., Lorincz, A. T., & ASC-US-COL Trial Group (2021). Effective methylation triage of HPV positive women with abnormal cytology in a middle-income country. *International journal of cancer*, *148*(6), 1383–1393. https://doi.org/10.1002/ijc.33314

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature methods*, *14*(4), 411–413. https://doi.org/10.1038/nmeth.4189

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S., McCulle, S. L., Karlebach, S., Gorle, R., Russell, J., Tacket, C. O., Brotman, R. M., Davis, C. C., Ault, K., Peralta, L., & Forney, L. J. (2011). Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl 1*(Suppl 1), 4680–4687. https://doi.org/10.1073/pnas.1002611107

Rebolj, M., Rimmer, J., Denton, K., Tidy, J., Mathews, C., Ellis, K., Smith, J., Evans, C., Giles, T., Frew, V., Tyler, X., Sargent, A., Parker, J., Holbrook, M., Hunt, K., Tidbury, P., Levine, T., Smith, D., Patnick, J., Stubbs, R., … Kitchener, H. (2019). Primary cervical screening with high risk human papillomavirus testing: observational study. *BMJ (Clinical research ed.)*, *364*, l240. https://doi.org/10.1136/bmj.l240

Reinson, T., Henno, L., Toots, M., Ustav, M., Jr, & Ustav, M. (2015). The Cell Cycle Timing of Human Papillomavirus DNA Replication. *PloS one*, *10*(7), e0131675. https://doi.org/10.1371/journal.pone.0131675

Reshmi, G., Surya, R., Jissa, V. T., Babu, P. S. S., Preethi, N. R., Santhi, W. S., … Pillai, M. R. (2011). Genomics C – T variant in a miRNA target site of BCL2 is associated with increased risk of human papilloma virus related cervical cancer — An in silico approach. *Genomics*, *98*(3), 189–193. https://doi.org/10.1016/j.ygeno.2011.06.005

Rincon-Orozco, B., Halec, G., Rosenberger, S., Muschik, D., Nindl, I., Bachmann, A., Ritter, T. M., Dondog, B., Ly, R., Bosch, F. X., Zawatzky, R., & Rösl, F. (2009). Epigenetic silencing of interferon-kappa in human papillomavirus type 16-positive cells. *Cancer research*, *69*(22), 8718–8725. https://doi.org/10.1158/0008-5472.CAN-09-0550

Ronco, G., Dillner, J., Elfström, K. M., Tunesi, S., Snijders, P. J., Arbyn, M., Kitchener, H., Segnan, N., Gilham, C., Giorgi-Rossi, P., Berkhof, J., Peto, J., Meijer, C. J., & International HPV screening working group (2014). Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet (London, England)*, *383*(9916), 524–532. https://doi.org/10.1016/S0140-6736(13)62218-7

Sahasrabuddhe, V. V., Luhn, P., & Wentzensen, N. (2011). Human papillomavirus and cervical cancer: biomarkers for improved prevention efforts. *Future microbiology*, *6*(9), 1083–1098. https://doi.org/10.2217/fmb.11.87Vink, F. J., Meijer, C. J. L. M., Clifford, G. M., Poljak, M., Anja, O., Petry, K. U., … Pino, M. (2020). Short Report FAM19A4 / miR124-2 methylation in invasive cervical cancer : A retrospective cross-sectional worldwide study, *0*(2019). https://doi.org/10.1002/ijc.32614

Sankaranarayanan, R., Basu, P., Kaur, P., Bhaskar, R., Singh, G. B., Denzongpa, P., & Grover, R. K. (2019). Policy Review Current status of human papillomavirus vaccination in India ' s cervical cancer prevention efforts. *Lancet Oncology*, *20*(11), e637–e644. https://doi.org/10.1016/S1470-2045(19)30531-5

Savage D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annual review of microbiology*, *31*, 107–133. https://doi.org/10.1146/annurev.mi.31.100177.000543

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1412–1417. https://doi.org/10.1073/pnas.0510310103

Schaerli, P., Willimann, K., Ebert, L. M., Walz, A., & Moser, B. (2005). Cutaneous CXCL14 targets blood precursors to epidermal niches for Langerhans cell differentiation. *Immunity*, *23*(3), 331–342. https://doi.org/10.1016/j.immuni.2005.08.012

Schiffman, M., Wentzensen, N., Wacholder, S., Kinney, W., Gage, J. C., & Castle, P. E. (2011). Human papillomavirus testing in the prevention of cervical cancer. *Journal of the National Cancer Institute*, *103*(5), 368–383. https://doi.org/10.1093/jnci/djq562

Schmees, C., Prinz, C., Treptau, T., Rad, R., Hengst, L., Voland, P., Bauer, S., Brenner, L., Schmid, R. M., & Gerhard, M. (2007). Inhibition of T-cell proliferation by Helicobacter pylori gamma-glutamyl transpeptidase. *Gastroenterology*, *132*(5), 1820–1833. https://doi.org/10.1053/j.gastro.2007.02.031

Schmitz, M., Eichelkraut, K., Schmidt, D., Zeiser, I., Hilal, Z., Tettenborn, Z., Hansel, A., & Ikenberg, H. (2018). Performance of a DNA methylation marker panel using liquid-based cervical scrapes to detect cervical cancer and its precancerous stages. *BMC cancer*, *18*(1), 1197. https://doi.org/10.1186/s12885-018-5125-8

Schütze, D. M., Kooter, J. M., Wilting, S. M., Meijer, C. J., Quint, W., Snijders, P. J., & Steenbergen, R. D. (2015). Longitudinal assessment of DNA methylation changes during HPVE6E7-induced immortalization of primary keratinocytes. *Epigenetics*, *10*(1), 73–81. https://doi.org/10.4161/15592294.2014.990787

Secher, T., Samba-Louaka, A., Oswald, E., & Nougayrède, J. P. (2013). Escherichia coli producing colibactin triggers premature and transmissible senescence in mammalian cells. *PloS one*, *8*(10), e77157. https://doi.org/10.1371/journal.pone.0077157

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60. https://doi.org/10.1186/gb-2011-12-6-r60

Seta, F., Campisciano, G., Zanotta, N., Ricci, G., & Comar, M. (2019). The Vaginal Community State Types Microbiome-Immune Network as Key Factor for Bacterial Vaginosis and Aerobic Vaginitis. *Frontiers in microbiology*, *10*, 2451. https://doi.org/10.3389/fmicb.2019.02451

Shurin, G. V., Ferris, R. L., Tourkova, I. L., Perez, L., Lokshin, A., Balkir, L., Collins, B., Chatta, G. S., & Shurin, M. R. (2005). Loss of new chemokine CXCL14 in tumor tissue is associated with low infiltration by dendritic cells (DC), while restoration of human CXCL14 expression in tumor cells causes attraction of DC both in vitro and in vivo. *Journal of immunology (Baltimore, Md. : 1950)*, *174*(9), 5490–5498. https://doi.org/10.4049/jimmunol.174.9.5490

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature methods*, *14*(4), 407–410. https://doi.org/10.1038/nmeth.4184

Singh, V., Miranda, T. B., Jiang, W., Frankel, A., Roemer, M. E., Robb, V. A., Gutmann, D. H., Herschman, H. R., Clarke, S., & Newsham, I. F. (2004). DAL-1/4.1B tumor suppressor interacts with protein arginine N-methyltransferase 3 (PRMT3) and inhibits its ability to methylate substrates in vitro and in vivo. *Oncogene*, *23*(47), 7761–7771. https://doi.org/10.1038/sj.onc.1208057

Sir, D., Huang, S., Afonina, I., Miller, B. A., & Beckmann, A. M. (1997). Human papillomavirus types 52 and 58 are prevalent in cervical cancers from Chinese women. *International journal of cancer*, *70*(4), 408–411. https://doi.org/10.1002/(sici)1097-0215(19970207)70:4<408::aid-ijc6>3.0.co;2-#

Smith, J. S., Lindsay, L., Hoots, B., Keys, J., Franceschi, S., Winer, R., & Clifford, G. M. (2007). Human papillomavirus type distribution in invasive cervical cancer and high-grade

cervical lesions: A meta-analysis update. *International Journal of Cancer*, *121*(3), 621–632. https://doi.org/10.1002/ijc.22527

Snellenberg, S., Strooper, L. M. A. D., Hesselink, A. T., Meijer, C. J. L. M., Snijders, P. J. F., Heideman, D. A. M., & Steenbergen, R. D. M. (2012). Development of a multiplex methylation-specific PCR as candidate triage test for women with an HPV-positive cervical scrape. *BMC Cancer*, *12*. https://doi.org/10.1186/1471-2407-12-551

Snijders, P. J., Steenbergen, R. D., Heideman, D. A., & Meijer, C. J. (2006). HPV-mediated cervical carcinogenesis: concepts and clinical implications. *The Journal of Pathology*, *208*(2), 152–164. https://doi.org/10.1002/path.1866

Snoek, B. C., Splunter, A., Bleeker, M., Ruiten, M., Heideman, D., Rurup, W. F., Verlaat, W., Schotman, H., Gent, M. V., Trommel, N., & Steenbergen, R. (2019). Cervical cancer detection by DNA methylation analysis in urine. *Scientific reports*, *9*(1), 3088. https://doi.org/10.1038/s41598-019-39275-2

Srinivasan, S., Morgan, M. T., Fiedler, T. L., Djukovic, D., Hoffman, N. G., Raftery, D., Marrazzo, J. M., & Fredricks, D. N. (2015). Metabolic signatures of bacterial vaginosis. *mBio*, *6*(2), e00204-15. https://doi.org/10.1128/mBio.00204-15

Stanczuk, G., Baxter, G., Currie, H., Lawrence, J., Cuschieri, K., Wilson, A., & Arbyn, M. (2016). Clinical validation of hrHPV testing on vaginal and urine self-samples in primary cervical screening (cross-sectional results from the Papillomavirus Dumfries and Galloway-PaVDaG study). *BMJ open*, *6*(4), e010660. https://doi.org/10.1136/bmjopen-2015-010660

Steenbergen, R. D. M., Snijders, P. J. F., Heideman, D. A. M., & Meijer, C. J. L. M. (2014). Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. *Nature Reviews Cancer*, *14*(6), 395–405. https://doi.org/10.1038/nrc3728

Steinbach, A., & Riemer, A. B. (2018). Immune evasion mechanisms of human papillomavirus: An update. *International journal of cancer*, *142*(2), 224–229. https://doi.org/10.1002/ijc.31027

Strichman-Almashanu, L. Z., Lee, R. S., Onyango, P. O., Perlman, E., Flam, F., Frieman, M. B., & Feinberg, A. P. (2002). A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome research*, *12*(4), 543–554. https://doi.org/10.1101/gr.224102

Strooper, L. M. A. De, Verhoef, V. M. J., Berkhof, J., Hesselink, A. T., Bruin, H. M. E. De, Kemenade, F. J. Van, … Heideman, D. A. M. (2016). Gynecologic Oncology Validation of the FAM19A4 / mir124-2 DNA methylation test for both lavage- and brush-based self-samples to detect cervical (pre) cancer in HPV-positive women. *Gynecologic Oncology*, *141*(2), 341–347. https://doi.org/10.1016/j.ygyno.2016.02.012

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 10.3322/caac.21660. Advance online publication. https://doi.org/10.3322/caac.21660

Suzuki, H., Toyota, M., Kondo, Y., & Shinomura, Y. (2009). Inflammation-related aberrant patterns of DNA methylation: detection and role in epigenetic deregulation of cancer cell transcriptome. *Methods in molecular biology (Clifton, N.J.)*, *512*, 55–69. https://doi.org/10.1007/978-1-60327-530-9_5

Szarewski, A., Ambroisine, L., Cadman, L., Austin, J., Ho, L., Terry, G., Liddle, S., Dina, R., McCarthy, J., Buckley, H., Bergeron, C., Soutter, P., Lyons, D., & Cuzick, J. (2008). Comparison of predictors for high-grade cervical intraepithelial neoplasia in women with abnormal smears. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *17*(11), 3033–3042. https://doi.org/10.1158/1055-9965.EPI-08-0508

Tidy, J. A. . et al. (2016). NHS Cervical Screening Programme Colposcopy and Programme Management. *Public Health England*, (20).

Tjalma, W. A. A., & Depuydt, C. E. (2013). European Journal of Obstetrics & Gynecology and Reproductive Biology Cervical cancer screening : which HPV test should be. *European Journal of Obstetrics and Gynecology*, *170*(1), 45–46. https://doi.org/10.1016/j.ejogrb.2013.06.027

Tournier, B., Chapusot, C., Courcet, E., Martin, L., Lepage, C., & Faivre, J. (2012). Why do results conflict regarding the prognostic value of the methylation status in colon cancers ? the role of the preservation method. *BMC Cancer*, *12*(1), 12. https://doi.org/10.1186/1471-2407-12-12

Tran, Y. K., Bo, O., Gorse, K. M., Wieland, I., Green, M. R., & Newsham, I. F. (1999). Advances in Brief A Novel Member of the NF2 / ERM / 4 . 1 Superfamily with Growth Suppressing Properties in Lung Cancer 1, 35–43.

Tshomo, U., Franceschi, S., Dorji, D., Baussano, I., Tenet, V., Snijders, P. J., Meijer, C. J., Bleeker, M. C., Gheit, T., Tommasino, M., & Clifford, G. M. (2014). Human papillomavirus infection in Bhutan at the moment of implementation of a national HPV vaccination programme. *BMC infectious diseases*, *14*, 408. https://doi.org/10.1186/1471-2334-14-408

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., … Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, *5*(3), 247–252. https://doi.org/10.1038/nmeth.1185

Vasiljević, N., Wu, K., Brentnall, A. R., Kim, D. C., Thorat, M. A., Kudahetti, S. C., Mao, X., Xue, L., Yu, Y., Shaw, G. L., Beltran, L., Lu, Y. J., Berney, D. M., Cuzick, J., & Lorincz, A. T. (2011). Absolute quantitation of DNA methylation of 28 candidate genes in prostate cancer using pyrosequencing. *Disease markers*, *30*(4), 151–161. https://doi.org/10.3233/DMA-2011-0790 Wylie, K. M., Mihindukulasuriya, K. A., Zhou, Y., Sodergren, E., Storch, G. A., & Weinstock, G. M. (2014). Metagenomic analysis of double-stranded DNA viruses in healthy adults, 1–10. https://doi.org/10.1186/s12915-014-0071-7

Vasiljević, N., Scibior-Bentkowska, D., Brentnall, A. R., Cuzick, J., & Lorincz, A. T. (2014). Credentialing of DNA methylation assays for human genes as diagnostic biomarkers of cervical intraepithelial neoplasia in high-risk HPV positive women. *Gynecologic oncology*, *132*(3), 709–714. https://doi.org/10.1016/j.ygyno.2014.02.001

Verlaat, W., Leeuwen, R. W. Van, Novianti, P. W., Schuuring, E., Meijer, C. J. L. M., Zee, A. G. J. Van Der, … Dna, H. (2018). Host-cell DNA methylation patterns during high-risk HPV-induced carcinogenesis reveal a heterogeneous nature of cervical pre-cancer. *Epigenetics*, *13*(7), 769–778. https://doi.org/10.1080/15592294.2018.1507197

Verlaat, W., Snijders, P. J., van Moorsel, M. I., Bleeker, M., Rozendaal, L., Sie, D., Ylstra, B., Meijer, C. J., Steenbergen, R. D., & Heideman, D. A. (2015). Somatic mutation in PIK3CA is

a late event in cervical carcinogenesis. *The journal of pathology. Clinical research*, *1*(4), 207–211. https://doi.org/10.1002/cjp2.27

Verlaat, W., Snoek, B. C., Wilting, S. M., Peter, J. F., Novianti, P. W., Splunter, A. P. Van, … Steenbergen, R. (2018). Identification and validation of a 3-gene methylation classifier for HPV-based cervical screening on self-samples. *Clinical Cancer Research*. https://doi.org/10.1158/1078-0432.CCR-17-3615

Vink, M. A., Bogaards, J. A., Meijer, C. J. L. M., & Berkhof, J. (2015). Primary human papillomavirus DNA screening for cervical cancer prevention: Can the screening interval be safely extended? *International Journal of Cancer*, *137*(2), 420–427. https://doi.org/10.1002/ijc.29381

Vinokurova, S., & von Knebel Doeberitz, M. (2011). Differential methylation of the HPV 16 upstream regulatory region during epithelial differentiation and neoplastic transformation. *PloS one*, *6*(9), e24451. https://doi.org/10.1371/journal.pone.0024451

Virtanen, S., Kalliala, I., Nieminen, P., & Salonen, A. (2017). Comparative analysis of vaginal microbiota sampling using 16S rRNA gene analysis. *PloS one*, *12*(7), e0181477. https://doi.org/10.1371/journal.pone.0181477

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr, & Kinzler, K. W. (2013). Cancer genome landscapes. *Science (New York, N.Y.)*, *339*(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Vuyst, H., Ndirangu, G., Moodley, M., Tenet, V., Estambale, B., Meijer, C. J., Snijders, P. J., Clifford, G., & Franceschi, S. (2012). Prevalence of human papillomavirus in women with invasive cervical carcinoma by HIV status in Kenya and South Africa. *International journal of cancer*, *131*(4), 949–955. https://doi.org/10.1002/ijc.26470

Wallin, K. L., Wiklund, F., Angström, T., Bergman, F., Stendahl, U., Wadell, G., Hallmans, G., & Dillner, J. (1999). Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. *The New England journal of medicine*, *341*(22), 1633–1638. https://doi.org/10.1056/NEJM199911253412201

Wang, K. D., Xu, D. J., Wang, B. Y., Yan, D. H., Lv, Z., & Su, J. R. (2018). Inhibitory Effect of Vaginal Lactobacillus Supernatants on Cervical Cancer Cells. *Probiotics and antimicrobial proteins*, *10*(2), 236–242. https://doi.org/10.1007/s12602-017-9339-x

Wang, S., Sun, H., Jia, Y., Tang, F., Zhou, H., Li, X., Zhou, J., Huang, K., Zhang, Q., Hu, T., Yang, R., Wang, C., Xi, L., Deng, D., Wang, H., Wang, S., Ma, D., & Li, S. (2015). Association of 42 SNPs with genetic risk for cervical cancer: an extensive meta-analysis. *BMC medical genetics*, *16*, 25. https://doi.org/10.1186/s12881-015-0168-z

Wentzensen, N., Sun, C., Ghosh, A., Kinney, W., Mirabello, L., Wacholder, S., Shaber, R., LaMere, B., Clarke, M., Lorincz, A. T., Castle, P. E., Schiffman, M., & Burk, R. D. (2012). Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *Journal of the National Cancer Institute*, *104*(22), 1738–1749. https://doi.org/10.1093/jnci/djs425

WHO. (2013). *WHO Guidelines for Screening and Treatment of Precancerous Lesions for Cervical Cancer Prevention*. World Health Organization.

Wiik, J., Sengpiel, V., Kyrgiou, M., Nilsson, S., Mitra, A., Tanbo, T., Monceyron Jonassen, C., Møller Tannæs, T., & Sjøborg, K. (2019). Cervical microbiota in women with cervical intra-epithelial neoplasia, prior to and after local excisional treatment, a Norwegian cohort

study. *BMC women's health*, *19*(1), 30. https://doi.org/10.1186/s12905-019-0727-0

Williams, V. M., Filippova, M., Filippov, V., Payne, K. J., & Duerksen-Hughes, P. (2014). Human papillomavirus type 16 E6* induces oxidative stress and DNA damage. *Journal of virology*, *88*(12), 6751–6761. https://doi.org/10.1128/JVI.03355-13

Wilting, S. M., & Steenbergen, R. D. M. (2016). Molecular events leading to HPV-induced high grade neoplasia. *Papillomavirus Research*, *2*, 85–88. https://doi.org/10.1016/j.pvr.2016.04.003

Winer, R. L., Kiviat, N. B., Hughes, J. P., Adam, D. E., Lee, S. K., Kuypers, J. M., & Koutsky, L. A. (2005). Development and duration of human papillomavirus lesions, after initial infection. *The Journal of infectious diseases*, *191*(5), 731–738. https://doi.org/10.1086/427557

Yeh, P. T., Kennedy, C. E., de Vuyst, H., & Narasimhan, M. (2019). Self-sampling for human papillomavirus (HPV) testing: a systematic review and meta-analysis. *BMJ global health*, *4*(3), e001351. https://doi.org/10.1136/bmjgh-2018-001351

Yim, E. K., & Park, J. S. (2007). Biomarkers in cervical cancer. *Biomarker insights*, *1*, 215–225

Zeng, R., Liu, Y., Jiang, Z. J., Huang, J. P., Wang, Y., Li, X. F., Xiong, W. B., Wu, X. C., Zhang, J. R., Wang, Q. E., & Zheng, Y. F. (2018). EPB41L3 is a potential tumor suppressor gene and prognostic indicator in esophageal squamous cell carcinoma. *International journal of oncology*, *52*(5), 1443–1454. https://doi.org/10.3892/ijo.2018.4316

Zhang, C., Liu, Y., Gao, W., Pan, Y., Gao, Y., Shen, J., & Xiong, H. (2018). The direct and indirect association of cervical microbiota with the risk of cervical intraepithelial neoplasia. *Cancer medicine*, *7*(5), 2172–2179. https://doi.org/10.1002/cam4.1471

Zielinski, G. D., Snijders, P. J., Rozendaal, L., Voorhorst, F. J., van der Linden, H. C., Runsink, A. P., de Schipper, F. A., & Meijer, C. J. (2001). HPV presence precedes abnormal cytology in women developing cervical cancer and signals false negative smears. *British journal of cancer*, *85*(3), 398–404. https://doi.org/10.1054/bjoc.2001.1926.