HOST SOMATIC VARIATION BETWEEN WOMEN LIVING WITH HIV WITH CERVICAL INTRAEPITHEIAL LESIONS (CIN3) AND THEIR HIV NEGATIVE COUNTERPARTS

by

Nosipho Mabizela (MBZNOS005)

Submitted to the University of Cape Town

In fulfilment of the requirements for the degree

MSc (Med) in Human Genetics

Faculty of Health Sciences

Department of Pathology, Division of Human Genetics

Pharmacogenomics and drug metabolism research group

Supervisors: Professor Collet Dandara and Dr Nyarai Soko, UCT/SAMRC Platform for Pharmacogenomics Research and Translation (PREMED) Unit, South Africa Medical Research Council, Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa

Co-Supervisors: Prof Richard Naidoo, Division of Anatomical Pathology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Faculty of Health Sciences, South Africa. **Dr Hue-Tsi Wu**, PathCare Claremont, Cape Town, South Africa.

DECLARATION

I, **Nosipho Mabizela**, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 16 June 2024

ABSTRACT

Despite the use of antiretroviral therapy, cervical cancer remains a leading malignancy in women with HIV, who face a six-fold increased risk. Infection with HIV and HPV has been linked to accelerated cervical cancer development. However, there are limited studies on the role of host somatic variations in HIV-positive and HIV-negative women on cervical cancer. Understanding these variations may help identify potential genetic factors that contribute to accelerated cervical cancer development and differential response to treatment. This knowledge is important in targeting interventions and improving outcomes for women with HIV and cervical cancer. Therefore, this study aims to investigate host somatic genetic variation between cervical biopsies obtained from HIV-positive or HIV-negative women with histologically confirmed CIN3 to determine potential differences in genomic landscapes and HPV infection between HIV-positive and HIV-negative women.

The matched case-control study utilized archived cervical biopsies from 88 women (44 HIV-positive, 44 HIV-negative) attending Groote Schuur Hospital Cancer Clinic between 2020 and 2022. The cases and controls were carefully age matched. HPV infection and type were confirmed using the Anyplex™ II HPV28 Detection kit. In cervical cancer, six hotspot regions in the four commonly mutated genes (*TP53*, *PIK3CA*, *PTEN*, and *EGFR*) were genotyped using Polymerase Chain Reaction and validated using Sanger Sequencing. Missense variant pathogenicity was assessed using SIFT, Polyphen-2, and ClinVar tools.

The median age was [37 years (IQR:34-41)] for HIV-positive women and [35 years (IQR:32- 43)] for HIV-negative women. In the HIV-negative cohort the women reported tobacco smoking ($p<0.0001$), menstruation irregularities ($p=0.005$), and contraception usage ($p=0.019$). These parameters were statistically significant when compared to HIV-positive cohort. Common HPV types identified were HPV 16 (43/88, 49%), 35 (12/88, 14%), and 58 (10/88, 11%). A total of 232 genetic variants were identified, with HIV-positive women having a significantly higher burden of pathogenic variants (31%) compared to 15% among the HIV-negative ($p=0.0406$). Identified mutations included stop-gain, missense, synonymous, and intron variants. The genes *TP53* and *PIK3CA* had more stop-gain variants among HIV-positive women (4/5) compared to HIV-negative

women with 1/5 of the 5 mutations. These damaging variants were more prevalent in women under 50 in both cohorts.

In conclusion, younger women (<50 years) showed predominantly damaging variants, indicating more aggressive cancer, and a possible reason for early onset in the younger cohort. HIV-positive women displayed a higher mutation burden in *PIK3CA* and pathogenic variants in *TP53,* emphasizing the need to further explore these genes in gene expression studies

Keywords: HIV-positive, HIV-negative, *TP53*, *PIK3CA*, *EGFR*

ACKNOWLEDGMENTS

I am humbled by the number of people who have assisted me over the past two years and made this research possible and successful. I want to express my utmost gratitude and appreciation to:

- God for giving me all the strength, wisdom, and direction, and I thank myself for all the hard work and for going beyond my comfort zone to ensure this project went as planned and never gave up, even when there were challenges.
- My parents (S. Mhlongo, D. Mabizela and my granny E. Mhlongo), thank you so much for the unconditional love, support, and care. All your kind words, love, care and support kept me going through this journey.
- My supervisors, Prof Collet Dandara and Dr Nyarai Soko, and co-supervisors, Prof Richard Naidoo and Dr Hue-Tsi Wu, thank you for the opportunity to pursue my studies under your guidance. Thank you for believing and having faith in me, for the patience you had in me, and for the endless support and encouragement you gave me until the end.
- My friends (Z. Memela, K. Mcinga, P. Mshunqwane and M. Mokoena), I am out of words for you. Thank you so much for making this journey so bearable and enjoyable. All the venting sessions we had every time my things were not going according to plan, all the effort you guys made, again and again, to try and understand what I was doing, all the undying comfort and support throughout the ups and downs of this journey. Thank you!
- The Platform for Pharmacogenomics Research and Translation Unit for being such a supportive and friendly group. Thank you for always being there to help troubleshoot when things were not going well in the lab.
- The HUMGEN family for providing such a welcoming working environment!
- National Research Foundation (NRF) for awarding me a scholarship, which made it possible to support myself and complete my studies and thank you to the UCT VC Research Scholarship in recognition of my exceptional academic record.

Table of Contents

List of Abbreviations

ART: Antiretroviral Therapy

HIV: Human Immunodeficiency Virus

AIDS: Acquired Immunodeficiency Syndrome

HPV: Human Papillomavirus

LMICs: Low-income and middle-income countries

CIN: Cervical Intraepithelial Neoplasia

LSIL: Low-grade Squamous Intraepithelial Lesion

HSIL: High-grade squamous Intraepithelial Lesion

HR-HPV: High-risk HPV

LR-HPV: Low-risk HPV

ORF: Open-Reading Frames

DDR: DNA Damage Repair

SMG: Significantly Mutated Genes

TCGA: The Cancer Genome Atlas

TP53: Tumour Protein 53

Rb: Retinoblastoma

MDM2: Mouse Double Minute 2 homologue protein

ARF: Alternative Reading Frame

ATR: Ataxia Telangiectasia and Rad3-related protein

ATM: Ataxia-Telangiectasia Mutated

CHEK1: Checkpoint Kinase 1

CHEK2: Checkpoint Kinase 2

CDKN1A: Cyclin-Dependent Kinase Inhibitor1

GADD45A: Growth Arrest and DNA-Damage inducible Alpha

PUMA: p53 upregulated modulator of apoptosis

BAX: Apoptosis regulator

PIK3CA: Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha

ABD: Adaptor-Binding Domain

RBD: Ras Binding domain

PI3K/Akt/mTOR: Phosphatidylinositol 3-kinase/protein kinase B/mammalian target of rapamycin

PIP2: Phosphatidylinositol-4,5-bis-phosphate

PIP3: Phosphatidylinositol-3,4,5-triphosphate

S6k: Ribosomal protein S6 kinase1

4EBP1: Eukaryotic translation initiation factor 4E binding protein 1

PTEN: Phosphatase and tensin homolog

EGFR: Epidermal growth factor receptor

DNA: Deoxyribonucleic Acid

HREC: Human Research Ethics Committee

FFPE: Formalin-fixed paraffin-embedded

RNA: Ribonucleic acid

RCF: Relative Centrifugal Force

RT-PCR: Real-Time Polymerase Chain Reaction

DPO: Dual Priming Oligonucleotides

TOCE: Tagging Oligonucleotide Cleavage and Extension

PCR: Polymerase Chain Reaction

IARC: International Agency Research on Cancer

FastAP: FastAp Thermosensitive Alkaline Phosphatase

ExoI: Exonuclease I

dNTP: Deoxynucleotide triphosphate

EDTA: Ethylenediaminetetraacetic acid

SIFT: Sorting Intolerant from Tolerant

PolyPhen2: Polymorphism Phenotyping v2

SD: Standard Deviation

IQR: Interquartile Range

LIST OF FIGURES

Figure 1.9: *EGFR* signaling pathway. The *EGFR* signaling pathway involves ligand-induced dimerization and activation, leading to two different downstream signaling pathways the Ras- Raf-MEK-MAPK and PI3K-AKT pathways, which in turn activate various biological processes. Meanwhile, *PTEN* exerts a negative regulatory effect on the PI3K-AKT pathway by converting PIP3 back to PIP2. Image adapted from Alyssa M. Krasinskas 2011. (118)

Figure 2.1: Steps to identify variants after sequencing.

Figure 3.1: **Distribution of HPV genotypes in the study cohort (n=87).** Thirteen women were HPV negative (HIV-positive=7 and HIV-negative=6).

Figure 3.2: Number of HPV types in women with confirmed CIN3. The figure demonstrates the number of women that were HPV negative, with one HPV type, those with multiple infections and the unspecified HPV type.

Figure 3.3: Distribution of HPV genotypes in women with confirmed CIN3 stratified by HIV status. The figure demonstrates the distribution of HPV mono-infection and multiple co-infections stratified by HIV status.

Figure 3.4: **A GeneMANIA gene–gene interaction network of the hotspot genes (*TP53, PIK3CA, PTEN and EGFR*).** Different colors of the network lines indicate different interactions (Physical interaction, co-expression, prediction, co-localization, genetic interaction, pathway, and shared protein domain). The thickness of the lines around each gene indicates the strength of interactions.

Figure 3.5: A representative gene for PCR amplification. The image indicates PCR fragments run at 100V for 1 hour on a 1.5% agarose gel for *PIK3CA* exon 9. The first lane contains the molecular weight marker (MWM), specifically the GeneRuler™ 100bp Ladder (Thermo Fisher Scientific), the second lane is the non-template control (NTC), and the third lane is the positive control (PC). The following lanes are individual samples. The amplicon was 248bp. The other genes were done the same way (Appendix I); the amplicon sizes for them were as follows: *TP53* exon 4: 353 bp, *PIK3CA* exon 20: 231 bp, *PTEN* exon 5: 236 bp, *EGFR* exon 19: 399 bp and *EGFR* exon 20:374 bp.

Figure 3.6: A representative of the Sanger sequencing results from *EGFR* exon 20. The electropherogram was visualized using the DNAStar Seqman Pro alignment tool. The region in the orange box indicates a synonymous variant a change of guanine to adenine that results in protein change Q787Q. **A.** is the GG, a wildtype genotype for patient 19021, **B.** is the GA heterogeneous genotype for patient 19049, **C.** is the AA mutant genotype for patient 19038.

Figure 3.7: Overall type of mutations identified for each gene. Missense mutations were common in all genes except for *EGFR*. Stop-gain were identified in *TP53* and *PIK3CA*.

Figure 3.8: Oncoprint shows the distribution of the total variants in the HIV-positive and HIV-negative cohorts. Each oncoprint box represents a mutation, and different colors represent different mutations, while each row represents a gene and exon. The key shows the total number of variants identified for HIV-positive and HIV-negative cohorts for each mutation type and the description of different colors in the oncoprint.

Figure 3.9: Oncoprint showing the distribution of total variants (combined, missense, stop gain, synonymous and intronic variants), number of HPV types detected per patient, HIV status and age range for the study cohort. The first top row shows the sample ID's, followed by the HIV status, HPV groupings and age range. Each oncoprint box represents each patient, and the description for each color is explained in the key below the oncoprint. Different shades of colors per gene show samples with multiple mutations. The column with only grey boxes and no other color means no variants were identified.

Figure 3.10: Overall number of pathogenic variants and their nature for each gene and exon, stratified by HIV status. Each box represents a woman with a pathogenic variant. The boxes with asterisks represent the most likely damaging variants, while those without asterisks are probably damaging.

Figure 3.11: Overall number of pathogenic variants per age group (<50 years old and ≥50 years old) per HIV status. The boxes with asterisks are most likely damaging variants, while those without are probably damaging. **A)** HIV-positive, **B)** HIV-negative

Figure 3.12: Overall number of most likely damaging variants per gene, stratified by HIV status. *TP53*: p=0.936, *PIK3CA*: p=0.007, *EGFR*: p=0.414.***TP53***: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol- 4,5- bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

List of Tables

CHAPTER 1: INTRODUCTION

1.1 Background

Despite the introduction and use of antiretroviral therapy (ART), the occurrence of cervical cancer among women living with human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) remains notably high. Women with HIV face a six-fold increased risk of developing invasive cervical cancer (1), and their progression from human papillomavirus (HPV) infection to invasive cancer is faster compared to HIV-negative women (2, 3). Cervical cancer is the fourth most diagnosed malignancy and cause of death in women globally (4), and HPV infection is the primary etiologic factor in developing cervical cancer. However, additional risk factors play a role in cancer progression, including HIV, thus making women HIV-positive more susceptible to cervical cancer compared to HIV-negative women. In addition, data on the impact of HIV on HPV infections and HPV-related cancers at the genetic level is limited.

The human host variation also plays a massive role in the cancer progression in both HIV-positive and HIV-negative women. Thus, analyzing cancerous tumours at a genetic level, looking at the commonly mutated genes in cervical cancer enables a more comprehensive understanding of the biology of anogenital cancers driven by HPV in women living with HIV. Understanding the initiation and development of precancerous lesions into cervical cancer is important. This knowledge allows the development of prevention, management, and therapeutic strategies. A deeper understanding of somatic differences between HIV-positive and HIV-negative women may potentially shed light on the differences HIV introduces into the cervical carcinogenesis process and hence give direction to prevention, management, and therapeutic strategies for HIV-positive women co-infected with HPV.

The literature review will comprise several sections. The first section delves into cervical cancer epidemiology and etiology. The second section explores HPV epidemiology and the human host's response to HPV infection. The third section addresses HIV epidemiology, the human host's response post-infection, and susceptibility to other diseases due to being HIV-positive. The fourth section will focus on HIV/HPV co-infection and its link to cervical cancer susceptibility. The fifth section examines key genes involved in cervical cancer (*TP53*, *PIK3CA*, *PTEN*, and *EGFR*) and their

impact on cancer progression. Thereafter, the research question, rationale, aim, hypothesis, and objectives will be provided.

1.2 Literature Review

1.2.1 Cervical cancer

1.2.1.1 Epidemiology of cervical cancer

Cervical cancer is a gynecologic cancer that develops in the cervix. Despite introducing screening and vaccination programs, it continues to be a prevalent malignancy globally, especially in low and middle-income countries (LMICs) (5). In 2020, 604 127 new cases and 341 831 deaths were recorded worldwide (5). Cervical cancer is the fourth most diagnosed malignancy and cause of death in women globally (4). In developing countries, cervical cancer is the second most diagnosed cancer and the cause of death of women after breast cancer (4). In South Africa, cervical cancer is the second most diagnosed cancer in women, with an incidence rate of between 22.8 to 27 per 100 000 women, 5743 new cases, and 3027 deaths per annum (6). Cervical cancer arises in the transformation zone, the region between the ectocervix squamous epithelium and the endocervix columnar epithelium, where continuous metaplastic changes occur due to mutations in an individual's genome (7). Thus, cervical cancer is a malignant neoplasm originating from cells found at the squamocolumnar junction of the cervix. Invasive cervical cancer starts from precancerous lesions called cervical intraepithelial neoplasia (CIN), which is divided into three categories: CIN1 (mild dysplasia), CIN2 (moderate dysplasia), and CIN3 (severe dysplasia) (**Figure 1.1**) (8). CIN 1 is classified as a low-grade squamous intraepithelial lesion (LSIL), whereas CIN 2 and CIN 3 are classified as high-grade squamous intraepithelial lesions (HSIL).

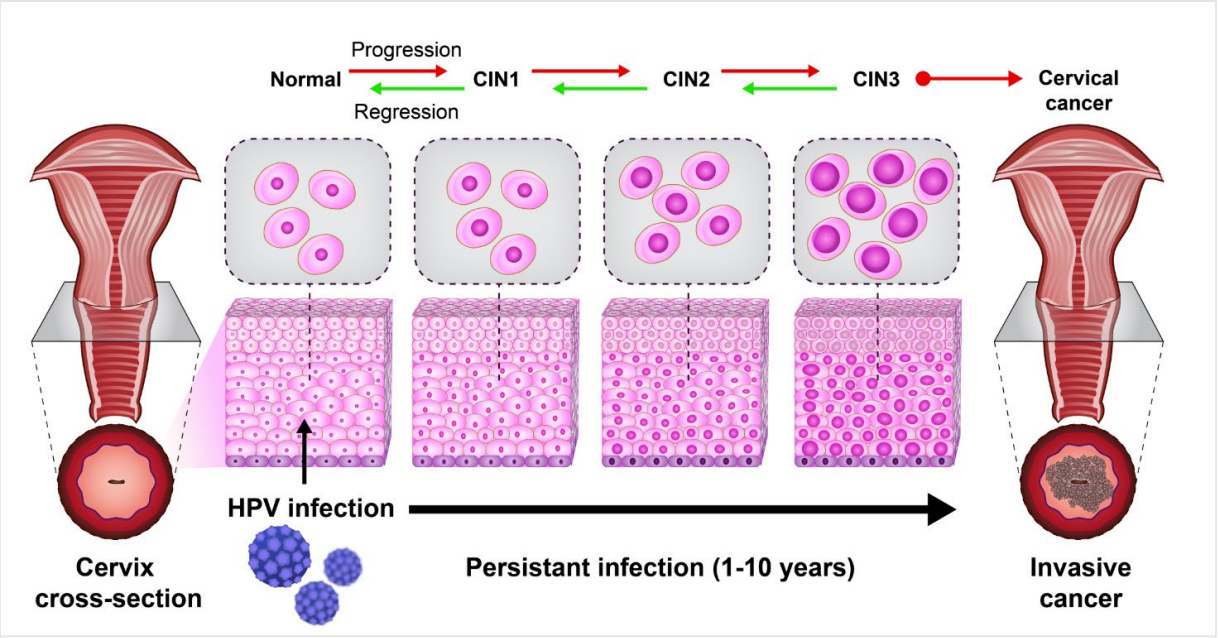


Figure 1.1: The development of cervical cancer. Initially, the epithelial cells transform to precancerous lesions, namely CIN1 (mild dysplasia), CIN2 (moderate dysplasia), and CIN3 (severe dysplasia and carcinoma in situ). In CIN 1, also called LSIL, the lesions commonly regress compared to CIN 2 and CIN 3, also referred to as HSIL. HSILs are due to persistent infection of human papillomavirus, which is the main etiologic factor for developing cervical cancer. (Adapted from Yousefi et al., 2022)(8)

1.2.1.2 Etiology of cervical cancer

Human papillomavirus (HPV) is a major etiologic factor in developing cervical cancer and its precursor, cervical intraepithelial neoplasia (CIN). Persistent infection with oncogenic HPV types causes the progression of CIN to invasive cervical cancer (9). However, oncogenic HPV types alone do not develop cervical cancer (10). Several risk factors influence the progression of precancerous lesions to invasive cervical cancer. However, these risk factors serve as cervical cancer co-factors that may promote cancer development with HPV infection. The main risk factors that play a role in susceptibility to cervical cancer include the early age of sexual debut, multiple sexual partners, cigarette smoking, long-term use of oral contraceptives, and immunosuppression (HIV infection) (11-13). Other risk factors include an individual's lifestyle, co-infection with other sexually transmitted diseases, host genetic variability, including chromosomal alterations and changes in candidate genes, and being infected with multiple HPV types (10).

1.2.2 Human Papillomavirus (HPV)

1.2.2.1 Epidemiology of HPV

Human Papillomavirus (HPV) is the most common sexually transmitted infection worldwide (14), affecting more than half of the sexually active population (15). There are more than 200 different types of HPV that have been identified, and they can be divided into two categories based on their propensity to cause cancer: high risk and low risk. The development of various anogenital cancers, including those of the cervix, vulva, vagina, penis, and anus, is linked to high-risk HPVs (HR-HPV), also known as oncogenic HPV, where HPV16 and HPV18 are the most common HPVs in cervical cancer, accounting for nearly 70% of cases (15). Low-risk HPVs (LR-HPV), also known as non-oncogenic HPV types, are linked to the development of genital warts and low-grade cervical lesions, with HPV6 and HPV11 being the most prevalent types in such diseases (16).

1.2.2.2 HPV viral structure

HPV is a small, non-enveloped, double-stranded DNA virus with a genome of approximately 8 kb in size (17, 18). These viruses infect mucosal and cutaneous epithelial cells (17). The HPV genome encodes eight open-reading frames (ORFs). The ORFs are divided into three functional sections: the early (E) region, which encodes at least seven viral proteins (E1, E2, and E4-E7) expressed in the early phase of infection; the late (L) region, which encodes the two viral structural proteins L1 and L2, that form the viral capsid in the late phase of infection, and the upstream regulatory region (URR) , also known as long control region (LCR)(17, 19, 20).

1.2.2.3 HPV life cycle

HPV relies on the host machinery for their replication and other factors, including transcriptional factors, microRNAs (miRNAs), kinases, apoptotic caspases, epigenetic enzymes, and DNA damage signaling (21). Hence, to replicate successfully, HPV uses the host cells' DNA Damage Repair (DDR) machinery. Upon HPV infection, the virus activates the DDR to amplify the viral genome and recruit the DDR proteins to viral replication sites (22, 23). HPV recruits DNA repair factors required for viral genome replication by expressing their viral proteins (24). The early HPV genes E1 and E2 activate the DDR and recruit repair proteins to the viral replication centers. The E6 and

E7 are crucial oncogenes that target hinge proteins involved in cell cycle control. The E6 oncogene targets the p53 protein, while the E7 oncogene specifically targets the retinoblastoma protein (pRb).

Additionally, E6 and E7 are required for both HPV genome maintenance and amplification and to prevent the DDR response from halting cell cycle progression or apoptosis. In the late regulation of viral function, the viral genes E4 and E5 play a role in virion release and prompt the loss of surface MHC-I expression in the epithelial cells, resulting in the evasion of immune surveillance. The L1 and L2 are late gene products that play a role in assembling the mature viral capsid. The LCR regulates the transcription of the E6 and E7 viral oncogenes through the transcription factors of the virus and the host cells (19). At least 1000 copies of the virus are produced during viral DNA replication in each cell. These virions increase the expression of L1 and L2 capsid proteins and the assembly of infectious viruses (25, 26).

1.2.2.4 HPV entry into the human host and human host response

HPV enters epithelium through micro-abrasions by entering the single-layered squamous cellular junction cells between the endo- and ectocervix (20, 27). The immune system plays a role in eradicating HPV infection; however, HPV infection can persist or be cleared up over several months. Therefore, some HR-HPVs cause transient infections, which do not result in cancer (28). However, over time, if changes occur to the viral genome or the infected host cell, the transient infection can become persistent due to the failure of the immune system to recognize and clear the virus, resulting in the progression of LSIL to HSIL and eventually to invasive cervical cancer (25). Persistent HR-HPV infections increase the risk of cervical cancer, and the persistence is higher in immunocompromised individuals, such as women living with HIV (WLWH) (29, 30).

1.2.3 Human Immunodeficiency Virus (HIV)

1.2.3.1 Epidemiology of HIV

The human immunodeficiency virus (HIV) remains one of the most devastating infectious diseases to emerge in recent history, affecting millions of people worldwide, despite the notable advancements in the development and scale-up of HIV treatment and prevention (31, 32). HIV is the causative agent of acquired immunodeficiency syndrome (AIDS), which leads to millions of deaths globally and AIDS-defining diseases. In 2022, an estimated 39.0 million people were living with HIV, 630 000 died from HIV-related illnesses, and 1.3 million acquired HIV (33).

1.2.3.2 HIV and host response

HIV targets the immune system and weakens people's defense against numerous infections and some types of cancer that people with healthy immune systems can easily fight off. The immune response is how the body recognizes and protects itself against bacteria, viruses, and other potential pathogens that appear foreign and harmful to it by inducing the production of numerous cytokines and interferons to eliminate these pathogens (34). The immune system recognizes and destroys or attempts to destroy these substances that contain antigens. Hence, various genes are important for defense against invading foreign organisms or viruses in the human body.

1.2.3.3 Susceptibility to other diseases due to being HIV-positive

To assist the human host response in fighting the HIV infection, antiretroviral therapy (ART) was introduced in 1987 and the global epidemiology of HIV infection has changed markedly because of the expanding access to ART (31). ART inhibits HIV viral replication, reconstitutes CD4 T-cell count and immunity, and lowers the risk of opportunistic infections such as Kaposi's sarcoma, candidiasis, and tuberculosis. (35). However, despite ART's introduction, the burden of human papillomavirus (HPV) associated premalignant and malignant lesions remains high in HIV-infected persons (2). Therefore, being HIV positive increases the incident risk and persistent HPV infection, rapid progression to HPV-associated premalignant and malignment lesions (2, 3). This is because HIV induces a loss of CD4 T cells, which also results in the depletion of these cells in the mucosal sites which are essential to control HIV and HPV infections (30). Previous studies have shown a

six-fold higher risk in WLWH to develop invasive cervical cancer compared to HIV-negative women (1). Thus, being HIV positive makes an individual more susceptible to other diseases, including HPV-related diseases such as cervical cancer. HIV induces a loss of CD4 T cells, which also results in the depletion of these cells in the mucosal sites which are essential to control HIV and HPV infections (30).

1.2.4 HIV/HPV co-infection and susceptibility to cervical cancer

HPV is the main etiologic factor for cervical cancer, and the main co-risk factor is HIV infection. These two infections increase susceptibility to cervical cancer. HIV and HPV are sexually transmitted viruses that cause chronic infections and diseases (36, 37), but the direct interaction of these viruses is unlikely since HIV targets the CD4 T-lymphocyte and HPV infects epithelial cells (38). However, it has been discovered that HIV can indirectly interfere with the HPV life cycle by transcribing early HPV genes or enhancing the expression of viral oncoproteins E6 and E7 (38). Notably, the HIV Trans-Activator of Transcription (*tat*) protein has been identified as a molecular link between the two viruses. Laboratory studies have shown that the *tat* protein can enhance the expression of HPV oncogenes E6 and E7 by activating the HPV long control region (39). Other previous studies have identified that HIV may influence the natural history of HPV through HIV-related immunodeficiency, increasing the susceptibility to HPV infection and possible reactivation of latent HPV infections (40, 41). Observations from immunosuppressed individuals have shown strong associations between low CD4 cell count, HPV persistence, and cervical cancer, demonstrating that HIV infection decreases cell-mediated immunity and enhances HPV pathogenesis (41-43). Therefore, women infected with HPV and co-infected with HIV have a substantial risk of developing cervical cancer. The risk of developing cervical cancer in these women is six-fold higher than in HIV-negative women (1).

Various studies have demonstrated differences in susceptibility in HIV-positive women and HIV-negative women. The morbidity of cervical cancer in HIV-positive women is 1.5 to 8 times higher compared to HIV-negative women (44-47). After 12 months of HPV infection, it has been reported that infection is cleared in approximately 66% of women, and approximately 90% of HPV infections are cleared after 24 months (48, 49). However, in immunocompromised individuals,

HPV infection is persistent. Konopnicki et al., identified that a CD4 count greater than 500 cells/µl for 18 months among HIV-positive women was significantly associated with a decreased risk of oncogenic HPV infection with an odds ratio of 0.88 (95% CI 0.82-0.94; p=0.0002)(50). Another study showed that HIV-positive women with baseline CD4+ count <200 (cells/mL) or infected with HPV-16/18 had higher rates of HPV progression and lower rates of HPV regression (51). Therefore, HIV co-infection has been associated with higher rates of HPV infection and has been reported to reduce HPV clearance, increase HPV persistence, and increase precancerous lesions, which increases the risk of progression to invasive cervical cancer. (39, 43). Another factor that plays a role in cancer progression is the host somatic variations that occur in tumour suppressor genes and the oncogenes.

1.2.5 Genes involved in cervical cancer

Somatic genomic alterations induced by HPV infection are considered a significant contributing factor to the development of cervical cancer and other HPV-associated cancers (52); thus, understanding what happens at the molecular level of cells is important. A group of genes are essential for controlling the normal growth, proliferation, and apoptosis of cells. These include tumor suppressor genes, oncogenes, and DNA repair genes. When mutations occur in these groups of genes, it leads to the dysregulation of normal cell growth, including overexpression of other proteins. Multiple genes have been described as significantly mutated genes (SMG) in cancers by The Cancer Genome Atlas (TCGA) (13). The significantly mutated genes in cervical cancer include *TP53*, *PIK3CA, PTEN* and *EGFR*. These genes are reported as significantly mutated in numerous studies (53, 54), including the TCGA.

1.2.5.1 Tumour protein p53 (*TP53*)

1.2.5.1.1 TP53 structure and function

The *TP53* gene is a tumor suppressor gene located on chromosome 17p13.1, and it encodes a tumor protein 53 (p53), which is a transcription factor responsible for initiating the transcription processes of several target genes involved in cellular processes when there is a cellular stress

signal (55, 56). The gene comprises 11 exons and 10 introns; the exons are located in different regions of the gene (**Figure 1.3**) (57). The p53 protein is also referred to as the guardian of the genome because it plays a vital role in preventing metastasis by suppressing angiogenesis (by inhibiting the growth of new blood vessels that will supply nutrients and energy to tumors), regulating cell proliferation (by ensuring cells with damaged DNA are not dividing), promoting apoptosis (by eliminating cells that have damaged DNA and irreparable damaged DNA), regulation of the cell cycle (by preventing the replication of damaged DNA allowing time for the cells to repair DNA), DNA damage repair, and maintaining genome stability (55, 58, 59). *TP53* is regarded as one of the most common mutated genes in human cancer(59). Up to 90% of *TP53* mutations are non-synonymous, involving single base substitution mutation, and are primarily found within the DNA binding domain region(60). Most mutations implicated in carcinogenesis are found between exons 4-9 (**Figure 1.2**), which encode the DNA binding region. The DNA binding region is important in binding to specific DNA sequences and integrates many cellular signals via protein-protein interactions to initiate the required cellular response (61). Inactivation and malfunctioning of the p53 protein leads to the loss of its tumor suppressor function and acquiring new functions that result in the progression of malignant tumors (62). Thus, it has been associated with the development and progression of many human cancers, including cervical cancer through the loss of its tumor suppressor function (63).

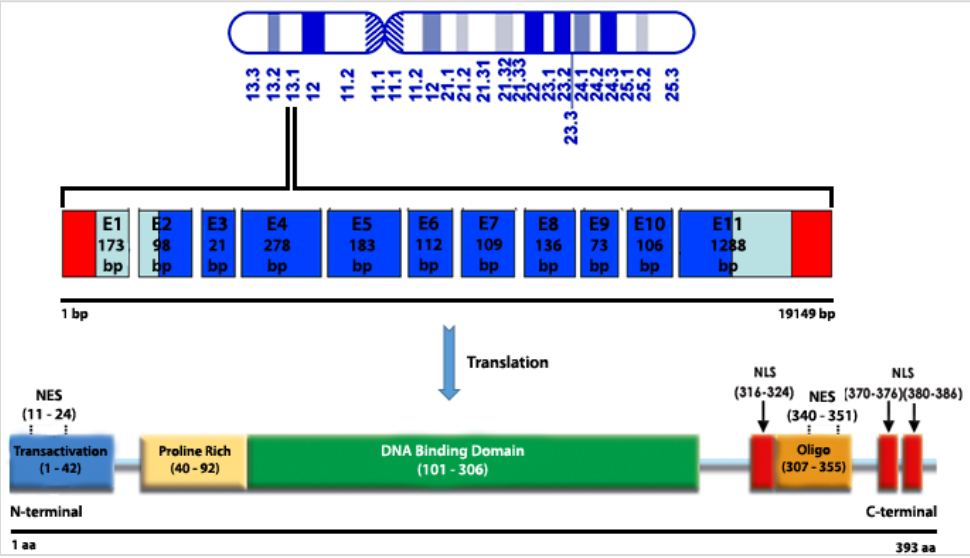


Figure 1.2: ***TP53* gene structure.** Image adapted from Pessoa et al., 2014 (64).

1.2.5.1.2 p53 and cancer progression

The *p53* gene is a major tumor suppressor, and the loss of its normal function causes carcinogenesis in many human cancers. To maintain its primary function, many molecules, including regulators of *p53* and *p53* target genes, play a role in its biological function and pathway. Another critical factor in the functioning of p53 is HPV viral oncoproteins that regulate the host tumor suppressor proteins(59). Human papillomavirus (HPV) is one of the main etiological factors in the development of cervical cancer, and the high-risk HPV types have been highly associated with the risk of developing intraepithelial lesions, squamous cell carcinoma, and adenocarcinoma of the cervix (65, 66). However, most HPV infections induce low-grade squamous epithelial lesions that spontaneously regress in more than 90% of the infected individuals(67). In Oncogenic HPVs, the early genes *E6* and *E7* play a role in transforming infected epithelial cells mainly through the inactivation of p53 and retinoblastoma (Rb) proteins, respectively, and related pathways (68). The HPV oncoprotein E6 plays a role in inactivating p53, resulting in tumor suppressor protein malfunctioning(59). This is done through the interaction of p53 and E6-associated protein (E6AP), which then binds to p53, triggering the degradation of p53 (**Figure 1.3**) through a ubiquitin degradation pathway (69, 70). The altered activity of p53 will result in the destruction of different processes, such as transcription and apoptosis, and disrupt cell cycle control, resulting in tumor progression(71).

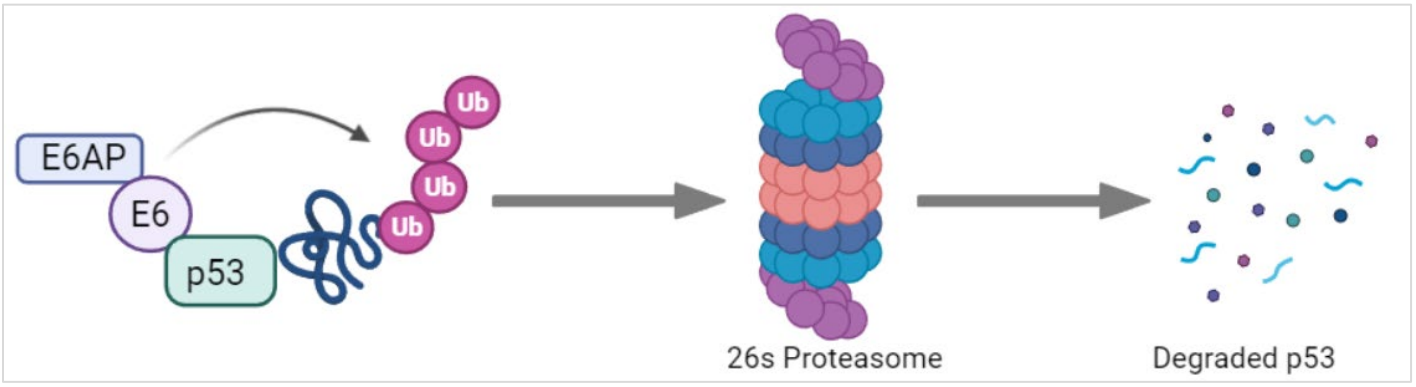


Figure 1.3: The ubiquitin pathway. Image adapted from Idres et al., 2022, with modifications (72). Image created with biorender.com

1.2.5.1.3 p53 and DNA damage repair

As previously stated, p53 has many protein regulators and target genes. In cells without DNA damage, p53 is bound to a regulator protein called Mouse double minute 2 homologue protein (MDM2). The MDM2 plays a role in maintaining the low levels of p53 by continuously degrading p53 in the ubiquitin-mediated degradation (73, 74), which will separate the p53-MDM2 complex and result in an MDM2 and a degraded p53. When there is DNA damage, the degradation of p53 stops to ensure p53 is available to perform its functions. In response to oncogene stress, the Alternative reading frame (ARF) tumor suppressor protein will be activated to inhibit MDM2[24], thus preventing the degradation of p53(73). In response to DNA damage, the DNA damage kinases ataxia telangiectasia and Rad3-related protein (ATR) and Ataxia-telangiectasia mutated (ATM) are activated and phosphorylate Checkpoint kinase 1 (CHEK1) or Checkpoint kinase 2 (CHEK2), respectively (75). The CHEK1/CHEK2 phosphorylates the p53, activating the transcription of target genes involved in specific cell responses according to the DNA damage type, thereby controlling the cell's fate (76). The cell cycle arrest genes such as cyclin-dependent kinase inhibitor1 (CDKN1A) encoding protein P21 and Growth arrest and DNA-damage inducible alpha (GADD45A) will be activated (77). Once the cell is arrested at G1, the DNA will be repaired by DNA damage genes. If DNA cannot be repaired, p53 prevents the cell from dividing. It signals it for apoptosis, which will activate the apoptosis genes such as p53 upregulated modulator of apoptosis (PUMA), Apoptosis regulator (BAX), and Noxa (77) (**Figure 1.4**). Thus, p53 can initiate DNA repair, cell-cycle arrest, and, most importantly, apoptosis when DNA repair is unsuccessful. This is important in human cancers because it suppresses tumors by stopping the proliferation of damaged DNA that can cause tumors. However, suppose there are mutations in the *TP53* gene. In that case, p53 will have aberrated binding in the DNA binding domain, resulting in the altered function of p53, which will fail to initiate the cell cycle, DNA repair, and apoptosis, resulting in damaged cells proliferating uncontrollably, leading to tumour progression.

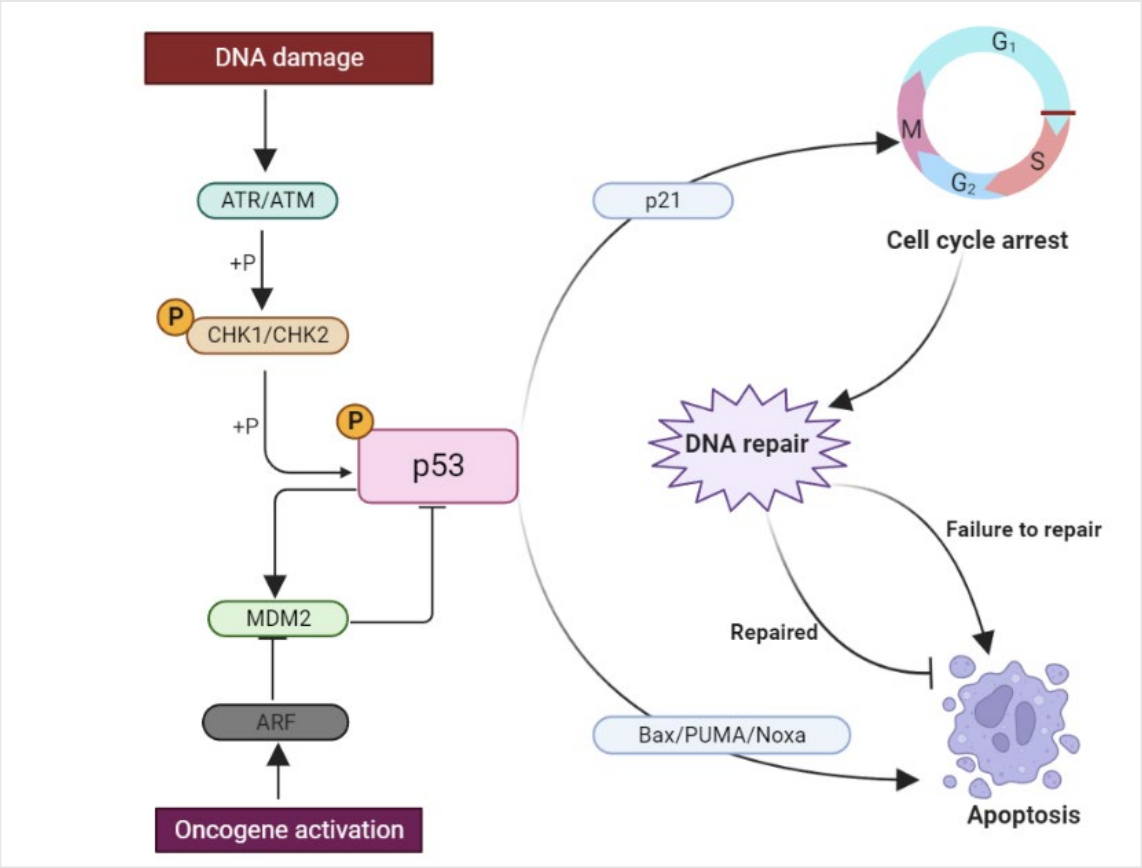


Figure 1.4: Interactions of p53 with target genes and proteins in DNA damage response. The ATM (ataxia telangiectasia mutated) and ATR (ataxia-telangiectasia-mutated-and-Rad3-related kinase) phosphorylate Chk1/2 (checkpoint kinase 1&2), which in turn phosphorylate p53 to activate it, leading to transcription of the target genes for cell cycle control, DNA repair, or apoptosis. Image created with biorender.com

1.2.5.2 Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha (*PIK3CA*)

1.2.5.2.1 *PIK3CA* structure and function

Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha (*PIK3CA*), is an oncogene and is one of the frequently mutated genes in cervical cancer (78). *PIK3CA* is located on chromosome 3q26.3 and comprises 20 exons, with exons 9 and 20 considered mutational hot spots harboring more somatic mutations that play a role in various stages of cancers (79, 80). Exons 9 and 20 are located on the helical and the kinase domains, respectively. The helicase domain plays a role in unwinding the DNA double helix during DNA replication and repair, while

the kinase domain is *PIK3CA* is a kinase in the family of phosphatidylinositol 3-kinase (PI3K), and it encodes the p110α protein, its catalytic subunit (81). The p110α subunit consists of five domains, namely: ABD (adaptor-binding domain, RBD (Ras Binding domain), C2 domain, and helical and kinase domains (**Figure 1.5**)(82). The PI3K family is vital in regulating various biological processes, including controlling cell growth, proliferation, cell survival and apoptosis (83, 84).

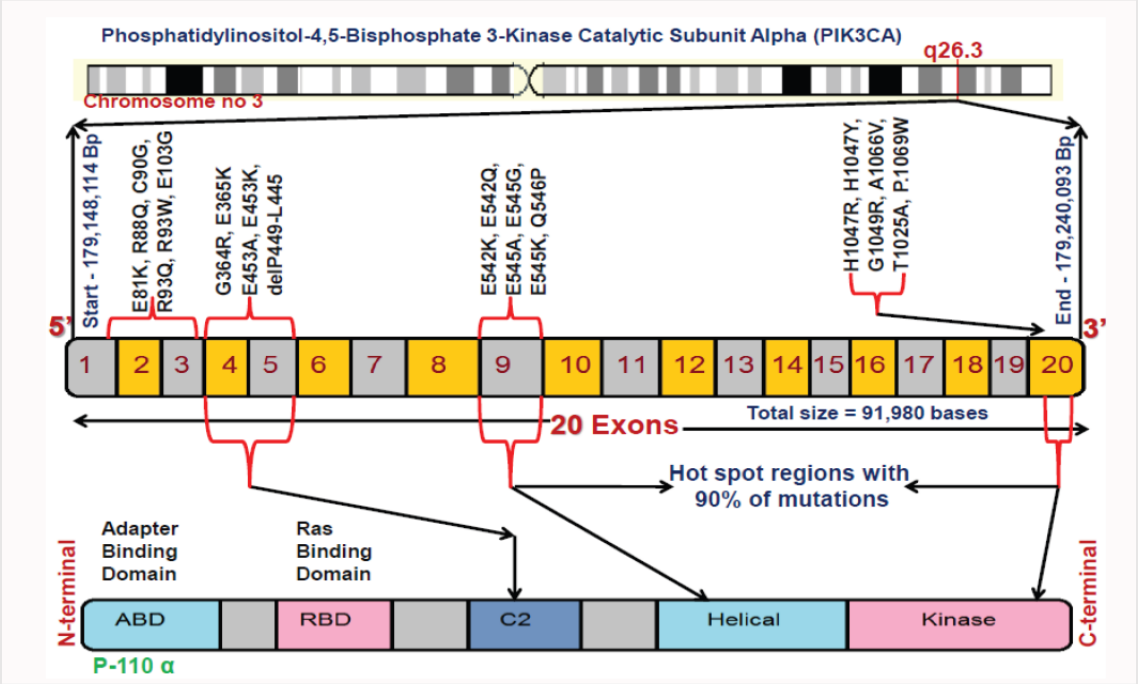


Figure 1.5: ***PIK3CA* gene structure**. *PIK3CA* comprises of 20 exons and five different protein domains with the helical and kinase domains being the hotspot domains. Image adapted from Akram-Husain and Ramakrishnan, 2016 (82).

1.2.5.2.2 PI3K/Akt/mTOR pathway

Cancer signaling pathways form a tangled network and play a huge role in cancer progression and targeted therapies. The phosphatidylinositol 3-kinase/protein kinase B/mammalian target of the rapamycin (PI3K/Akt/mTOR) pathway is one of human carcinogenesis's most important signaling pathways (85). The growth factor receptor tyrosine kinase (RTKs) sends activating signals to PI3K to initiate the PI3K/Akt/mTOR signaling pathway (**Figure 1.6**)(86). The activated PI3K converts phosphatidylinositol-4,5-bis-phosphate (PIP2) to phosphatidylinositol-3,4,5-

triphosphate (PIP3). This phosphorylation activates kinase AKT, a serine-threonine kinase. Akt then phosphorylates and leads to mTOR activation. Activated mTOR phosphorylation leads to phosphorylation and activation of two downstream signaling molecules, ribosomal protein S6 kinase1 (S6k) and eukaryotic translation initiation factor 4E binding protein 1 (4EBP1)(86). Activating these signaling molecules leads to enhanced translation of proteins that contribute to cell growth, proliferation, apoptosis, and survival (**Figure 1.6**) (87). A tumor suppressor phosphatase and tensin homolog (*PTEN*) can inhibit this signaling pathway, which dephosphorylates PIP3, reversing Akt activation and preventing downstream signaling (86). Another important factor that plays a role in this pathway is HPV infection. HPV oncoproteins E6 and E7 are known to play a role in suppressing pRb and p53, respectively. However, in addition to the suppression, HPV oncoproteins interact with other pathways, including the PI3K pathway (86). Active HPV infection interacts with *PIK3CA* genes to induce host genetic instability, deregulation of proliferation, survival, and resistance to apoptosis, eventually leading to tumorigenesis by activation and alteration of the downstream compounds of the PI3K pathway (88, 89).

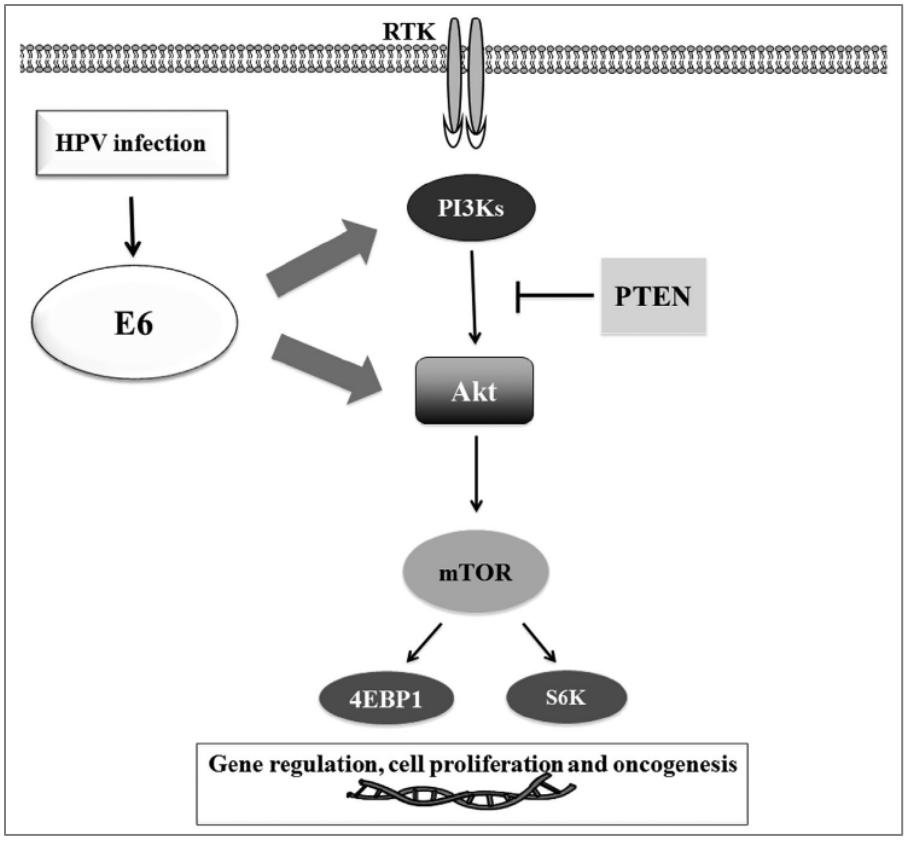


Figure 1.6: **PI3K/Akt/mTOR pathway.** RTK sends activating signals to PI3K to initiate the signaling pathway. E6 HPV oncoprotein also induces signaling pathway through PI3K activation. PI3K=phosphoinositide 3 kinase; *PTEN*= phosphatase and tensin homologue deleted on chromosome ten; Akt=protein kinase B; mTORC= mammalian target of rapamycin complex; 4EBP1=eukaryotic initiation factor 4E-binding protein 1; S6K=s6 kinase; RTK= receptor tyrosine kinase. Image adapted from Bahrami et al., 2017 (86).

1.2.5.2.3 Effects of mutations on PI3K and cancer progression

The activated PI3K/AKT/mTOR signaling pathway regulates downstream cellular functions, including cell cycle, apoptosis, protein synthesis, DNA damage repair, and angiogenesis (90). In most human cancers, it has been reported that PI3K/AKT/mTOR pathway is disrupted by alterations of the PI3K pathway (85). Abnormal activation of this pathway results in aberrant cell-cycle progression, altered adhesion, apoptosis inhibition, motility, and angiogenesis, indicating the significance of the PI3K pathway in carcinogenesis (91). *PIK3CA* is the most frequently mutated oncogene in human cancers and in cervical cancer, *PIK3CA* mutations are found in 14–23% of the cases (92). Mutations increase PIK3CA kinase activity, resulting in the overactivation of PI3K/AKT/mTOR signaling pathway. The overexpression of mutant PIK3CA proteins is believed

to result in tumor progression and resistance to standard therapy (93, 94). Another factor that increases the activation of this pathway is the tumor suppressor gene *PTEN*. *PTEN* plays a role in inhibiting the activation of Akt (**Figure 1.6**). However, *PTEN* mutations lead to overexpression of *PIK3CA* oncogene due to disrupting its inhibitory effect (95).

1.2.5.3 Phosphatase and tensin homolog (*PTEN*)

1.2.5.3.1 Structure and function

Phosphatase and tensin homolog (*PTEN*) is a tumor suppressor gene located at chromosome 10q23.31 (96). It encodes a 403 amino acid protein called phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase with lipid and protein phosphatase activities. *PTEN* gene has 9 exons, two major functional domains (a phosphatase domain and a C2 domain) and three structural regions [a short N-terminal, a C-terminal tail containing PEST sequences and a PDZ-interaction motif] (**Figure 1.7**) (97). The phosphatase domain plays a role in *PTEN*'s enzymatic activity as a phosphatase by dephosphorylating PIP3 to PIP2, thereby counteracting the activation of the PI3K pathway involved in cell growth and survival. Therefore, phosphatase activity plays a role in regulating cell proliferation and preventing tumor formation. In contrast, the C2 domain regulates cell signaling by targeting *PTEN* to the cell membrane, where it can interact with PIP3 and carry out its lipid phosphatase activity. The main function of *PTEN* is the regulation of the PI3K/Akt/mTOR, whereby it inhibits the PI3K pathway by dephosphorylating phosphatidylinositol 3,4,5-triphosphate (PIP3) to PI-4,5-bisphosphate (PIP2) thus directly antagonizing the PI3K function and inhibiting the activation of downstream signaling events (96). Therefore, *PTEN* plays a role in inhibiting cell cycle progression, induction of cell death, modulation of arrest signal, and stimulation of angiogenesis (98).

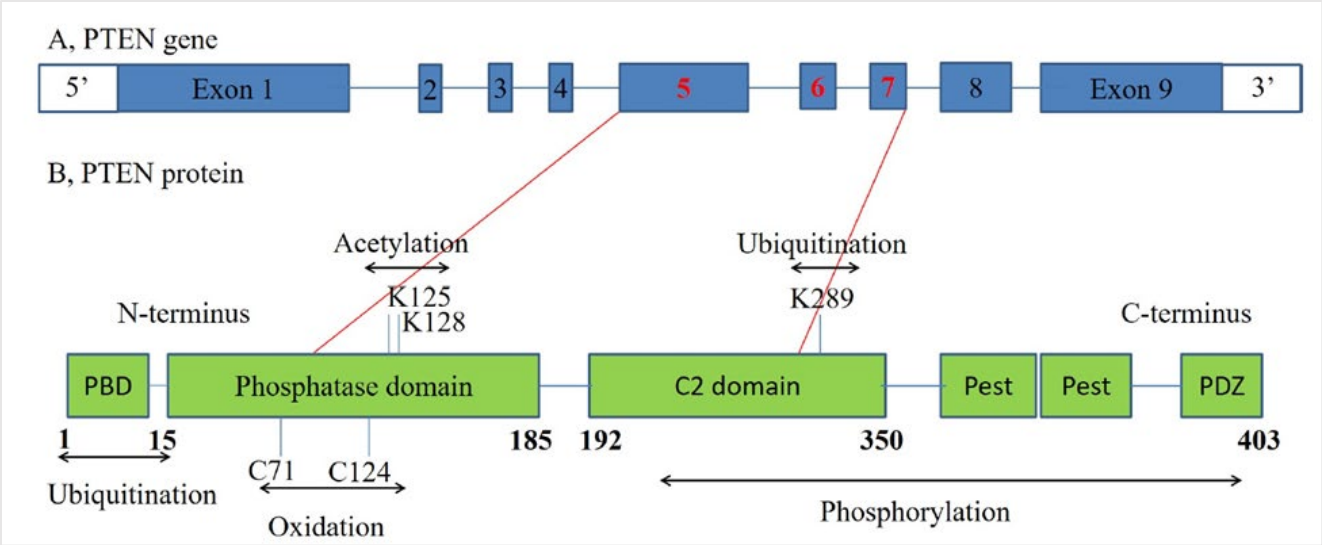


Figure 1.7: *PTEN* gene structure. The tumor suppressor gene *PTEN* comprises of 9 exons. The protein name is phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase. It encodes a polypeptide chain consisting of 403 amino acids and has different protein domains with phosphate domain (encoded by exon 5) being the hotspot domain as it enzymatic function of the protein. Image adapted from Szegedi et al., 2023 (99).

1.2.5.3.2 Effects of PTEN mutations on cancer progression

PTEN inactivation is a frequent event in many cancer types and can occur through various genetic alterations, including point mutations, large chromosomal deletions, and epigenetic mechanisms (96). The frequent somatic mutations caused by this tumor suppressor gene have even been reported in various sporadic tumors, including endometrial cancer, breast cancer, prostate cancer, malignant melanoma, and thyroid tumors (100-105). Mutations or the loss of *PTEN* results in an increased activity of PIP3 and persistent activation of PI3K effectors, which will impact cancer development and may promote tumorigenesis due to uncontrolled cell proliferation and genomic instability (106, 107).

1.2.5.4 The epidermal growth factor receptor (*EGFR*)

1.2.5.4.1 Structure and Functions

The epidermal growth factor receptor (*EGFR*), located on chromosome 7, is an oncogene with a genomic span of 110 kb distributed across 28 exons (108-110). The exons are categorized into distinct domains, with exons 1-16 in the extracellular domain, exon 17 in the transmembrane

domain, and exons 18-28 in the intracellular domain (111) (**Figure 1.8**) and consists of different functions within these domains. The hotspot regions of mutations in *EGFR* are exons 18-21 located within the tyrosine kinase domain (**Figure 1.8**) (112). *EGFR* plays a role in regulating multiple functions such as proliferation, differentiation, and acceleration of the transformation of malignant cells (113). These functions are activated when *EGFR* is activated through dimerization by activating the tyrosine kinase domain (114, 115).

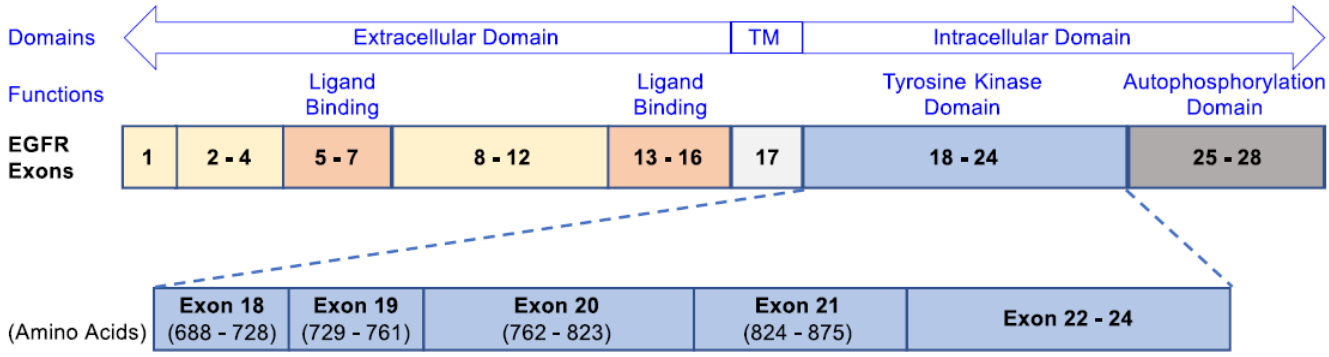


Figure 1.8: *EGFR* gene structure. *EGFR* different protein domains, functions and amino acids of the hotspot regions. Image modified from Wang et al., 2022. (116)

1.2.5.4.2 Pathway and effect of mutations on cancer progression

Activation of *EGFR* leads to the activation of many signaling pathways, including Ras/Raf/Mitogen-activated protein kinase (Ras/MAPK) pathway and phosphatidylinositol 3-kinase/AKT (PI3K/AKT) pathway (**Figure 1.9**) (116). To activate these pathways, epidermal growth factor (EGF) attaches to the extracellular domain of the EGFR, resulting in the conformational changes of EGFR that induce the tyrosine kinase domain to dimerize (116). These pathways then activate many biological downstream processes, such as cell survival, proliferation, and metastasis. When these tightly regulated pathways go awry due to mutations, it can lead to malignant transformation and tumor progression by promoting increased cell proliferation, prolonged cell survival, angiogenesis, resistance to apoptosis, invasion, and metastasis (117, 118).

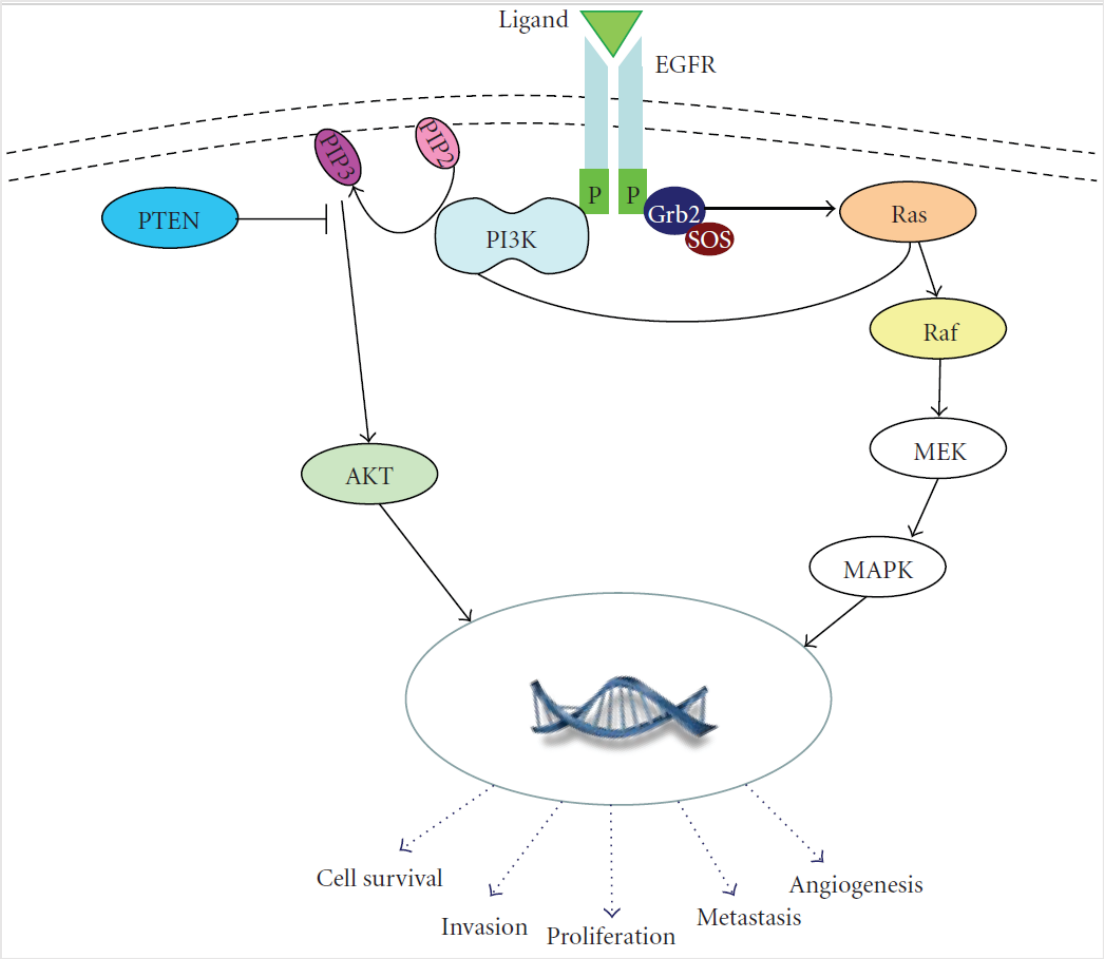


Figure 1.9: *EGFR* signaling pathway. The *EGFR* signaling pathway involves ligand-induced dimerization and activation, leading to two different downstream signaling pathways the Ras-Raf-MEK-MAPK and PI3K-AKT pathways, which in turn activate various biological processes. Meanwhile, *PTEN* exerts a negative regulatory effect on the PI3K-AKT pathway by converting PIP3 back to PIP2. Image adapted from Alyssa M. Krasinskas 2011. (119)

From evidence reported in literature, *TP53*, *PIK3CA*, *PTEN* and *EGFR* are frequently mutated genes in cervical cancer and mutations in specific exons have a significant impact on cancer progression. Therefore, *TP53* exon 4, *PIK3CA* exon 9 and 20, *PTEN* exon 5, and *EGFR* exon 19 and 20 will be the focus of this study. **Table 1.1** shows the location and the function of the hotspot regions selected in selected genes.

Table 1.1: A thorough literature search guided the selection of tumour hotspots**.**

Gene	Exon	Location	Function
TP53	4	DNA binding domain	Bind specific DNA sequences and integrate many cellular signals via protein-protein interactions to initiate the required cellular response (61).
PIK3CA	9	Helical domain	It plays a role in unwinding the DNA double helix during DNA replication and repair (81).
PIK3CA	20	Kinase domain	Responsible for the phosphorylation of phosphatidylinositol-4,5-bisphosphate (PIP2) into phosphatidylinositol-3,4,5-trisphosphate (PIP3) (81).
PTEN	5	Phosphatase domain	It plays a role in *PTEN*'s enzymatic activity as a phosphatase which metabolizes PIP3 to PIP2 in PI3K/AKT/mTOR pathway (120, 121)
EGFR	19 and 20	Tyrosine domain	Plays a role in activating signaling pathways promoting proliferation (112).

TP53: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

1.2.6 Research question

It has been observed that HIV-positive women are six time more likely to progress to invasive cervical cancer than HIV-negative women (1). Several reasons have been cited and these include a weak immune system due to HIV, alteration of the HPV infection natural history, which may result in rapid progression to high-grade lesions that regress more slowly and interaction between viral proteins, with HIV proteins enhancing the effectiveness of HPV proteins, possibly causing disruptions in the normal cell cycle (122). However, reports have shown instances of HPV and HIV con-infection, but their interaction has not been properly assessed on its effects on cervical progression. Thus, the knowledge gap to be evaluated is whether the HIV/HPV co-infection influences the prognosis of cervical cancer positively or negatively. In addition, what are the host genetic drivers that influence such infections. Therefore, this study's research question was whether there are host somatic differences among precancerous cervical biopsies obtained from

HIV-positive and HIV-negative women, and how these differences compare among HIV-positive and HIV-negative women.

1.2.7 Rationale

A growing voice shows that HIV infection contributes to rapidly progressing cervical cancer in women living with HIV compared to their HIV-negative counterparts. Despite apparent "immune restoration" rendered by antiretroviral therapy, women living with HIV have a 6-fold increased risk of developing invasive cervical cancer (1); in addition, they progress faster to invasive cervical cancer after HPV infection. Globally, studies investigating somatic alterations between HIV-positive and HIV-negative women with precancerous lesions or invasive cervical cancer are limited. In African populations, these studies are rare and yet Southern African populations have the highest HIV and cervical cancer burden compared to other regions (1). Cancers arise due to changes in the cell genome which result in changes in gene expression that ultimately lead to uncontrolled proliferation of cancerous cells and the development of tumors. Examining cancerous tumors at the genetic level makes it possible to understand better the biology of HPV-driven anogenital cancers in women living with HIV. Therefore, there is a need to understand what happens at the molecular level that causes fast progression to invasive cervical cancer in HIV-positive women compared to HIV-negative counterparts. Hence, this study will identify a cohort of HIV-positive and HIV-negative women with histologically confirmed cervical intraepithelial neoplasia (CIN3) in order to compare these two groups. This understanding will add in the application of the knowledge, which could improve patient care and diagnosis by providing prognosis markers for women living with HIV who are co-infected with HPV.

1.2.8 Aim

The study aimed to investigate the host somatic genetic variation between cervical biopsies obtained from HIV-positive and HIV-negative women with histologically confirmed CIN3 to determine potential differences in the genomic landscapes.

1.2.9 Hypotheses

1.2.9.1 Null hypothesis

There are no host somatic genetic variations between HIV-positive and HIV-negative women with histologically confirmed CIN3 biopsies.

1.2.9.2 Alternate hypothesis

There are differences in the host somatic genetic variation between HIV-positive and HIV-negative women with histologically confirmed CIN3 biopsies.

1.2.10 Objectives

1. To recruit a cohort of HIV-positive and HIV-negative women with histologically confirmed cervical intraepithelial neoplasia (CIN3).
2. To screen for HPV status of the CIN3 cohort
3. To evaluate somatic genetic variation focusing on T*P53*, *PIK3CA*, *PTEN*, and *EGFR* in CIN3 biopsies obtained from HIV-positive and HIV-negative women and HPV status.

CHAPTER 2: METHODS AND MATERIALS

This research project was carried out in the Pharmacogenomics and Drug Metabolism Research Group, Division of Human Genetics, University of Cape Town. The Pharmacogenomics and Drug Metabolism Research Group is funded by (i) the South African Medical Research Council (SAMRC), (ii) the National Research Foundation (NRF) of South Africa and the University of Cape Town. In this study, the aim was to investigate host somatic genetic variation between cervical biopsies obtained from HIV-positive or HIV-negative women co-infected with HPV to determine potential differences in cervical cancer outcomes and HPV infection between HIV-positive and HIV-negative women. Therefore, this chapter will introduce all the methods used in this study.

2.1 Study design and setting

This was a matched case-control study comparing HIV-positive and HIV-negative patients with histologically confirmed precancerous lesions, matched for age. Cervical biopsies obtained from the Division of Anatomical Pathology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, were employed in this study. This study selected women attending the Groote Schuur Hospital Cancer Clinic between 2020 and 2022. The cases were archived cervical tissue from HIV-positive women with histologically confirmed cervical intraepithelial neoplasia 3(CIN3). Cases were matched to controls which consisted of archived cervical tissue from HIV-negative women with histologically confirmed CIN3. The cases to controls were matched at a ratio of 1:1 based on age. In instances where exact age matching was not feasible, a 2-5-year range was applied.

2.2 Characteristics of the study population

Archived cervical biopsies for women who met the inclusion criteria were selected from the Division of Anatomical Pathology, Department of Pathology, Faculty of Health Sciences, University of Cape Town. The study population were women with histologically confirmed cervical intraepithelial neoplasia 3 (CIN3). To be eligible for this study, women had to be ≥ 18 years old, have histologically confirmed CIN3, have available stored biopsies collected at diagnosis and cervical biopsies collected prior to treatment (as radiotherapy and chemotherapy may affect deoxyribonucleic acid (DNA)). Women < 18 years old, had no confirmed CIN3, did not

have archived cervical biopsies, and women who had been treated with chemotherapy or radiotherapy prior to tumor resection were excluded from the study.

2.3 Sample size determination

This study compared differences in somatic polymorphisms between HIV-positive women (cases) and HIV-negative women (controls). Both cases and controls had histologically confirmed CIN3. This was a matched pair case-control study where the measured outcome is polymorphism differences in somatic genes between the case and control pairs. An immunogen study estimated that the differences in polymorphisms were, on average, 30% between the women with HIV/HPV co-infection when compared to women infected by HPV (Human Papilloma Virus) alone (49). On the assumption that the study will have a power of 80% of finding significance and using a two-sided test with a type 1 error, the sample size was calculated as follows:

$$n = (r+1)/r.\ (SD^2(Z\beta + Z\alpha\beta)^2)/d^2$$

$$= (1+1)/1.\ (0.502)^2(0.84+1.96)2/\ (0.30)^2$$

$$= 43.55$$

$$=\mathbf{44\ pairs}$$

Where n = number of pairs, r = ratio of control to cases, SD = standard deviation, d = expected mean difference between cases and controls, Zβ = standard normal variate for power, Zαβ = standard normal variate at 5% type 1 error.

2.4 Ethical clearance

This was a sub-study of the study entitled **"Investigating the role of host genetic variations in HIV/HPV co-infected women with pre-cancerous lesions and/or invasive cervical cancer."** The parent study was reviewed and approved (HREC Ref: 244/2022) by the Human Research Ethics Committee (HREC) of the Faculty of Health Sciences, University of Cape Town (Appendix A). Ethical clearance for the sub-study was sought from the Human Research Ethics Committee (HREC) of the Faculty of Health Sciences (HREC Ref:560/2022) (Appendix B).

2.5 Obtaining cervical biopsies (Pathology tissue requirements)

Formalin-fixed paraffin-embedded (FFPE) tumor blocks or unstained sections were used in the study. Tissues were assessed for histological type, subtype, grade, and p16

immunohistochemistry. The accepted tumors were from untreated patients diagnosed with cervical precancerous lesions and confirmed by pathology. Tissues with at least 80% tumour were used in the analysis. Lastly, tumors from patients with previous malignancies were accepted; however, any previous malignancies were noted.

2.6 Sample selection

Clinical reports were used to create a preliminary cohort of HIV-positive and HIV-negative women with histologically confirmed pre-cancerous lesions. Histological reports were used to determine the grade of the precancer lesion. The viral load report was used to determine the HIV status, and for those women who did not have viral load reports, pap smear reports, which record HIV status, were used to specify their HIV status. Once women with CIN3 and specified HIV status were identified, a pathologist requested the slides from the biorepository and checked whether each slide had sufficient CIN3. Then, the slides with the greatest CIN3 tissue and their blocks were requested for sectioning. The exclusion criteria for the slides not selected for sectioning were too little tissue, mostly CIN2, and lost slides.

2.7 Tissue sectioning

The FFPE blocks to be sectioned were placed in ice to cool them down and harden the wax for easy sectioning. The old blade was removed from the microtome, and the surface of the microtome was cleaned with 70% ethanol. A new blade (LS35 low profile disposable microtome blades, CPS, South Africa) was inserted in the microtome. Before cutting, forceps were cleaned using xylene to remove any residual wax from the previous sample sectioning. The microtome (Leica RM2125 RTS, Biosystems, United States) thickness was set to 4µm. The FFPE block was trimmed to remove the rough surface and flatten the block. The block was then cut, and forceps were used to collect the paraffin sections; then, four curls were placed in 1.5ml epi tubes and labeled accordingly.

2.8 DNA extraction

Genomic DNA was extracted from FFPE curls following a manufacturer's protocol from Zymo genomic DNA extraction FFPE kit (Zymo Research, California, USA). The protocol had three steps.

The first step was deparaffinization, which removed the paraffin wax in FFPE tissue. The 400 µl deparaffinization solution was added in Epi tubes with four tissue sections and incubated and incubated at 55°C for 1 minute, then vortexed briefly. The deparaffinization solution was removed from the sample by using a micropipette. The deparaffinization step was repeated twice to remove as much paraffin wax as possible. The second step was tissue digestion, which released the DNA from the cells. To the deparaffinized tissue sample, 45µl of H_2O, 45µl of 2X digestion buffer, and 10 µl of Proteinase K were added, then incubated overnight at 55°C. The proteinase K plays a role in tissue degradation by digesting proteins and harmful enzymes. This allows DNA to be in solution and available for purification. To inactivate the enzyme, the digestion solution was incubated for 20 minutes at 94°C. Then, 5µl of RNase A mix was added and incubated at room temperature for 5 minutes. RNase A removes RNA (ribonucleic acid) contamination; incubation with RNase A removes any remaining enzymes. The third step was DNA purification. The 350µl of genomic lysis buffer was added to the solution and mixed thoroughly by vortexing. To precipitate the DNA, 135µl of chilled isopropanol was added, then centrifuged at 10 000 rcf for 1 minute. The supernatant was transferred to a Zymo-Spin™ IICR column in a collection tube and centrifuged at 10 000 rcf for 1 minute. The wash buffers were used to wash impurities in the DNA. 400µl of genomic DNA wash 1 was added to a spin column in a new collection tube, then centrifuged at 10 000 rcf for 1 minute, and the flow-through was discarded. 700µl of Genomic DNA wash 2 was added to a spin column and centrifuged at ≥12 000 rcf for 1 minute, and the flow-through was discarded. 200µl of Genomic DNA wash 2 was added to a spin column and centrifuged at ≥12 000 rcf for 1 minute, and the flow-through was discarded. The collection tube was discarded, and the Zymo-SpinTM IICR column was transferred to a 1.5 microcentrifuge tube. 50µl of DNA elution buffer was added and incubated at room temperature for 5 minutes, followed by centrifugation at 14 000 rcf for 30 seconds to elute the DNA. The eluted DNA was ready for molecular-based applications or stored at ≤ 20°C for future use.

2.9 DNA Quality and Quantity assessment

Nanodrop and integrity gel were used to assess the quality and quantity of the extracted DNA.

2.9.1 Nanodrop

Nanodrop® ND-1000 spectrophotometer with the operating software version 3.81 (Thermo Fisher Scientific, Waltham, Massachusetts, United States) was used to determine the concentration and purity of DNA. Blanking was done using 2 µl elution buffer, and 1.5 µl of DNA was loaded onto the Nanodrop pedestal to determine the DNA concentration and purity. The concentration (ng/µl) and the absorbance ratios A260/280 and A260/230 were recorded. These ratios help determine the sample's purity levels or whether there is any contamination in a sample (Appendix C).

2.9.2 Integrity gel

DNA integrity was done to determine the quality of the DNA. A 1% agarose gel electrophoresis was prepared with 1X TBE (Tris/Borate/EDTA) buffer. 3 µl of DNA was mixed with 3 µl loading dye and loaded in each well of the agarose gel, and 8 µl of molecular weight marker (MWM) (GeneRular 100 bp Plus DNA Ladder, Thermo Fisher Scientific, Waltham, Massachusetts, United States) (Appendix E) was loaded. For all gel electrophoresis, 8 µl of MWM was used to determine the size of the amplified fragment. The gel ran for 60 minutes at 100 volts, and DNA bands on the agarose gel were visualized with a UVPro transilluminator (UVitec Limited, Cambridge, UK) (Appendix D).

Following the assessment of DNA quality and quantity, working solutions were prepared. Samples with DNA concentrations above 100 ng/µl were diluted with distilled water to a working sample concentration of 50 ng/µl. While those with concentrations less than 100 ng/µl, were not diluted. The stock DNA was stored in a -20 °C freezer, while the working solution was stored in a 4 °C refrigerator. The experiments were conducted using prepared working solutions. Stock solutions were used for samples that posed challenges in the amplification process.

2.10 HPV Genotyping

The genotyping of the high-risk and low-risk HPV subtypes was performed using the Anyplex™ II HPV28 Detection kit (Seegene, Seoul, South Korea), following the manufacturer's protocol. The Anyplex™ II HPV28 detection kit targets the HPV L1 gene using the primers provided in the Anyplex kit. This is a multiplex, real-time polymerase chain reaction (RT-PCR) assay that uses the Dual Priming Oligonucleotides (DPO™) and tagging oligonucleotide cleavage and extension

(TOCE™) technologies, which allows for simultaneous detection of 19 high-risk HPV genotypes namely HPV type 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 69, 73, 82, and 9 low-risk HPV genotypes namely HPV type 6, 11, 40, 42, 43, 44, 54, 61, 70 and determines an internal control (human beta-globin) in a single reaction.

A total reaction of 20 µl was used. Each reaction contained 5 µl of L1 primers (the TOCE Oligo mix), 5 µl of DNA polymerase, 5 µl of RNase-free water, and 250 ng of the sample's nucleic acid. A negative control and 3 positive controls (DNA mixture of pathogen clones) were run for each test to confirm the amplification. The CFX96™ real-time thermocycler (Bio-Rad, Applied Biosystems, California, USA) was used following the cycling conditions in the Seegene protocol (Appendix F). Fluorescence was detected using the endpoint cyclic melting temperature analysis. Seegene Viewer v2.0 program (Seegene, Seoul, South Korea) was used to analyze and interpret fluorescence data. Each result was considered valid when the internal control was detected in the sample (IC+). A positive result (+) indicated HPV DNA presence, while a negative result (-) indicated the absence of the viral DNA.

2.11 Genomic characterization

In this study, commonly mutated genes in cervical cancer were screened for genetic alterations using a sequencing approach. Somatic mutations of *TP53* (exon 4), *PIK3CA* (exon 19 and 20), *PTEN* (exon 5), and *EGFR* (exon 19 and 20) in 88 tumor samples were analyzed using polymerase chain reaction (PCR) and Sanger sequencing. The study explicitly reports alterations detected in the hotspot regions of these genes, as identified in the existing literature.

2.11.1 Primer design

Primers for polymerase chain reaction (PCR) amplification for *EGFR* were designed using PrimerQuest (https://eu.idtdna.com/Primerquest/Home/Index). Properties of a good primer set considered included a maximum annealing temperature difference of 5°C between forward and reverse primer, length of the primers between 18 to 25 nucleotides, a GC content between 40-60 %, and a ΔG value between 0 and -9 kcal/mole. OligoAnalyzer (https://eu.idtdna.com/calc/analyzer) was used to check for self-dimer, hetero-dimer, and hairpin based on their ΔG values. PrimerBLAST (https://www.ncbi.nlm.nih.gov/tools/primer-

blast/) was used to determine primer specificity to the target region. Primers used for *PIK3CA* and *PTEN* were taken from the literature, while primers for *TP53* were taken from the International Agency Research on Cancer (IARC) (IARC protocol, 2019) (**Table 2.2**). The primer sequences are shown in **Table 2.1**. All the primers were ordered from Inqaba Biotechnical Industries (Pretoria, South Africa).

Table 2.1: List of primer sequences and their annealing temperatures

Genes	Exons	Primer sequences	Amplicon size	Tm ^{0}C	Reference
PIK3CA	9	forward 5´-GGGAAAAATATGACAAAGAAAGC-3´, reverse 5´-GAGATCAGCCAAATTCAGTT-3',	248	55	(123)
	20	forward 5´-TTGATGACATTGCATACATTCG-3', reverse 5'- GGTCTTTGCCTGCTGAGAGT-3'.	236	55	
PTEN	5	forward 5'-TGTGAAGATCTTGACCAATGGC-3' reverse 5'-AAATTCTCAGATCCAGGAAGAGG-3'	231	60	(105)
EGFR	19	forward 5'-CCAGCAATATCAGCCTTAGGTG 3' reverse 5'- CCACTAGAGCTAGAAAGGGAAAG 3'	399	60.9	Designed
	20	forward 5'- TAAACGTCCCTGTGCTAGGT -3' reverse 5'- CCGTATCTCCCTTCCCTGATTACC-3'	374	60.9	
	20 (new set)	forward 5'- TGACTCCGACTCCTCCTTTA-3' reverse 5'- ATCTCCCTTCCCTGATTAC-3'	502	60	(124)
TP53	4	forward 5'-TGCTCTTTTCACCCATCTAC-3' reverse 5'- ATACGGCCAGGCATTGAAGT-3'	353	55	(125)

TP53: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

2.11.2 PCR amplification

PCR was used to amplify the region of interest for each gene. The region-specific primers that bind to the DNA template were optimized to an optimal annealing temperature (**Table 2.1**). A PCR mix was made containing 1 X of Green Go Taq Reaction Buffer (Promega Cooperation, Madison, USA), 0.4 mM of deoxyribonucleoside triphosphate (dNTPs)(Kappa Biosystems, Cape Town, South Africa), 0.03 U Go Taq Polymerase (Promega Illinois, USA), 1.5 mM magnesium chloride (Thermo Scientific, Waltham, USA), 0.4 µM forward primer, 0.4 µM reverse primer

(Inqaba Biotech, SA), and was made up to 25 µl with nuclease-free water (Thermo Scientific, Waltham, USA). A positive DNA control (DNA from the laboratory) and a no template control (reaction mixture with all reagents but no template DNA) were included in each run. A no-template control was used to monitor any contamination in the samples. In each PCR mix made, 200 ng of DNA was added.

PCR was done using an Applied Biosystems SimpliAmp^{TM} Thermal Cycler (Thermo Scientific, USA). The PCR cycling conditions were an initial denaturation at 95°C for 3 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing for 30 seconds (**Table 2.1**), and extension at 72°C for 1 minute, then final extension at 72°C for 5 minutes. For samples that were challenging to amplify, these cycling conditions were used: initial denaturation at 95°C for 10 minutes followed by 35 cycles of denaturation at 95°C for 1 minute, annealing for 1 minute (**Table 2.1**), and extension at 72°C for 1 minute, then final extension at 72°C for 10 minutes. The size of the PCR products was checked using a 1.5% agarose gel. 5 µl of the PCR product mixed with 2 µl of Gel Red (Biotium, CA, USA) was loaded into an agarose gel and electrophoresed for 60 minutes at 100V. A 100 bp Gene Ruler (Thermo Scientific, Wilmington, USA) was used as the molecular weight marker to determine the amplicon size. The fragments were visualized with a UVItech transilluminator (UVTECH, Cambridge, UK) using UVPro software.

2.11.3 PCR optimization

There are instances in which a particular DNA region provides difficulty to amplify by PCR, especially for FFPE samples, due to the degradation of the DNA. The challenges encountered during amplification process included the absence of bands in the gels, indicating a failure to amplify the target DNA segment effectively. Thus, various optimization steps were undertaken. Firstly, the DNA concentration was increased by changing the working solution DNA and using the stock solution DNA. Some of the samples not amplified were amplified after adjusting the DNA concentration. However, other samples were still not amplified. Other common optimization methods used included magnesium concentration optimization, denaturing and

annealing times and the number of cycles. Despite all these optimization steps, some samples were still not amplified.

The next step was to use PCR-enhancing agents added to the amplification mix to help obtain the expected PCR product. Betaine (N,N,N-trimethyglycine) is one of the commonly used additives, and it was used here. Betaine is an enhancer solution known to help amplify challenging targets by minimizing polymerization interruptions caused by secondary structures, thus preventing the polymerase dissociation from the DNA strand. Some of the samples were amplified; however, most of those not amplified still gave problems. Therefore, after all these optimization steps, the samples that were not amplified for each exon of interest were excluded for that exon.

However, for *EGFR* exon 20, in a total sample size of 88 women (44: HIV-positive and 44: HIV-negative women), only 55 samples were amplified. To try to increase the sample size for this exon, new primers were adapted from the literature (**Table 2.1**) and ordered from Inqaba Biotechnical Industries. The new primers used for PCR amplification were unsuccessful in amplifying these samples. Therefore, all those samples not amplified for *EGFR* exon 20 were excluded during analysis.

2.11.4 Post PCR clean-up

Post-PCR clean-up was performed to remove excess deoxynucleotide triphosphate (dNTPs) and unincorporated primers in the PCR product before sequencing. Two enzymes were used: the FastAp™ Thermosensitive Alkaline Phosphatase (FastAP) (Thermo Fisher Scientific, Waltham, USA) and Exonuclease I (ExoI) (Thermo Fisher Scientific, Waltham, USA). A clean-up master mix with 4 U ExoI, 1 U of FastAP, and 13.8µl of nuclease-free water was added to 5µl of the PCR product to make a total of 20 µl of the clean-up reaction. The clean-up reaction mix was incubated at 37^0C for 1 hour, followed by the inactivation of the enzymes at 75^0C for 15 minutes in the Applied Biosystems SimpliAmp™ Thermal Cycler.

2.11.5 Sequencing

A single-direction sequencing was used using either a forward or reverse primer. The cleaned-up PCR products were sequenced using the Big-Dye Terminator V3.1 cycle sequencing kit (Life Technologies, CA, USA), which consists of the Big-dye Terminator mix and Big-dye Terminator

buffer. A sequencing reaction mix of 1 µM reverse primer/ forward primer, 1 X of Bigdye Terminator mix, and 1 X of Bigdye Terminator buffer was added to 5 µl of cleaned-up products. The Big dye terminator sequencing conditions were initial denaturation at 98^0C for 5 minutes, followed by 25 cycles of denaturation at 96^0C for 30 seconds, annealing at 50^0C for 15 seconds, extension at 60^0C for 4 minutes, then final extension at 60^0C for 4 minutes. The sequencing reaction was done in an Applied Biosystems SimpliAmp™ Thermal Cycler.

2.11.6 Post-sequencing clean-up

The sequencing products were cleaned to remove the incorporated primers, dNTPs, and excess dyes from the sequencing using the ethanol precipitation method that uses a 20 µl sequencing reaction. Firstly, the 10 µl of the sequencing products were made to a volume of 20 µl by adding 10 µl of sabax H_2O. The sequencing products were transferred to a 96-well plate, and 5µl of the 125mM ethylenediaminetetraacetic acid (EDTA), which helps stabilize the products and remove incorporated dyes, was added. The mixture was mixed using a pipette, and then 60 µl of 100% ethanol was added to the wells, which helped precipitate the DNA. The plate was sealed with foil and vortexed for 2 to 3 seconds, and left on an ice block in the freezer for 15 minutes, followed by centrifugation for 45 minutes at 1870 relative centrifugal field (rcf). After the centrifugation, the foil was removed, and the plate was dried by placing it upside down on a paper towel. Centrifugation for 1 minute at 180 rcf followed. After centrifugation, 60 µl of 70% ethanol was added to each well to remove excess contaminants. The plate was sealed with foil and centrifuged for 15 minutes at 1870 rcf followed by removing the foil as before, and centrifugation was done for 1 minute with the plate upside down on a paper towel. The plate was airdried in a dark drawer for 15 minutes, and 10 µl of Hi-Di™ Formamide (Applied Biosystems, CA, USA) was added to the wells to resuspend samples and vortexed. The plate was then covered with septa.

2.11.7 Capillary electrophoresis

The sequencing products were denatured to create single DNA strands. For denaturation, the plate was placed on a GeneAmp PCR System 9700 (Applied Biosystems) and denatured at 95^0C for 5 minutes, followed by snap cooling by placing the plate on an ice block for 2 minutes to ensure no re-annealing occurred. After snap cooling, the samples were loaded onto the 3130xl DNA Analyzer (Applied Biosystems, CA, USA) for capillary electrophoresis.

2.12 Analysis

2.12.1 Data analysis

Data from the somatic gene sequencing using Sanger sequencing was analyzed using the DNASTAR Lasergene SeqMan Pro alignment tool (v16.0.0), Blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&BLAST_SPEC=&LINK_LOC=blasttab&LAST_PAGE=blastn) and dbSNP (https://www.ncbi.nlm.nih.gov/snp/). The sequence from DNASTAR for each sample was copied and pasted into the Blastn query box. Once the sequence had been submitted, the alignment view pairwise with dots for identifiers was chosen, and the differences between the query (reference sequence) and subject (sample sequence) were highlighted in red (**Figure 2.1**). Finally, the identified variant was searched on dbSNP to get complete information about the variant. For each gene, the obtained sequences were compared with the NCBI Reference Sequence: NG_017013.2 in the case of the *TP53* gene, NCBI Reference Sequence: NG_012113.2 in the case of the *PIK3CA* gene, NCBI Reference Sequence: NG_007466.2 in the case of the *PTEN* gene and with the NCBI Reference Sequence: NG_007726.3 in the case of the *EGFR* gene. GeneMANIA (https://genemania.org/) was used to visualize the gene-gene interaction network and predict the function of genes. In GeneMANIA, homo sapiens were selected as the target organism.

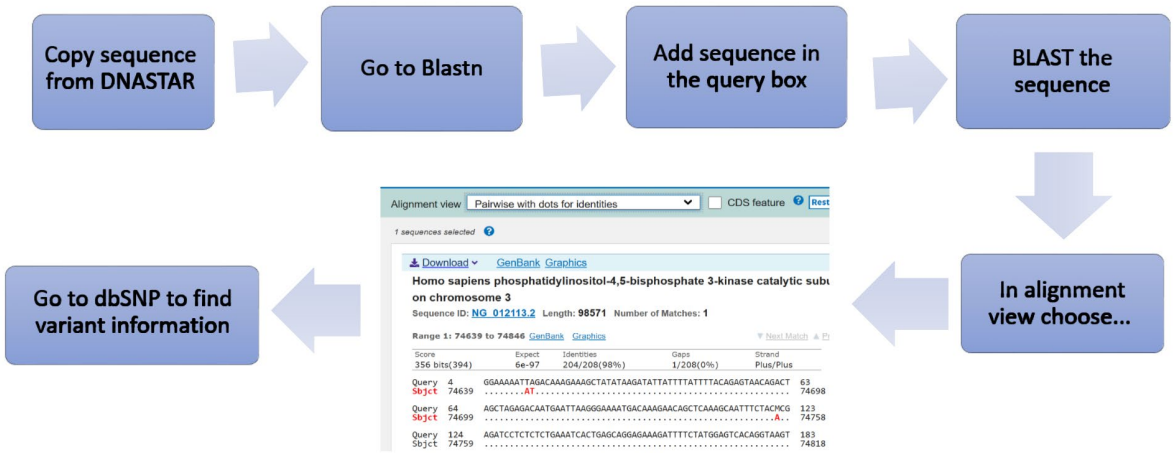


Figure 2.1: Steps to identify variants after sequencing.

2.12.2 Pathogenicity testing

The missense variants are considered non-synonymous, meaning they cause a protein change when they occur. However, pathogenicity testing is required to check whether the protein change is pathogenic or benign. Therefore, for the missense variants, pathogenicity testing tools Sorting Intolerant from Tolerant (SIFT) (https://sift.bii.a-star.edu.sg/www/Extended_SIFT_chr_coords_submit.html), and Polymorphism Phenotyping v2 (PolyPhen2) (http://genetics.bwh.harvard.edu/pph2/bgi.shtml) were used. ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), a public archive that provides human genetic variant interpretations and supporting evidence of the significance of each variant to the disease, was used. SIFT is a pathogenicity test tool that looks at the sequence homology and the physical properties of amino acids to predict potential amino acid substitution impact on the protein function, and it computes score that ranges from 0-1 (126). Variants with a SIFT score greater than or equal to 0.05 were tolerated, while those with a SIFT score less than 0.05 were reported as deleterious (126). Polyphen-2 predicts how amino acid substitution affects the structure and function of a protein (127). It computes scores ranging from 0-1, with a score less than or equal to 0.446 were reported as benign, those greater than 0.446 and less than or equal to 0.908 were reported as possibly damaging, and those greater than 0.908 were reported as probably damaging (127).Thirdly, ClinVar, a public archive that provides human genetic variant interpretations and supporting evidence of the significance of each variant to the disease was used (128).

2.12.3 Statistical analysis

Statistical analysis was performed using the STATA® SE-64 software program (version 15.0 for Windows). The categorical data were reported as N (%) [where N = the number of individuals in the category and % = the frequency of individuals within the category]. The continuous data were reported as mean ± standard deviation (SD) or median ± interquartile range (IQR) depending on whether the data is normally distributed or not normally distributed. The Shapiro-Wilks test was used to test for the normality of the data. Chi-square was used to compare the differences in sociodemographic and clinical data between HIV-positive and HIV-negative women. The p-value < 0.05 was considered statistically significant.

CHAPTER 3: RESULTS

In this study, we sought to investigate whether there are differences in host somatic genetic variation between cervical biopsies obtained from HIV-positive women and their HIV-negative counterparts. Both groups were infected with HPV. The recruited participants attended the Groote Schuur Hospital Cancer Clinic between 2020 and 2022. This chapter presents the findings on the host somatic variations in four frequently mutated genes associated with cervical cancer: *TP53*, *PIK3CA*, *PTEN,* and *EGFR*. The analysis targets six distinct hotspots within these genes, providing valuable insights into the differential genetic landscape between HIV-positive and HIV-negative participants.

3.1 Clinical and demographic characteristics

Eighty-eight samples (n=88) were retrieved for this study consisting of 44 HIV-positive women (case group) and 44 HIV-negative women (control group). Clinical and demographic characteristics of the study are presented in **Table 3.1,** with data available for 82 women. The clinical files for 6 patients could not be located from the hospital archives, hence, these patients were excluded from the clinical and demographic analyses. Therefore, there were 43 HIV-positive women and 39 HIV-negative women. More HIV-negative women reported smoking history (20 HIV-negative versus 5 HIV-positive; $p< 0.0001$), menstruation irregularity (26 HIV-negative versus 9 HIV-positive; $p= 0.005$), and use of contraception (30 HIV-negative versus 24 HIV-positive; $p= 0.019$) when compared to their HIV-positive counterparts. For other variables, there were no statistically significant differences between HIV-positive and HIV-negative women. Most of the HIV positive women had low viral load with 44% less than detectable and 37% with <200 copies/ml. For CD4 count, at least 23% (n=10) of HIV-positive women had a CD4 count greater than 500 cells/mm^3, and a minority 19% (n=8) of the HIV-positive cohort had a lower-than-expected CD4 count of less than 200 cells/mm^3. These results of the viral load and the CD4 count show good adherence to ART in women included in this study.

Table 3.1 Demographic characteristics of HIV-positive and HIV-negative women with CIN3

Characteristic	HIV-positive N = 43	HIV-negative N = 39	P-value
Median age (IQR)	37 (34-41)	35 (32-43)	0.382
Ethnicity			
African	6 (14%)	5 (13%)	0.131
Mixed Ancestry	3 (7%)	2 (5%)	
Unknown	34 (79%)	32 (82%)	-
Comorbidities			
Yes	16 (37%)	18 (46%)	0.45
No	13 (30%)	14 (36%)	
Unknown	14 (33%)	7 (18%)	-
Family history of cancer			
Yes	1 (2%)	7 (18%)	0.715
No	1 (2%)	4 (10%)	
Unknown	41 (98%)	28 (72%)	-
Smoking History			
Yes	5 (12%)	20 (51%)	<0.0001*
No	33 (52%)	19 (49%)	
Unknown	5 (12%)	0	-
Alcohol History			
Yes	6 (14%)	6 (15%)	0.784
No	11 (26%)	9 (23%)	
Unknown	26 (60%)	24 (62%)	-
BMI			
Average 20-24	11 (26%)	5 (13%)	0.243
Above average 25-29	1 (2%)	2 (5%)	
Unknown	31 (72%)	32 (82%)	-
Menarche			
10-15 years	6 (14%)	12 (31%)	0.119
16-20 years	4 (9%)	1 (2%)	
>20 years	0	1 (2%)	
Unknown	33 (77%)	25 (65%)	-
Menstruation irregularity			
Monthly	15 (35%)	3 (8%)	0.005*
Skipped for 1-3 months	1 (8%)	5 (13%)	
Amenorrhea	8 (19%)	21 (53%)	

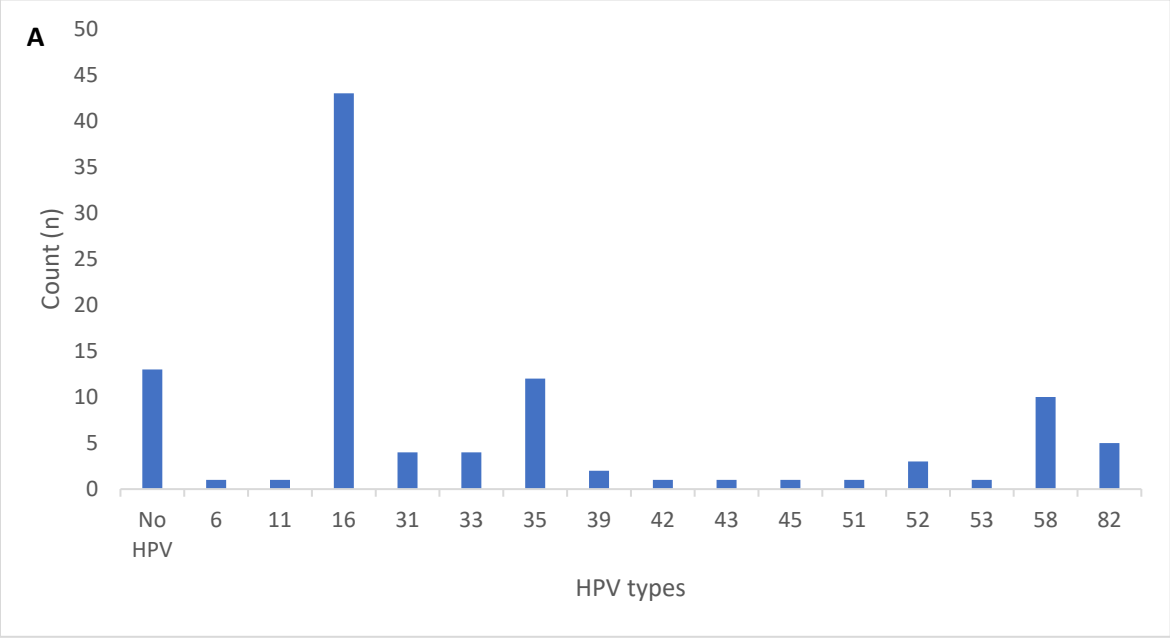
Menopause	6 (14%)	3 (8%)	
Unknown	13 (24%)	7 (18%)	-
Type of contraceptive			
Oral	0	1 (2%)	0.019*
Implant	2 (5%)	5 (13%)	
Injectable	18 (42%)	19 (49%)	
Tubal ligation	4 (9%)	5 (13%)	
None	15 (35%)	2 (5%)	
Unknown	4 (9%)	7 (18%)	-
Duration of contraceptive use			
1-5 years	2 (5%)	1 (2%)	0.902
6-10 years	2 (5%)	1 (2%)	
>10 years	3 (7%)	3 (8%)	
Unknown	36 (83%)	35 (88%)	-
Parity			
Primiparity	5 (12%)	2 (5%)	0.179
Multiparity	38 (88%)	37 (95%)	
Age at first conception			
15-20 years	6 (14%)	5 (13%)	0.186
21-25 years	4 (9%)	4 (10%)	
26-30 years	0	2 (5%)	
Unknown	33 (77%)	28 (72%)	-
Viral load (copies/ml), n (%)			
Less than detectable	19 (44%)		-
<200	16 (37%)		
200-999	3 (7%)		
>1000	5 (12%)		
CD4 count (cells/mm^3), n (%)			
<200	8 (19%)		-
200-500	12 (28%)		
>500	10 (23%)		
Unknown	13 (30%)		

SD: Standard deviation; **IQR**: interquartile range; **BMI**: body mass index,

***Represents statistical significance

3.2 HPV analysis

The 88 CIN3 patients were screened for high-risk HPV (HR-HPV), and low-risk HPV (LR-HPV), and only 87 samples were analyzed, as one sample was invalid during genotyping. 11 HR-HPV types (16, 31, 33, 35, 39, 45, 51, 52, 53, 58 and 82) and of 9 LR-HPV types (6, 11, 42 and 43) were detected in the study cohort. HPV 16 was the most common HPV type (n=43/87, 49%) identified in the study cohort, followed by HPV 35 (n=12/87, 14%) and HPV 58 (n=10/87, 11%) (**Figure 3.1A**). When stratified by HIV status, most different types of HPV were identified in the HIV-positive cohort (Figure 3.1B).



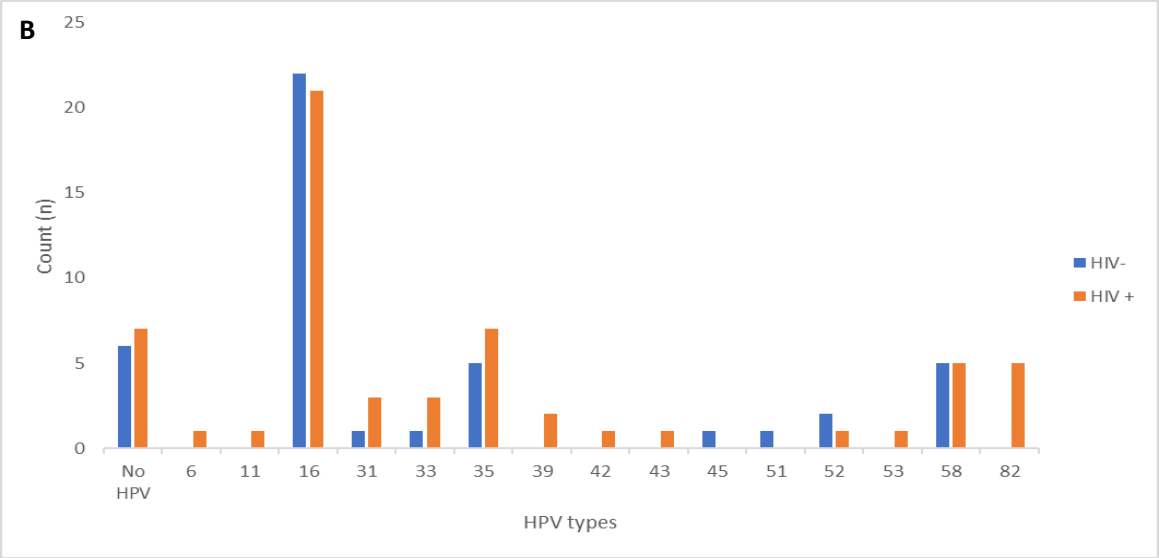


Figure 3.1: **Distribution of HPV genotype (n=87). A**: Distribution of HPV genotypes in the study cohort. **B**: Distribution of HPV genotypes stratified by HIV status.

About 70% (n=62) of the study cohort had a single HPV infection (**Figure 3.2**), 15% (n=13) of the study cohort had no HPV DNA, and 14% (n=12) had at least 2 HPV co-infections. One sample failed during genotyping and was excluded from further HPV analysis. HIV-positive women had significantly higher HPV co-infections compared to HIV-negative women (p=0.002) **(Figure 3.3).**

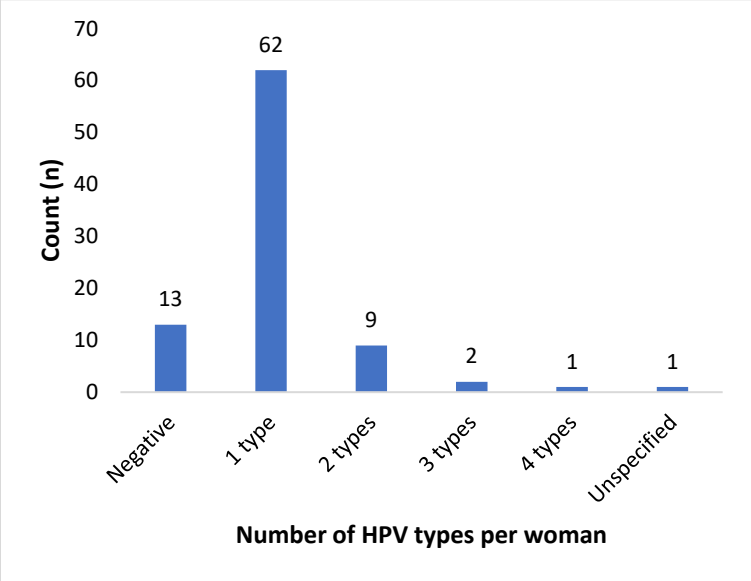


Figure 3.2: Number of HPV types in women with confirmed CIN3. The figure demonstrates the number of women that were HPV negative, with one HPV type, those with multiple infections and the unspecified HPV type.

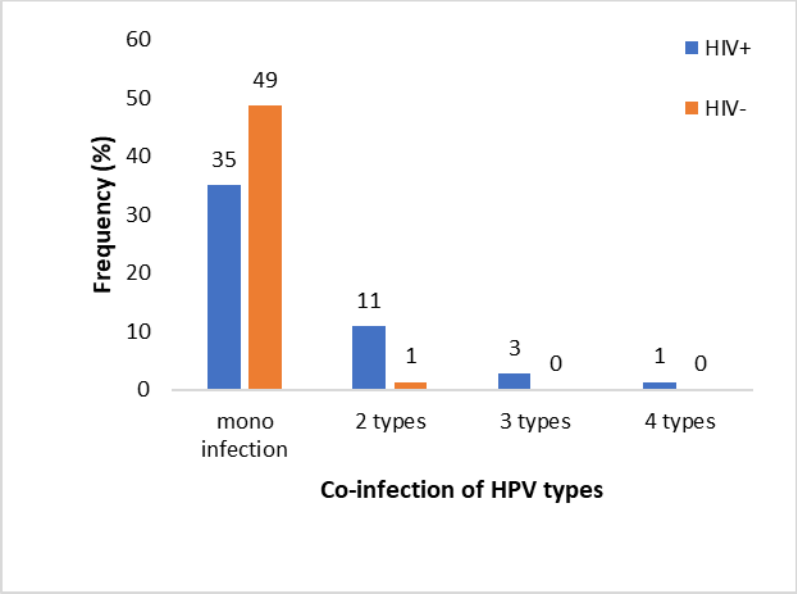


Figure 3.3: Distribution of HPV genotypes in women with confirmed CIN3 stratified by HIV status. The figure demonstrates the distribution of HPV mono-infection and multiple co-infections stratified by HIV status.

3.3 Selection of genes of interest

GeneMANIA (https://genemania.org/) is useful in determining the interactions between genes. Previous studies have identified various significantly mutated genes in cervical carcinogenesis. The Cancer Genome Atlas (TCGA) study identified 14 significantly mutated genes in cervical cancer, including *PIK3CA*, *EP300*, *FBXW7*, *HLA-B*, *PTEN*, *NFE2L2*, A*RID1A*, *KRAS*, *MAPK1 SHKBP1, ERBB3, CASP8, HLA-A,* and *TGFBR2* (80, 129). While a study by Qiu et al. identified variants in the significantly mutated genes, including *PIK3CA*, *MTOR*, *KMT2D*, *FAT1*, *MDC1*, and *TP53* in cervical cancer patients (130). Using GeneMANIA (**Figure 3.4**), we confirmed that *PIK3CA, TP53, PTEN* and *EGFR* played a pivotal role in cervical carcinogenesis pathways and, therefore, selected these four genes and their hotpots in an effort to explore potential somatic mutation variation between cervical cancer biopsies obtained from HIV positive women when compared to HIV negative women.

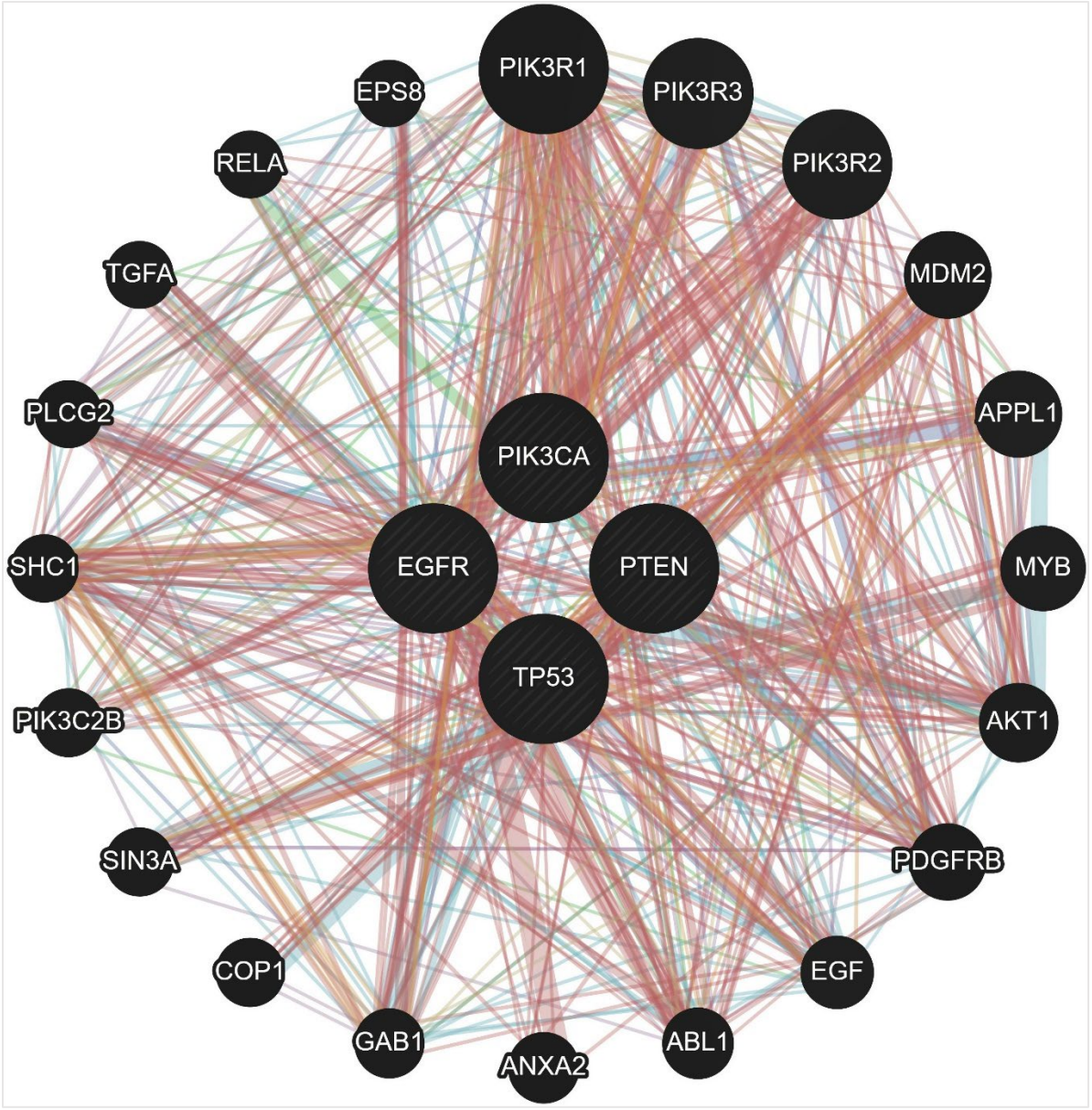


Figure 3.4: **A GeneMANIA gene-gene interaction network of the hotspot genes (*TP53, PIK3CA, PTEN* and *EGFR*).** Different colors of the network lines indicate different interactions (Physical interaction, co-expression, prediction, co-localization, genetic interaction, pathway, and shared protein domain). The thickness of the lines around each gene indicates the strength of interactions.

3.4 Genomic characterization

In this study, six hotspot regions in the four commonly altered genes, *TP53* (exon 4), *PIK3CA* (exon 19 and 20), *PTEN* (exon 5), and *EGFR* (exon 19 and 20) in cervical cancer were amplified using PCR and genotyped via Sanger sequencing.

3.4.1 PCR amplification

Each hotspot region in each gene was amplified for this study. **Figure 3.5** shows a representative figure of PCR of an amplified region. A 100bp ladder molecular weight marker (MWM) was used to measure the size of each PCR product (Appendix E). The no template control (NTC) was used to check for contamination and amplification in all hotspot regions. Meanwhile, positive control (PC) was amplified in all the runs, which identifies the validity of the run. The bands of the samples were clear and of expected fragment sizes. Due to the degradation of FFPE DNA, some samples presented with faint bands, but some could be resolved in subsequent mutation characterization with some failing.

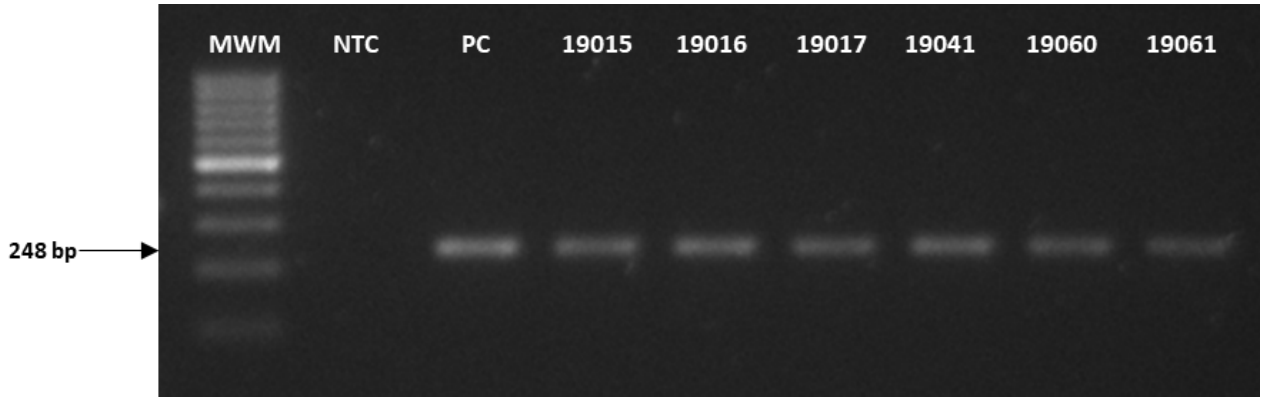


Figure 3.5: A representative gene for PCR amplification. The image indicates PCR fragments run at 100V for 1 hour on a 1.5% agarose gel for *PIK3CA* exon 9. In the first lane is the molecular weight marker (MWM), the GeneRuler™ 100bp Ladder (Thermo Fisher Scientific), the second lane is the non-template control (NTC), and the third lane is the positive control (PC). The following lanes are individual samples. The amplicon was 248bp. The other genes were done the same way (Appendix I); the amplicon sizes for them were as follows: *TP53* exon 4: 353 bp, *PIK3CA* exon 20: 231 bp, *PTEN* exon 5: 236 bp, *EGFR* exon 19: 399 bp and *EGFR* exon 20:374 bp.

3.4.2 Sanger sequencing

The variants for each hotspot were identified using Sanger Sequencing. The electropherogram of the patient's DNA was compared to the electropherogram with no mutations (reference sequences). **Figure 3.6A-C** shows the electropherogram of a patient's sequence against the reference sequence. The letter "r" in the electropherogram shows that the substitution is a purine (A or G). The identified mutation was a change from guanine to adenine, a synonymous

substitution as there was no change in the resultant amino acid glutamine at position 787 (Q787Q). All the other variants were identified in a similar fashion.

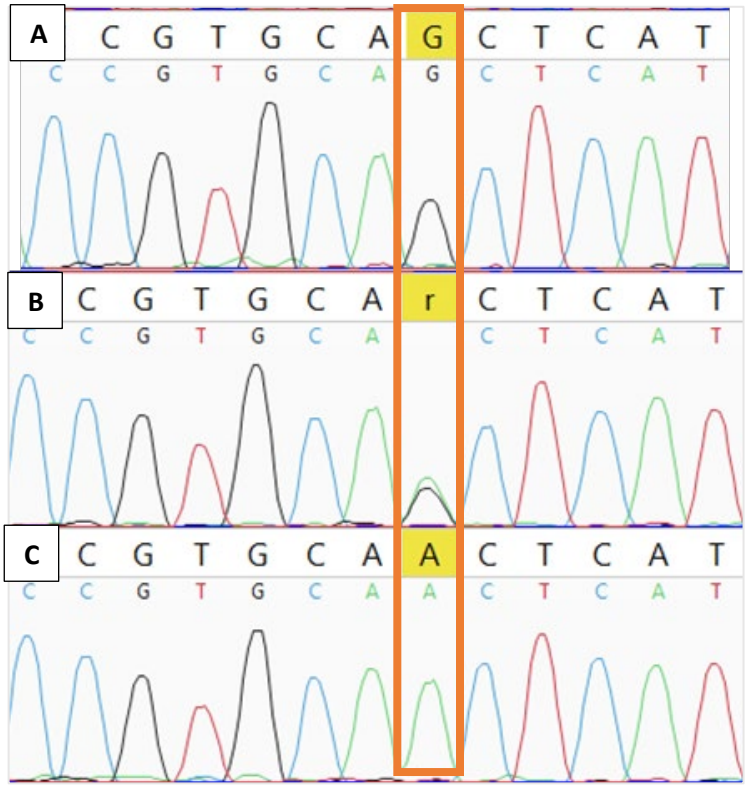


Figure 3.6: A representative of the Sanger sequencing results from *EGFR* exon 20. The electropherogram was visualized using the DNAStar SeqMan Pro alignment tool. The region in the orange box indicates a synonymous variant, a change of guanine to adenine that results in protein change Q787Q. **A.** is the GG, a wildtype genotype for patient 19021, **B.** is the GA heterogeneous genotype for patient 19049, **C.** is the AA mutant genotype for patient 19038.

3.5 Genomic landscapes

3.5.1 Total variants in the cohort

The sequencing results identified four types of mutations: stop-gain, missense, synonymous and intronic mutations. *TP53* had a higher frequency of missense and stop-gain mutations than all other genes (**Figure 3.7**). *PIK3CA* also presented with all four types of mutations, most of which were missense mutations. *PTEN* had no stop-gain nor intronic mutations, while *EGFR* had no stop-gain mutations but exhibited a higher frequency of intron mutations.

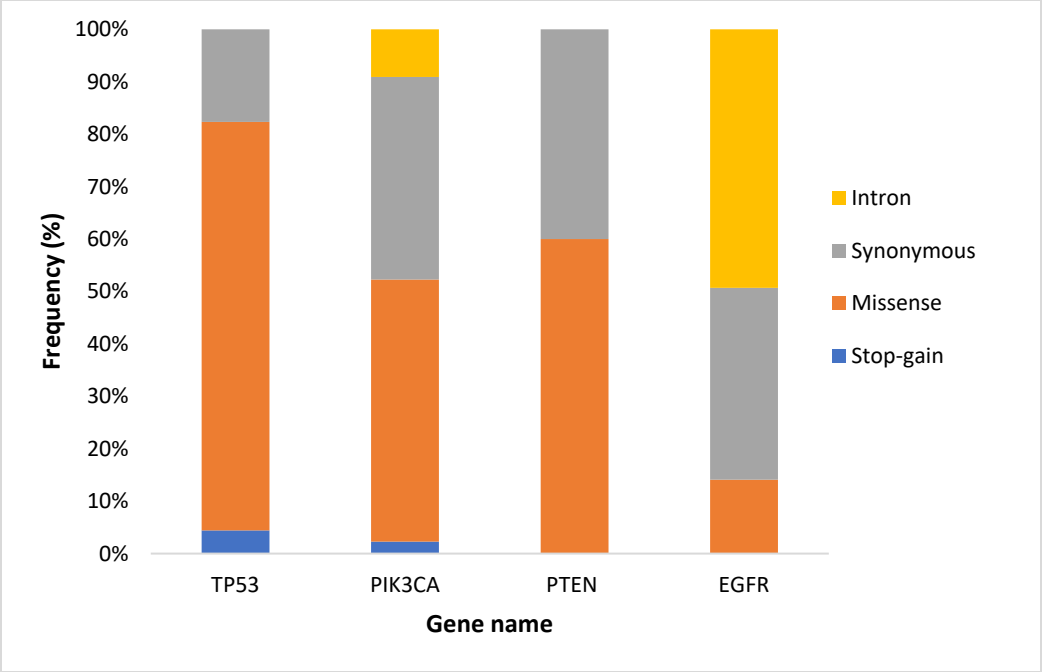
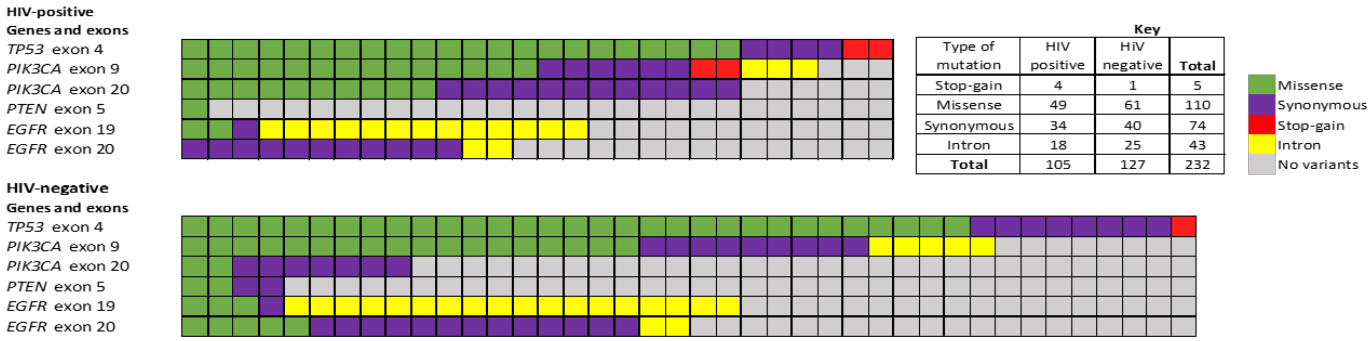


Figure 3.7: Overall type of mutations identified for each gene. Missense mutations were common in all genes except for *EGFR*. Stop-gain variants were identified in *TP53* and *PIK3CA*.

Evaluating individual hotspot regions, *TP53* exon 4 was sequenced in 77 samples, and 64% of the samples had variants. *PIK3CA* exon 9, was sequenced in 80 samples and 40% of the samples had variants, while *PIK3CA* exon 20 (76 samples), 26% had variants. *PTEN* exon 5 (76 samples), *EGFR* exon 19 (78 samples), and *EGFR* exon 20 (55 samples), frequency of samples with variants were 4%, 38%, and 47%, respectively (Appendix G and H). In total, 232 variants were identified (**Figure 3.8**). HIV-positive women had less variants (n=105) compared to their HIV-negative counterparts (n=127).

Three stop-gain variants (Q100*, Q68*, W91*) were identified in *TP53* exon 4 and two (R516*, in two patients) in *PIK3CA* exon 9, with more identified in HIV-positive women. HIV-positive women had higher mutation burden for *PIK3CA* exon 20, compared to HIV-negative women. Missense variants were identified across all the exons; however, a higher number was identified in *TP53* exon 4 for both cohorts. Similarly, synonymous variants were identified across all exons; however, more variants were identified in *EGFR* exon 20. Lastly, the primers for *PIK3CA* exon 9, *EGFR* exon

19 and 20 overlapped in the intronic regions, leading to the identification of intronic variants within these specified exons (**Figure 3.8**) (Appendix J).



Type of mutation	HIV positive	HIV negative	Total
Stop-gain	4	1	5
Missense	49	61	110
Synonymous	34	40	74
Intron	18	25	43
Total	105	127	232

Figure 3.8: Oncoprint shows the distribution of the total variants in the HIV-positive and HIV-negative cohorts. Each oncoprint box represents a mutation, and different colors represent different mutations, while each row represents a gene and exon. The key shows the total number of variants identified for HIV-positive and HIV-negative cohorts for each mutation type and the description of different colors in the oncoprint.

The mutation burden in the study cohort is illustrated in **Figure 3.9** below. Among HIV-positive patients, 14% (n=6/44) had no variants in the sequenced hotspots, while for HIV-negative women, it was 18% (n=8/44). At least 23% (n=10/44) of HIV-positive women had only one variant in any of the hotspot regions, compared to HIV-negative women (9%, n=4/44). Some patients had a higher mutation burden (≥3 variants). However, a higher mutation burden did not necessarily mean having more pathogenic variants. Few cases had a higher mutation burden that consisted of pathogenic variants. In a 29-year-old HIV-negative woman with HPV 33 (Sample ID: 63), one stop-gain variant was identified in *TP53* (W91*), and two other benign variants (P47S and A69T), as well as one synonymous variant (P82P). The second woman was a 37-year-old HIV-positive woman, HPV negative (Sample ID: 1); two stop-gain variants were identified in *TP53* (Q100* and E68*), and two benign variants (P72R and P47S). A 40-year-old HIV-positive woman whose HPV negative (Sample ID: 7) had a stop-gain variant R516* in *PIK3CA* exon 9, a benign missense variant N526K and a synonymous variant T536T. Another interesting case identified was a 53-year-old HIV-positive woman with HPV 82 (Sample ID: 31) who had variants in *PIK3CA* and *EGFR*. In *PIK3CA* exon 9, a stop-gain (R516*), a pathogenic variant (D520V) and a benign variant N526K were

identified. In *PIK3CA* exon 20, two pathogenic variants, N1068I and H1069Q, and three synonymous variants, G1050G, Q1042Q and T1052T, were identified. Meanwhile, for *EGFR* exon 19, only one intron variant was identified, and for *EGFR* exon 20, one intronic and synonymous variant (Q734Q) was identified.

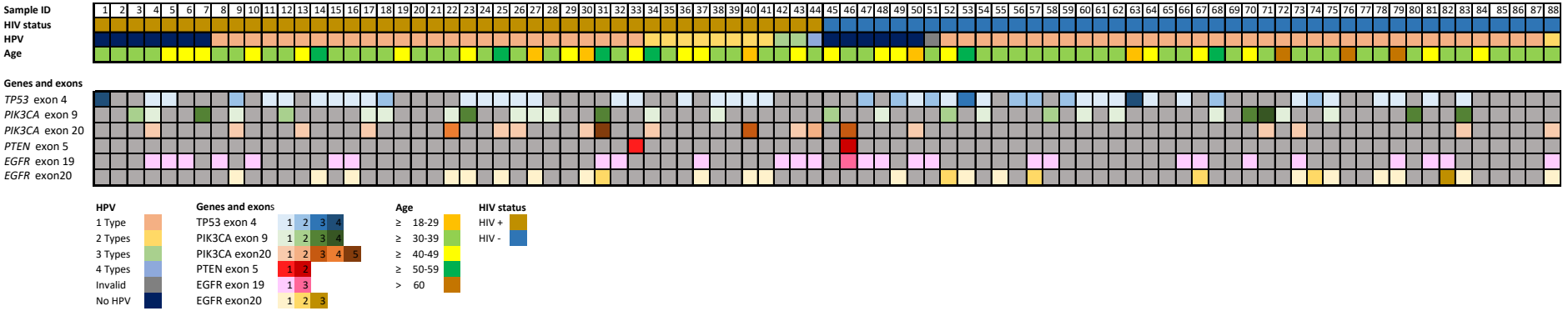


Figure 3.9: Oncoprint showing the distribution of total variants (combined, missense, stop gain, synonymous and intronic variants), number of HPV types detected per patient, HIV status and age range for the study cohort. The first top row shows the sample ID's, followed by the HIV status, HPV groupings and age range. Each oncoprint box represents each patient, and the description for each color is explained in the key below the oncoprint. Different shades of colors per gene show samples with multiple mutations. The column with only grey boxes and no other color means no variants were identified.

3.6 Pathogenicity testing

Missense variants were further analyzed to determine whether the amino acid substitution is damaging to the protein function. Sorting Intolerant from Tolerant (SIFT), Polymorphism Phenotyping v2 (Polyphen-2), and ClinVar were the tools used to test for the pathogenicity of the missense variants. The pathogenicity analysis results, including stop-gain variants, are presented in **Tables 3.2 and 3.3**.

Table 3.2: Pathogenicity testing for missense and stop-gain variants identified in the HIV-positive cohort.

			SIFT		Polyphen2			
Gene	**No. of samples:**	**protein change (Nucleotide change)**	**Score**	**Pathogenicity**	**Score**	**Pathogenicity**	**ClinVar**	**Decision**
***TP53* exon 4**	1	P67L (G/A)	0.76	Tolerated	0.031	Benign	Uncertain significance	Unlikely damaging
	18	P72R (G/C)	0.23	Tolerated	0.083	Benign	Benign	Unlikely damaging
	1	P47S (G/A)	0.9	Tolerated	0.017	Benign	Benign	Unlikely damaging
	1	R65G (T/C)	0.46	Tolerated	0	Benign	Uncertain significance/Likely benign	Unlikely damaging
	1	P82A (G/C)	0.24	Tolerated	0.001	Benign	Uncertain significance/Likely benign	Unlikely damaging
	1	Q100* (G/A)		Damaging due to stop			Pathogenic/Likely pathogenic	Most likely damaging
	1	Q68* (C/A)		Damaging due to stop			not reported in ClinVar	Most likely damaging
***PIK3CA* exon 9**	2	S541F (G/A)	0.37	Tolerated	0.929	possibly damaging	not reported on ClinVar	Probably damaging
	1	K532R (A/G)	1	Tolerated	0	benign	not reported on ClinVar	Unlikely damaging
	1	L531V (C/G)	0.01	Damaging	0.059	benign	not reported on ClinVar	Probably damaging
	1	S535F (C/T)	0.71	Tolerated	0.037	benign	not reported on ClinVar	Unlikely damaging
	1	L551V (C/G)	0.36	Tolerated	0.614	Possibly damaging	not reported on ClinVar	Probably damaging
	1	H554P (A/C)	0.15	Tolerated	0.748	possibly damaging	not reported on ClinVar	Probably damaging
	2	Q530H (G/C)	0.04	Damaging	0.854	possibly damaging	not reported on ClinVar	Most likely damaging

	1	E545A (A/C)	0.01	Damaging	0.997	probably damaging	not reported on ClinVar	Most likely damaging
	3	N526K (T/A)	0.96	Tolerated	0.002	benign	not reported on ClinVar	Unlikely damaging
	1	D520V (A/C)	0.01	Damaging	0.034	benign	not reported on ClinVar	Probably damaging
	2	R516* (A/C)		Damaging due to stop			not reported on ClinVar	Most likely damaging
***PIK3CA* exon 20**	3	H1060Q (C/G)	0	Damaging	1	probably damaging	not reported on ClinVar	Most likely damaging
	3	T1061P (T/G)	0.04	Damaging	0.982	probably damaging	not reported on ClinVar	Most likely damaging
	1	D1056N (G/A)	1	Tolerated	0.933	possibly damaging	not reported on ClinVar	Probably damaging
	1	K1054E (A/G)	0.02	Damaging	0.996	probably damaging	not reported on ClinVar	Most likely damaging
	1	Q1064H (G/T)	0.39	Tolerated	0	benign	not reported on ClinVar	Unlikely damaging
	1	M1040T (T/A)	1	Tolerated	0.001	benign	not reported on ClinVar	Unlikely damaging
	1	N1068I (A/T)	0.01	Damaging (low confidence)	0.995	probably damaging	Uncertain significance	Most likely damaging
***PTEN* exon 5**	1	D162N (G/A)	0	Damaging	0.033	Benign	not reported on ClinVar	Probably damaging
***EGFR* exon 19**	2	A690P (G>C)	0.01	Damaging	1	probably damaging	Uncertain significance	Most likely damaging

SIFT: Sorting Intolerant from Tolerant, **PolyPhen2**: Polymorphism Phenotyping v2, ***TP53***: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

Table 3.3: Pathogenicity testing for missense and stop-gain variants identified in HIV-negative cohort.

			SIFT		Polyphen2			
Gene	**No. of samples**	**protein change (Nucleotide change)**	**Score**	**Pathogenicity**	**Score**	**Pathogenicity**	**ClinVar**	**Decision**
***TP53* exon 4**	1	P67L (G/A)	0.76	Tolerated	0.031	Benign	Uncertain significance	Unlikely damaging
	20	P72R (G/C)	0.23	Tolerated	0.083	Benign	Benign	Unlikely damaging
	3	P47S (G/A)	0.9	Tolerated	0.017	Benign	Benign	Unlikely damaging
	1	A69T (C/T)	0.59	Tolerated	0.003	Benign	Uncertain significance/Likely benign	Unlikely damaging
	2	P85S (G/A)	0.82	Tolerated	0.002	Benign	Uncertain significance	Unlikely damaging
	1	Y107H (A/G)	0.82	Tolerated	0.868	Possibly damaging	Benign/Uncertain significance	Probably damaging
	1	H115R (T/C)	0.01	Damaging	0.021	Benign	Likely benign	Probably damaging
	1	T123I (G/A)	0	Damaging	0.971	Possibly damaging	Uncertain significance	Most likely damaging
	1	D42G (T/C)	0.31	Tolerated	0.029	Benign	Uncertain significance/Likely benign	Unlikely damaging
	1	W91* (C/T)		Damaging due to stop			Pathogenic/Likely pathogenic	Most likely damaging
***PIK3CA* exon 9**	7	S541F (G/A)	0.37	Tolerated	0.929	Possibly damaging	not reported on ClinVar	Probably damaging
	2	K532R (A/G)	1	Tolerated	0	benign	not reported on ClinVar	Unlikely damaging
	1	L531V (C/G)	0.01	Damaging	0.059	benign	not reported on ClinVar	Probably damaging
	3	S535F (C/T)	0.71	Tolerated	0.037	benign	not reported on ClinVar	Unlikely damaging
	1	L551V (C/G)	0.36	Tolerated	0.614	Possibly damaging	Not reported on ClinVar	Probably damaging
	1	E542D (A/T)	0.32	Tolerated	0.102	benign	not reported on ClinVar	Unlikely damaging
	1	E529D (A/C)	0.2	Tolerated	0	benign	not reported on ClinVar	Unlikely damaging
	1	Q546P (A/C)	0.07	Tolerated	1	probably damaging	Pathogenic /likely pathogenic	Most likely damaging
***PIK3CA* exon 20**	1	H1060Q (C/G)	0	Damaging	1	probably damaging	not reported on ClinVar	Most likely damaging

	1	T1061P (T/G)	0.04	Damaging	0.982	probably damaging	not reported on ClinVar	Most likely damaging
***PTEN* exon 5**	1	Q149L (A/T)	0.23	Tolerated	0.624	Possibly damaging	Not reported	Probably damaging
	1	G143V (G/T)	0.08	Tolerated	0.839	Possibly damaging	Uncertain significance	Probably damaging
***EGFR* exon 19**	1	R695K (G/A)	0.11	Tolerated	0.99	probably damaging	not reported on ClinVar	Probably damaging
	1	E696Q (G/C)	0.01	Damaging	0.991	probably damaging	Likely pathogenic	Most likely damaging
	1	K692R (A/G)	0	Damaging	1	probably damaging	Likely pathogenic	Most likely damaging
***EGFR* exon20**	2	H720R (A/G)	0	Damaging	0.409	benign	Likely pathogenic	Most likely damaging
	3	D717N (G/A)	0.28	Tolerated	0.267	benign	Uncertain significance	Unlikely damaging

SIFT: Sorting Intolerant from Tolerant, **PolyPhen2**: Polymorphism Phenotyping v2, ***TP53***: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

Variants were categorized based on their reported damaging status in SIFT or PolyPhen2 or ClinVar. Those reported as damaging in one tool were classified as probably damaging, while those with high confidence of damage reported by at least two tools were classified as most likely damaging. Variants not reported as damaging in any tool were classified as unlikely damaging. From the above two tables, when considering both probably damaging and most likely damaging variants across all genes, 25 and 23 variants were identified in HIV-positive and HIV-negative women, respectively. However, there were significantly (p=0.0406) more likely damaging variants among the HIV-positive group (31%, n=17), compared to HIV-negative women (15%, n=9). Thus, the prevalence of most likely damaging variants was greater in HIV-positive women.

3.6.1 Genomic landscapes of pathogenic variants

After pathogenicity testing, 48 pathogenic variants were identified. HIV-positive women had more damaging variants (25 versus 23) compared to the HIV-negative women. However, the two groups had no statistical significance (p=0.4321). A higher mutation burden of stop-gain variants was identified in HIV-positive women. Additionally, HIV-positive women had more most likely damaging variants compared to HIV-negative women (17 HIV-positive women, 9 HIV-negative) (p=0.0406).

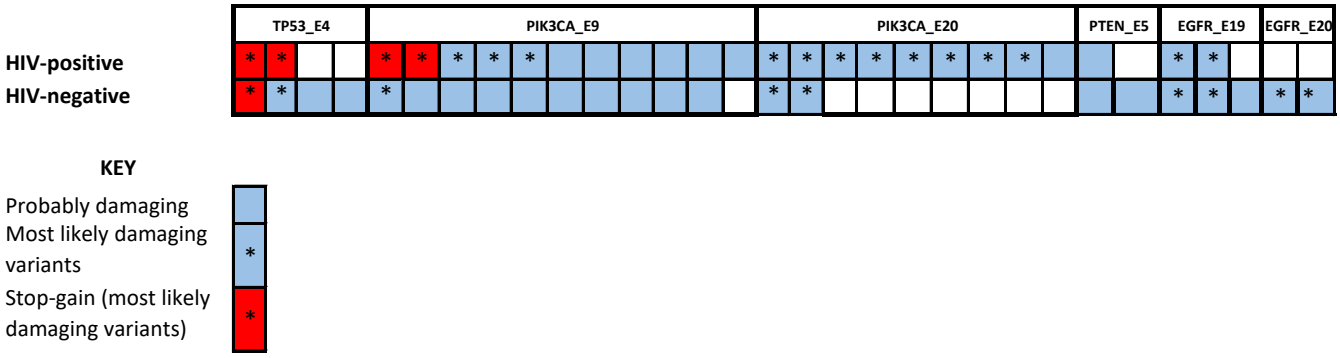


Figure 3.10: Overall number of pathogenic variants and their nature for each gene and exon, stratified by HIV status. Each box represents a woman with a pathogenic variant. The boxes with asterisks represent the most likely damaging variants, while those without asterisks are probably damaging.

The distribution of damaging variants categorized by age is illustrated in **Figure 3.11** below, where women <50 years considered young and those ≥50 years considered old. Among HIV-positive women, 24% (n=6) of variants were observed in the older cohort, while in HIV-negative women it was 8% (n=2) (p=0.1552). Notably, all stop-gain variants were identified in the younger cohort for both HIV-positive and HIV-negative women, highlighting a major occurrence of damaging variants in the younger age in both HIV-positive and HIV-negative.

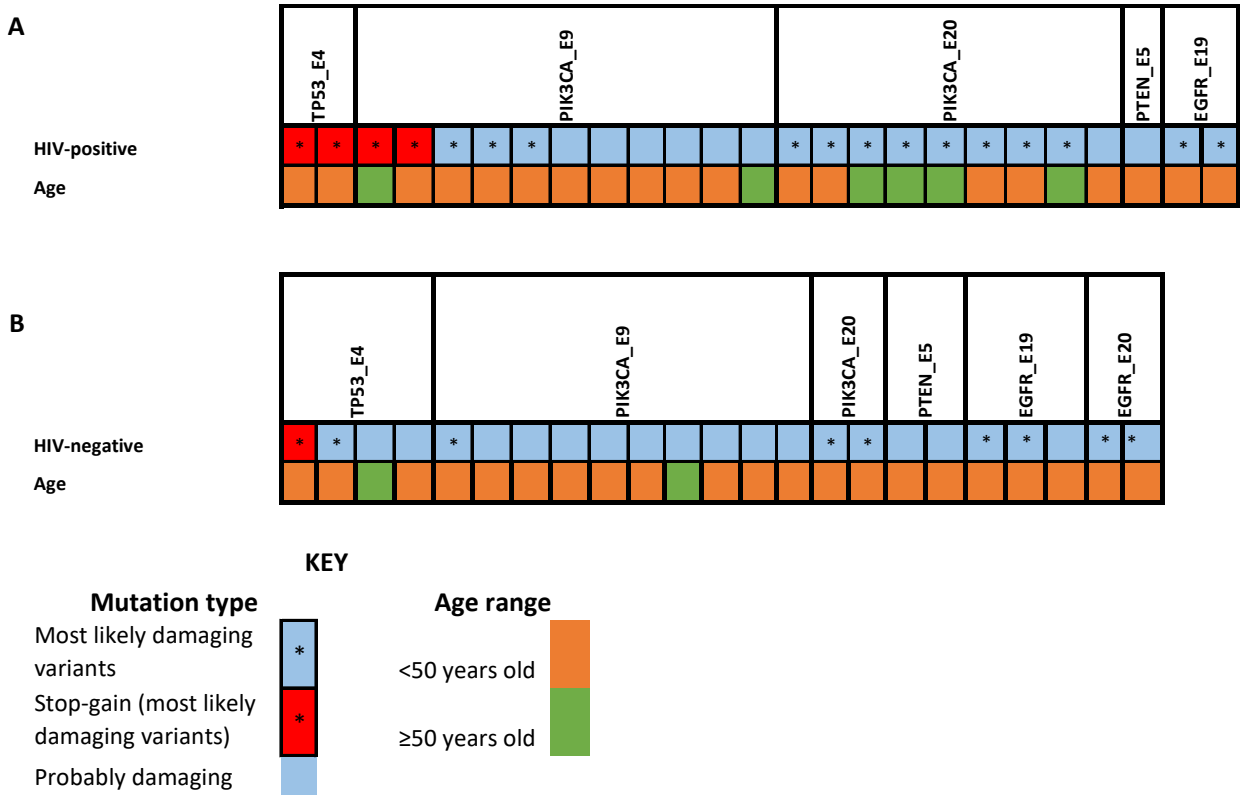


Figure 3.11: Overall number of pathogenic variants per age group (<50 years old and ≥50 years old) per HIV status. The boxes with asterisks represent most likely damaging variants, while those without asterisks are probably damaging. **A)** HIV-positive, **B)** HIV-negative

Three genes, T*P53*, *PIK3CA* and *EGFR* had the most likely damaging variants (**Figure 3.12**). *PIK3CA* had a higher mutation burden in HIV-positive women when compared to HIV-negative women. However, for HIV-negative women, *EGFR* had more variants. A significant difference in the mutation burden was observed in *PIK3CA (*p=0.0071).

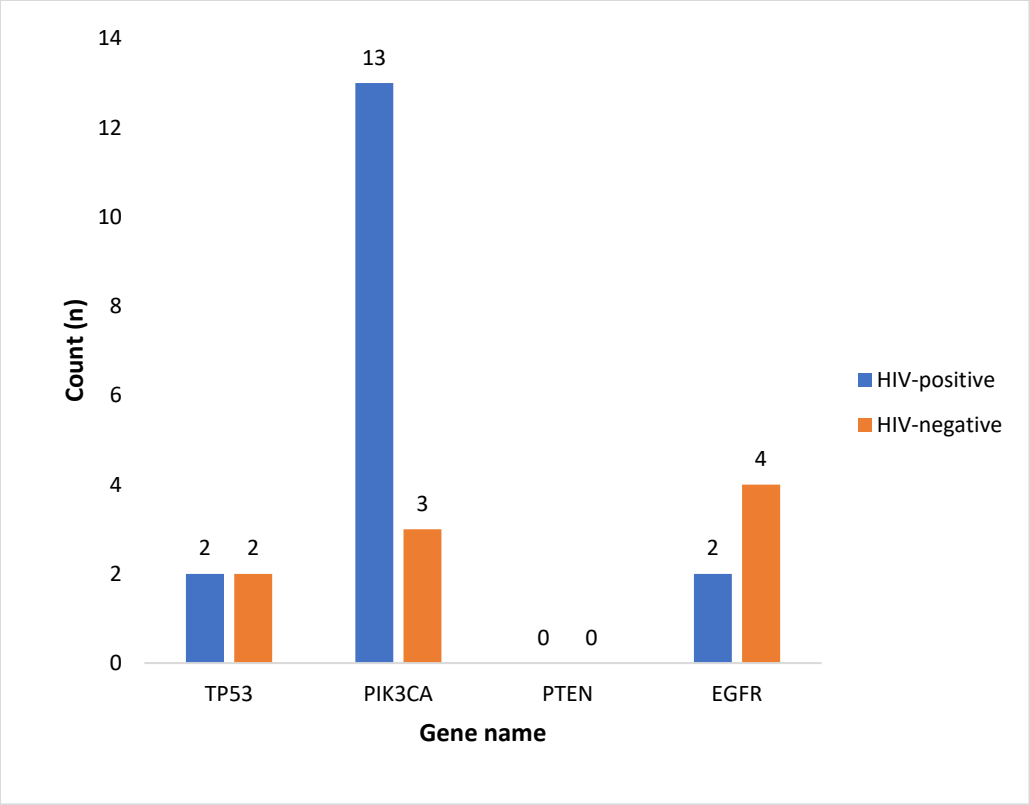


Figure 3.12: Overall number of most likely damaging variants per gene, stratified by HIV status. *TP53*: p=0.936, *PIK3CA*: p=0.007, *EGFR*: p=0.414. ***TP53***: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol-4,5- bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

CHAPTER 4: DISCUSSION

In this chapter, we will discuss the results obtained from the study, which aimed to investigate host somatic genetic variation between CIN3 cervical biopsies of HIV-positive and HIV-negative women and attempt to explain what these results may mean for cervical cancer among HIV infected women

4.1 Clinical and demographic characteristics

The study cohort size (n=88) consisted of 44 HIV-positive women and 44 HIV-negative women. The median age observed in our study cohort of HIV-positive and HIV-negative women with cervical intraepithelial lesions (CIN3) was 36 (IQR:33-43), which is lower than the median age reported in other studies conducted on women with cervical cancer. A study by Taku et al., and Kuguyo et al., reported a median age of 40 (IQR: 33-48) and 51(IQR:42-62), respectively (131, 132). Other studies have shown that patients with cervical cancer tend to be relatively young, with the median age at diagnosis of 47 years; however, almost 50% of cases diagnosed in this study were under the age of 35 years (133). This means that in our South African cohort, there is earlier onset of cervical cancer whose molecular drivers need to be decoded and understood in the context of high HIV prevalence.

We report a greater burden of HPV co-infection among the HIV positive group compared to those who did not have HIV. These results keep up with the results from previous studies, a study by Mcharo et al., also found a higher proportion of multiple concurrent HPV types in HIV-positive women compared to HIV-negative women (134). Women who are HIV-positive have an increased risk of multiple HPV infection compared to HIV-negative women, making them more susceptible to developing squamous intraepithelial lesions due to immunosuppression and HPV co-infection (135-138). However, literature also indicates that HPV infection alone is not adequate to form HPV-associated cancers (139). Other co-factors have been reported as necessary, such as smoking history, hormonal contraceptives, high parity, immunodeficiencies, and host genetic makeup. In this study, co-factors for the development of cervical cancer were analyzed. There are significant differences between HIV-positive and HIV-negative women for the presence of smoking history ($p< 0.0001$), menstruation irregularity ($p= 0.005$), and use of contraception ($p= 0.019$).

History of smoking is one of the risk factors for various cancers, including cervical cancer(140). Smoking has been independently shown to increase the risk of developing cervical cancer, and individuals who smoke are said to have a four times higher risk of developing cervical cancer compared to nonsmokers, even after adjusting for other risk factors such as sexual behavior and history of infections.(141, 142). These findings suggest that smoking drives cervical cancer by impairing the immune system and damaging the cervical epithelium through increased modification of DNA in the cervical epithelium(143), which lowers the amount of intraepithelial Langerhans cells (144). In this study, we found out that there were more smokers among women who were HIV-negative, and we think that this could be due to adjusted behavior among those with HIV who might have stopped or reduced smoking habits as part of their treatment compliance. Thus, this difference in the two groups is difficult to explain unless replicated again in another cohort.

HIV-negative women were more likely to have irregular menstruation patterns and use contraception. Irregular menstruation and the use of contraceptives have been associated with the development of cervical cancer. This association may suggest that the use of contraceptives leads to hormonal imbalances, which may lead to irregular menstruation, which in turn can promote the development of abnormal cervical cells that will increase the risk of cervical cancer (145). This is due to the estrogen and progesterone present in these hormonal contraceptives, which can alter gene expression and thereby increase the incidence of cervical cancer(146). Also, women using hormonal contraceptives have been reported to have an increased chance of engaging in unprotected sex, placing them at a higher risk of HPV infection, a major etiology of cervical cancer (147).

Cervical cancer is the most common cancer in women living with HIV and is considered an AIDS-defining illness (148). Our study showed that HIV-positive women had a higher risk of developing cervical cancer compared to HIV-negative women. This finding is similar to a study by Bosch et al., that reported HIV-positive women have a two-to-twenty-two-fold increased risk of developing cervical cancer compared to HIV-negative women (149). While a study by Stelzle et

al., reported a six-times more occurrence of cervical cancer in HIV-positive women compared to HIV-negative women (1). This increased risk is due to factors such as recurring oncogenic HPV infection, higher viral loads and decreased CD4 cell counts (150). Studies have shown that higher viral load in HIV-positive women is associated with increased risk of cervical dysplasia and progression to invasive cervical cancer (151), which is a result of immune suppression that leads to failure to fight off AIDS-defining illness or opportunistic diseases. Thus, it is important to monitor and manage viral load in HIV-positive women to facilitate early detection and prevention of cervical cancer. In this study, the commonly detected viral load was less than detectable (44%), with only 5% of the study participants having a viral load greater than 1000 copies/ml.

Another important factor that is a driver of carcinogenesis is CD4 count. A lower CD4 count has been associated with cervical cancer in women living with HIV (152). Konopnicki et al. showed that a CD4 count greater than 500 cells/µl for 18 months among HIV-positive women was significantly associated with a reduced risk of oncogenic HPV acquisition, which is a driver of the development of cervical cancer (50). While another study showed that HIV-positive women with baseline CD4 count <200 (cells/mL) had higher rates of HPV progression and lower rates of HPV regression (51). In this study, 19% of HIV-positive women had less than 200 cells/mm^3, and the majority had a CD4 count greater than 500 cells/mm^3. The study findings demonstrate ART adherence among HIV-positive women, evidenced by lower viral load and higher CD4 count in majority of the women. This suggests immune similarities between HIV-positive and HIV-negative women, with any differences potentially attributed to the co-infection of HIV and HPV and the effect of HIV/HPV co-infection on the cervical tissue.

4.2 HPV types

HPV is a common sexually transmitted infection that can cause various types of cancers, including cervical cancer. Therefore, this study identified the prevalence and distribution of high-risk HPV (HR-HPV) and low-risk HPV (LR-HPV) types among HIV-positive and HIV-negative women with CIN3. The screening results showed that 13 women in the cohort tested negative for HPV. However, they were not excluded in the study because the screening kit only targeted 28 specific

HPV types. It is important to note that there are over 200 types of HPV, therefore, these women may still have HPV strains not covered by the kit. Among the 19 HR-HPV types of the selected kit assay, 11 were detected in this study cohort and these findings indicate a diverse range of HR-HPV genotypes present in the study cohort. While among the nine LR-HPV types, only four were detected. The most common HPV genotype detected in the study cohort was HPV16, followed by HPV35 and HPV58. HPV16 was the most prevalent in the study cohort in women with CIN3, which aligns with the previous study by Stoler et al., which reported HPV16 as the most prevalent genotype in CIN3 cases (153).

The prevalence of HPV genotypes may differ based on race and geographical location (154, 155). A meta-analysis of HPV prevalence in 5 continents displayed that HPV16 and HPV18 were the most frequent types worldwide, accounting for approximately 70.9% of cervical cancers (156, 157). However, in the current study, no HPV18 was identified. The samples from this study were squamous cell carcinomas and not adenocarcinomas; a study by Bulk et al., identified that HPV18 is more commonly found in adenocarcinoma, whereas HPV16 is associated with squamous cell carcinoma (158). Hence, no HPV 18 was identified in this study. After HPV16, HPV35 was the second most prevalent type.

In sub-Saharan Africa, HPV35 has a prevalence of up to 40% among women with cervical intraepithelial neoplasia (CIN) or cervical cancer (131, 159-163). A recent study suggested a strong link between HPV35 and cervical carcinogenesis, particularly in women of African ancestry (164). A study conducted in Mozambique found that HPV35 was the most prevalent genotype among women with cervical neoplasia (165). While Mbulawa et al., 2021 identified HPV35 as the significant HPV type among females in the Eastern Cape Province of South Africa (166). Among women of African ancestry in Southern Africa, HPV35 has been detected in approximately 10% of cervical cancer cases, which is higher than the worldwide prevalence of 2% (164). Therefore, it is crucial to study further and understand the impact of HPV35 on cervical cancer cases in African women. Also, the high prevalence of HPV35 in African women calls for targeted prevention efforts and consideration of its inclusion in HPV vaccines (167). Furthermore, our study found that HIV-positive women had a higher rate of HPV co-infections compared to HIV-negative women ($p=0.002$) (**Figure 3.3**). This finding suggests that HIV-positive women may be more susceptible to

multiple strains of HPV due to the compromised immune system, making it easier for multiple strains of HPV to establish and persist in the body (168, 169).

4.3 Genomic landscapes

Cervical cancer is a gynecological cancer that develops in the cervix, and the main etiologic factor in its development is infection with HPV (53, 170). However, oncogenic mutations can trigger neoplastic transformation, leading to abnormal protein or changing the protein's expression level coded by the mutated gene (53). Thus, various studies have been done to understand the genomic landscapes of cervical cancer. These landscapes have included significantly mutated genes in cervical cancer, such as *TP53*, *PIK3CA*, *PTEN*, and *EGFR*. A study by Sharmin et al., identified mutational profiles in *EGFR*, *KRAS,* and *PIK3CA* (53). Another study by Jiang et al. created a genomic landscape that looked at the top ten frequently mutated genes among patients with gynecological cancer, including *TP53*, *PIK3CA,* and *PTEN* (54). These genes play a crucial role in the development and progression of cervical cancer; thus, understanding their alterations is needed and can potentially lead to the development of new drug targets.

Genomic profiles of four frequently mutated genes (*TP53*, *PIK3CA*, *PTEN,* and *EGFR*) in cervical cancer were performed in this study, comparing HIV-positive and HIV-negative women with CIN3. In total, 232 mutations were found in the cohort. In general, HIV-positive women had a lower mutation burden (n=105), compared to their HIV-negative counterparts (n=127). However, mutation burden refers to the total number of genetic mutations present in a tumor, regardless of whether such mutations have detrimental effects or not, thus, provides crude insights into the genomic instability of cancer cells (171). However, a more nuanced approach needs to look at the burden of damaging mutations. There is a lack of data in the literature that compares mutation burden amongst HIV-positive and HIV-negative women who have CIN or cervical cancer. In our cohort, one possible reason for the differences observed in mutation burden among HIV positive and HIV negative people could be the use of ART in HIV-positive women. ART plays a role in reducing viral replication and has transformed HIV infection into a manageable chronic condition (172) and potentially directly contributing to the genomic stability of cancer

cells through a reducing mutation burden in HIV-infected cells, thus reducing incidence of cervical lesions. However, in our study, we report that the burden of damaging mutations was higher among the HIV-positive women, pointing to a possible contributory role of the HIV-positivity and use of ART in the negative sense. HIV-positive individuals often experience persistent HPV infections because HIV weakens the immune system, making it harder for the body to clear the HPV virus. As a result, HPV infections tend to persist and have a slower clearance rate compared to HIV-negative counterparts. The prolonged presence of HPV in HIV-positive individuals creates an extended timeframe and environment for mutations to occur, potentially increasing the development of precancerous lesions and eventually cervical cancer. The evidence of this study is contradictory to other previous studies which reported the association of ART use with reduction of the occurrence of cervical lesions (173-177). While other studies support our finding that the use of ART may not have any beneficiary effect on the incidence of cervical lesions (173, 174, 177-181).

Eighty-eight (38% of total identified mutations) were found in *PIK3CA*, seventy-one (31%) in *EGFR*, sixty-eight (29%) in *TP53* and five (2%) in *PTEN*. The higher mutation burden was identified in *PIK3CA,* followed by *EGFR*. Similarly, a study by Sharmin et al., found a mutation rate of 23.9% and 52.17% for *EGFR* and *PIK3CA*, respectively (53). However, a study by Wright et al., found a lower *EGFR* mutation frequency compared to the current study (3.8% versus 31%), and a lower mutation frequency for *PIK3CA* compared to our study (31.3% versus 38%)(78). *PIK3CA* frequency of mutations in the current study was also higher than that identified in a study by Spaans et al., (38% versus 20%) (182). *TP53* also had a higher frequency of mutations when compared to a study by Wang et al., (29% versus 12%), while *PTEN* had a lower frequency of mutations in this study (2% versus 16%) (183).

When looking at the frequencies of mutations stratified by HIV status, *PIK3CA* mutations were more frequent in HIV-positive women (53% compared to 47% HIV-negative). This finding contradicts a previous study that identified a higher proportion of *PIK3CA* mutations in HIV-negative women compared to HIV-positive women (45% compared to 29% HIV-positive), resulting in a 1.3 times higher expression of *PIK3CA* in HIV-negative women (184). Conversely,

higher mutation frequencies for *TP53*, *EGFR*, and *PTEN* were observed in HIV-negative women (*TP53*: 41% versus. 59%, *EGFR*: 41% versus. 59%, *PTEN*: 20% versus. 80%), respectively. Comparing and discussing the somatic profiles of HIV-positive and HIV-negative women with CIN3 is important to understand the cancer progression between these two groups. However, so far, there is a lack of literature to support the findings of this study.

The majority of the mutations identified were missense (n=110), followed by synonymous (n=74), intronic (n=43) and lastly, stop-gain (n=5). Stop-gain mutations were identified in *TP53* (n=3) and *PIK3CA* (n=2). This may imply that these mutations are more likely to have a significant functional impact on the protein due to the stop-gain, as it results in a truncated protein. This could potentially lead to aberrant cell signaling and impaired tumor suppression, resulting in cancer development. On the other hand, *PTEN* showed no stop-gain or intronic mutations. This suggests that *PTEN* may not undergo significant alterations in its protein function, potentially maintaining its enzymatic activity and role as a tumor suppressor to prevent carcinogenesis. The *EGFR* gene showed a higher frequency of intronic mutations. The intronic mutations may not disturb the protein-coding sequence but could have functional consequences by altering gene regulation or splicing patterns.

4.4 Genomic landscapes of Pathogenic variants

The identified missense variants were tested for pathogenicity to determine whether base substitution causes a benign or pathogenic change to the protein. Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping v2 (Polyphen-2) were used. For this study, variants identified as pathogenic in at least two tools were classified as most likely damaging, while those pathogenic in one tool were classified as probably damaging. Variants that were benign in all the tools were classified as unlikely damaging. Several studies have looked at cervical cancer screening in HIV-positive and HIV-negative women (185, 186), and others have looked at the role of HPV in cervical cancer progression (187). However, no studies have looked at the somatic differences between HIV-positive and HIV-negative women. Even though it is known that HIV-positive women have a higher risk of developing cervical cancer due to the increased risk of getting HPV infection and faster progression to cervical cancer compared to HIV-negative women

(1), it is still not known what happens in the host somatic genome. Therefore, this study aimed to fill this gap. Genomic landscapes for the genes significantly mutated in cervical cancer were performed, and pathogenicity testing was performed. The high burden of most likely damaging variants in HIV-positive women suggests the increased susceptibility to CIN3 and cervical cancer in these women compared to HIV-negative women. This is supported by a study by Mpata and Nkosi that stated women living with HIV are 2-10 times more likely to develop precancerous lesions (188). These pathogenic mutations may be a driver of carcinogenesis in these women as they alter the function of the protein, thus increasing susceptibility to development of cancer. However, a study by Gagliardi et al., that characterized the genomic, transcriptomic and epigenomic landscapes of 118 cervical cancers from an understudied population of HIV-positive and HIV-negative Ugandan patients identified no differences in mutation burden between HIV-positive and HIV-negative women in the Ugandan population (184).

It has been noted that HIV-positive women have more pathogenic variants than HIV-negative women. Another factor to note that has been identified in this study is the age distribution of the pathogenic variants. The findings suggest a potential link between age and the occurrence of pathogenic variants. Most of the pathogenic variants were identified in women younger than 50 (**Figure 3.11**). Women younger than 50 years old were classified as young, and those older than 50 years of age were classified as old. The reason for more pathogenic variants in younger women may include changes in sexual behavior and the burden of associated sexually transmitted infections, though the exact causes remain unclear (189, 190). In both groups, majority of the pathogenic variants were identified in a younger cohort, with only 24% and 8% of these variants identified in an older cohort for HIV-positive and HIV-negative women, respectively. All the stop-gain variants were identified in a younger cohort of women, which indicates that younger individuals may be more susceptible to having pathogenic variants.

Overall, the most likely damaging variants were found in *PIK3CA* for HIV-positive women (n=13) than HIV-negative (n=3) (p=0.0071). *PIK3CA* is a significantly mutated gene in cervical cancer reported in multiple studies, with a frequency of 5.7-35.7% (191). The mutations found in *PIK3CA* can result in neoplastic transformation, which results in cancer progression by altering the PI3K/AKT/mTOR pathway (129). Thus, understanding the function and mutations of *PIK3CA* is

important to identify whether it is an effective target of molecular targeted therapy or as a predictive biomarker in both gene therapy and radiation therapy (192). However, these studies have looked at cervical cancer and few have looked at cervical intraepithelial neoplasia (CIN1, CIN2 and CIN3). Verlaat et al., looked at comprehensive mutation profiling of squamous precursor lesions of cervical cancer and identified that somatic mutations in *PIK3CA* are present at a later stage during cervical carcinogenesis (191). The mutations were mostly identified in cervical carcinoma; a minority was found in CIN3, and none were found in CIN2 or CIN1. While other previous studies that comprised limited sample sizes have reported no *PIK3CA* mutations in CIN cases (193, 194). In the current study, we report a higher number of *PIK3CA* pathogenic mutations in CIN3 compared to other sequenced genes (*TP53, PTEN* and *EGFR*).

PTEN and *TP53* are one of the significantly mutated genes in cervical cancer, and they are either mutated or functionally inactive in cervical cancer (195-197). *PTEN* did not have any most likely damaging variants for both HIV-positive and HIV-negative women. This could be that *PTEN* variants may not have a significant impact on the development or progression of CIN3 and only come into play at a later stage once cervical cancer has developed and advanced. Mutations in *PTEN*, even if they arise later in cervical cancer, remain significant in driving cancer progression and affecting treatment outcomes. *PTEN* is involved in various signaling pathways including the PI3K pathway, its mutations lead to dysregulation of cellular processes that promote cancer progression through disrupting these pathways. Furthermore, various biological mechanisms result in the development of precancerous lesions, and *PTEN* variants may not be involved in this development. For *TP53*, there was no significant difference in the number of most likely damaging variants in *TP53* for both HIV-positive and HIV-negative women in this study (p=0.9362). Mutations of *TP53* in CIN3 cases have been reported. A study by Tornesello et al., identified 13% of *TP53* mutations in CIN3 cases; however, there was a higher frequency of mutation in squamous cell cervical cancer (16%) (194). Thus, suggesting that the majority of these mutations manifest at advanced stages of cancer.

Abnormal activation of *EGFR* (mutations/amplification/overexpression) has been reported in different human cancers, including cervical cancer (198). In the present study, *EGFR* had most likely damaging variants present in both HIV-positive and HIV-negative women; however, there

was no statistical significance between them (p=0.4141). Previous studies have reported about 70% of *EGFR* alterations in cervical squamous carcinoma cases (199). However, *EGFR* mutations have been reported as being rare in high-grade cervical lesions and invasive cervical cancer (200). *EGFR* plays a role through dimerization that activates a tyrosine kinase domain to regulate multiple functions such as cell growth, differentiation, gene expression, and development (201). Thus, suggesting that its mutations may have a significant impact at the early stages of cervical lesions by contributing to the initiation and the development of cancer (202). This is also supported by previous studies that did not identify *EGFR* mutations in cervical cancer cases (203, 204).

4.5 Study limitations and recommendations

The first limitation of the study is the lack of information on the ethnicity of most of the cohort participants. The Second limitation of this study was that it only focused on a small set of genes. The recommendations for future studies are to explore a broader genomic landscape between HIV-positive and HIV-negative women. *PIK3CA* and *TP53* pathogenic variants were more prominent in HIV-positive women. These two genes alone could drive the differences at a molecular level and should be sequenced further beyond the hotspot regions. Also, differential expression studies must be done to understand the impact of these pathogenic variants on the protein level. This understanding will help us understand the role of host genetics in the pathology of cervical cancer in HIV-positive women that may lead to accelerated progression. In the long run, this knowledge may uncover diagnostic, prognostic, and therapeutic biomarkers in women co-infected with HIV and HPV.

4.6 Strengths of the study

This study stands out for several reasons. This is the first study investigating host somatic variation among HIV-positive versus HIV-negative women with histologically confirmed CIN3. Our study identified pathogenic variants using pathogenicity testing tools. Furthermore, it explored various genes known to be frequently mutated in cervical cancer, broadening our understanding of the

genetic landscape in this African cohort. Additionally, it had comprehensive HIV data, including CD4 count and the viral load. The strengths of the study make significant contributions to the field, offering both novel insights and a comprehensive approach to understanding the interplay between genetic factors, HIV status, and cervical cancer progression.

4.7 Conclusion

In this study, we were particularly interested in the four genes significantly mutated in cervical cancer and selected mutation hotspots only for these genes because these regions tend to accumulate a high frequency of mutations compared to the other regions of the target genes. Therefore, in this study, we checked the profiles of the mutations that were previously reported and added new data to the literature by comparing these profiles between HIV-positive and HIV-negative women to understand the faster progression from precancerous lesions to invasive cervical cancer in HIV-positive women. The results from this study indicate that HIV-positive women have a higher number of pathogenic variants compared to HIV-negative women. These variants include a higher mutation burden of stop-gain variants in HIV-positive women. This suggests that HIV infection has an impact on the host genetic profiles, resulting in increased genetic alterations that may be the cause of rapid progression to invasive cervical cancer in these women.

Another important finding was that these pathogenic variants were identified more in the younger cohort, women younger than 50. This finding implies that age may be considered as a potential factor influencing the accumulation of pathogenic variants in both HIV-positive and HIV-negative women. Understanding the relationship between age and the occurrence of pathogenic variants in these women is important as it may help in finding therapies that can prevent the effect of these mutations at an early age and possibly lead to more older women that have less aggressive cancer. Different genes have different mutation profiles; thus, studying each gene in depth is essential to understanding the effect of mutations in carcinogenesis. Finally, with respect to HIV status, HPV and host genetics, our study highlights possible interaction between HIV and HPV on the occurrence of CIN3 in younger women and postulates host genetics (mutation burden) as a possible enabler. Understanding this interaction is helpful in coming up with

molecular-based interventions. It may uncover diagnostic, prognostic, and therapeutic biomarkers in women co-infected with HIV and HPV that may enhance improved treatment and management of cervical cancer in women living with HIV and AIDS.

REFERENCES

1. Stelzle D, Tanaka LF, Lee KK, Ibrahim Khalil A, Baussano I, Shah ASV, et al. Estimates of the global burden of cervical cancer associated with HIV. The Lancet Global Health. 2021;9(2):e161-e9.
2. Brickman C, Palefsky JM. Human papillomavirus in the HIV-infected host: epidemiology and pathogenesis in the antiretroviral era. Curr HIV/AIDS Rep. 2015;12(1):6-15.
3. Mbulawa ZZ, Marais DJ, Johnson LF, Coetzee D, Williamson AL. Impact of human immunodeficiency virus on the natural history of human papillomavirus genital infection in South African men and women. J Infect Dis. 2012;206(1):15-27.
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
5. Singh D, Vignat J, Lorenzoni V, Eslahi M, Ginsburg O, Lauby-Secretan B, et al. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. Lancet Glob Health. 2023;11(2):e197-e206.
6. WHO. Cervical cancer 2021 [Available from: https://www.westerncape.gov.za/general-publication/cervical-cancer.
7. Okunade KS. Human papillomavirus and cervical cancer. J Obstet Gynaecol. 2020;40(5):602-8.
8. Yousefi Z, Aria H, Ghaedrahmati F, Bakhtiari T, Azizi M, Bastan R, et al. An Update on Human Papilloma Virus Vaccines: History, Types, Protection, and Efficacy. Front Immunol. 2022;12:805695.
9. Castellsague X. Natural history and epidemiology of HPV infection and cervical cancer. Gynecol Oncol. 2008;110(3 Suppl 2):S4-7.
10. de Freitas AC, Gurgel AP, Chagas BS, Coimbra EC, do Amaral CM. Susceptibility to cervical cancer: an overview. Gynecol Oncol. 2012;126(2):304-11.
11. Collins S, Rollason TP, Young LS, Woodman CB. Cigarette smoking is an independent risk factor for cervical intraepithelial neoplasia in young women: a longitudinal study. Eur J Cancer. 2010;46(2):405-11.
12. Ribeiro AA, Costa MC, Alves RR, Villa LL, Saddi VA, Carneiro MA, et al. HPV infection and cervical neoplasia: associated risk factors. Infect Agent Cancer. 2015;10:16.
13. Frumovitz M, Sun CC, Schover LR, Munsell MF, Jhingran A, Wharton JT, et al. Quality of life and sexual functioning in cervical cancer survivors. J Clin Oncol. 2005;23(30):7428-36.
14. Hellner K, Dorrell L. Recent advances in understanding and preventing human papillomavirus-related disease. F1000Res. 2017;6.
15. Castro-Munoz LJ, Manzo-Merino J, Munoz-Bello JO, Olmedo-Nieva L, Cedro-Tanda A, Alfaro-Ruiz LA, et al. The Human Papillomavirus (HPV) E1 protein regulates the expression of cellular genes involved in immune response. Sci Rep. 2019;9(1):13620.
16. Lacey CJ, Lowndes CM, Shah KV. Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease. Vaccine. 2006;24 Suppl 3:S3/35-41.
17. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology. 2004;324(1):17-27.
18. Van Doorslaer K, Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, et al. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res. 2017;45(D1):D499-D506.
19. zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer. 2002;2(5):342-50.
20. Doorbar J. The papillomavirus life cycle. J Clin Virol. 2005;32 Suppl 1:S7-15.
21. McKinney CC, Hussmann KL, McBride AA. The Role of the DNA Damage Response throughout the Papillomavirus Life Cycle. Viruses. 2015;7(5):2450-69.
22. Moody CA, Laimins LA. Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. PLoS Pathog. 2009;5(10):e1000605.

23. Sakakibara N, Chen D, McBride AA. Papillomaviruses use recombination-dependent replication to vegetatively amplify their genomes in differentiated cells. PLoS Pathog. 2013;9(7):e1003321.
24. Gautam D, Moody CA. Impact of the DNA Damage Response on Human Papillomavirus Chromatin. PLoS Pathog. 2016;12(6):e1005613.
25. Graham SV. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. Clin Sci (Lond). 2017;131(17):2201-21.
26. Bordignon V, Di Domenico EG, Trento E, D'Agosto G, Cavallo I, Pontone M, et al. How Human Papillomavirus Replication and Immune Evasion Strategies Take Advantage of the Host DNA Damage Repair Machinery. Viruses. 2017;9(12).
27. Herfs M, Yamamoto Y, Laury A, Wang X, Nucci MR, McLaughlin-Drubin ME, et al. A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. Proc Natl Acad Sci U S A. 2012;109(26):10516-21.
28. Stanley M. Pathology and epidemiology of HPV infection in females. Gynecol Oncol. 2010;117(2 Suppl):S5-10.
29. Koshiol J, Lindsay L, Pimenta JM, Poole C, Jenkins D, Smith JS. Persistent human papillomavirus infection and cervical neoplasia: a systematic review and meta-analysis. Am J Epidemiol. 2008;168(2):123-37.
30. Mbuya W, McHaro R, Mhizde J, Mnkai J, Mahenge A, Mwakatima M, et al. Depletion and activation of mucosal CD4 T cells in HIV infected women with HPV-associated lesions of the cervix uteri. PLoS One. 2020;15(10):e0240154.
31. Wilkins T. HIV-1-epidemiology-pathophysiology-and-transmission. Nursing times. 2020;116(7):39.
32. Akullian A, Vandormael A, Miller JC, Bershteyn A, Wenger E, Cuadros D, et al. Large age shifts in HIV-1 incidence patterns in KwaZulu-Natal, South Africa. Proc Natl Acad Sci U S A. 2021;118(28).
33. WHO. HIV and AIDS 2022 [Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids.
34. Zhou R, Liu L, Wang Y. Viral proteins recognized by different TLRs. J Med Virol. 2021;93(11):6116-23.
35. Saharia KK, Koup RA. T cell susceptibility to HIV influences outcome of opportunistic infections. Cell. 2013;155(3):505-14.
36. Burchell AN, Winer RL, de Sanjose S, Franco EL. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. Vaccine. 2006;24 Suppl 3:S3/52-61.
37. Maria J. Wawer, Ronald H. Gray, Nelson K. Sewankambo, David Serwadda, Xianbin Li, Oliver Laeyendecker, et al. Rates of HIV-1 Transmission per Coital Act, by Stage of HIV-1 Infection, in Rakai, Uganda. The Journal of Infectious Diseases. 2005;191(9):1403-9.
38. Du P. Human Papillomavirus Infection and Cervical Cancer in HIV+ Women. Cancer Treat Res. 2019;177:105-29.
39. Vernon SD, Hart CE, Reeves WC, Icenogle JP. The HIV-1 tat protein enhances E2-dependent human papillomavirus 16 transcription. Virus Res. 1993;27(2):133-45.
40. Palefsky J. Biology of HPV in HIV infection. Adv Dent Res. 2006;19(1):99-105.
41. Guidry JT, Scott RS. The interaction between human papillomavirus and other viruses. Virus Res. 2017;231:139-47.
42. De Vuyst H, Lillo F, Broutet N, Smith JS. HIV, human papillomavirus, and cervical neoplasia and cancer in the era of highly active antiretroviral therapy. European Journal of Cancer Prevention. 2008;17(6):545-54.
43. Denny LA, Franceschi S, de Sanjose S, Heard I, Moscicki AB, Palefsky J. Human papillomavirus, human immunodeficiency virus and immunosuppression. Vaccine. 2012;30 Suppl 5:F168-74.

44. Hawes SE, Critchlow CW, Faye Niang MA, Diouf MB, Diop A, Toure P, et al. Increased risk of high-grade cervical squamous intraepithelial lesions and invasive cervical cancer among African women with human immunodeficiency virus type 1 and 2 infections. J Infect Dis. 2003;188(4):555-63.
45. Chaturvedi AK, Madeleine MM, Biggar RJ, Engels EA. Risk of human papillomavirus-associated cancers among persons with AIDS. J Natl Cancer Inst. 2009;101(16):1120-30.
46. Holmes RS, Hawes SE, Toure P, Dem A, Feng Q, Weiss NS, et al. HIV infection as a risk factor for cervical cancer and cervical intraepithelial neoplasia in Senegal. Cancer Epidemiol Biomarkers Prev. 2009;18(9):2442-6.
47. Clifford GM, Polesel J, Rickenbach M, Dal Maso L, Keiser O, Kofler A, et al. Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy. J Natl Cancer Inst. 2005;97(6):425-32.
48. Giuliano AR, Lu B, Nielson CM, Flores R, Papenfuss MR, Lee JH, et al. Age-specific prevalence, incidence, and duration of human papillomavirus infections in a cohort of 290 US men. J Infect Dis. 2008;198(6):827-35.
49. Giuliano AR, Lee J-H, Fulp W, Villa LL, Lazcano E, Papenfuss MR, et al. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. The Lancet. 2011;377(9769):932-40.
50. Konopnicki D, Wit SD, Clumeck N. HPV and HIV coinnfection- complex interaction resulting in epidemiology, clinical and therapeutic implications. Future Virology. 2013;8(9):903–15.
51. Whitham HK, Hawes SE, Chu H, Oakes JM, Lifson AR, Kiviat NB, et al. A Comparison of the Natural History of HPV Infection and Cervical Abnormalities among HIV-Positive and HIV-Negative Women in Senegal, Africa. Cancer Epidemiol Biomarkers Prev. 2017;26(6):886-94.
52. Wang L, Zhou Y, Cao C, Lin S, Zhi W, Zhang D, et al. The exon 12-containing LHX6 isoforms promote cervical cancer cell proliferation by regulating the MAPK signaling pathway. Cancer Med. 2022;11(19):3657-73.
53. Sharmin S, Zohura FT, Islam MS, Shimonty A, Khan MA, Parveen R, et al. Mutational profiles of marker genes of cervical carcinoma in Bangladeshi patients. BMC Cancer. 2021;21(1):289.
54. Jiang C, Lu Y, Liu H, Cai G, Peng Z, Feng W, et al. Clinical characterization and genomic landscape of gynecological cancers among patients attending a Chinese hospital. Front Oncol. 2023;13:1143876.
55. Li T, Kon N, Jiang L, Tan M, Ludwig T, Zhao Y, et al. Tumor suppression in the absence of p53-mediated cell-cycle arrest, apoptosis, and senescence. Cell. 2012;149(6):1269-83.
56. Peng L, Xu T, Long T, Zuo H. Association Between BRCA Status and P53 Status in Breast Cancer: A Meta-Analysis. Med Sci Monit. 2016;22:1939-45.
57. Liu X, Feng D, Huo X, Xiao X, Chen Z. Association of intron microsatellite status and exon mutational profiles of TP53 in human colorectal cancer. Exp Ther Med. 2019;18(6):4287-94.
58. Nassar A, Abouelhoda M, Mansour O, Loutfy SA, Hafez MM, Gomaa M, et al. Targeted next generation sequencing identifies somatic mutations in a cohort of Egyptian breast cancer patients. J Adv Res. 2020;24:149-57.
59. Khan MA, Tiwari D, Dongre A, Sadaf, Mustafa S, Das CR, et al. Exploring the p53 connection of cervical cancer pathogenesis involving north-east Indian patients. PLoS One. 2020;15(9):e0238500.
60. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat. 2007;28(6):622-9.
61. Natan E, Baloglu C, Pagel K, Freund SM, Morgner N, Robinson CV, et al. Interaction of the p53 DNA-binding domain with its n-terminal extension modulates the stability of the p53 tetramer. J Mol Biol. 2011;409(3):358-68.
62. Fang L, Du WW, Lyu J, Dong J, Zhang C, Yang W, et al. Enhanced breast cancer progression by mutant p53 is inhibited by the circular RNA circ-Ccnb1. Cell Death Differ. 2018;25(12):2195-208.

63. Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. Nature Reviews Molecular Cell Biology. 2008;9(5):402-12.
64. Pessôa IA, da Silva FPE, Anselmo NP, de Oliveira EHC. Alterations in TP53 gene – Implications in Tumorigenesis Process and Prognosis in Central Nervous System Cancer. Tumors of the Central Nervous System - Primary and Secondary2014.
65. Albosale AH, Garbuzova OA, Kovalenko KA, Mashkina EV. Polymorphism of TP53 gene and the risk of high human papillomavirus load in cervical epithelial cells. Gene Reports. 2022;26.
66. Maria Lina Tornesello CA, Luigi Buonaguro, Simona Losito, Stefano Greggi, Buonaguro FM. <TP53 and PIK3CA gene mutations in adenocarcinoma, squamous cell carcinoma and high-grade intraepithelial neoplasia of the cervix.pdf>. Journal of Translational Medicine. 2014;12:255.
67. Steenbergen RD, Snijders PJ, Heideman DA, Meijer CJ. Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. Nat Rev Cancer. 2014;14(6):395-405.
68. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer. 2010;10(8):550-60.
69. Scheffner M, Vierstra JMHRD, Howley PM. The HPV-16 E6 and EG-AP Complex functions as ubiquitin protein ligase in the ubiquitination of p53. 1993;75(3):495–505.
70. Garima, Pandey S, Pandey LK, Saxena AK, Patel N. The Role of p53 Gene in Cervical Carcinogenesis. J Obstet Gynaecol India. 2016;66(Suppl 1):383-8.
71. Martinez J, Georgoff I, Martinez J, Levine A. Cellular localization and cell cycle regulation by a temperature sensitive p53 protein. Genes Development. 1991;5(2):151-9.
72. Idres YM, McMillan NAJ, Idris A. Hyperactivating p53 in Human Papillomavirus-Driven Cancers: A Potential Therapeutic Intervention. Mol Diagn Ther. 2022;26(3):301-8.
73. Feroz W, Sheikh AMA. Exploring the multiple roles of guardian of the genome: P53. Egyptian Journal of Medical Human Genetics. 2020;21(1).
74. Mary B. The Role p53 Protein in DNA Repair. p53 - A Guardian of the Genome and Beyond2022.
75. Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nat Rev Mol Cell Biol. 2013;14(4):197-210.
76. Bode AM, Dong Z. Post-translational modification of p53 in tumorigenesis. Nat Rev Cancer. 2004;4(10):793-805.
77. Hafner A, Bulyk ML, Jambhekar A, Lahav G. The multiple mechanisms that regulate p53 activity and cell fate. Nat Rev Mol Cell Biol. 2019;20(4):199-210.
78. Wright AA, Howitt BE, Myers AP, Dahlberg SE, Palescandolo E, Van Hummelen P, et al. Oncogenic mutations in cervical cancer: genomic differences between adenocarcinomas and squamous cell carcinomas of the cervix. Cancer. 2013;119(21):3776-83.
79. Kelley LC, Shahab S, Weed SA. Actin cytoskeletal mediators of motility and invasion amplified and overexpressed in head and neck cancer. Clin Exp Metastasis. 2008;25(4):289-304.
80. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, et al. Landscape of genomic alterations in cervical carcinomas. Nature. 2014;506(7488):371-5.
81. Ma YY, Wei SJ, Lin YC, Lung JC, Chang TC, Whang-Peng J, et al. PIK3CA as an oncogene in cervical cancer. Oncogene. 2000;19(23):2739-44.
82. Akram RH, Ramakrishnan V. PIK3CA gene mutations in major gynecologic cancers. Clin Oncol. 2016;1:1110.
83. Cantley LC. The phosphoinositide 3-kinase pathway. Science. 2002;296(5573):1655-7.
84. Courtney KD, Corcoran RB, Engelman JA. The PI3K pathway as drug target in human cancer. J Clin Oncol. 2010;28(6):1075-83.
85. German S, Aslam HM, Saleem S, Raees A, Anum T, Alvi AA, et al. Carcinogenesis of PIK3CA. Hereditary cancer in clinical practice. 2013;11:1-3.

86. Bahrami A, Hasanzadeh M, Hassanian SM, ShahidSales S, Ghayour-Mobarhan M, Ferns GA, et al. The Potential Value of the PI3K/Akt/mTOR Signaling Pathway for Assessing Prognosis in Cervical Cancer and as a Target for Therapy. J Cell Biochem. 2017;118(12):4163-9.
87. Bahrami A, Khazaei M, Hasanzadeh M, ShahidSales S, Joudi Mashhad M, Farazestanian M, et al. Therapeutic Potential of Targeting PI3K/AKT Pathway in Treatment of Colorectal Cancer: Rational and Progress. J Cell Biochem. 2018;119(3):2460-9.
88. Keysar SB, Astling DP, Anderson RT, Vogler BW, Bowles DW, Morton JJ, et al. A patient tumor transplant model of squamous cell cancer identifies PI3K inhibitors as candidate therapeutics in defined molecular bins. Mol Oncol. 2013;7(4):776-90.
89. Zhang L, Wu J, Ling MT, Zhao L, Zhao KN. The role of the PI3K/Akt/mTOR signalling pathway in human cancers induced by infection with human papillomaviruses. Mol Cancer. 2015;14:87.
90. Pergialiotis V, Nikolaou C, Haidopoulos D, Frountzas M, Thomakos N, Bellos I, et al. PIK3CA Mutations and Their Impact on Survival Outcomes of Patients with Cervical Cancer: A Systematic Review. Acta Cytol. 2020;64(6):547-55.
91. Vivanco I, Sawyers CL. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. Nat Rev Cancer. 2002;2(7):489-501.
92. Arafeh R, Samuels Y. PIK3CA in cancer: The past 30 years. Semin Cancer Biol. 2019;59:36-49.
93. McCubrey JA, Steelman LS, Abrams SL, Lee JT, Chang F, Bertrand FE, et al. Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. Adv Enzyme Regul. 2006;46:249-79.
94. Sarris EG, Saif MW, Syrigos KN. The Biological Role of PI3K Pathway in Lung Cancer. Pharmaceuticals (Basel). 2012;5(11):1236-64.
95. Broderick DK, Di C, Samuels TJPYR, Cummins JM, Fults REMDW, Velculescu VE, et al. Mutations of PIK3CA in Anaplastic Oligodendrogliomas, High-Grade astrocytomas and medulloblastomas. Cancer research. 2004;64(15):5048–50.
96. Molinari F, Frattini M. Functions and Regulation of the PTEN Gene in Colorectal Cancer. Front Oncol. 2013;3:326.
97. Lee JO, Yang H, Georgescu MM, Di Cristofano A, Maehama T, Shi Y, et al. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. Cell. 1999;99(3):323-34.
98. Sansal I, Sellers WR. The biology and clinical relevance of the PTEN tumor suppressor pathway. J Clin Oncol. 2004;22(14):2954-63.
99. Szegedi K, Szabo Z, Kallai J, Kiraly J, Szabo E, Bereczky Z, et al. Potential Role of VHL, PTEN, and BAP1 Mutations in Renal Tumors. J Clin Med. 2023;12(13).
100. Jin G, Kim MJ, Jeon HS, Choi JE, Kim DS, Lee EB, et al. PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. Lung Cancer. 2010;69(3):279-83.
101. Minaguchi T, Yoshikawa H, Oda K, Ishino T, Yasugi T, Onda T, et al. PTEN mutation located only outside exons 5, 6, and 7 is an independent predictor of favorable survival in endometrial carcinomas. Clin Cancer Res. 2001;7(9):2636-42.
102. Yang J, Ren Y, Wang L, Li B, Chen Y, Zhao W, et al. PTEN mutation spectrum in breast cancers and breast hyperplasia. J Cancer Res Clin Oncol. 2010;136(9):1303-11.
103. Aguissa-Toure AH, Li G. Genetic alterations of PTEN in human melanoma. Cell Mol Life Sci. 2012;69(9):1475-91.
104. Baig RM, Mahjabeen I, Sabir M, Masood N, Hafeez S, Malik FA, et al. Genetic changes in the PTEN gene and their association with breast cancer in Pakistan. Asian Pac J Cancer Prev. 2011;12(10):2773-8.
105. Malentacchi F, Turrini I, Sorbi F, Projetto E, Castiglione F, Fambrini M, et al. Pilot investigation of the mutation profile of PIK3CA/PTEN genes (PI3K pathway) in grade 3 endometrial cancer. Oncol Rep. 2019;41(3):1560-74.

106. Zhang S, Yu D. PI(3)king apart PTEN's role in cancer. Clin Cancer Res. 2010;16(17):4325-30.
107. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor. Nat Rev Mol Cell Biol. 2012;13(5):283-96.
108. Qureshi R, Arora H, Biswas S, Perwez A, Naseem A, Wajid S, et al. Mutation analysis of EGFR and its correlation with the HPV in Indian cervical cancer patients. Tumour Biol. 2016;37(7):9089-98.
109. Davies RL, Grosse VA, Kucherlapati R, Bothwell M. Genetic analysis of epidermal growth factor action: assignment of human epidermal growth factor receptor gene to chromosome 7. Proc Natl Acad Sci U S A. 1980;77(7):4188-92.
110. Kondo I, Shimizu N. Mapping of the human gene for epidermal growth factor receptor (EGFR) on the p13 leads to q22 region of chromosome 7. Cytogenet Cell Genet. 1983;35(1):9-14.
111. Roskoski R, Jr. The ErbB/HER family of protein-tyrosine kinases and cancer. Pharmacol Res. 2014;79:34-74.
112. Hsu WH, Yang JC, Mok TS, Loong HH. Overview of current systemic management of EGFR-mutant NSCLC. Ann Oncol. 2018;29(suppl_1):i3-i9.
113. Dewi IGASM, Sriwidyani NP, Ekawati NP. The role of epidermal growth factor receptor as progression factor in cervical intraepithelial neoplasia and squamous cell carcinoma. Bali Medical Journal. 2021;10(1):238-42.
114. Soonthornthum T, Arias-Pulido H, Joste N, Lomo L, Muller C, Rutledge T, et al. Epidermal growth factor receptor as a biomarker for cervical cancer. Ann Oncol. 2011;22(10):2166-78.
115. Li Q, Tang Y, Cheng X, Ji J, Zhang J, Zhou X. EGFR protein expression and gene amplification in squamous intraepithelial lesions and squamous cell carcinomas of the cervix. Int J Clin Exp Pathol. 2014;7(2):733-41.
116. Wang J, Lam D, Yang J, Hu L. Discovery of mobocertinib, a new irreversible tyrosine kinase inhibitor indicated for the treatment of non-small-cell lung cancer harboring EGFR exon 20 insertion mutations. Med Chem Res. 2022;31(10):1647-62.
117. Spano JP, Fagard R, Soria JC, Rixe O, Khayat D, Milano G. Epidermal growth factor receptor signaling in colorectal cancer: preclinical data and therapeutic perspectives. Ann Oncol. 2005;16(2):189-94.
118. Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. FEBS J. 2010;277(2):301-8.
119. Krasinskas AM. EGFR Signaling in Colorectal Carcinoma. Patholog Res Int. 2011;2011:932932.
120. Liu T, Wang Y, Wang Y, Chan AM. Multifaceted Regulation of PTEN Subcellular Distributions and Biological Functions. Cancers (Basel). 2019;11(9).
121. Chalhoub N, Baker SJ. PTEN and the PI3-kinase pathway in cancer. Annu Rev Pathol. 2009;4:127-50.
122. Clarke B, Chetty R. Postmodern Cancers: the role of HIV in uterine cervical cancer. Mol Path. 2002(55):19-24.
123. Arsenic R, Treue D, Lehmann A, Hummel M, Dietel M, Denkert C, et al. Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer. BMC Clin Pathol. 2015;15:20.
124. Jain D, Ramachandrappa VS, Singh V, Malik PS, Madan K, Faruq M, et al. Use of Exfoliative Specimens and Fine-Needle Aspiration Smears for Mutation Testing in Lung Adenocarcinoma. Acta Cytol. 2017;61(6):455-61.
125. IARCprotocol. Detection of TP53 mutations by Sanger sequencing. 2019.
126. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452-7.
127. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.

128. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980-5.
129. Chung TK, Van Hummelen P, Chan PK, Cheung TH, Yim SF, Yu MY, et al. Genomic aberrations in cervical adenocarcinomas in Hong Kong Chinese women. Int J Cancer. 2015;137(4):776-83.
130. Qiu L, Feng H, Yu H, Li M, You Y, Zhu S, et al. Characterization of the Genomic Landscape in Cervical Cancer by Next Generation Sequencing. Genes (Basel). 2022;13(2).
131. Taku O, Mbulawa ZZA, Phohlo K, Garcia-Jardon M, Businge CB, Williamson AL. Distribution of Human Papillomavirus (HPV) Genotypes in HIV-Negative and HIV-Positive Women with Cervical Intraepithelial Lesions in the Eastern Cape Province, South Africa. Viruses. 2021;13(2).
132. Kuguyo O, Dube Mandishora RS, Thomford NE, Makunike-Mutasa R, Nhachi CFB, Matimba A, et al. High-risk HPV genotypes in Zimbabwean women with cervical cancer: Comparative analyses between HIV-negative and HIV-positive women. PLoS One. 2021;16(9):e0257324.
133. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393(10167):169-82.
134. Mcharo R, Lennemann T, France J, Torres L, Gari M, Mbuya W, et al. HPV Type Distribution in HIV Positive and Negative Women With or Without Cervical Dysplasia or Cancer in East Africa. Frontiers in Oncology. 2021;11.
135. Teixeira MF, Sabido M, Leturiondo AL, de Oliveira Ferreira C, Torres KL, Benzaken AS. High risk human papillomavirus prevalence and genotype distribution among women infected with HIV in Manaus, Amazonas. Virol J. 2018;15(1):36.
136. Moscicki AB, Ellenberg JH, Crowley-Nowick P, Darragh TM, Xu J, Fahrat S. Risk of high-grade squamous intraepithelial lesion in HIV-infected adolescents. J Infect Dis. 2004;190(8):1413-21.
137. Herrero R, Castle PE, Schiffman M, Bratti MC, Hildesheim A, Morales J, et al. Epidemiologic profile of type-specific human papillomavirus infection and cervical neoplasia in Guanacaste, Costa Rica. J Infect Dis. 2005;191(11):1796-807.
138. Clifford GM, Gonçalves MAG, Franceschi S, Grp HHS. Human papillomavirus types among women infected with HIV:: a meta-analysis. Aids. 2006;20(18):2337-44.
139. Schwartz RH. Acquisition of immunologic self-tolerance. Cell. 1989;57(7):1073-81.
140. Banura C, Mirembe FM, Katahoire AR, Namujju PB, Mbidde EK. Universal routine HPV vaccination for young girls in Uganda: a review of opportunities and potential obstacles. Infect Agent Cancer. 2012;7(1):24.
141. Winkelstein W, Jr. Smoking and cervical cancer--current status: a review. Am J Epidemiol. 1990;131(6):945-57; discussion 58-60.
142. Brisson J, Roy M, Fortier M, Bouchard C, Meisels A. Condyloma and intraepithelial neoplasia of the uterine cervix: a case-control study. Am J Epidemiol. 1988;128(2):337-42.
143. Simons AM, Phillips DH, Coleman DV. Damage to DNA in cervical epithelium related to smoking tobacco. BMJ. 1993;306(6890):1444-8.
144. Barton SE, Maddox PH, Jenkins D, Edwards R, Cuzick J, Singer A. Effect of cigarette smoking on cervical epithelial immunity: a mechanism for neoplastic change? Lancet. 1988;2(8612):652-4.
145. Syahadat DS, Eviyulianti NM, Rau MJ, Mantao E, Krisnasari S. The Risks of Sexual and Reproductive Activity on the Occurrence of Cervical Cancer in Central Sulawesi Province: Case Study of patients of Undata Hospital. Journal of Health and Nutrition Research. 2022;1(3):171-7.
146. Ness RB, Grisso JA, Klapper J, Schlesselman JJ, Silberzweig S, Vergona R, et al. Risk of ovarian cancer in relation to estrogen and progestin dose and use characteristics of oral contraceptives. SHARE Study Group. Steroid Hormones and Reproductions. Am J Epidemiol. 2000;152(3):233-41.
147. McFarlane-Anderson N, Bazuaye PE, Jackson MD, Smikle M, Fletcher HM. Cervical dysplasia and cancer and the use of hormonal contraceptives in Jamaican women. BMC Womens Health. 2008;8:9.

148. WHO. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach, 2nd ed 2016 [Available from: https://www.who.int/publications/i/item/9789241549684.
149. Bosch FX, Broker TR, Forman D, Moscicki AB, Gillison ML, Doorbar J, et al. Comprehensive control of human papillomavirus infections and related diseases. Vaccine. 2013;31 Suppl 7(Suppl 7):H1-31.
150. Ports KA, Haffejee F, Mosavel M, Rameshbabu A. Integrating cervical cancer prevention initiatives with HIV care in resource-constrained settings: A formative study in Durban, South Africa. Glob Public Health. 2015;10(10):1238-51.
151. Abraham AG, Strickler HD, D'Souza G. Invasive Cervical Cancer Risk Among HIV-Infected Women Is a Function of CD4 Count and Screening. Jaids-J Acq Imm Def. 2013;63(5):E163-E.
152. Kelly H, Weiss HA, Benavente Y, de Sanjose S, Mayaud P, Art, et al. Association of antiretroviral therapy with high-risk human papillomavirus, cervical intraepithelial neoplasia, and invasive cervical cancer in women living with HIV: a systematic review and meta-analysis. Lancet HIV. 2018;5(1):e45-e58.
153. Stoler MH, Wright TC, Jr., Parvu V, Yanson K, Cooper CK, Andrews J. Stratified risk of high-grade cervical disease using onclarity HPV extended genotyping in women, >/=25 years of age, with NILM cytology. Gynecol Oncol. 2019;153(1):26-33.
154. Smith JS, Lindsay L, Hoots B, Keys J, Franceschi S, Winer R, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. Int J Cancer. 2007;121(3):621-32.
155. Lin H, Ma YY, Moh JS, Ou YC, Shen SY, ChangChien CC. High prevalence of genital human papillomavirus type 52 and 58 infection in women attending gynecologic practitioners in South Taiwan. Gynecol Oncol. 2006;101(1):40-5.
156. Bruni L, Diaz M, Castellsague X, Ferrer E, Bosch FX, de Sanjose S. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. J Infect Dis. 2010;202(12):1789-99.
157. Crow JM. HPV: The global burden. Nature. 2012;488(7413):S2-3.
158. Bulk S, Berkhof J, Bulkmans NW, Zielinski GD, Rozendaal L, van Kemenade FJ, et al. Preferential risk of HPV16 for squamous cell carcinoma and of HPV18 for adenocarcinoma of the cervix compared to women with normal cytology in The Netherlands. Br J Cancer. 2006;94(1):171-5.
159. Ramogola-Masire D, McGrath CM, Barnhart KT, Friedman HM, Zetola NM. Subtype distribution of human papillomavirus in HIV-infected women with cervical intraepithelial neoplasia stages 2 and 3 in Botswana. Int J Gynecol Pathol. 2011;30(6):591-6.
160. Richter K, Becker P, Horton A, Dreyer G. Age-specific prevalence of cervical human papillomavirus infection and cytological abnormalities in women in Gauteng Province, South Africa. S Afr Med J. 2013;103(5):313-7.
161. Denny L, Adewole I, Anorlu R, Dreyer G, Moodley M, Smith T, et al. Human papillomavirus prevalence and type distribution in invasive cervical cancer in sub-Saharan Africa. Int J Cancer. 2014;134(6):1389-98.
162. Clifford GM, de Vuyst H, Tenet V, Plummer M, Tully S, Franceschi S. Effect of HIV Infection on Human Papillomavirus Types Causing Invasive Cervical Cancer in Africa. J Acquir Immune Defic Syndr. 2016;73(3):332-9.
163. Rad A, Sorbye SW, Dreyer G, Hovland S, Falang BM, Louw M, et al. HPV types in cervical cancer tissue in South Africa: A head-to-head comparison by mRNA and DNA tests. Medicine (Baltimore). 2017;96(47):e8752.
164. Pinheiro M, Gage JC, Clifford GM, Demarco M, Cheung LC, Chen Z, et al. Association of HPV35 with cervical carcinogenesis among women of African ancestry: Evidence of viral-host interaction with implications for disease intervention. Int J Cancer. 2020;147(10):2677-86.

165. Gravitt PE, Kamath AM, Gaffikin L, Chirenje ZM, Womack S, Shah KV. Human papillomavirus genotype prevalence in high-grade squamous intraepithelial lesions and colposcopically normal women from Zimbabwe. Int J Cancer. 2002;100(6):729-32.
166. Mbulawa ZZA, Somdyala NI, Mabunda SA, Williamson AL. High human papillomavirus prevalence among females attending high school in the Eastern Cape Province of South Africa. PLoS One. 2021;16(6):e0253074.
167. Happel AU, Balle C, Havyarimana E, Brown B, Maust BS, Feng C, et al. Cervicovaginal Human Papillomavirus Genomes, Microbiota Composition and Cytokine Concentrations in South African Adolescents. Viruses. 2023;15(3).
168. Adebamowo SN, Olawande O, Famooto A, Dareng EO, Offiong R, Adebamowo CA, et al. Persistent Low-Risk and High-Risk Human Papillomavirus Infections of the Uterine Cervix in HIV-Negative and HIV-Positive Women. Front Public Health. 2017;5:178.
169. de Oliveira GR, Vieira VC, Avila EC, Finger-Jardim F, Caldeira TD, Gatti FA, et al. Human papillomavirus type distribution and HPV16 intratype diversity in southern Brazil in women with and without cervical lesions. Mem Inst Oswaldo Cruz. 2017;112(7):492-8.
170. Snijders PJ, Steenbergen RD, Heideman DA, Meijer CJ. HPV-mediated cervical carcinogenesis: concepts and clinical implications. J Pathol. 2006;208(2):152-64.
171. Yarchoan M, Albacker LA, Hopkins AC, Montesion M, Murugesan K, Vithayathil TT, et al. PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. JCI Insight. 2019;4(6).
172. Ward AR, Mota TM, Jones RB. Immunological approaches to HIV cure. Semin Immunol. 2021;51:101412.
173. Kahesa C, Mwaiselage J, Wabinga HR, Ngoma T, Kalyango JN, Karamagi CA. Association between invasive cancer of the cervix and HIV-1 infection in Tanzania: the need for dual screening. BMC Public Health. 2008;8:262.
174. Firnhaber C, Westreich D, Schulze D, Williams S, Siminya M, Michelow P, et al. Highly active antiretroviral therapy and cervical dysplasia in HIV-positive women in South Africa. J Int AIDS Soc. 2012;15(2):17382.
175. Blitz S, Baxter J, Raboud J, Walmsley S, Rachlis A, Smaill F, et al. Evaluation of HIV and highly active antiretroviral therapy on the natural history of human papillomavirus infection and cervical cytopathologic findings in HIV-positive and high-risk HIV-negative women. J Infect Dis. 2013;208(3):454-62.
176. Chen YC, Li CY, Liu HY, Lee NY, Ko WC, Ko NY. Effect of antiretroviral therapy on the incidence of cervical neoplasia among HIV-infected women: a population-based cohort study in Taiwan. AIDS. 2014;28(5):709-15.
177. Sekirime WK, Gray R. HIV infection among Uganda women with cervical cancer: a retrospective study. Gynecol Obstet Invest. 2007;63(4):222-8.
178. McKenzie KP, Rogers RK, Njoroge JW, John-Stewart G, Richardson BA, Mugo NR, et al. Cervical squamous intraepithelial lesions among HIV-positive women on antiretroviral therapy in Kenya. Curr HIV Res. 2011;9(3):180-5.
179. Mogtomo ML, Malieugoue LC, Djiepgang C, Wankam M, Moune A, Ngane AN. Incidence of cervical disease associated to HPV in human immunodeficiency infected women under highly active antiretroviral therapy. Infect Agent Cancer. 2009;4:9.
180. Odida M, Sandin S, Mirembe F, Kleter B, Quint W, Weiderpass E. HPV types, HIV and invasive cervical carcinoma risk in Kampala, Uganda: a case-control study. Infect Agent Cancer. 2011;6(1):8.
181. ter Meulen J, Eberhardt HC, Luande J, Mgaya HN, Chang-Claude J, Mtiro H, et al. Human papillomavirus (HPV) infection, HIV infection and cervical cancer in Tanzania, east Africa. Int J Cancer. 1992;51(4):515-21.

182. Spaans VM, Trietsch MD, Peters AA, Osse M, Ter Haar N, Fleuren GJ, et al. Precise Classification of Cervical Carcinomas Combined with Somatic Mutation Profiling Contributes to Predicting Disease Outcome. PLoS One. 2015;10(7):e0133670.
183. Wang M, Fan W, Ye M, Tian C, Zhao L, Wang J, et al. Molecular profiles and tumor mutational burden analysis in Chinese patients with gynecologic cancers. Sci Rep. 2018;8(1):8990.
184. Gagliardi A, Porter VL, Zong Z, Bowlby R, Titmuss E, Namirembe C, et al. Analysis of Ugandan cervical carcinomas identifies human papillomavirus clade-specific epigenome and transcriptome landscapes. Nat Genet. 2020;52(8):800-10.
185. Saasa-Modise ML, Musonda JM, Sikwese-Musonda J, Maseko NJ, Hlophe L, Kubeka G. Cervical cancer screening in women living with HIV attending primary care clinics in a health district, South Africa: a descriptive cross-sectional study. Pan Afr Med J. 2022;43:32.
186. Sarah Maria N, Olwit C, Kaggwa MM, Nabirye RC, Ngabirano TD. Cervical cancer screening among HIV-positive women in urban Uganda: a cross sectional study. BMC Womens Health. 2022;22(1):148.
187. Wang X, Huang X, Zhang Y. Involvement of Human Papillomaviruses in Cervical Cancer. Front Microbiol. 2018;9:2896.
188. Mpata PC, Nkosi ZZ. Experiences of cervical cancer screening in HIV-positive women in Zimbabwe. Curationis. 2021;44(1):e1-e7.
189. Castanon A, Leung VM, Landy R, Lim AW, Sasieni P. Characteristics and screening history of women diagnosed with cervical cancer aged 20-29 years. Br J Cancer. 2013;109(1):35-41.
190. Foley G, Alston R, Geraci M, Brabin L, Kitchener H, Birch J. Increasing rates of cervical cancer in young women in England: an analysis of national data 1982-2006. Br J Cancer. 2011;105(1):177-84.
191. Verlaat W, Snijders PJ, van Moorsel MI, Bleeker M, Rozendaal L, Sie D, et al. Somatic mutation in PIK3CA is a late event in cervical carcinogenesis. J Pathol Clin Res. 2015;1(4):207-11.
192. Mohseni M, Park BH. PIK3CA and KRAS mutations predict for response to everolimus therapy: now that's RAD001. J Clin Invest. 2010;120(8):2655-8.
193. Cui B, Zheng B, Zhang X, Stendahl U, Andersson S, Wallin KL. Mutation of PIK3CA: possible risk factor for cervical carcinogenesis in older women. Int J Oncol. 2009;34(2):409-16.
194. Tornesello ML, Annunziata C, Buonaguro L, Losito S, Greggi S, Buonaguro FM. TP53 and PIK3CA gene mutations in adenocarcinoma, squamous cell carcinoma and high-grade intraepithelial neoplasia of the cervix. J Transl Med. 2014;12:255.
195. Tornesello ML, Buonaguro L, Buonaguro FM. Mutations of the TP53 gene in adenocarcinoma and squamous cell carcinoma of the cervix: a systematic review. Gynecol Oncol. 2013;128(3):442-8.
196. Tommasino M, Accardi R, Caldeira S, Dong W, Malanchi I, Smet A, et al. The role of TP53 in Cervical carcinogenesis. Hum Mutat. 2003;21(3):307-12.
197. Cheung TH, Lo KW, Yim SF, Chan LK, Heung MS, Chan CS, et al. Epigenetic and genetic alternation of PTEN in cervical neoplasm. Gynecol Oncol. 2004;93(3):621-7.
198. Kato S, Okamura R, Mareboina M, Lee S, Goodman A, Patel SP, et al. Revisiting Epidermal Growth Factor Receptor (EGFR) Amplification as a Target for Anti-EGFR Therapy: Analysis of Cell-Free Circulating Tumor DNA in Patients With Advanced Malignancies. JCO Precis Oncol. 2019;3.
199. Chen Q, Huang Y, Shao L, Han-Zhang H, Yang F, Wang Y, et al. An EGFR-Amplified Cervical Squamous Cell Carcinoma Patient with Pulmonary Metastasis Benefits from Afatinib: A Case Report. Onco Targets Ther. 2020;13:1845-9.
200. Orekoya AA, Abdus-Salam AA, Oyesegun AR, Ntekim A, Folasire AM, Okolo CA, et al. EGFR and HER2 expression in cervical cancer patients in Ibadan, Nigeria. Ecancermedicalscienc. 2023;17.
201. Soonthornthum T, Arias-Pulido H, Joste N, Lomo L, Muller C, Rutledge T, et al. Epidermal growth factor receptor as a biomarker for cervical cancer. Annals of Oncology. 2011;22(10):2166-78.
202. Badawy O, Loay I, Negmaldin D. P16 and EGFR analyses in cervical carcinoma on paraffin material in Egyptian females: the NCI experience. Egyptian Journal of Pathology. 2023;43(1):86-93.

203. Arias-Pulido H, Joste N, Chavez A, Muller CY, Dai D, Smith HO, et al. Absence of epidermal growth factor receptor mutations in cervical cancer. Int J Gynecol Cancer. 2008;18(4):749-54.
204. Schrevel M, Gorter A, Kolkman-Uljee SM, Trimbos JB, Fleuren GJ, Jordanova ES. Molecular mechanisms of epidermal growth factor receptor overexpression in patients with cervical cancer. Mod Pathol. 2011;24(5):720-8.

APPENDICES

Appendix A: Ethics Approval Letter from the University of Cape Town for the main study

HUMAN RESEARCH ETHICS COMMITTEE
17 JUL 2023
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938) This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.6.2024
☐ Not approved	See attached comments		
Signature Chairperson of the HREC/ Designee		Date Signed	17/7/2023

Note: Please email this form and supporting documents (if applicable) in a combined pdf-file to hrec-enquiries@uct.ac.za.
Please clarify your plan for research-related activities during COVID-19 lockdown.
Please use the latest form found on our website:
http://www.health.uct.ac.za/fhs/research/humanethics/forms

Comments to PI from the HREC

Principal Investigator to complete the following:

1. Protocol information

Date (when submitting this form)			
HREC REF Number	HREC 244/2022	Current Ethics Approval was granted until	30 June 2023
Protocol title	Investigating the role of host genetic variation in HIV/HPV co-infected women with precancerous lesions and invasive cervical cancer		
Protocol number (if applicable)			
Are there any sub-studies linked to this study?	√ Yes	☐ No	
If yes, could you please provide the HREC Reference number for all sub-studies? **Note:** A separate FHS016 must be submitted for each sub-study.	HREC 560/2022 (granted until 30 September 2023)		
Principal Investigator	Professor Collet Dandara		
Department / Office Internal Mail Address	Department of Pathology, Division of Human Genetics collet.dandara@uct.ac.za		

Appendix B: Ethics Approval Letter from the University of Cape Town for the current study

HUMAN RESEARCH ETHICS COMMITTEE
17 JUL 2023
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938) This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.9.2024
☐ Not approved	See attached comments		
Signature Chairperson of the HREC/ Designee		Date Signed	17/7/2023

Note: Please email this form and supporting documents (if applicable) in a combined pdf-file to hrec-enquiries@uct.ac.za.
Please clarify your plan for research-related activities during COVID-19 lockdown.
Please use the latest form found on our website:
http://www.health.uct.ac.za/fhs/research/humanethics/forms

Comments to PI from the HREC

Principal Investigator to complete the following:

1. Protocol information

Date (when submitting this form)			
HREC REF Number	HREC 560/2022	Current Ethics Approval was granted until	30 September 2023
Protocol title	Host somatic variation in HIV/HPV co-infected women with precancerous lesions and cervical cancer		
Protocol number (if applicable)			
Are there any sub-studies linked to this study?	☐ Yes	√ No	
If yes, could you please provide the HREC Reference number for all sub-studies? **Note:** A separate FHS016 must be submitted for each sub-study.			
Principal Investigator	Professor Collet Dandara		

Appendix C: DNA quantification data from Nanodrop

Sample ID	ng/uL	A260	260/280	260/230	Constant
19001	282.6	5.652	1.78	1.7	50
19002	233.93	4.679	1.82	1.36	50
19003	80.49	1.61	1.82	1.97	50
19004	125	2.5	1.74	1.94	50
19005	271.52	5.43	1.75	1.2	50
19006	219.83	4.397	1.86	1.55	50
19007	81.64	1.633	1.87	0.83	50
19008	293.16	5.863	1.75	1.95	50
19009	147.62	2.952	1.8	0.44	50
19011	121.78	2.436	1.78	2.18	50
19012	108.62	2.172	1.64	0.13	50
19013	404.07	8.081	1.7	1.85	50
19014	207.93	4.159	1.82	1.41	50
19015	283.65	5.673	1.84	1.56	50
19016	117.91	2.358	1.8	0.27	50
19017	208.8	4.176	1.78	0.32	50
19018	193.34	3.867	1.53	0.52	50
19019	189.67	3.793	1.76	0.33	50
19020	196.34	3.927	1.84	2.24	50
19021	230.76	4.615	1.79	2.26	50
19022	87.41	1.748	1.82	1.39	50
19023	58.77	1.175	1.81	2.27	50
19024	239.31	4.786	1.77	2.32	50
19025	123.88	2.478	1.61	0.15	50
19026	499.22	9.984	1.85	1.09	50
19027	53.51	1.07	1.79	0.81	50
19028	182.38	3.648	1.68	2.54	50
19029	283.44	5.669	1.7	0.33	50
19030	108.52	2.17	1.82	1.27	50
19031	157.91	3.158	1.64	1.07	50
19032	148.52	2.97	1.5	1.35	50
19033	232.66	4.653	1.67	2.17	50
19034	224.33	4.487	1.82	1.32	50
19035	82.74	1.655	1.72	2.51	50
19036	236.79	4.736	1.74	1.18	50
19037	276.46	5.529	1.86	1.36	50
19038	66.62	1.332	1.81	1.98	50
19039	140.19	2.804	1.86	2.13	50
19040	265.34	5.307	1.83	2.25	50

19041	237.59	4.752	1.85	2.29	50
19042	122.45	2.449	1.81	1.82	50
19043	67.66	1.353	1.77	1.77	50
19044	349.74	6.995	1.63	0.91	50
19045	300.25	6.005	1.69	2.44	50
19046	99.16	1.983	1.85	2.05	50
19047	74.26	1.485	1.83	0.76	50
19048	39.37	0.787	1.92	2.05	50
19049	104.08	2.082	1.86	1.52	50
19050	92.25	1.845	1.72	2.2	50
19051	274.82	5.496	1.82	2.02	50
19052	166.67	3.333	1.85	2.13	50
19053	274.96	5.499	1.71	2.12	50
19054	238.18	4.764	1.73	0.34	50
19055	250.54	5.011	1.77	1.52	50
19056	198.87	3.977	1.73	1.75	50
19057	351.7	7.034	1.73	1.74	50
19058	183.4	3.668	1.57	0.96	50
19059	196.3	3.926	1.82	2.19	50
19060	130.37	2.607	1.82	1.29	50
19061	52.86	1.057	1.65	0.38	50
19062	125.1	2.502	1.67	1.7	50
19063	73.76	1.475	1.84	0.33	50
19064	101.11	2.022	1.82	2.25	50
19065	323.05	6.461	1.62	1.26	50
19066	284.84	5.697	1.84	2.15	50
19067	271.66	5.433	1.77	2.18	50
19068	458.79	9.176	1.8	2.11	50
19069	178.79	3.576	1.79	1.8	50
19070	145.72	2.914	1.68	1.88	50
19071	211.24	4.225	1.72	2.42	50
19072	63.62	1.272	1.52	2.11	50
19073	72.47	1.449	1.52	1.69	50
19074	187.88	3.758	1.82	1.82	50
19075	191.73	3.835	1.85	2.46	50
19076	49.32	0.986	1.79	1.81	50
19077	83.4	1.668	1.87	1.07	50
19078	91.68	1.834	1.77	1.98	50
19079	107.35	2.147	1.63	1.08	50
19080	126.91	2.538	1.88	1.32	50
19081	157.75	3.155	1.82	1.9	50
19082	107.42	2.148	1.55	0.51	50

19083	181.08	3.622	1.82	2.23	50
19084	127.61	2.552	1.82	2.17	50
19085	103.98	2.08	1.84	1.63	50
19086	143.24	2.865	1.83	1.14	50
19087	205.78	4.116	1.86	2.16	50
19088	236.71	4.734	1.81	1.81	50
19089	154.68	3.094	1.8	2.05	50

Appendix D: Integrity gel representative

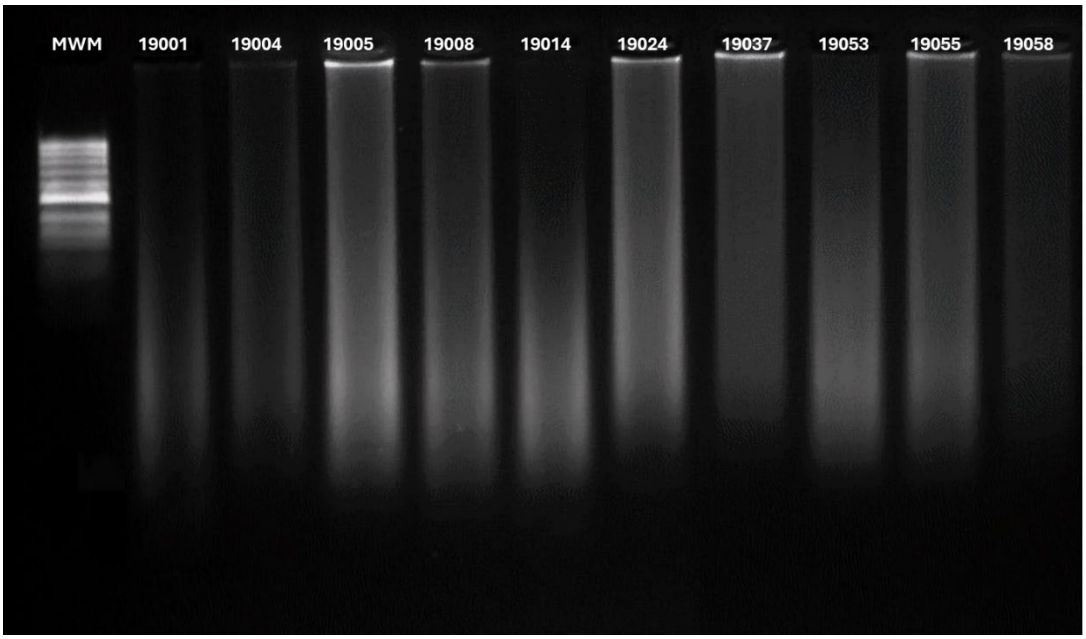
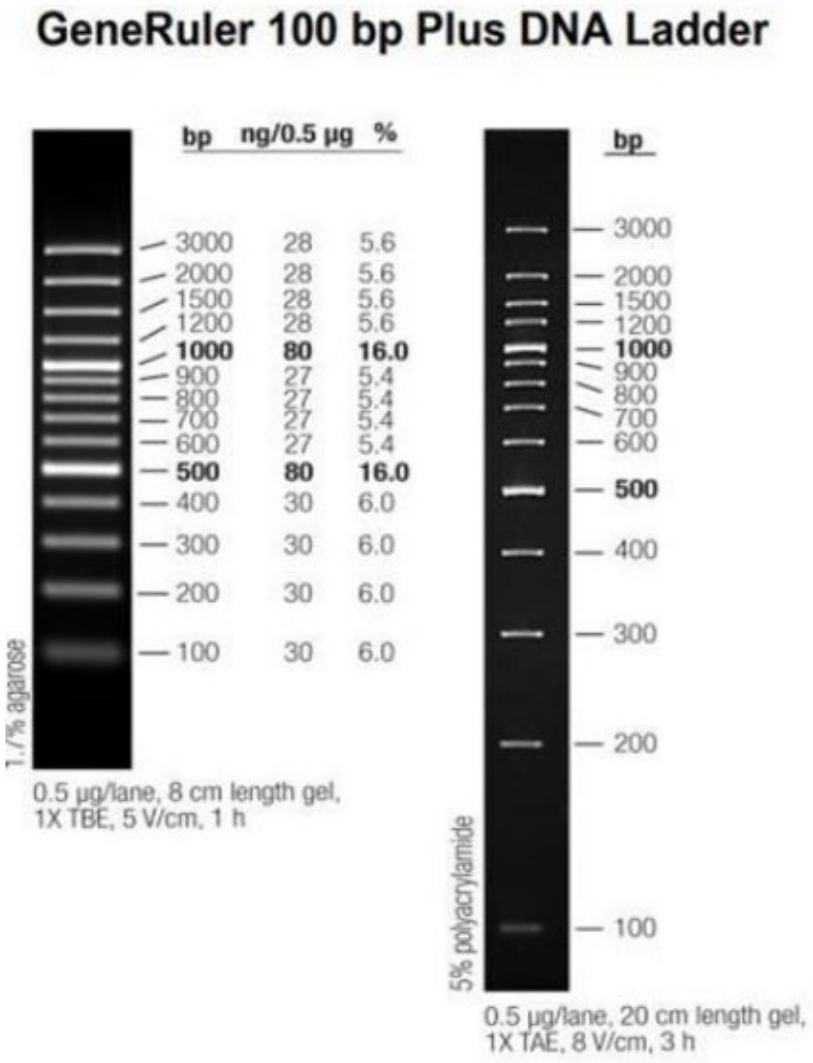


Figure E: Integrity gel. Run at 100V for 1 hour on a 1% agarose gel. The first lane is the molecular weight marker (MWM), the GeneRular™ 100bp Ladder (Thermo Fisher Scientific). The following lanes are individual sample ID's.

Appendix E: GeneRular 100 bp Plus DNA Ladder (Thermo Fisher Scientific), loaded onto the gels



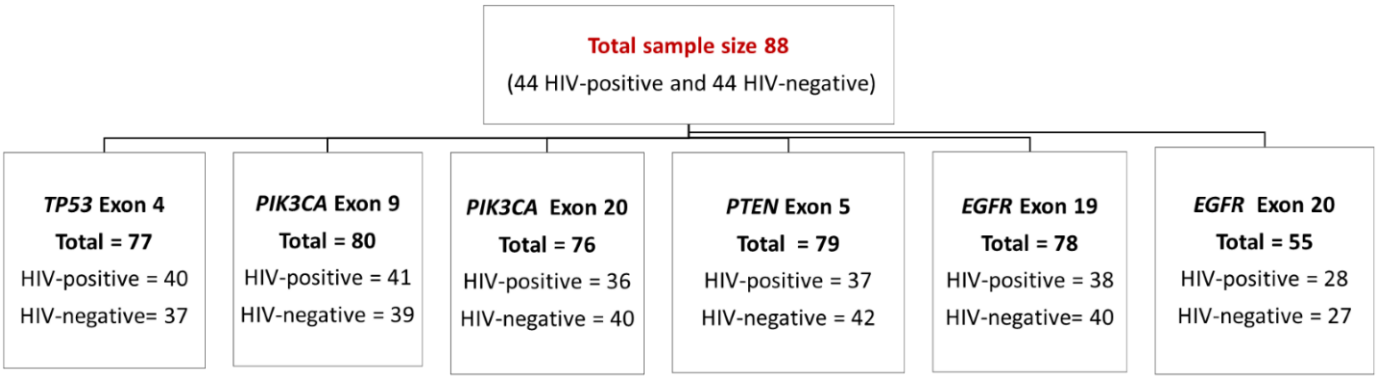
Appendix F: Cycling conditions for HPV genotyping (Table adapted from the manufacturer's protocol Seegene Anyplex™ II HPV28 Detection kit)

i) cyclic-CMTA (Melt analysis of three times)

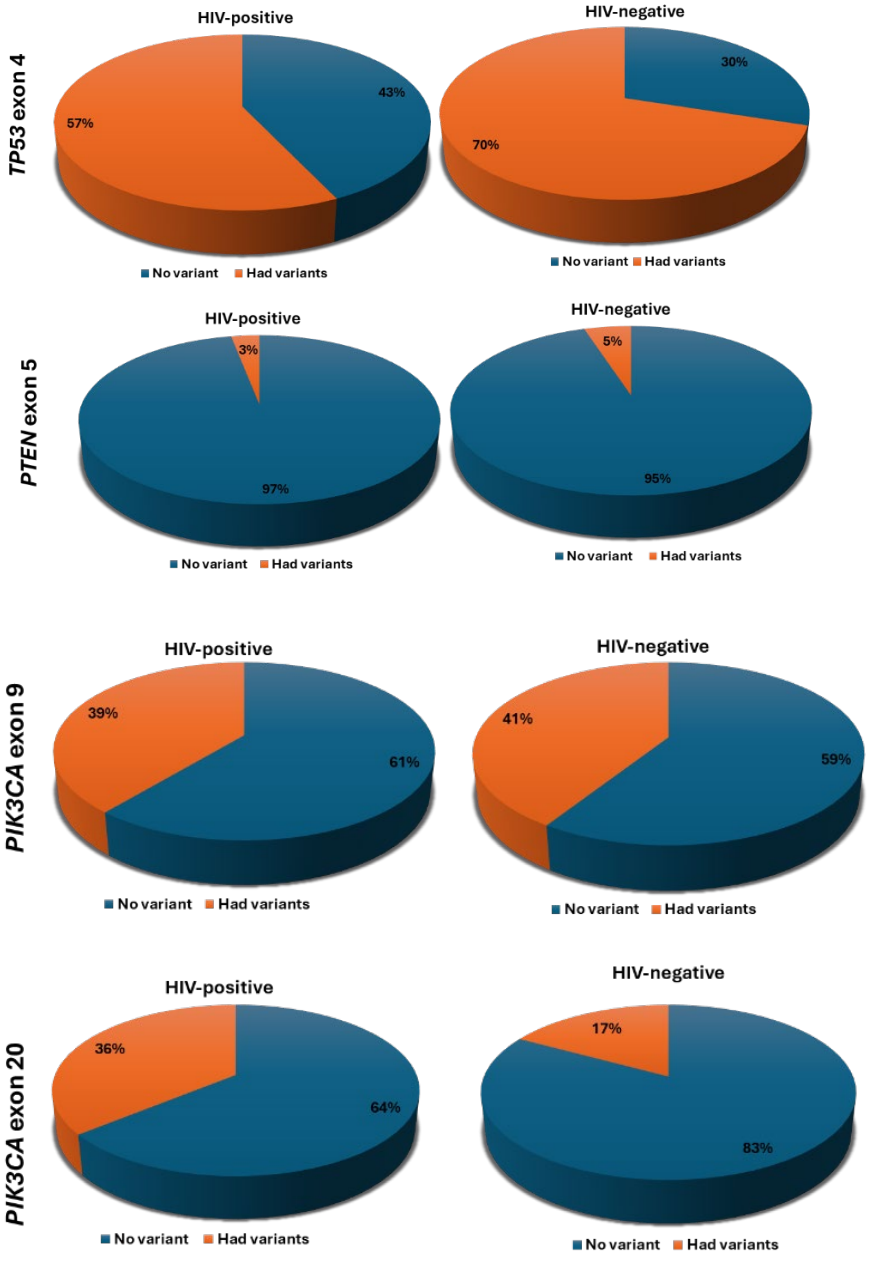
Step	Temperature	Duration	No. of cycles
1	50°C	4 min	
2	95°C	15 min	
3	95°C	30 sec	30
4	60°C	1 min	
5	72°C	30 sec	
6	GOTO 3, 29 more times		
7	55°C	30 sec	
8*	Melting curve 55°C ~ 85°C (5 s / 0.5°C)		
9	95°C	30 sec	10
10	60°C	1 min	
11	72°C	30 sec	
12	GOTO 9, 9 more times		
13	55°C	30 sec	
14*	Melting curve 55°C ~ 85°C (5 s / 0.5°C)		
15	95°C	30 sec	10
16	60°C	1 min	
17	72°C	30 sec	
18	GOTO 15, 9 more times		
19	55°C	30 sec	
20*	Melting curve 55°C ~ 85°C (5 s / 0.5°C)		

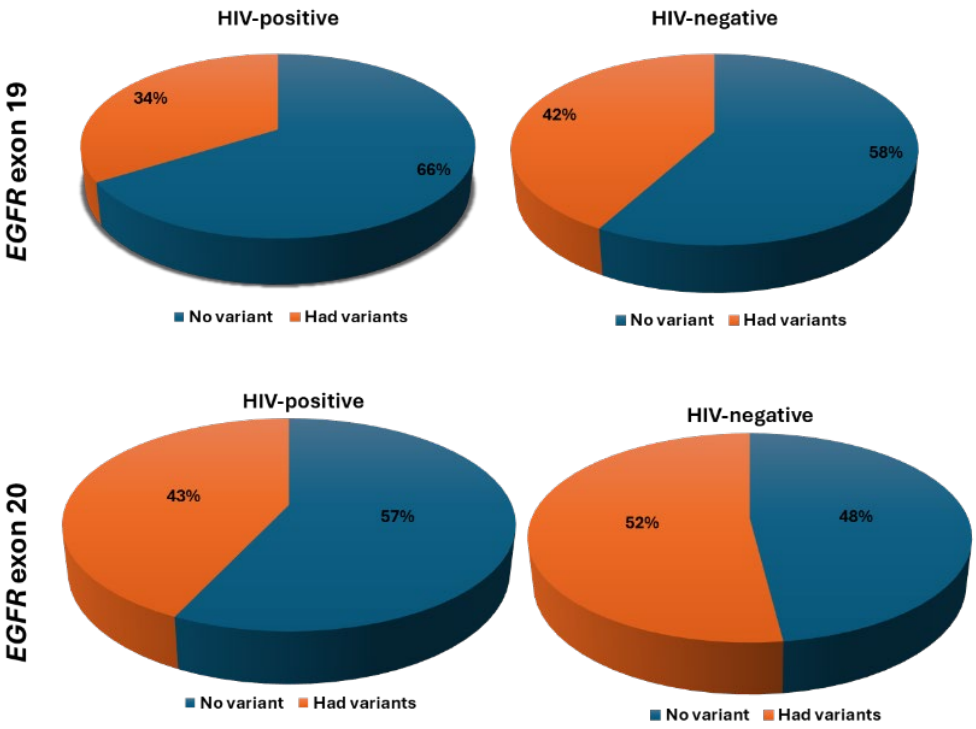
***Note: Plate Read at Steps 8, 14 and 20.** Fluorescence is detected at Melting.

Appendix G: The total number of samples successfully sequenced for all the genes of interest and each exon, stratified by HIV status



Appendix H: Distribution of samples that had variants and no variants for each exon stratified by HIV status.




EGFR exon 19
HIV-positive
34%
66%
No variant
Had variants
HIV-negative
42%
58%
No variant
Had variants
EGFR exon 20
HIV-positive
43%
57%
No variant
Had variants
HIV-negative
52%
48%
No variant
Had variants

Appendix I: PCR amplification for all the regions of interest

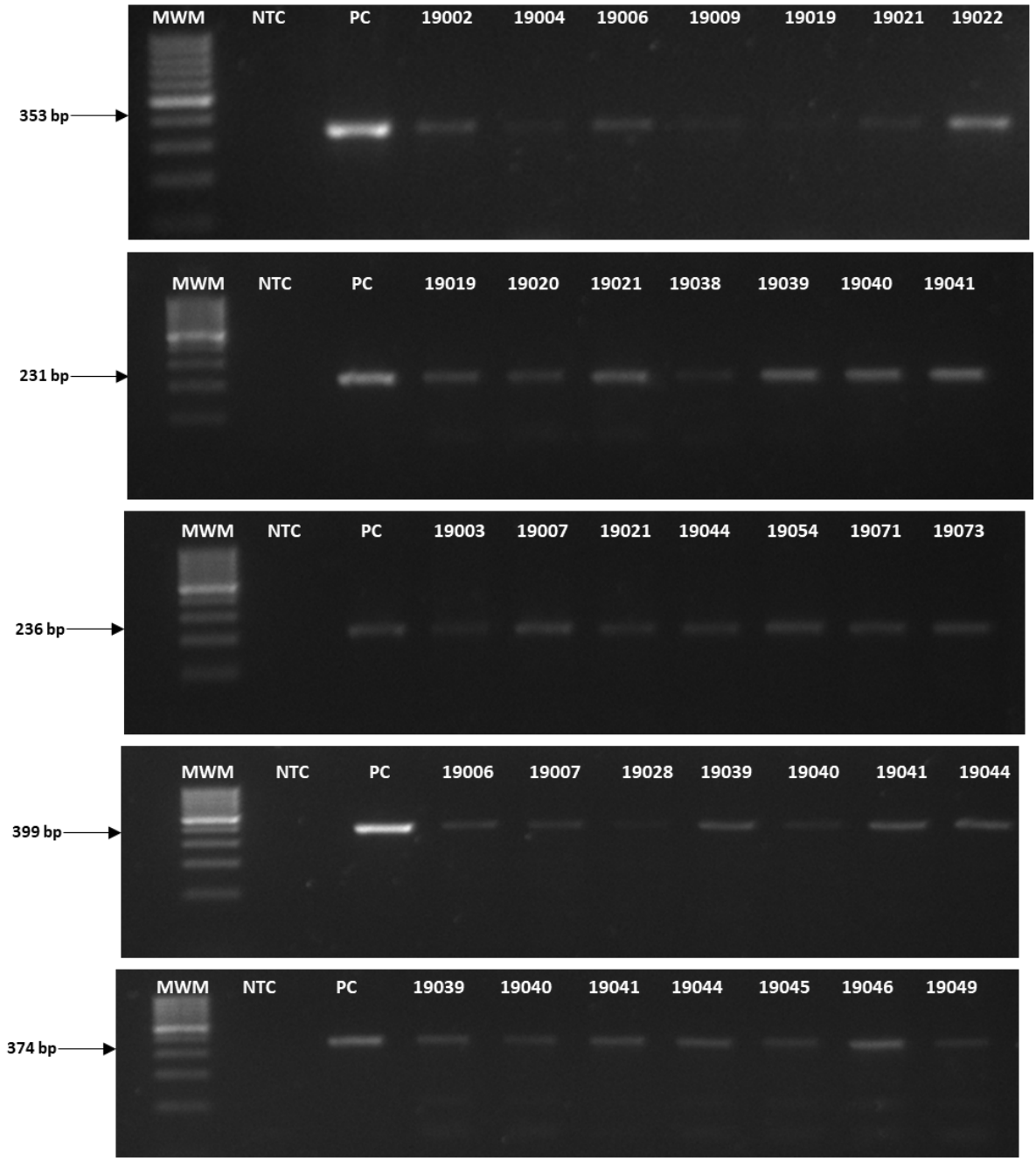


Figure H: PCR amplification. Each gel indicates PCR fragments run at 100V for 1 hour on a 1.5% agarose gel. The first lane is the molecular weight marker (MWM), the GeneRular™ 100bp Ladder (Thermo Fisher Scientific), the second lane is the non-template control (NTC), and the third lane is the positive control (PC). The following lanes are individual samples. **A:** *TP53* exon 4= 353 bp, **B:** *PIK3CA* exon 20= 231 bp, **C:** *PTEN* exon 5= 236 bp, **D:** *EGFR* exon 19= 399 bp, **E:** *EGFR* exon 20=374 bp. Full names: ***TP53***: Tumour Protein 53, ***PIK3CA***: Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha, ***PTEN***: Phosphatase and tensin homolog, ***EGFR****:* Epidermal growth factor receptor

Appendix J: List of different variants identified in a cohort for each gene/exon

Gene/Exon	Mutation	Allele change	Amino Acid change	rs number
TP53 exon 4	synonynous	C/A	P82P	rs372397095
	Synonymous	G/C	P64P	rs1555526723
	Synonymous	C/T	P36P	rs1800370
	Synonymous	A/G	D41D	rs369129220
	Synonymous	G/T	G112G	rs876659177
	Missense	G/A	P47S	rs1800371
	Missense	C/T	A69T	rs1567556618
	Missense	C/G	P72R	rs1042522
	Missense	A/G	Y107H	rs368771578
	Missense	G/T	P67Q	rs1555526709
	Missense	T/C	H115R	rs730881996
	Missense	G/A	T123I	rs1555526486
	Missense	C/A	P82A	rs1555526664
	Missense	A/G	R65G	rs1555526721
	Stop-gain	C/T	W91*	rs876660548
	Stop-gain	C/T	Q100*	rs1567555994
	Stop-gain	C/T	E68*	rs869312782
PIK3CA exon 9	Synonymous	T/G	T536T	rs1724887126
	Synonymous	G/A	L540L	rs2108408275
	Synonymous	T/G	S536S	rs1724887126
	Missense	A/C	E545A	rs121913274
	Missense	T/G	L551V	rs988134846
	Missense	G/A	L551L	rs2108408282
	Missense	T/A	E542D	rs2108408314
	Missense	G/A	S541F	rs2108408282
	Missense	G/A	S535F	rs2108408188
	Missense	T/G	K532R	rs1266238180
	Missense	G/A	L531I	rs2108408134
	Missense	C/G	L531V	rs2108408134
	Missense	T/G	E529D	rs1366386478
	Missense	T/G	Q546P	rs39751720
	Missense	A/C	D520V	rs2108407987
	Missense	T/A	N526K	rs2108408075
	Missense	G/C	Q530H	rs2108408130
	Missense	A/C	H554P	rs2108408475
	Stop gained	A/C	R516*	rs2108407927
	intron	C/T	-	rs895647434

	intron	C/T	-	rs988134846
	intron	A/C	-	rs2108408513
	intron	G/A	-	rs2108408683
	intron	T/A	-	rs2108408665
	intron	G/A	-	rs104886001
	intron	T/G	-	rs776839472
	intron	G/A	-	rs104886001
***PIK3CA* exon 9**	Synonymous	A>G	Q1042Q	rs1725282114
	Synonymous	A>G	T1052T	rs2108429959
	Synonymous	G>C	G1050G	rs1576949990
	Missense	C>G	H1060Q	rs748925418
	Missense	T>G	T1061P	rs2108430073
	Missense	T>G	K1054E	rs2108429976
	Missense	G>T	Q1064H	rs2108430107
	Missense	A>G	M1040T	rs1725282028
	Missense	G>A	D1056N	rs2108429997
	Missense	T>G	T1061P	rs2108430073
	Missense	A>T	N1068I	rs1725284914
***PTEN* exon 5**	Synonymous	C/A	G143G	rs1589646539
	Synonymous	A/T	A148A	rs778663292
	Missense	G/T	G143V	rs786202047
	Missense	A/T	Q149L	rs2132243781
	Missense	G/A	D162N	rs2132244111
***EGFR* exon 19**	Synonymous	G/A	P708P	rs764064214
	Synonymous	T/G	G735G	rs2128954605
	Missense	G/C	A690P	rs759256622
	Missense	A/G	K692R	rs121913433
	Missense	G/C	A690P	rs759256622
	Missense	G/A	R748I	rs1379373864
	Missense	G/A	E749Q	rs1057520037
	intron	T/A	-	rs2128955101
	intron	T/C	-	rs17290378
	intron	A/G	-	rs2017000
	intron	T/C	-	rs1197399144
	intron	T/C	-	rs2128954920
	intron	C/T	-	rs2128955021
	intron	T/A	-	rs2128954937
	intron	C/T	-	rs17290371
	intron	T/A	-	rs1444138271
	intron	T/A	-	rs2128954937

	intron	G/A	-	rs2128954924
	intron	T/C	-	rs2128955090
	intron	C/T	-	rs2128955051
	intron	T/C	-	rs1343138617
***EGFR* exon 20**	Synonymous	G/A	Q734Q	rs1050171
	Missense	G/A	D770N	rs2128958340
	Missense	A/G	H773R	rs121913432
	Intron	T/C	-	rs10241451