GENOMIC EXPRESSION PROFILE OF HUMAN PAPILLOMAVIRUS 16 AND 18 – ASSOCIATED PRE-CANCEROUS LESIONS AND CANCER OF THE CERVIX

SHANDRA DEVI A/P BALASUBRAMANIAM

UNIVERSITI SAINS MALAYSIA

2022

GENOMIC EXPRESSION PROFILE OF HUMAN PAPILLOMAVIRUS 16 AND 18 – ASSOCIATED PRE-CANCEROUS LESIONS AND CANCER OF THE CERVIX

by

SHANDRA DEVI A/P BALASUBRAMANIAM

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

June 2022

ACKNOWLEDGEMENT

First, I would like to thank God for the blessings and for providing me the strength to overcome the challenges in this journey to successfully complete this research. I am very thankful and would like to pay full appreciation to my supervisor Prof. Dr. Gurjeet Chatar Singh for her guidance, patience, valuable advice, and close supervision. She was a very patient and supportive supervisor to me. Her continuous support and guidance had driven me successfully to complete my research and publish my papers. I also would like to extend my gratitude to my both co-supervisor, Assoc. Prof. Dr. Venugopal Balakrishnan and Assoc. Prof. Dr. Oon Chern Ein for their scientific advice and knowledge. They also gave many insightful discussions and suggestions to me. I would like to take this opportunity to acknowledge the Research University Grant (No. 1001/CIPPM/812116) from Universiti Sains Malaysia for providing funds during the research project.

Next, I would like to thank the Institute for Research in Molecular Medicine (INFORMM) for providing a place and comfortable environment for me to carry out my research. I would also extend my thanks to all the admiration staff, science officers, and lab assistants who have always been pleasant and have assisted me in so many ways. My special thanks to En. Farid from Institut Perubatan dan Pergigian Termaju (IPPT), USM, Bertam for his assistance and guidance during my research work. Besides that, I would like to thank Ms. Teng, an application specialist for Human Transcriptome Array, Affymetrix for her guidance and assistance during my research work. Furthermore, I would like also to thank Ms. Karen, an application specialist from Nanostring for her guidance and assistance during my research work at Universiti Malaya.

I would like to acknowledge my appreciation to my friends Ashwaq and Eshtiyag who has been very caring and accompanied me during my lab works. They had been supportive to me to complete my research. Furthermore, not to forget, my friends Baharak, and Kalaivani Muniandy who were there throughout my research. A big thanks to my lab mates Richard, Wei Kang, Ashwaq, and the intern students for the outing that we had together.

Apart from that, I would like to thank my family especially my beloved parents Mr. Balasubramaniam and Mrs. Parwathy for their concern, moral support, encouragement, prayers, and blessings that kept me going on all the time to complete my research work. I must express my profound gratitude to my husband Mr. M.Shashitharan for his unfailing support and continuous encouragement. I am glad that he would stand on my side and love me unconditionally throughout my research journey.

With a special mention, I would like to express my gratitude to my late uncle Mr. Jeyaseelan who had played a fatherly role to me and given me the courage and strength for my research. I would also like to express my gratitude to my beloved brother, Mr. Indra Perumal for being supportive and having confidence in me to finish my study. My sisters (Mrs. Mary and Mrs. Vijaya Vadivu), brother-in-law (Mr. Sivanaya), nieces, and nephews unconditionally loved me and pushed me farther than I thought I could go.

Not to forget, my parents-in-law, Mr. Mahalingam and Mrs. Linggamuthu who also had supported me throughout my research. Honestly, I never thought I would complete my research as I got married during my research. I was worried to balance

my research and marriage life. Huge thanks to my husband and in-laws who gave me the strength to complete my research. Next, I must thank my brother in-law (M.Lexmanraj), sister-in-law and her husband, Mrs. M.Shoba, and Dr. TuneshKumar who always lend their sympathetic ear during my difficult times and always motivate and refresh me in preparation to face the next challenge.

Last but not least, I would like to thank my college friends, Sugitha, Ivy, Ranjini, and Nittiyaa for their support and encouragement throughout my research. Not to forget, my entire Penang State Puteri MIC Team especially Vijaya Letchumi and Uma Mageswary for their unconditional moral support. Next, huge thanks to Prof. Liong Min Tze who gave me moral support and guidance in completing my thesis. Lastly, I would like to extend my heartfelt thanks to Dr. Sutha Devaraj for motivating me in my fnal submission. Very big thanks to all who had contributed either directly or indirectly to the success of this project.

SHANDRA DEVI A/P BALASUBRAMANIAM

P-NFD0001/16 (R)

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

%	Percentage
µl	Microlitre
µM	Micrometre
cDNA	Complementary deoxyribonucleic acid
CIN	Cervical intraepithelial neoplasia
DNA	Deoxyribonucleic acid
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
g	Gram
GO	Gene ontology
HGCIN	High grade cervical intraepithelial neoplasia
HGSIL	High-grade squamous intraepithelial lesion
HPV	Human papillomavirus
HTA	Human transcriptome array
IHC	Immunohistochemistry
Kb	kilo base pair
L	Litre
LCM	Laser capture microdissection
LGCIN	Low grade cervical intraepithelial neoplasia
LGSIL	Low-grade squamous intraepithelial lesion
M	Molar
mg	Milligram
MHC	Major histocompatibility class
mL	Millilitre
mM	Millilmolar

mRNA	messenger RNA
ng	Nanogram
nm	nano metre
nM	Nanomolar
°C	Degree Celsius
PCA	Principal component array
PCR	Polymerase chain reaction
pH	Power of hydrogen
QC	Quality control
q-PCR	Quantitative polymerase chain reaction
RIN	RNA integrity number
RNA	Ribonucleic acid
RPM	Revolutions per minute
RT-PCR	Real time polymerase chain reaction
SCC	Squamous cell carcinoma
WHO	World Health Organization

LIST OF APPENDICES

PROFIL EKSPRESI GENOMIK *HUMAN PAPILLOMAVIRUS* 16 DAN 18 – YANG BERKAITAN DENGAN LUKA PRA-KANSER DAN KANSER SERVIKS

ABSTRAK

Kanser serviks adalah salah satu kanser yang biasa berlaku pada wanita dan kebanyakkanya disebabkan oleh jangkitan *human papillomavirus* (HPV). Jangkitan dengan *human papillomavirus* (HPV) berisiko tinggi dan penyatuan genom HPV ke dalam kromosom hos sel epitelium serviks adalah peristiwa awal penting dalam perkembangan luka serviks neoplastic. Oleh itu, objektif utama kajian ini adalah untuk memaparkan tandatangan genomik yang terlibat dalam patogenesis jangkitan *human papillomavirus* (HPV) 16 dan 18 pada luka pra-kanser (CIN) dan kanser sel skuamosa serviks (SCC), berbanding dengan serviks normal. Perbezaan profil ekspresi mRNA ditentukan untuk menilai ekspresi gen hos yang dikawal atas dan dikawal ke bawah serta untuk menilai laluan molekul yang terganggu yang terlibat dalam karsinogenesis serviks. Dalam kajian ini, 29 tisu *formalin-fixed paraffin-embedded* (FFPE) dikumpulkan, termasuk CIN kelas rendah (LGCIN), CIN kelas tinggi (HGCIN), SCC, dan serviks normal disaring untuk HPV 16 atau HPV 18 positif dengan menggunakan kaedah imunohistokimia (IHC) dan *qRT-PCR.* Daripada jumlah ini, 9 sampel tisu HPV-positif (3 LGCIN, 3 HGCIN, 3 SCC) dan 3 sampel tisu serviks normal (HPV-negatif) telah dipisahkan secara mikro untuk mendapatkan kawasan yang diminati dalam epitelium skuamosa sebelum pengekstrakan RNA. Pemprofilan transkriptomik telah dilakukan menggunakan kaedah *The Affymetrix, GeneChip Human Transcriptome Array 2*.0 (HTA 2.0). Manakala, *nCounter® PanCancer Pathway Array, Nanostring* digunakan untuk nCounter® PanCancer Pathway Array,

Nanostring untuk mengenal pasti gen yang diekspresikan secara berbeza (DEG) dan laluan yang berkaitan secara signifikan dalam setiap peringkat perkembangan kanser serviks. Gen yang berkaitan dengan tindak balas sel terhadap rangsangan dan kolagen yang berkaitan dengan *extracellular matrix* (ECM) secara signifikan diatur ke bawah semasa peralihan dari serviks normal ke LGCIN. Sebagai tambahan *MAPK*, kesalahan transkripsi dan laluan *JAK-STAT* terlibat, sementara *IL1B* dapat mendorong keradangan dan secara tidak langsung mengaktifkan *MMP9*, mengakibatkan pemecahan kolagen dan migrasi sel. Peralihan ke HGCIN telah mengganggu pengaturan morfogenesis, tindak balas terhadap luka, pengembangan organisme multisel, pembezaan sel, dan lekatan sel. Gen histone disregulasi dalam HGCIN. Semasa peralihan HGCIN, kitaran sel - laluan *apoptosis*, dengan peningkatan *E2F1* dan *MCM2*, serta gen pembaikan DNA *BRCA2-BRIP1* dan *FANCA*, sangat penting. Akhirnya, di SCC, gen yang terlibat dalam organisasi kromosom, pemasangan kompleks protein DNA, kitaran sel, dan percambahan sel adalah proses biologi yang paling banyak terlibat jika dibandingkan dengan serviks normal. Gen lokus berat imunoglobulin diatur secara signifikan. Selanjutnya, jika dibandingkan dengan serviks normal, kitaran sel dan laluan pemberian isyarat, serta peningkatan regulasi *c-MYC*, tampaknya penting dalam SCC. Deregulasi gen yang berkaitan dengan imun, termasuk pengubahsuaian histone dan gen rantai berat imunoglobulin, dijumpai di HGCIN dan SCC, yang menunjukkan imun keluar dari tindak balas anti-tumor hos. Semasa peringkat awal dan akhir karsinogenesis serviks, matriks ekstraselular memainkan peranan penting. Laluan dan gen yang deregulasi memberikan pemahaman yang lebih baik mengenai mekanisme molekul yang terlibat dalam perkembangan kanser serviks yang berkaitan dengan HPV.

GENOMIC EXPRESSION PROFILE OF

HUMAN PAPILLOMAVIRUS 16 AND 18 – ASSOCIATED

PRE-CANCEROUS LESIONS AND CANCER OF THE CERVIX

ABSTRACT

Cervical cancer is one of the most common cancers in women and is caused by high-risk human papillomavirus (HPV) infection. The integration of HPV into host cervical epithelial cells causes genetic alterations and consequent changes in gene expression affecting downstream molecular pathways, leading to the development of cervical cancer. This study aimed to profile the genomic signatures involved in the pathogenesis of HPV 16 and 18 associated pre-cancerous lesions (cervical intraepithelial neoplasia, CIN) and squamous cell carcinoma (SCC) of the cervix, in comparison to the normal cervix. The differential mRNA expression profiles were determined, to evaluate the expression of up-regulated and down-regulated host genes and to evaluate the perturbed molecular pathways involved in cervical carcinogenesis. In this study, 29 formalin-fixed paraffin-embedded (FFPE) tissues including low-grade CIN (LGCIN), high-grade CIN (HGCIN), SCC, and normal cervix were screened for HPV 16 and HPV 18 using immunohistochemistry and qRT-PCR method. Of these, 9 HPV-positive (3 LGCIN, 3 HGCIN, 3 SCC) and 3 normal cervix (HPV-negative) tissue samples were microdissected to obtain regions of interest in the squamous epithelium before RNA extraction. Transcriptomic profiling was performed using the Affymetrix, GeneChip Human Transcriptome Array 2.0 (HTA 2.0) and the nCounter® PanCancer Pathway Array, Nanostring to identify differentially expressed genes (DEGs) and significantly associated pathways in each stage of cervical cancer development. The results showed that genes associated with cellular response to

stimulus and collagen in association with the extracellular matrix (ECM) were significantly downregulated during the transition from normal cervix to LGCIN. In addition, the mitogen-activated protein kinase (MAPK), transcriptional misregulation and JAK-STAT pathways are implicated, while *IL1B* may promote inflammation and indirectly activate *MMP9*, resulting in collagen breakdown and cell migration. In the progression to HGCIN the pathways involved were regulation of morphogenesis, response to wounding, multicellular organism development, cell differentiation, and cell adhesion. Histone genes were dysregulated while in the cell cycle – apoptosis pathway, there was upregulation of *E2F1* and *MCM2*, as well as the DNA repair genes *BRCA2-BRIP1* and *FANCA*. Finally, in SCC, genes involved in chromosome organisation, DNA protein complex assembly, cell cycle, and cell proliferation are the most involved biological processes when compared to normal cervix. The immunoglobulin heavy locus genes were significantly upregulated. The cell cycle and signalling pathways, as well as *c-MYC* upregulation, appear to be essential in SCC. Immune-related gene deregulation, including histone modification and immunoglobulin heavy chain genes, was found in HGCIN and SCC, implying immune escape from the host anti-tumour response. During the early and late stages of cervical carcinogenesis, the ECM plays an important role. The dysregulated pathways and genes provide a better understanding of the molecular mechanisms involved in the development of HPV-associated cervical cancer.

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Human papillomaviruses (HPVs) are small double stranded-circular-stranded DNA viruses with genomes containing 8 kb of DNA sequences. Papillomaviruses are the main etiological factors of cervical cancer. The virus mainly affects the epithelia of skin, anogenital and oropharyngeal sites. To date, more than 130 HPV genotypes that have been identified. HPV has been detected in many organs and can cause a wide variety of infective lesions, from benign warts (condyloma acuminata), low and high grade intraepithelial or dysplastic lesions (pre-cancerous) and cancers.

A subgroup of HPV types, referred, as high-risk HPV genotypes are associated with intraepithelial lesions that have a high potential for progression to invasive carcinoma (Chang et al., 1996; Daniela et al., 2011). Infections by high-risk HPV genotypes typically have been associated with cervical cancers (Walbomers et al., 1999). HPV type 18 and 16 are the most prevalent genotypes worldwide in squamous cell carcinoma and adenocarcinoma (Mounz et al., 2003). The National Cancer Registry report indicated that cervical cancer is the one of the most common cancer in females in Malaysia, accounting for 12.9% of all female cancers (WHO, 2020). HPV type 16 and 18 are also the common predisposing factor for cervical squamous cell carcinoma and adenocarcinomas in Malaysia (Cheah et al., 2011). In contrast, low-risk HPV types; HPV 6 and HPV 11 are more often associated with benign tumours such as urogenital warts and only rarely found in malignant tumours (Chang et al., 1996).

In the development of cervical cancer, the infection of high-risk HPV types is to overcome the transcriptional control of viral gene expression in the infected

keratinocytes. The viral genes E6 and E7 play the major role to affect the cellular control function of host cells. The HPVs integrate the HPV genome into the chromosome of host cells, which then makes the E6, and E7 proteins are the only viral proteins that are consistently expressed in cervical cancers (Margaret et al., 2010). Once these viral genes interrupt the cellular mechanism of the host proteins, the function of cell proliferation, apoptosis, differentiation, metabolism, epigenetic reorganization and genomic instability (McLauglin et al., 2009) are affected.

In general, the HPV E6 and E7 oncoproteins play a major role in inactivating the tumor suppressor genes that operate the cell cycle. The E7 oncoprotein binds to retinoblastoma (pRb) gene and activates the E2F transcription factor, which triggers the overexpression of proteins necessary for DNA replication (Munger et al., 2004). The E6 oncoprotein targets the p53 for proteolytic degradation and induces activation of telomerase (Thomas et al., 1999). Constant activity of the viral proteins E6 and E7 lead to genomic instability, accumulation of oncogene mutations, further loss of cell-growth control and cancer (Nubia et al., 2006).

Majority of immunocompetent patients infected by HPV will clear the infection through normal immunological mechanisms. Generally, CIN is considered to progress in severity over time, passing from CIN 1 to CIN2/3 and then to invasive carcinoma (Theodoros et al., 2005). However, in some cases, the CIN 1 can revert to normal. The differences in host-viral interactions are thought to play an important role in this mechanism.

A literature search revealed scarce data on the gene signature profiles elucidating the molecular events that lead to either progression of pre-cancerous lesions to cancer using FFPE tissue. This study aims to elucidate the genomic

signatures in normal to HPV-associated pre-cancerous lesions and cancer of the cervix using FFPE cervical tissue.

1.2 Rationale of the study

HPV, being the most common etiologic agent in the development of cervical cancer highlights the importance of this study. The pathogenesis of persistent high-risk HPV infection involves the overexpression of viral oncoproteins that are able to inhibit a variety of host cellular proteins that affect various biological processes including cell proliferation, cell cycle, and apoptosis. The combination of viral and host cellular changes induce the transformation of HPV- infected cervical epithelial cells into precancerous lesion and further progress to invasive cancer. Elucidation of the genomic expression profiles of different stages of HPV-associated cervical lesions allows in-depth understanding of the molecular pathways involved in cervical carcinogenesis. The results will also provide information for the future development of targeted therapy against precancerous cervical lesions.

1.3 General Objective

To profile genomic signatures involved in the pathogenesis of human papillomavirus (HPV) 16 and 18 infections in pre-cancerous lesions and cancer of the cervix.

1.4 Specific Objectives

1) To determine the differential mRNA expression profiles in normal cervix, HPV-associated pre-cancerous lesions and squamous cell carcinoma of the cervix.
2) To evaluate the expression of the up-regulated and down-regulated host genes in cervical carcinogenesis.
3) To predict the putative molecular pathways involved in cervical carcinogenesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction on Cancer

Cancer is a disease in which abnormal cells grow and spread uncontrollably. Cells in all types of cancer begin to divide uncontrollably and spread into neighbouring tissues. Cancer can develop anywhere in the human body. In a normal state, human cells grow and divide to form new cells as needed by the body. Older, dead, or damaged cells are replaced by new ones in a normal process based on physiological demands. Cancer develops when the cellular reproduction process becomes uncontrollable. Cancer, in other terms, is a condition characterised by uncontrollable, uncoordinated, and unfavourable cell division. Cancer cells, unlike normal cells, continue to grow and divide throughout their lives, multiplying into even more harmful cells (Ferlay et al., 2019).

Cancer cells divide and replicate, resulting in the formation of a clump of cancer cells known as a tumour. Tumours are classified as benign or malignant. The majority of benign tumours are not cancerous. They do not spread to surrounding epithelial tissues. Furthermore, the majority of benign tumours are not life threatening. Malignant tumours, on the other hand, grow at an exponential rate to adjacent epithelial tissues or other organs through a variety of processes. Metastasis refers to the process by which cancer cells spread from the initial tumour site to other parts of the body (Davoli & de Lange, 2011).

There are over 100 different types of cancer. Cancer types are usually named after the organs or tissues where the cancers form. Lung cancer, for example, begins in lung cells, and brain cancer begins in brain cells. Cancers can also be classified

based on the type of cell that gave rise to them, such as an epithelial cell or a squamous cell (Davoli & de Lange, 2011).

According to World Health Organisation (WHO) estimates from 2015, cancer is the first or second leading cause of death for people under the age of 70 in 91 of 172 countries, and it ranks third or fourth in another 22 countries. Cancer is a major health issue in Malaysia. Malaysian National Cancer Registry 2018 reported 43 837 new cases and 26 395 deaths with 11% increase in new cancer cases and nearly 30% more deaths from cancer compared to the 2007–2011 report. The most common types of cancer in Malaysia are breast cancer, colorectal cancer, lung cancer, and liver cancer, which contribute to about half of the country's total cancer cases. Among women, cervical and breast cancers are the leading cause of death (MNCR, 2018).

2.2 Cervical Cancer Demographics

Cervical cancer is a type of malignant epithelial tumour that develops in the uterine cervix. Cervical cancer is the fourth most frequent disease in women globally, with 527,624 new cases and 265,672 deaths in 2012 (Bruni, 2017). It has been stated that low and middle-income countries account for around 85% of all cervical cancer fatalities worldwide (WHO, 2016) with a death rate that is 18 times greater than in affluent ones (Small et al., 2017). Death rates are quite high in various nations in Sub-Saharan Africa, Latin America, and Asia (Yang et al., 2004).

Cervical cancer is one of the most prevalent cancer among Malaysian women (Bruni et al., 2017). There were 4,696 cervical cancer cases been reported annually and the ministry of health Malaysia had stated an average of 2000-3000 hospital admission of cervical cancer with late stages of disease annually in the country (Devi et al., 2008). The recent Malaysian consensus reported there were approximately 2,145

new cervical cancer cases and 624 deaths are diagnosed annually (Bruni et al., 2017). Furthermore, Malaysia has a mortality rate that is twice that of the Netherlands, the United Kingdom, and Finland (Zaridah, 2014).

2.3 Risk Factors of Cervical Cancer

Numerous studies have found that women with lower socioeconomic status, bad personal and sexual hygiene, smoking, intercourse at a young age, or having multiple sexual partners are at a higher risk of developing cervical cancer (Small et al., 2017, Min et al., 2017). However, HPV infection is the major cause of cervical cancer developmen (Bosch et al., 2002). Cervical cancer does not affect every woman who has HPV infection. Cervical cancer carcinogenesis is a dynamic process that occurs more frequently in younger women and less frequently in elderly women. Cervical cancer is avoidable if detected at an early stage. The cofactors of the disease are the primary underlying variables that accelerate the progression of cancer.

2.4 Introduction to Human Papillomavirus (HPV)

HPV is one of the leading causes of sexually transmitted disease in both men and women worldwide, particularly in developing countries. The majority of HPV infections are transient, and studies have found that the most of sexually active people are exposed to and infected with the virus at some point in their lives. Human papillomaviruses (HPV) are small non-enveloped double-stranded DNA viruses with genomes containing 8 kb of DNA sequences. There have been over 200 HPV genotypes identified to date, which are divided into mucosal and cutaneous HPV (De Villiers et al., 2004). Only HPVs with the ability to infect basal epithelial cells of the skin or inner-lining tissues are classified as cutaneous or mucosal types. HPVs of the

cutaneous type infect the keratinized surface of the skin, primarily on the hands and feet. In the meantime, mucosal type HPVs infect the lining of the mouth, throat, respiratory tract, or anogenital tract epithelium. The high-risk HPVs are linked to mucosal infection, while the low-risk HPVs are linked to cutaneous lesions. Low-risk HPV types, such as HPV 6, 11, 42, 43, and 44, are associated with benign lesions that typically form warts and are seldom found in malignant tumours (Egawa and Doorbar, 2017). High-risk HPVs, on the other hand, including HPV 16, 18, 31, 33, 34, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, and 70, are related with cervical cancer lesions, with HPV 16 being the most common high-risk HPV, followed by HPV 18 (Li et al., 2011). In the case of carcinogenesis, persistent infection with either HPV 16 or HPV 18 causes 70% of cervical cancers and 50% of CIN 3 lesions (Smith et al., 2007).

2.4.1 Structure of HPV

Human papillomaviruses (HPV) are a small type of non-enveloped double-stranded DNA virus. They have a spherical shape and diameter of 50-60 nm. The capsid of the virus is composed of L1 and L2 proteins as depicted in Fig. 2.1. The viral genome is composed of 8 kb of DNA sequences with three major regions. A schematic diagram of the HPV genome is shown in Figure 2.2 (Munoz et al., 2006). The first region, the noncoding region (URR), regulates E6 and E7 gene transcription. The early (E) region, which consists of six open reading frames (E1, E2, E4, E5, E6, and E7) that encode structural proteins important in viral replication and oncogenesis, is the second region. The late region (L), which encodes the structural proteins L1 and L2, is the third region. The regulation of viral gene expression is complicated and is governed by cellular and viral transcription factors. Table 2.1 shows the key roles of

HPV viral proteins in initiation of infection and subsequent cancer progression (de Sanjosé et al., 2017).

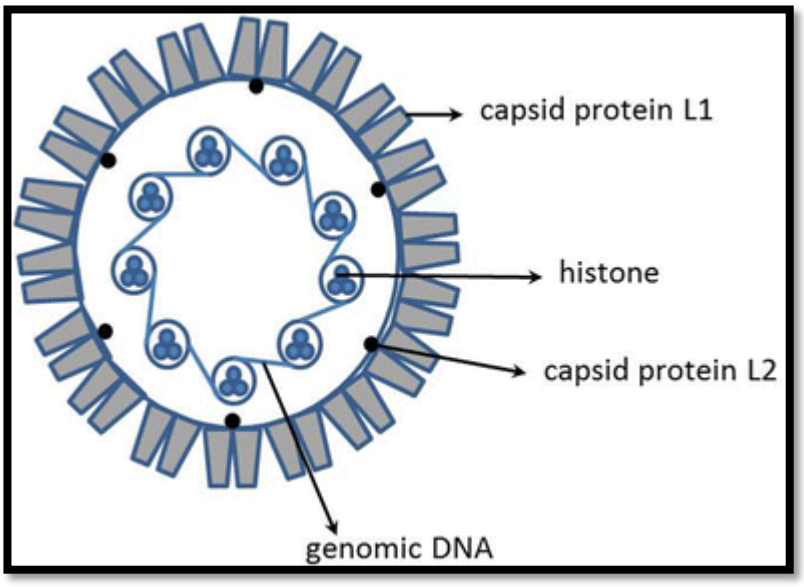

Figure 2.1 Structure of HPV virus comprising capsid proteins of L1 and L2, histone and genomic DNA. Figure adapted from: (Fernandes & Fernandes, 2012).

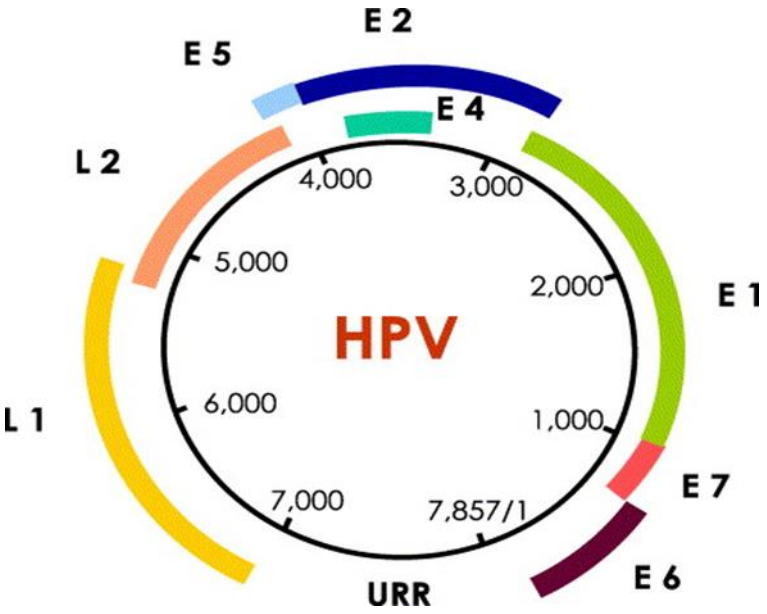

Figure 2.2 Schematic diagram of HPV genome

The schematic presentation of the HPV genome showing the arrangement of the early E or nonstructural genes, the capsid genes (L1 and L2) and the upstream regulatory region (URR). Figure adapted from: (Munoz et al., 2006).

In terms of taxonomy, currently, more than 200 types of HPV belong to the family *Papillomaviridae*. Differences in the *L1* gene sequence are the basis for the latest classification of HPV, according to which five genera of viruses are distinguished: α, β, γ, μ, and ν. Approximately 60 HPVs belonging to the α group show affinity for cervical epithelial cells. HPVs are classified into two groups depending on their oncogenic potential. The group of high-risk viruses includes 13 types including 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 66, and 68 which have the potential to cause cervical cancer. The group of low-risk viruses includes types 6, 11, 42, 43, and 44, and are responsible for the clinically overt form, genital warts (GW), and other HPVs of lesser clinical significance (Munoz e tal., 2003). Notably, infection with the most oncogenic types, 16 and 18, increases the risk of developing cervical cancer by more than 200-fold (Bosch et al., 2002).

Table 2.1 List of HPV viral proteins and their major functions (de Sanjose et al., 2017)

Viral protein	Protein Functions
E1	Viral DNA replication and transcription
E2	Viral DNA replication, apoptosis, transcription repressor of E6/E7
E4	Viral DNA replication
E5	Immune recognition (MHC)
E6	p53 degradation, alteration of cell cycle regulation, apoptosis resistance
E7	pRb degradation, re-entry into S phase cell cycle, p16 overexpression
L1	Major viral capsid protein
L2	Minor viral capsid protein

2.4.2 Life Cycle of HPV

The HPV life cycle starts with the infection of stem cells in the epithelium's basal layer. The virus makes a low-copy-number entry into the basal cells, where the viral DNA replicates and does not multiply until the host cell matures into a keratinocyte (Doorbar, 2006). Microabrasions in the epithelium are the mechanism by which HPV enters basal cells (Figure 2.3).

The viral load will grow as the viral DNA replicates. To integrate its genome into the host protein, the virus releases its viral protein. After entering the cells, the virus releases E1 and E2 viral proteins that aid in viral replication. In keratinocytes, high-risk HPV E2 proteins can act as transcriptional activators (Bodily and Laimins, 2011) but they also act as transcriptional repressors of viral gene expression. E2 proteins are also involved in the segregation of viral genomes during cell division by integrating viral genomes with mitotic chromosomes (You et al., 2004). E6 and E7 proteins are activated by high E1 and E2 gene expression, which then modulates the host cell cycle regulators to maintain long-term replication competence

The E6 and E7 oncoproteins are the most important in the progression of cervical cancer. E6 protein inhibits p53 by degrading the p53 gene, causing cells to divide uncontrollably. The E7 protein, on the other hand, would then bind to the pRB, retinoblastoma gene and cause the E2F, transcription factor gene, to be released, causing the cells to enter the S phase of the cell cycle. The HPV needs to get rid of the host immune system in order to maintain and differentiate in the infected host cells. As a result, the E5 oncoprotein interacts with cell membrane growth factors, causing the MHC class 1 molecules to be regulated (Goldstein et al., 1992). Several more viral genomes are released as complete virions at this point. When there is a persistent HPV

infection, the L1 and L2 genes are highly expressed, allowing the complete virions to be released. The virus enters into the epithelial tissue through micro abrasion and infects the basal cells at low copy number. Upon infection, the viral oncoproteins E1, E2, E4, E5, E6, and E7 are triggered. These oncoproteins interact with the host cellular function and disrupt the normal function of the host cells leading to proliferation of abnormal cells. The L1 and L2 help in releasing the complete virions out. Mention Figure in text

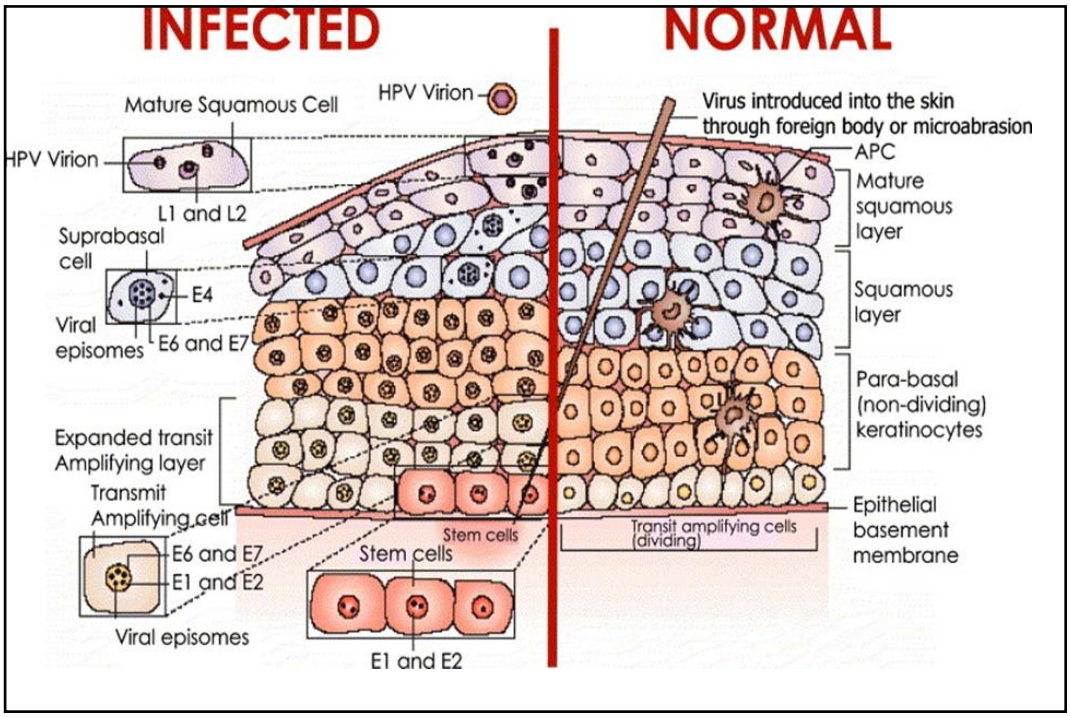

Figure 2.3 The life cycle of HPV in squamous epithelium. Figure adopted from: (Tomaic, 2016).

2.5 Pathological Changes in the Cervix

The cervix is the lower, cylindrical part of the uterus that connects to the vagina via the endocervical canal (Bermudez et al., 2015). The endocervical canal is lined with stratified squamous epithelium that covers the ectocervix and columnar epithelium that covers the endocervix. The transition zone between these cells is known as the squamo-columnar junction. Pre-malignant transformation and dysplasia of cells most commonly occur near the squamo-columnar junction and are typically caused by high-risk HPV infection, primarily genotypes 16 and 18. CIN refers to premalignant alterations and dysplasia of squamous cells in the cervical epithelium. If CIN is not treated at an early stage, it can progress to carcinoma *in situ* and invasive carcinoma if the HPV is able to deactivate the host cellular functions (Small et al., 2017).

According to studies, invasive disease develops as a result of the progression of mild dysplasia to severe dysplasia and then to carcinoma *in situ* (Holowaty et al., 1999). CIN is a term used to describe these precursor lesions, which represent varying degrees of disordered cell maturation in the cervical epithelium. CIN is divided into three types: CIN 1, CIN 2, and CIN 3. CIN 1 is defined as mild dysplasia; one-third of the epithelial tissue is abnormal, CIN 2 is known as moderate dysplasia; two-thirds of the epithelial tissue is aberrant, and CIN 3 is known as severe dysplasia. The Bethesda system is intended to make cytological diagnostics simpler. As a consequence, lesions with CIN 1 are categorised as low-grade squamous intraepithelial lesions (LGSIL), while lesions with CIN 2 or CIN 3 are classed as high-grade squamous intraepithelial lesions (HGSIL) in this system (Anderson et al., 1991, Arends et al., 1998). According to a study conducted by Holowaty et al, both mild and moderate dysplasia (CIN 1 and CIN 2) are more likely to regress than advance (Holowaty et al., 1999). The risk of

mild dysplasia cases progressing to severe dysplasia is only 1%, but the risk of moderate dysplasia (CIN 2) progressing is 16% -25%, and the progression could take 2 to 5 years. Without treatment, the transition from dysplasia to invasive carcinoma can take years or decades in most cases. Nevertheless, it has been reported that 10% of patients made this transition in less than a year (Small et al., 2017).

There is significant evidence that viral proteins in HPV promote dysplastic alterations in cells, as well as the transformation of precancerous to cancerous lesions (Holowaty et al., 1999). A person with mild dysplasia or CIN 1 (low-grade CIN) may recover from infection with the help of the host's immune system (Chabra, 2016). CINs are categorised histologically based on their severity. Fig. 2.4 depicts the distribution of epithelial cells in the cervix in normal, precancerous (CIN 1, CIN 2, and CIN 3) and cancerous situations (Balasubramaniam et al., 2019). In a normal cervix, epithelial cells are properly structured. Therefore, in CIN and cancer, HPV-infected cells become dysplastic. When the lower one-third of the epithelium reveals dysplasia, CIN 1 denotes mild dysplasia. CIN 2 or moderate dysplasia occurs when two-thirds of the epithelium is affected. When more than two-thirds or the entire thickness of the epithelium is affected, the condition is classified as severe dysplasia (CIN 3). CIN 2 and CIN 3 lesions are usually grouped together as high grade CIN.

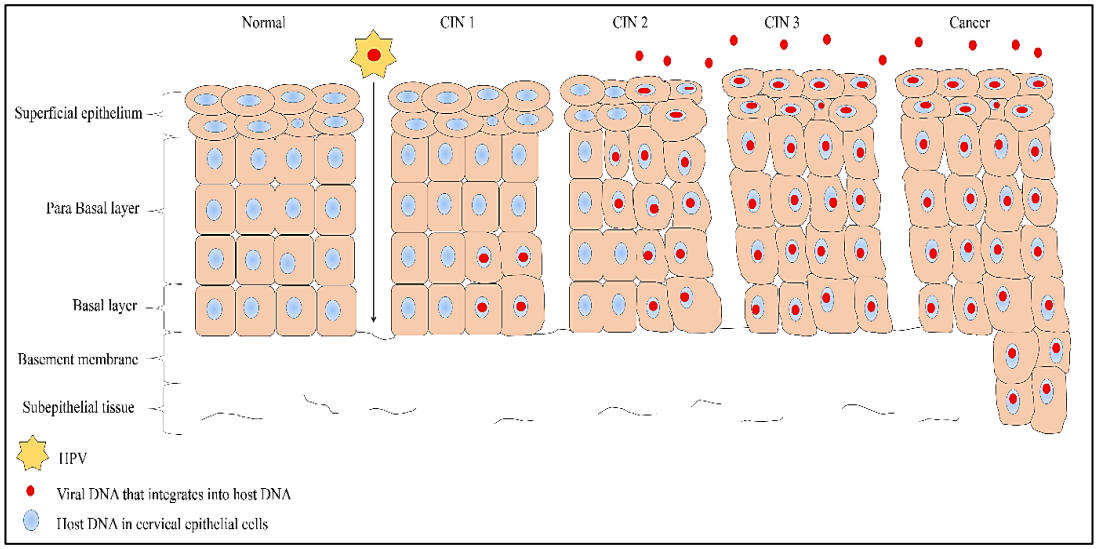

Figure 2.4 Distribution of normal and HPV-infected squamous epithelial cells in normal, precancerous lesions (CIN 1, CIN 2, and CIN 3) and cancer of the cervix.

The initial stage of carcinogenesis is controlled by viral HPV integration and host factors. HPV enters the basal epithelial cells through a micro-wound. Subsequently, the virus integrates its genome into the host genome in the nucleus through nuclear envelope breaks. Once it enters the nucleus, HPV takes over control of the host genome, self-replicates, and spreads across the epithelial cells. Further replication of the viral genome causes the host cells to grow irregularly and in a disorganized manner compared to normal cells. Subsequently, the virions are sloughed off with the dead squamous cells of the host epithelium, enabling further transmission. Figure adapted from: (Balasubramaniam et al., 2019).

2.6 Involvement of HPV Oncoprotein in Cervical Carcinogenesis

The integration of the HPV genome into a host chromosome is one of the major events in HPV-induced carcinogenesis (MuÈnger et al., 2001). When infected with high-risk HPV types, the E6 and E7oncoproteins play a significant role in basal or parabasal cell proliferation (Pagliarulo, 2014). Increased expression of the high-risk E7 protein causes significant epigenetic reprogramming of the cell, which is also thought to be necessary for E7's cell cycle entry and progression stimulation (Lees et al., 1993). This mechanism enables the transcription of genes, the products of which are essential for the cell to enter the S phase of the cell cycle (Sarris et al., 2014) via various mechanisms.

The E6 oncoproteins of high-risk HPVs inactivate p53 function (Mantovani and Banks, 2001). Degradation of p53, on the other hand, has an effect on p53's normal activities, which govern G1 arrest, apoptosis, and DNA repair (Chaurushiya and Weitzman, 2009) Increased mitogen-activated protein (MAP) kinase activity is induced by the E5 gene product, which improves cellular responses to growth and differentiation factors (Zhang et al., 2002). As a result, the host cell continues to proliferate and differentiates slowly. In fact, E5 is involved in the formation of koilocytes. The E2 gene product is a DNA binding protein that prevents the transcription of the E6 and E7 genes while allowing the E1 gene product to bind to the viral origin of replication, which is located with the LCR. The integration of viral DNA into the host cell chromosome inactivates E2 expression, resulting in increased expression of the E6 and E7 oncogenes and genetic destabilisation. The expression of the E4 gene was discovered in the late stages of infection. The gene promotes virion release into the environment by disrupting intermediate filaments of the keratinocyte cytoskeleton (Bogdanovic et al., 2009).

From CIN 1 to CIN 3, the expression of E6 and E7 oncoproteins increases in cervical cancer (Klaes et al., 1999). An increase in E6 and E7 activity causes genetic errors in the cells, which eventually leads to cancer progression. Reduced E6 and E7 activity in CIN 1 is insufficient for the oncoprotein to play a role in cancer progression. Nevertheless, the E7 and E6 proteins can promote cancer progression by deregulation of the cell cycle and triggering a mutation in the p53 protein. p53 and pRb are well-known cellular tumour suppressors that play roles in cell cycle progression, DNA repair, apoptosis, differentiation, senescence, and chromatin remodelling (Klaes et al., 1999). When these proteins are influenced by viral proteins, they aid in the progression of cancer. Low-risk HPV types do not promote cell proliferation, which is why infection does not result in cancer, as opposed to high-risk HPV types, which increase cell proliferation. Fig. 2.5 depicts a simplified mechanism for the role of E6 and E7 oncoproteins in cervical carcinogenesis.

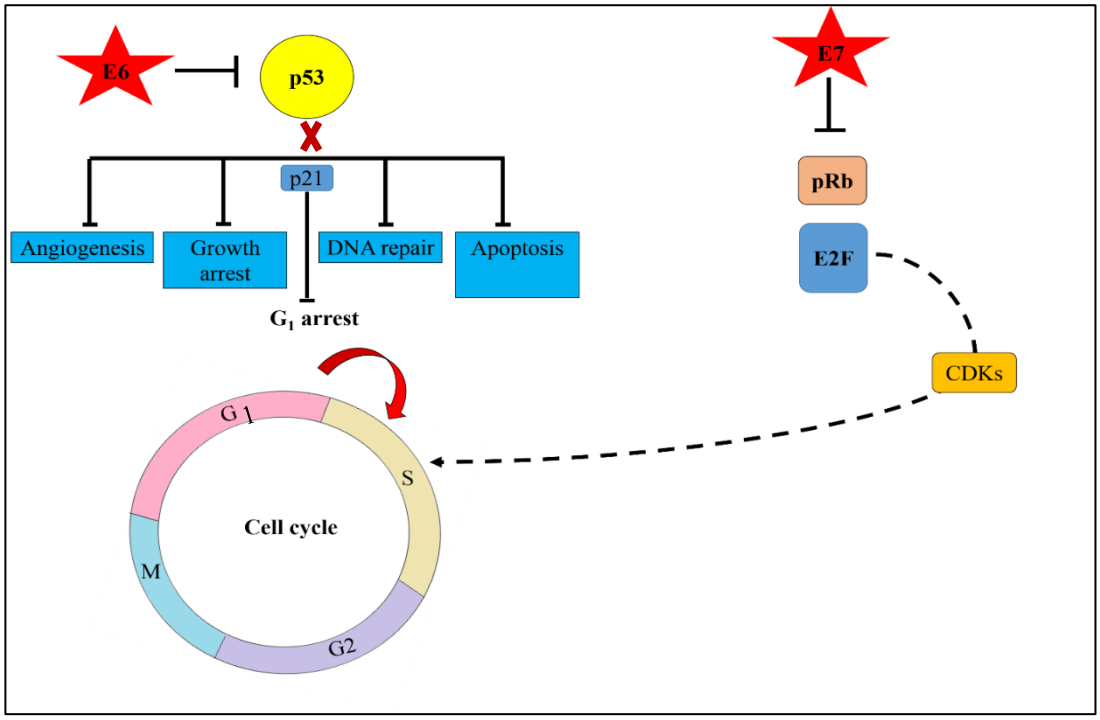

Figure 2.5 Schematic diagram illustrating the role of HPV oncoproteins in cervical carcinogenesis.

Upon infection, the viral oncoprotein E6 binds to p53, a tumour suppressor protein and disables its normal function. The host cell's ability to undergo DNA repair, apoptosis, or growth arrest and angiogenesis is disrupted. Activation of p53 will activate cyclin-dependent kinase inhibitor (p21) to inhibit cells to remain at G_1 arrest. However, upon HPV infection, the E6 degrades p53 which causes cells to enter S phase of cell cycle. Simultaneously, the E7 oncoprotein binds to retinoblastoma (pRb). Binding of E7 to pRb causes it to release E2F, a transcription factor that activates the cyclin dependent kinase (CDKs). This causes the cell cycle to lose control, allowing the cells to re-enter into S phase of the cell cycle. When infected cells differentiate and proliferate at a high level, it will promote the development of abnormal dysplastic cells. Figure adapted from: (Balasubramaniam et al., 2019).

2.7 Molecular Changes involved in Cervical Carcinogenesis

2.7.1 Genetic Changes in Precancerous Lesion, Low-Grade CIN (LGCIN)

An infected epithelial cell undergoes a lengthy process of transformation into a cancerous cell after being infected with HPV. This process is linked to the accumulation of DNA modifications in host cell genes, which can occur both epigenetically and as a result of genetic alterations in oncogenes and tumour suppressor genes. HPV enters the host basal squamous cells by a micro-wound or abrasion during infection. To begin the infection, the HPV must integrate into the host cell, where a series of genetic events occur at the basal epithelial cells, directly permitting viral replication. These events then create a favourable environment for neoplastic progression (Gius et al., 2007). During this time, the virus must avoid detection by the host immune system in order to continue replication in the basal epithelial cells. As a result, LGCIN is the site where the virus continues to infect cervical cells. It takes at least 10-12 years for a precancerous lesion to progress to invasive carcinoma (Wallin et al., 1999, Zielinski et al., 2001). Lesions that do not regress from LGCIN may progress to high-grade CIN within 2-3 years of infection (Winer et al., 2005).

In general, when infected, host cells activate the innate and adaptive immune systems, which are governed by the major histocompatibility complex (MHC) classes I and II. Antigen is presented to cytotoxic T cells with CD8+ receptors by MHC class I molecules, whereas helper T cells with CD4+ receptors are presented by MHC class II molecules. During HPV infection, the virus penetrates the epithelium, the outermost physical boundary against infections, and ultimately activates the innate immune system. By expressing toll-like receptors (*TLRs*), macrophages, Langerhans cells, and

natural killer cells attempt to suppress HPV (Deligeoroglou et al., 2013). *TLRs* bind to viral components and activate transcription factors such as nuclear factor kappa B *(NF-KB)* and interferon response factor-3 *(IRF3*), resulting in the production of cytokines (Carmody and Chen, 2007). TLRs indirectly activate MHC class I and class II in this scenario. It has been reported that high-risk HPV can express its viral protein E5 in order to inhibit MHC class I (Ashrafi et al., 2005). Furthermore, the E6 viral oncoprotein has the ability to block *TLR*, which activates *IRF3* (Park et al., 2000).

Toll-like receptors (*TLRs*) in host cells serve a basic role in pathogen recognition and innate immunity activation, as well as in the production of cytokines required for the establishment of efficient immunity. *TLR 3, 4, 7, 8*, and *9* have been shown to play an important role in antiviral immunity by stimulating the downstream production of interferons (*IFNs*) (Sato et al., 2009). During the early stages of cervical carcinogenesis, the differentiated cervical epithelium regulates *TLR* receptors, which in turn elicit antiviral responses via IFN-regulatory factor (*IRF*) (DeCarlo et al., 2011). According to one study, the expression of the *TLR*-9 gene varies depending on the stage of cervical cancer development. *TLR-9* gene expression was found to be lower in LGCIN than in HGCIN, with the highest expression in SCC group samples (Ghosh et al., 2015). However, continuous overexpression of E6 and E7 oncoproteins may down-regulate continuous overexpression of E6 and E7 oncoproteins, on the other hand, may down-regulate *TLR-9*, impairing the subsequent interferon response, resulting in immune evasion and prolonged infection (Hasan et al., 2007). It has been reported that the HPV E7 oncoprotein can bind to Histone deacetylase 1 (*HDAC1*) and inhibit histone acetylation, thereby downregulates *TLR9* signalling (Hasan et al., 2013). Another study found that HPV upregulates the epidermal growth factor receptor (*EGFR*) to promote the expression of the interferon-

related developmental regulator I (*IFRD1*), which reduces cytokine production by inhibiting *NF-KB* (Tummers et al., 2015).

HPV has the ability to bypass the innate immune system and enter epithelial cells. Dendritic cells engulf the HPV antigen as soon as they enter the cell and go through a maturation process. The antigen is then sent to MHC class I or II molecules on the cell surface by the phagolysosome. When CD4+ and T cells bind to the T-cell receptor (TCR), they will also bind to it. The antigen-presenting cells (APC) will then activate CD4+ and T cells, causing them to become cytotoxic. When APC is activated, pro-inflammatory and antiviral cytokines such as *IFN-γ* and *TNF-α* are activated. This stimulates macrophages, promoting inflammation or tumour immunity. Moreover, the interleukins are stimulated to respond to extracellular pathogens. The activation of APC, however, causes the production of Tregs (regulatory T cells). Tregs will produce interleukin (*IL-10*) and transforming growth factor beta (*TGF-β*), both of which will impede APC function. As a result, in HPV cancer progression, the number of Treg cells produced correlates with the transformation of cells from normal to precancerous lesion and cancer. Women with persistent HPV 16 infection have been found to have significantly more Tregs than women who are HPV-negative (Torres-Poveda et al., 2014). Furthermore, Treg-inducing molecules such as *TGF- β1* have been found to be higher in lesions progressing from LGCIN to invasive cervical cancer in some other study (Molling et al., 2007).

Gius discovered that HPV alters the cellular immune system *IFNAR1* (IL receptor I), *EMP1, L1RN,* and *IL1RN* (IL receptor antagonist) genes at the early stage of infection, implying that at the early stage of HPV infection, the virus changed expression of these genes to evade the host immune system and permissively enable progression of infected cells to become cancerous (Gius et al., 2007). Another research

reveals that IL1R2 is downregulated when cervical cancer progresses (Niu et al., 2017). According to a recent study, the Hippo-Yap pathway is involved in cervical cancer (He et al., 2019). The study discovered that the principal effector of the Hippo signalling pathway, *YAP1* (Yes-associated protein), interacts with the HPV E6 oncoprotein to initiate and progress cervical cancer. *YAP1*, also known as *YAP65*, is a transcription coregulator that helps in the regulation of cellular proliferation and apoptotic genes. It is a component of the hippo signaling pathway, which regulates various aspects of the development and growth of organs. It is also known to be an oncogene that is commonly expressed in various human cancers (Huang et al., 2005). Indeed, dysregulation of YAP -mediated transcriptional activity is associated with the development of abnormal cell growth and hyperactivation of *YAP* has been observed in many cancers (Harvey, Zhang & Thomas, 2013; Johnson & Halder, 2014; Shimomura et al., 2014). As such, *YAP1* represents a potential target for cancer treatment (Moroishi, Hansen & Guan, 2015). In this situation, the HPV oncoprotein attaches to *YAP1* and works synergistically to block *YAP1* degradation (Wang and Davis, 2019). Furthermore, the Cancer Genome Atlas (TCGA) Research Network discovered that YAP1 is one of the genes that is significantly expressed in squamous tumours, implying that the Hippo/Yap1 pathway may be implicated in the development of cervical cancer (Network, 2017). The overexpression of *YAP1* in cervical epithelial cells is sufficient to cause the formation of cervical squamous cell carcinoma (He et al., 2019). Furthermore, *YAP1* activation increases the susceptibility to HPV infection by upregulating putative HPV receptors and inhibiting innate immunity (He et al., 2015, He et al., 2019). *EGFR, ITGRA6,* and *SDC1* are the putative HPV receptors. The study also discovered that *YAP1* upregulation reduces the expression of *TLR 2* and *4*, which are important components of innate immunity (He

et al., 2019). Moreover, *YAP1* has been found to suppress TANK-binding kinase 1 (*TBK1*) activity, thereby negatively regulating the production of type I interferon (Zhang et al., 2018). *TBK1* is a molecule that is activated by pattern recognition receptors (*PRRs*) in the presence of a virus, resulting in the production of type I interferon. It is a key protein in signalling pathways such as cell proliferation, autophagy, and insulin signalling pathways (Gu et al., 2016). Although it has been reported that *YAP1* could be a prognostic biomarker in cervical carcinogenesis, the role of *YAP1* in cervical cancer development needs to be explored further. In conclusion, HPV-associated cancer develops as a result of evasion of the host immune system, which leads to increased host cellular dysfunction. Fig. 2.6 and Fig. 2.7 depict the essential events that occur when HPV E6 and E7 oncoproteins expression cause evasion of the host immune system. According to the literature study, the host immune system is affected during the early stages of cervical cancer, which leads to cancer progression.

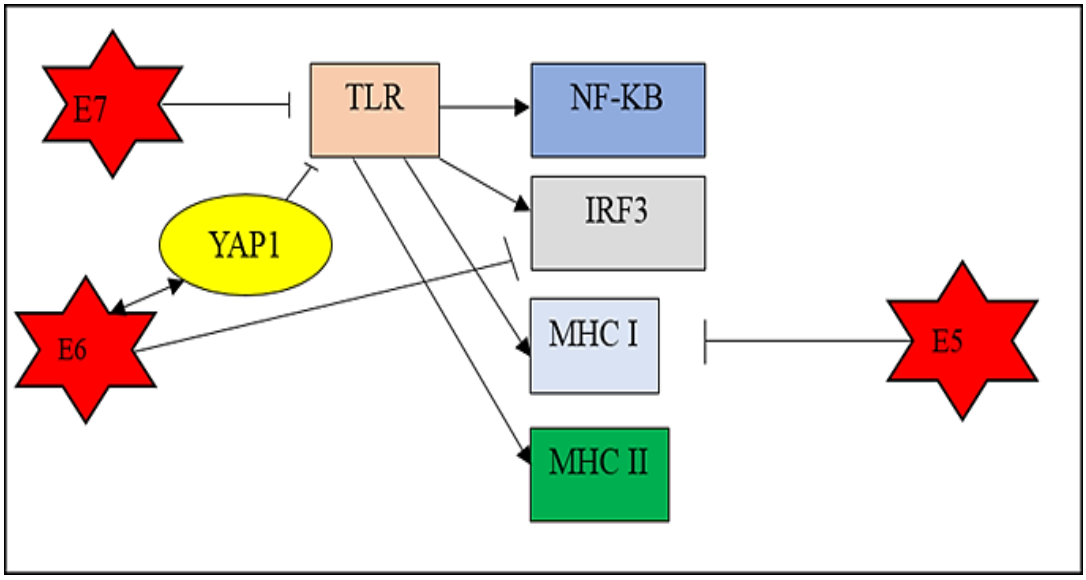

Figure 2.6 Mechanisms involved upon HPV infection. Upon HPV infection, the host immune system triggers the toll-like receptors (*TLR*) to activate the nuclear factor-kappa B (*NF-KB*) and interferon regulatory factor 3 (*IRF3*), to activate pro-inflammatory factors and antiviral cytokines. The *TLR* also activate the MHC class I and II. However, the HPV is able to express its viral oncoprotein E5 to inhibit the MHC class I mechanisms. The oncoprotein E6 can inhibit the production of *IRF3*. The oncoprotein E6 also binds to *YAP1* preventing degradation of *YAP1* and inhibiting the *TLR*. Figure adapted from: (Balasubramaniam et al., 2019).