ELUCIDATION AND COMPARISON OF PROTEOME PROFILES IN NORMAL, PRECANCEROUS AND CERVICAL CANCER TISSUES

GAAYATHRI A/P KUMARASAMY

UNIVERSITI SAINS MALAYSIA

2023

ELUCIDATION AND COMPARISON OF PROTEOME PROFILES IN NORMAL, PRECANCEROUS AND CERVICAL CANCER TISSUES

by

GAAYATHRI A/P KUMARASAMY

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

April 2023

ACKNOWLEDGEMENT

First and foremost, I would like to express my outmost gratitude to my supervisor Prof. Dr. Gurjeet Kaur and my co-supervisor Assoc. Prof. Dr. Mohd Nazri Ismail for their advice and guidance in completing the whole study. I also would like to thank Prof. Dr. Peter Hoffmann and Dr. Parul Mittal University of South Australia, for their technical support for mass spectrometry analysis. Not forgetting to mention Assoc. Prof. Dr. Sharifah Emilia Tuan Sharif from School of Medical Sciences, Universiti Sains Malaysia, for providing the formalin-fixed paraffin-embedded (FFPE) tissues for this study. At the same time, I would like to thank the Institute for Research in Molecular Medicine (INFORMM) and Analytical Biochemistry Research Centre (ABrC) for providing all the necessities, including laboratory and instruments throughout this study. Special thanks were also conveyed to my family and friends for their support and assistance. They have supported me throughout the journey of this study. Especially my parents who encouraged me to complete this study and not forgetting their financial support. I would like to thank the Lord for everything. Whatever happens, it's all because of Him. Finally, I would like to dedicate this thesis to my late father Kumarasamy @ Tony Cheng.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
ACN	Acetonitrile
ANXA2	Annexin A2
BSA	Bovine serum albumin
CID	Collision-induced dissociation
CRNN	Cornulin
DC	Detergent compatible
DDT	Dithiothreitol
ESI	Electrospray ionization
FFPE	Formalin-fixed paraffin-embedded
HPV	Human papillomavirus
HRP	Horseradish peroxidase
IAA	Iodoacetamide
IHC	Immunohistochemistry
LC	Liquid chromatography
LC-ESI-MS/MS	Liquid chromatography-electrospray ionisation- tandem mass spectrometry
MS/MS	Tandem mass spectrometry
PCR	Polymerase chain reaction
PI3K	Phosphoinositide 3-kinases
PTM	Post-translational modification
RC	Reducing agent
SCC	Squamous cell carcinoma
SIL	Squamous intraepithelial lesion

TFA	Trifluoracetic acid
Tims	Trapped ion mobility spectrometry
TM	Transmembrane
TOF	Time of flight

LIST OF SYMBOLS

°C	Degree Celsius
µg	Microgram
µL	Microlitre
Fmol	Femtomole
G	Gram
G	Relative centrifugal force
L	Litre
mA	Milliampere
Mg	Milligram
mL	Millilitre
mM	Millimolar
mW	Molecular weight
V	Volt
A	Alpha
B	Beta

LIST OF APPENDICES

KAJIAN DAN PERBANDINGAN PROFIL PROTEOME DALAM TISU NORMAL SERVIK, PRAKANSER DAN KANSER

ABSTRAK

Kanser serviks merupakan kanser keempat yang paling umum pada wanita di seluruh dunia dan terutamanya disebabkan oleh jangkitan virus papilloma manusia (HPV). Analisis proteomik menggunakan pendekatan spektrometri jisim yang sensitif dan berkapasiti tinggi dapat membantu memahami mekanisme molekul karsinogenesis kanser serviks dan penemuan biomarker baru untuk diagnosis awal. Sehingga kini, penanda protein yang spesifik untuk langkah-langkah berurutan dalam pembentukan kanser serviks belum dilaporkan. Tujuan kajian ini adalah untuk menentukan profil protein dan protein yang diekspresikan secara berbeza dalam serviks normal, lesi intraepitelial skuamosa (SIL) yang berkaitan dengan HPV yang merupakan prakanser, dan kanser serviks yang diekstrak daripada tisu tertanam formalin-terikat (FFPE). Komponen epitelium skuamosa telah dimikrodiseksi daripada enam kes serviks normal, lima SIL, dan enam SCC, kemudian dianalisis menggunakan spektrometri jisim *orbitrap* dan *trapped ion mass spectrometry-time of flight* (timsTOF). Selepas pengukuran kuantitatif protein menggunakan perisian MaxQuant dan Perseus (perubahan lipatan log2 $\geqslant$ 1.5 atau $\leqslant$ -1.5, nilai $p \leq 0.05$), analisis bioinformatik dilakukan menggunakan DAVID untuk analisis pengayaan gen, STRING untuk rangkaian interaksi protein-protein dan ConsensusPathDB untuk analisis pengayaan laluan. Annexin A2 dan cornulin kemudian dipilih untuk penilaian imunohistokimia pada tisu mikrotiap yang terdiri daripada 10 kes serviks normal, 21 SIL, dan 101 kes SCC. Sejumlah 3597 protein telah dikenal pasti yang terdiri daripada

589, 550, dan 1570 protein yang unik untuk kumpulan serviks normal, SIL, dan SCC, masing-masing, manakala 332 protein bersilang antara ketiga-tiga kumpulan. Di antara protein yang diekspresikan secara berbeza, 39 telah mengalami penurunan regulasi dalam SIL berbanding serviks normal, 51 peningkatan regulasi dalam SCC berbanding SIL, manakala dalam SCC berbanding serviks normal, terdapat 49 peningkatan dan 4 penurunan regulasi. Analisis bioinformatik menunjukkan bahawa proses pengikatan adalah fungsi molekul paling penting, manakala proses biologi seperti kebisuan kromatin dan perhimpunan nukleosom adalah signifikan untuk semua kumpulan histologi. Karsinogenesis virus dan nekroptosis adalah laluan penting dalam semua kumpulan histologi, mencerminkan peranan mereka dalam proliferasi sel, migrasi, dan metastasis. Penilaian imunohistokimia menunjukkan kehadiran protein annexin A2 dalam semua kumpulan histologi dan dianggap tidak signifikan secara statistik. Sebaliknya, ekspresi cornulin berbeza secara signifikan antara kumpulan dengan kecenderungan menurun dari serviks normal ke sel skuamosa karsinoma, menyokong peranannya sebagai gen penekan tumor dan berkaitan dengan laluan isyarat nekroptosis dan karsinogenesis virus. Laluan seperti karsinogenesis virus dan nekroptosis ditemui mengalami penurunan regulasi dalam SIL berbanding serviks normal dan regulasi meningkat dalam SCC berbanding SIL, menunjukkan kemungkinan laluan diaktifkan dan dimatikan berdasarkan ekspresi protein. Cornulin mungkin menjadi biomarker berpotensi untuk diagnosis awal kanser serviks dan memerlukan kajian lanjut.

ELUCIDATION AND COMPARISON OF PROTEOME PROFILES IN NORMAL, PRECANCEROUS AND CERVICAL CANCER TISSUES

ABSTRACT

Cervical cancer is the fourth most common cancer in women worldwide and is primarily caused by infection with the human papillomavirus (HPV). Proteomic analysis utilising sensitive, high throughput mass spectrometry approaches can aid in understanding the molecular mechanisms of cervical carcinogenesis and the discovery of novel biomarkers for early diagnosis. To date, the protein signatures specific to the sequential steps in cervical cancer development have not been reported. The aim of this study was to determine the protein profiles and differentially expressed proteins in the normal cervix, HPV-associated squamous intraepithelial lesion (SIL) which is precancerous and cervical cancer, extracted from formalin-fixed paraffin-embedded (FFPE) tissues. The squamous epithelial component was microdissected from six cases of normal cervix, five SIL, and six SCC, then subjected to orbitrap mass spectrometry and trapped ion mass spectrometry time of flight (timsTOF). After protein quantification using MaxQuant and Perseus software (log2 fold change $\geqslant 1.5$ or $\leqslant -1.5$, *p*-value of 0.05), bioinformatics analysis was carried out using DAVID for gene enrichment analysis, STRING for protein-protein interaction network and ConsensusPathDB for pathway enrichment analysis. Annexin A2 and cornulin were then chosen for immunohistochemical evaluation on tissue microarrays comprising 10 cases of normal cervixes, 21 SIL, and 101 SCC cases. A total of 3597 proteins were identified comprising 589, 550, and 1570 proteins unique to the normal cervix, SIL,

and SCC groups, respectively, while 332 proteins overlapped between the three groups. Among the differentially expressed proteins, 39 were downregulated in SIL compared to normal cervix, 51 were upregulated in SCC compared to SIL, while in SCC compared to normal cervix there were 49 upregulated and 4 downregulated proteins. Bioinformatics analyses revealed the binding process as the topmost molecular function whereas chromatin silencing, and nucleosome assembly biological processes were significant for all histological groups. Viral carcinogenesis and necroptosis were critical pathways in all histological groups, reflecting their role in cell proliferation, migration and metastasis. Immunohistochemical evaluation demonstrated the presence of annexin A2 protein in all histological groups and was considered statistically non-significant. In contrast, cornulin expression was significantly different between the groups with a decreasing trend from the normal cervix to squamous cell carcinoma, supporting its role as a tumour suppressor gene, and linked to necroptosis and viral carcinogenesis signalling pathways. Pathways such as viral carcinogenesis and necroptosis are found to be downregulated in SIL compared to normal cervix and upregulated in SCC compared to SIL, indicating the possibilities of being switched on and off based on the protein expression. Cornulin may be a potential biomarker for the early diagnosis of cervical cancer and warrants further studies.

CHAPTER 1

INTRODUCTION

1.1 Background

Cervical cancer is one of the most common gynaecological cancers in the world. It is ranked the fourth most common cancer among women after breast cancer, colorectal cancer, and lung cancer. In 2018, it caused an estimated 570 000 cases and was responsible for approximately 311 000 deaths (Arbyn et al., 2020).

Cervical cancer was the leading cause of cancer-related deaths in women among eastern, middle, southern, and western Africa. The highest number of cases and fatalities were recorded in China and India in 2020 and accounted for 35% of the global cervical cancer incidence and deaths (Arbyn et al., 2020). In Malaysia, cervical cancer is the third most common cancer among women. Deaths due to cervical cancer in young women are uncommon, however, the incidence rate of cervical cancer is higher among women above 30 years old and in their 60s, with 54.7% in women between the ages of 40 and 59 (Mustafa et al., 2022).

Infection with human papillomavirus (HPV) is the most important etiological factor (zur Hausen, 2002), with high-risk HPV types16 and HPV18 being most prevalent and detected in approximately 70% of cervical cancers (de Sanjose et al., 2010; Rogovskaya et al., 2013). The integration of HPV into the host genome permits the expression of the E6 and E7 viral oncogene proteins that alter the host genome, proteome, and intracellular signalling network of the cervical epithelium which promote oncogenesis.

Most HPV infections of the cervix are asymptomatic and spontaneously cleared by the immune system within six months to two years. However, persistent high-risk HPV infections may cause abnormal changes in epithelial cells of the cervical transformation zone, initiating the development of low- and high-grade squamous intraepithelial lesions (SIL) that may eventually progress to invasive cervical cancer (Muñoz et al., 2003; Woodman et al., 2007).

1.2 Problem statement and rationale of the study

Although the Papanicolaou (Pap) smear was introduced in 1969 by the Ministry of Health Malaysia as a screening tool for detection of precursor lesions and early cervical cancer, the uptake has been dismal with only 22% of eligible women undergoing Pap smear (Yunus et al., 2018). The HPV test has recently been recommended but this is not widely available in Malaysia. There is an urgent need to discover new biomarkers for the early detection of disease. This requires an in-depth understanding of the cellular processes and molecular mechanisms initiating cervical cancer, particularly focusing on how little changes in regulatory genes or proteins can interrupt a variety of cellular functions.

Proteins represent end products that control most of the cellular functions and biological processes in the human body. Sensitive high-throughput technologies using proteomics-based approach is the method of choice in cancer research for diagnostic, prognostic and therapeutic applications. The mass spectrometry (MS) technology is a powerful tool for proteomics research as it allows a complete characterization of proteins and gives a boost to cancer biomarker discovery.

To date, there is no report on the protein profiles in cervical carcinogenesis from the normal cervix to HPV-associated squamous intraepithelial lesion (SIL)

(precancerous) and cervical cancer using formalin-fixed paraffin-embedded (FFPE) tissues**.** The present study focuses on profiling the proteins and elucidating the pathways associated with cervical cancer development and progression as well as the discovery of novel diagnostic biomarkers.

1.3 Objectives of the study

General objective:

To elucidate the molecular mechanisms in cervical carcinogenesis via protein profiling in FFPE cervical tissues

Specific objectives:

1. To determine the protein profiles and significant differentially expressed proteins between normal cervix, HPV-associated squamous intraepithelial lesion (SIL) and squamous cell carcinoma (SCC).
2. To analyse the molecular functions, biological processes, protein interactions and pathways involved in each histological group.
3. To evaluate annexin A2 and cornulin proteins immunohistochemically and their association with clinicopathological parameters.
4. To postulate the molecular mechanisms of proteins implicated in cervical carcinogenesis.

1.4 Flow chart of study

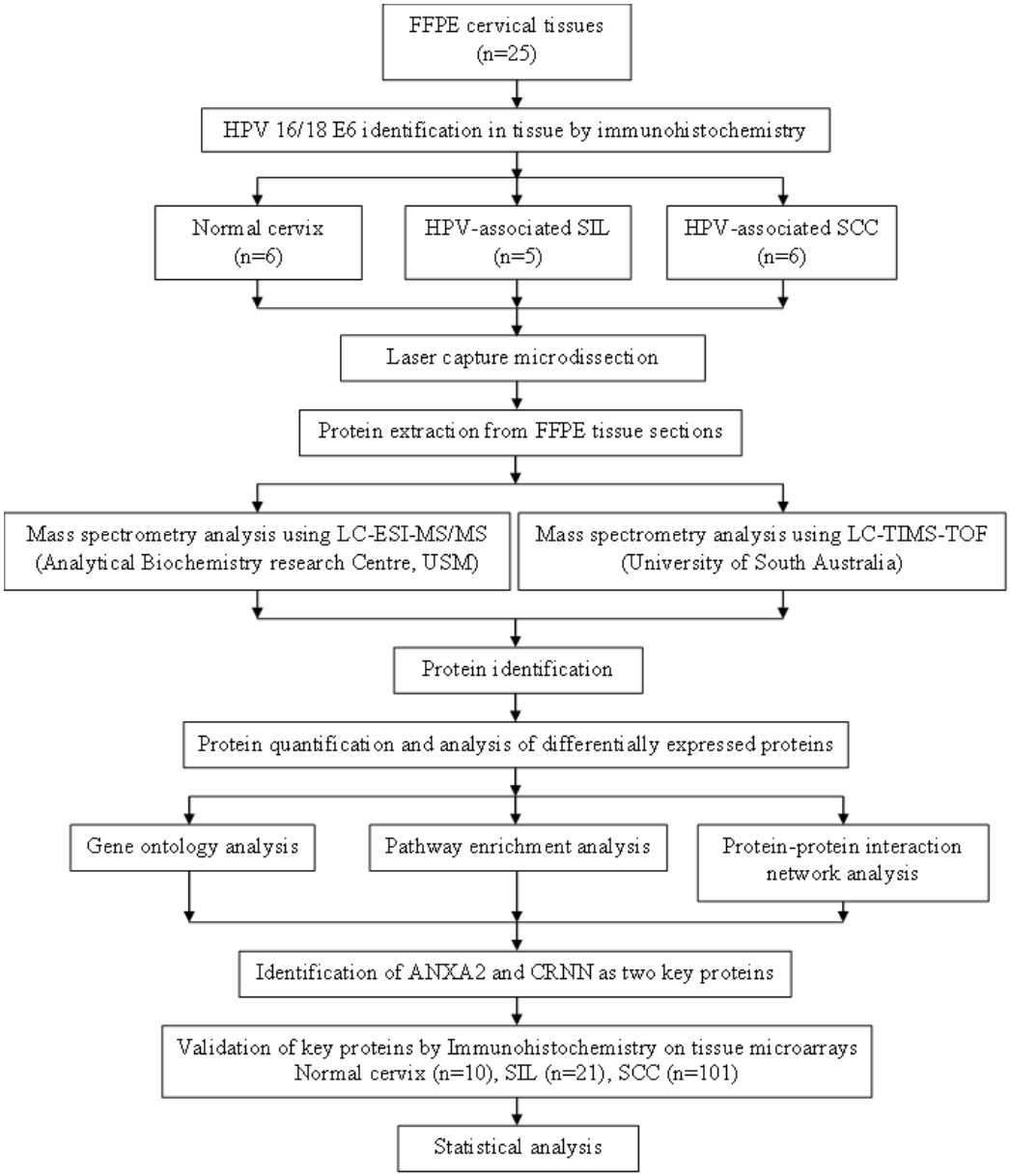

Figure 1.1 Summary of the overall approach of this study

*HPV-human papillomavirus, SIL-squamous intraepithelial lesion, SCC-squamous cell carcinoma, FFPE-formalin-fixed paraffin-embedded, LC-ESI-MS/MS-liquid chromatography electron spray ionisation mass spectrometry, LC-TIMS-TOF-liquid chromatography trapped ion mobility spectrometry time of flight, ANXA2-annexin A2, CRNN-cornulin.

CHAPTER 2

LITERATURE REVIEW

2.1 Cancer overview

Cancer is a leading cause of death, with an estimated 19.3 million new cases and 10 million deaths in 2020 (Sung et al., 2021). Cancer develops when genetic mutation disrupts the mechanism at the cellular level, transforming the normal cells into cancerous cells. The cancerous cells can reproduce indefinitely and invade neighbouring tissues due to the loss of control in normal cellular processes. Cancer involves a multi-stage process caused by various external elements such as radiation, chemicals, tobacco, and infectious organisms, as well as certain internal factors including inherited mutations, immune conditions, hormones, and random mutations (Mathur et al., 2015). Besides, there are many other factors such as diet, infection, obesity, lack of physical activity, and pollution that are known to increase the risk of cancer. These factors may initiate or promote carcinogenesis in the human body.

2.2 Gynecological cancers

Gynaecological cancers form a heterogeneous cluster of tumours originating in the organs of the female reproductive system (Brot & Soares, 2019). Globally, ovarian cancer (OC) is ranked first among gynaecological cancers with the highest incidence and morbidity rate, followed by cervical and endometrial cancer (EC) (Bray et al., 2018). Whereas, vaginal and vulvar cancers are rare (Brot & Soares, 2019). Most of these malignancies cause either vague or no symptoms, leading to late-stage diagnoses and poor patient outcomes. Cervical cancer is the only gynaecological

cancer that may be detected at an early stage through cervical exfoliative cytology (PAP smear) screening programs (Kerkar & Kulkarni, 2006).

2.3 Cervical cancer

Cervical cancer is the third most common gynaecological cancer in developing countries and represents a major global health crisis (Benard et al., 2014; Bray et al., 2018; Moshkovich et al., 2015). Cervical cancer develops in the cells within the cervix. The cervix is lined by stratified squamous epithelial cells that cover the ectocervix, and the mucus-secreting columnar epithelium that lines the endocervical canal. The cervix is illustrated in Figure 2.1. Squamous cell carcinomas are the most common type of ectocervical tumour, representing approximately 75% of all invasive cervical carcinoma cases. Endocervical tumours, on the other hand, are usually adenocarcinomas (Small et al., 2017).

2.4 Histological subtypes of cervical cancer

Cervical cancer can be classified into different subtypes based on its histological features. The most common histological types of cervical cancer are squamous cell carcinoma (SCC), which accounts for approximately 75-90% of cervical cancer, and adenocarcinoma (ADC), which accounts for approximately 10-25% of all cervical cancer cases (K. B. M. Lee et al., 2006). While, adenosquamous carcinoma (ASC) is a rare histological type that is composed of both glandular and squamous elements (Jung et al., 2017).

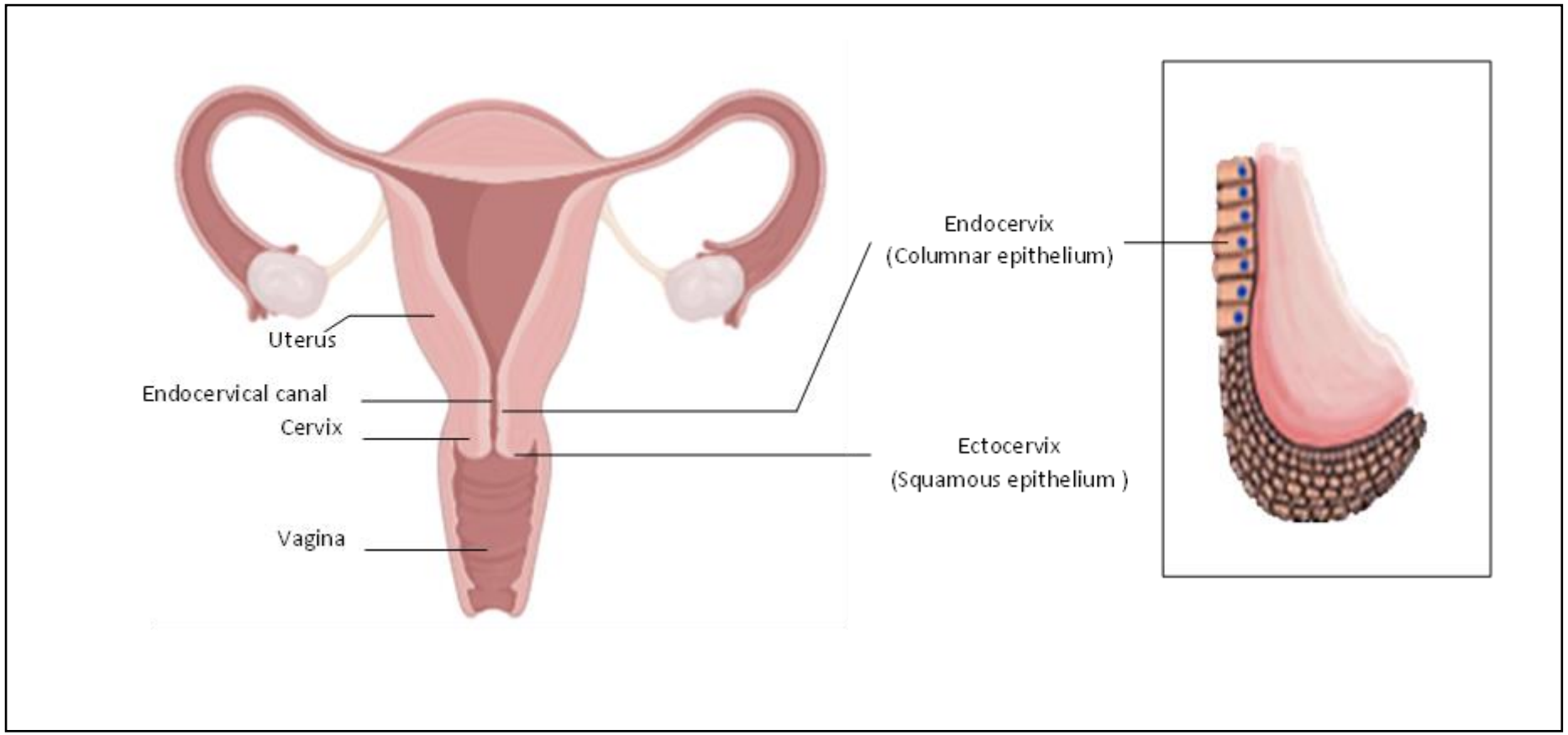

Figure 2.1 Schematic diagram of the cervix. The endocervix is the inner lining of the cervix that leads into the uterus. The ectocervix is the cervix's outer layer that connects to the vagina. The transformation zone is where the squamous cells connect to the columnar cells in the cervix.

Adapted from: Gynaecologic Pathology: An Atlas of Essential Pathology for Gynaecologists (Ivan et al., 2006).

2.5 Epidemiology of cervical cancer

Cervical cancer is one of the most common cancers among women (Mattiuzzi & Lippi, 2020). Over the years, the proportion of women under 35 years old diagnosed with cervical cancer has increased from 10% to 40% (Song et al., 2017). It is the most common malignancy among female patients in developing countries, with 452,000 new cases diagnosed in 2008 and ranked ninth in developed countries with 77,000 new cases (S. Zhang et al., 2020). Approximately 90% of cervical cancer deaths occur in developing or underdeveloped countries, and the mortality rate in low- and middle-income countries is 18 times greater than in wealthy countries (WHO, 2018). Cervical cancer incidence rates vary greatly between populations, ranging from 3.6 per 100,000 women in Switzerland to 75.9 per 100,000 in Malawi (over a 20-fold difference). High incidences are often observed among populations in Melanesia, Latin America and the Caribbean, South-Central Asia, and sub-Saharan Africa. Conversely, in developed countries such as Europe, North America, Australia/New Zealand, the Middle East, China, and Japan, the incidence rate is often low (Memon & Bannister, 2019). In Malaysia, cervical cancer is one of the common cancers, affecting approximately 12.9 % of women, with an age-standardised incidence rate of 19.7 per 100,000 females. Several risk factors can increase the likelihood of developing cervical cancer (Mustafa et al., 2022)

2.6 Risk factors for cervical cancer

Human papillomavirus (HPV) is the most important etiological factor for cervical cancer. Almost all cervical cancer cases are caused by HPV infection with high-risk subtypes 16 and 18 (Cohen, Jhingran, Oaknin, & Denny, 2019). The development of invasive cancer can take up to 20 years from the precursor lesion

initiated by sexually transmitted HPV (Y. Yuan et al., 2021). Having multiple sexual partners and sexual intercourse at a young age have shown to increase the risk of cervical cancer. When compared to controls, having more than six lifetime sexual partners significantly increases the relative risk of cervical cancer. Intercourse before the age of 20 confers a significantly increased risk of developing cervical cancer compared to intercourse after the age of 21 (Berrington De González & Green, 2007). Smoking is another risk factor for cervical cancer. It is reported that current and former smokers with HPV infection have two to three fold prevalence of high grade squamous intraepithelial lesion (HSIL) and invasive cancer (Lea & Lin, 2012). A poor socioeconomic level can also lead to the development of cervical cancer. People who are living in underdeveloped countries have limited access to cervical cancer screening and are lack of awareness, resulting in increased number of cases (Ghebre et al., 2017; Roura et al., 2014). For example, in Malaysia, the Pap smear test is provided free of charge at primary healthcare facilities throughout the country. Despite the free services, Malaysian women's Pap smear test uptake is low (Abdullah et al., 2011; Al-Naggar et al., 2010; Gan & Dahlui, 2013).

2.7 HPV-associated cervical cancer development

There are approximately 30 HPV types that primarily infect the cervix, vagina, vulva, penis, and anus (Okunade, 2020). Based on their association with cervical cancer and its precursor lesions, HPV types are classified as high-risk or low-risk. Types 6, 11, 42, 43, and 44 are low-risk HPV types, while types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82 are high-risk HPV types (Walboomers et al., 1999). Infections with high-risk HPV 16 and 18 are the primary cause of pre-cancerous and malignant cervical lesions. HPV infection is generally transmitted via sexual contact and 90% of the infections are cleared spontaneously due to immunological

responses in the host. However, 10% of infections may continue to persist, resulting in the formation of squamous intraepithelial lesion (SIL) (S. Zhang et al., 2020). SIL can be graded into low-grade squamous intraepithelial lesion (LSIL), and high-grade intraepithelial lesion (HSIL) (Chan et al., 2019). LSIL usually represents the infection stage rather than the stage of disease development, thus it does not always signify the disease progression. On the other hand, HSIL can potentially progress and develop invasive cervical cancer (Chan et al., 2019). According to statistics, more than 70-80% of LSIL spontaneously regress or become undetectable without treatment (Moscicki et al., 2010). The HSIL is a precancerous state with a 0.2% to 4% chance of progressing to aggressive cancer within a year (Gravitt et al., 2013; Moscicki et al., 2010). Over a 30-year period 30% of untreated HSIL may progress into invasive cancer, whereas treated HSIL has a 1% chance of becoming invasive (Gravitt et al., 2013; McCredie et al., 2008; Moscicki et al., 2010; Stanley, 2010). it typically takes 20 years for the HPV infection to develop into cervical cancer (Rodriguez et al., 2008). High-risk HPV DNA was found in 99.7% of cervical cancer specimens (Walboomers et al., 1999). Figure 2.2 illustrates the stages in the development of cervical cancer.

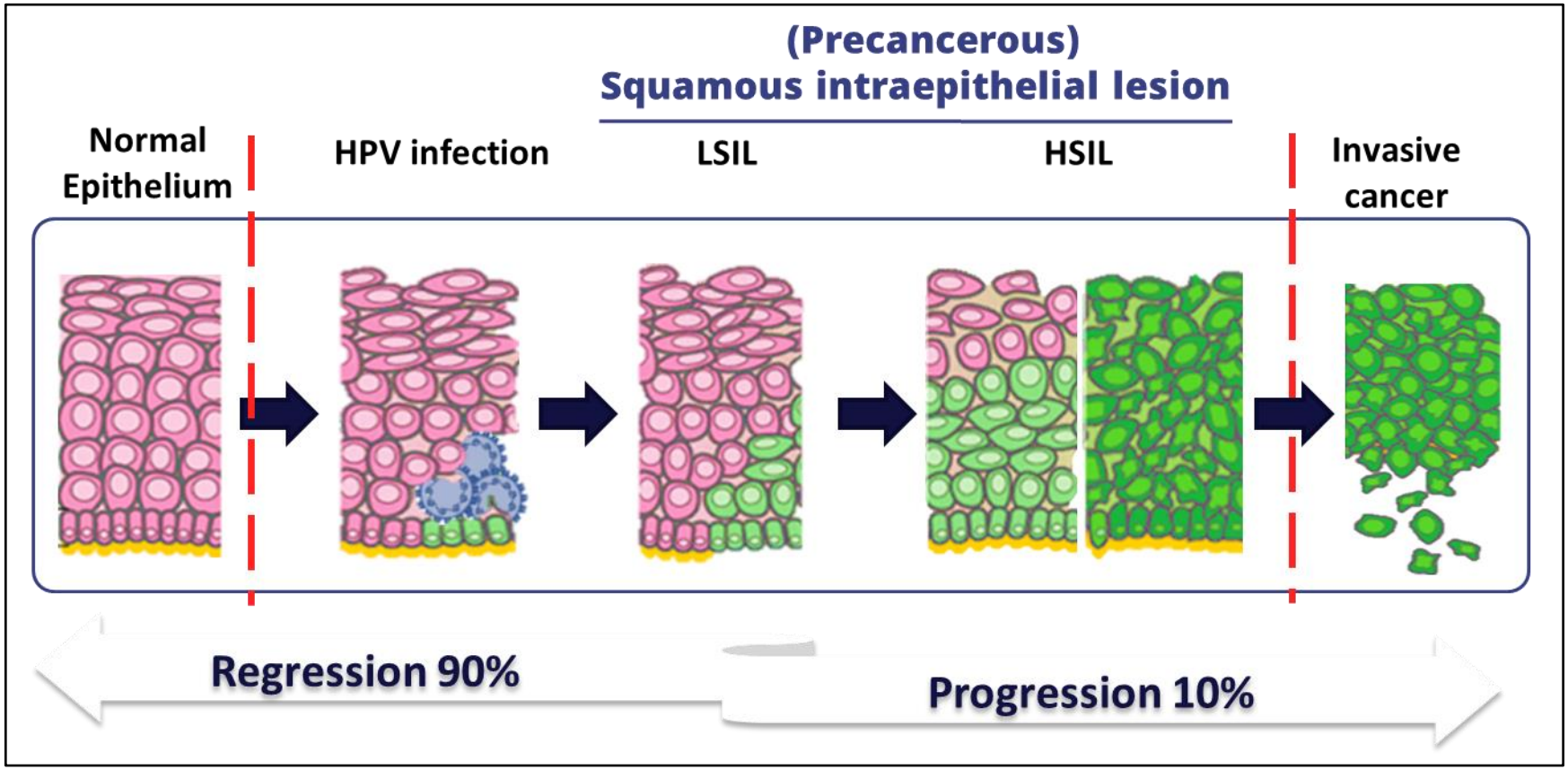

Figure 2.2 Schematic diagram showing stages in the progression of cervical cancer. Schematic diagram showing stages in the development of cervical cancer. The normal epithelium develops into squamous intraepithelial lesion after HPV infection. 10 % of the SIL can progress to invasive cancer, whereas 90 % of the infection clear spontaneously. Adapted from: https://teachmeobgyn.com/gynaecology/cervix/cervical-cancer/

- Squamous cells; - Epithelial basement membrane; - HPV virus; - transformed cells; - tumour cells

2.8 Cervical carcinogenesis and Human Papilloma Virus (HPV)

Cervical carcinogenesis is a multistage process characterised by the accumulation of DNA changes in host cell genes. These alterations include both epigenetic and genetic modification in oncogenes and tumour suppressor genes, which are important regulators of cell cycle progression, chromosomal stability, telomere activation, and apoptosis (Gupta et al., 2018). However, the integration of the viral genome into the host appears to be a crucial stage in the onset of tumorigenesis. The section below delves into details of HPV and its effect on cellular and epigenetic changes.

2.8.1 HPV classification

The Papillomaviridae family comprises 39 genera where HPVs are found in five of these genera known as alpha papillomaviruses, beta papillomaviruses, gamma papillomaviruses, mupapillomaviruses, and nupapillomaviruses. The HPVs can be classified according to three groups where group 1 carcinogens: carcinogenic to humans, Group 2A: probably carcinogenic to humans, and group 2B: possibly carcinogenic to humans (Xuelian Wang et al., 2018). The HPVs categorized in group-1 include HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, and HPV59. HPV68 is considered as high-risk type, thus categorized under group 2A. Approximately 96% of cervical malignancies are associated with one of the 13 HPVs in groups 1 and 2A (Arbyn et al., 2014). The group 2B HPVs which includes HPV26, HPV30, HPV34, HPV53, HPV66, HPV67, HPV 69, HPV70, HPV73, HPV82, HPV85, and HPV97 are associated with rare cases of cervical cancer. It is difficult to determine the carcinogenicity of group 2B HPVs

because fewer incidences of cervical cancer are linked to them (Xuelian Wang et al., 2018).

2.8.2 HPV genome

The papillomavirus genome is made up of a small double-stranded and highly conserved DNA which comprises approximately 8000 base pairs. The HPV genome consists of three regions, encoding eight genes, categorized as long control region (LCR) early region (E) and late region (L) based on their expression patterns (Okunade, 2020; Oyervides-Muñoz et al., 2018). The early region of 4000 base pairs encodes three regulatory genes (E1, E2 and E4) and three oncogenes (E5, E6 and E7) involved in viral replication and cell transformation. The second region with 3000 base pairs encodes two genes L1 and L2, that make up the viral capsid. Another 1000 base pairs consist of a long control region (LCR) which controls viral DNA replication and transcriptional regulatory elements (Chan et al., 2019; Jing et al., 2018). The genome of high-risk HPV is illustrated in Figure 2.3 and the functions of the proteins encoding the genes are listed in table 2.1. The functions of the proteins are explained in detail in subsection 2.6.3. The HPV serotypes are genetically distinct from one another, and the conventional classification system involves a whole genome in which the L1 nucleotide sequence differs from another HPV genome by at least 10%. The HPV types are classified numerically and chronologically based on the date of discovery (Cubie et al., 2011).

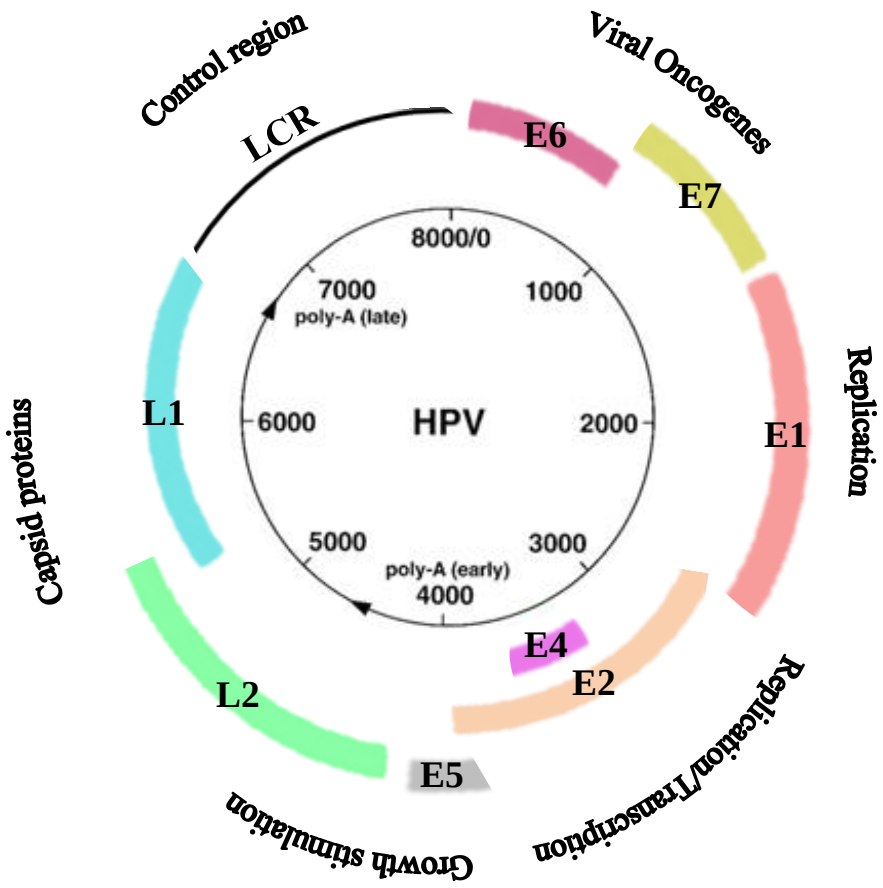

Figure 2.3 Human papillomavirus genome organisation. Schematic diagram illustrating the high-risk HPV genome showing the locations of the early (E1, E2, E4, E5, E6 and E7), late (L1 and L2), and long control region (LCR). Many early genes are involved in viral replication, transcriptional regulation, genome maintenance, and immune system evasion.

Adapted from (Antonishyn, 2008; Berg, 2010; Okunade, 2020)

Table 2.1 HPV proteins and their functions

Proteins	**Functions**
E1	Replication of viral genome and its maintenance
E2	Initiation of viral DNA replication Regulates transcription of E6 and E7
E4	Release of viral particles
E5	Enhance growth factor signalling pathways
E6	Inhibits p53 and causes loss in cell cycle regulation
E7	pRb mediated deregulation of the cell cycle
L1	Major Capsid protein
L2	Major capsid protein

2.8.3 HPV life cycle

Cervical carcinogenesis is closely associated with the events that occur during the HPV life cycle. Cervical cancer begins at the transformation zone, where continuous metaplastic changes take place between the squamous epithelium of the ectocervix and the columnar epithelium of the endocervix (Okunade, 2020). The basal layer cells in the ectocervical stratified squamous epithelium operate as stem cells and divide as they substitute the cells that normally slough off from the surface layer. When a basal cell splits via mitosis, two daughter cells are formed: one rises and becomes a terminally differentiated cell, while the other remains in the basal layer to maintain the pool of dividing cells. The virus's first targets are basal cells, which are susceptible due to micro-wounds. HPV virions enter cells via connecting with specific receptors, such as the alpha-6 integrin, which binds HPV-16. The early promoter is activated once the basal cells are exposed to infection, resulting in the expression of viral helicase E1. The E1 viral helicase along with E2 facilitates a transient cycle of replication known as "establishment replication," causing rapid amplification of 50-100 viral genome copies per cell (Anacker & Moody, 2017; McBride, 2017). Viral genomes are then maintained at low copy numbers in these undifferentiated cells by reproducing alongside cellular DNA. As infected basal cells proliferate, viral DNA is partitioned between daughter cells, one of which migrates out from the basal layer and begins terminal differentiation.

The productive phase of the viral life cycle is triggered by epithelial differentiation, resulting in the activation of the late promoter and the expression of late viral genes (E4, E5, L1, L2), as well as high levels of E1 and E2, which drive viral genome amplification to thousands of copies per cell. Because the immunogenic

capsid proteins L1 and L2 are exclusively expressed in highly differentiated suprabasal cells, virion assembly and release are confined to the epithelium's top layer (Narisawa-Saito & Kiyono, 2007). While normal epithelial cells exit the cell cycle following differentiation, the E6 and E7 proteins deactivate normal cell cycle checkpoints to force developing cells back into the cell cycle, which results in a G2 environment that provides cellular factors essential for productive viral replication (Moody, 2017). Persistent infection with high-risk human papillomavirus types 16, 18, 45, 31, 33, 52, 58, and 35 have been well described as a cause of cervical cancer (Muñoz, 2003). Remarkably, several studies have found that HPV16 is more common in squamous cell carcinomas, and HPV18 and HPV45 are more common in adenocarcinomas (Hadžisejdić et al., 2006; Raub et al., 2014; Tjalma et al., 2015).

2.8.4 HPV-related cervical carcinogenesis

Cervical carcinogenesis is usually associated with high-risk HPV infection. When HPV infection occurs, the DNA becomes mutated, resulting in viral DNA integration. The viruses can evade cellular and immune defence mechanisms while promoting cell proliferation and preventing cellular apoptosis (Chan et al., 2019). HPV16 oncogenic potential is regulated by viral transcriptional factors. The HPV16 genome can first manifest as an unintegrated tiny DNA molecule known as an episome, which leads to benign and precancerous cervix lesions. However, HPV16 has the ability to integrate its genome into the host genome, resulting in high-grade squamous intraepithelial lesion (HSIL) and cervical carcinoma (Lehoux et al., 2009). When the viral genome integrates with the host DNA genome, the E6 and E7 oncoproteins are upregulated and subsequently deregulate key proteins in cellular signalling pathways, inhibiting two essential tumour suppressor proteins, p53 and pRb (Oyervides-Muñoz et al., 2018). The E6 and E7 are small proteins with 150 and 100 amino acids,

respectively, that has no known enzymatic activity but can alter host cell activity when attached to cellular proteins (Chan et al., 2019).

The binding of E6 to p53 is regulated by E6-associated protein (E6AP), an E3 ubiquitin-protein ligase, forming a trimeric complex E6/E6AP/p53 (Xuelian Wang et al., 2018). This binding causes p53 degradation and leads to cell proliferation. E7, on the other hand, binds to pRb, causing it to be inactivated and degraded (Chan et al., 2019). The inactivation of pRb by phosphorylation resulted in uncontrolled cell proliferation or cell death (Gordon & Du, 2011). However, not all integrations are dependent on the expression of the E6 and E7 oncogenes (Groves & Coleman, 2015). Several studies reported that cervical cancer has driver mutations in proteins such as phosphatidylinositide 3-kinases catalytic subunit (PIK3CA), a critical protein in the PI3K pathway, Kirsten rat sarcoma viral oncogene homolog (KRAS), and epidermal growth factor receptor (EGFR) (Wright et al., 2013).

2.9 Cervical cancer screening

Cervical cancer has a long preclinical stage that can last decades without causing symptoms in women. An effective screening test is important to detect precancerous lesions before they progress to invasive cancer. Cervical cancer screening procedures include the standard Papanicolaou or Pap smear, visual inspection with acetic acid and Lugol's iodine (VIA/VILI), liquid-based cytology (LBC), and HPV testing. Since the 1950s, Pap smear has dramatically lowered the disease burden of cervical cancer in developed countries, particularly in the United States. However, the traditional Pap smear accuracy can be easily influenced by factors such as cytological room level, sample procedure, quality of slide, staining skills, competent technicians, and cytological personnel experience. The sensitivity of

cytology can be as high as 80%-90% in developed countries with high standard experimental settings and technical levels, whereas it can be as low as 30%-40% in middle and low-income countries (S. Zhang et al., 2020). In Malaysia, Pap smear coverage remains around 22% of estimated eligible women (Yunus et al., 2018).

To overcome the drawbacks of conventional Pap smear in cervical cancer screening, the liquid-based cytology was developed and approved by the Food and Drug Administration (FDA). In contrast to the traditional Papanicolaou smear, which involves smearing a cervical sample onto a slide and then applying a fixative, liquid-based cytology involves sampling and cell transfer to a liquid media, with automated processing. Despite several theoretical advantages of liquid-based cytology, such as improved cell collection and preparation, blood and debris filtering, and fewer unsatisfactory results, some studies show no significant difference in sensitivity or specificity for the detection of SIL when compared to the conventional Papanicolaou smear. Instead, the advantages of liquid-based cytology are limited to the utilization of a single specimen for concurrent HPV, gonorrhoea, and chlamydial infection testing in addition to cytology (Bedell et al., 2020). As a result, the American College of Obstetricians and Gynaecologists stated that both procedures are acceptable for cervical cancer screening (Martin-Hirsch & Wood, 2011).

The addition of HPV testing to cervical cytology is one of the most recent improvements to cervical cancer screening standards. The detection of high-risk HPV in cervical lesion biopsies and exfoliated cells has shifted towards polymerase chain reaction (PCR)-based systems (Harari et al., 2014) and, most recently, next-generation sequencing (NGS) assays (Conway et al., 2012). In 2014, the use of HPV in cervical screening was approved by FDA. Thereafter, HPV detection became important in the practice of cervical cancer screening (Bedell et al., 2020).

With the advancement in science and technology, artificial intelligence (AI)-based products have been widely used in the health care system. AI has also shown promise in cytology-based screening and colposcopy examinations based on picture pattern recognition in cervical cancer prevention and management (Bao, Bi, et al., 2020; Bao, Sun, et al., 2020). These AI-based technologies or systems can intelligently identify lesions and assist medical personnel in clinical examination and diagnosis, thereby alleviating diagnostic challenges in primary care clinics.

2.10 Diagnosis

Precancerous lesions of the cervix normally do not cause pain or any other symptoms and are undetectable unless a woman undergoes screening test. The symptoms only appear when the abnormal cervical cells become cancerous and start to invade the surrounding tissue (Mishra et al., 2011). At an early-stage cervical cancer, the symptoms are often asymptomatic, however, watery vaginal secretions and bleeding or intermittent drip bleeding after sexual intercourse may occurs at times. Meanwhile, invasive cervical cancer may have foul vaginal discharges, abnormal vaginal bleeding, or pelvic pain (J. L. Wang et al., 2013).

Cervical cancer symptoms necessitate a pelvic examination, visualization of the cervix and vaginal mucosa, and cervical cytology. Colposcopy and biopsy are performed on symptomatic patients with cytology indicative of invasion but no visible lesion (Cohen, Jhingran, Oaknin, Denny, et al., 2019). A cone biopsy is necessary if malignancy is suspected clinically or through cervical cytology but not proven by histopathological evaluation of cervical biopsies. The clinical staging of cervical cancer is based on tumour size and degree of pelvic extension (Pecorelli, 2009). According to International Federation of Gynaecology and Obstetrics (FIGO)

recommendations, staging involves physical examination, endoscopic procedures, and imaging modalities such as computed tomography, magnetic resonance imaging and positron emission tomography (PET), (Cohen, Jhingran, Oaknin, Denny, et al., 2019).

Clinical staging of cervical cancer is based on tumour size and the degree of pelvic extension. When tumour dimensions can be determined histologically, stage is assigned to microscopic lesions. The stage should always be assigned at the time of diagnosis and should never be changed. Staging of cervical tumours according to the International Federation of Gynaecology and Obstetrics (FIGO) and is listed in Table 2.2 (Pecorelli, 2009).

Table 2.2 Staging of cervical tumours according to the International Federation of Gynaecology and Obstetrics (FIGO)

Description	FIGO staging
Cervical carcinoma confined to the cervix (without extension to uterine corpus)	I
Invasive carcinoma diagnosed only by microscopy, stromal invasion with a maximum depth of 5·0 mm measured from the base of the epithelium, and horizontal spread of 7·0 mm or less; vascular space involvement, venous or lymphatic, does not affect classification	IA
Measured stromal invasion no greater than 3·0 mm and lateral spread no greater than 7·0 mm	IA1
Measured stromal invasion greater than 3·0 mm and no greater than 5·0 mm, and horizontal spread no greater than 7·0 mm	IA2
Clinically visible lesion confined to the cervix or microscopic lesion greater than IA2	IB
Clinically visible lesion no greater than 4·0 cm in greatest dimension	IB1
Clinically visible lesion greater than 4·0 cm in greatest dimension	IB2
Cervical carcinoma invades beyond the uterus but not the pelvic wall or lower third of vagina	II
Tumour without parametrial invasion	IIA
Clinically visible lesion no greater than 4·0 cm in greatest dimension	IIA1
Clinically visible lesion greater than 4·0 cm in greatest dimension	IIA2
Tumour with parametrial invasion	IIB
Tumour extends to pelvic wall, involves lower third of vagina, causes hydronephrosis, or a combination of all symptoms, or non-functioning kidney	III
Tumour involves lower third of vagina, without extending to the pelvic wall	IIIA
Tumour extends to pelvic wall, causes hydronephrosis or non-functioning kidney, or both	IIIB
Tumour invades mucosa of bladder or rectum, extends beyond the true pelvis, or both	IV
Tumour invades mucosa of bladder or rectum	IVA
Tumour extends beyond the true pelvis	IVB

2.11 Treatment

The proper treatment for cervical cancer is based on the International Federation of Gynaecology and Obstetrics (FIGO) clinical staging guidelines. With screening programs, cervical precancerous lesions can be discovered and treated early using low-cost technologies such as cryotherapy, loop electrosurgical excision technique, or thermocoagulation (Maza et al., 2017). Unfortunately, due to the prevalence of advanced stages of cervical cancer in low and middle-income countries, these efficient low-cost approaches are insufficient in a setting where 80% of women with precancerous lesions are untreated (Gage et al., 2003).

Mostly, early cervical malignancies are treated surgically, with procedures such as cervical conization, radical trachelectomy, simple or radical hysterectomy, and pelvic lymphadenectomy (Hill, 2020). Stage IA1 cervical cancer is usually treated using cone biopsy in a basic setting, whereas in a limited setting, both cone biopsies and pelvic lymphadenopathy are recommended. In enhanced settings, radical trachelectomy is preferred for stage IB1 patients desiring future fertility (Vu et al., 2018).

Radiation and chemoradiation are the standard treatment for advanced cervical cancer. The recent advancement in radiation therapy for the treatment of bulky cervical cancer includes the use of image-guided radiation rather than intracavitary vaginal brachytherapy (Hill, 2020). In contrast to external beam radiotherapy, brachytherapy uses a radiation source that is inserted into the uterus and vagina, allowing a larger dose of radiation to reach the cervix without damaging nearby tissues (Cohen, Jhingran, Oaknin, Denny, et al., 2019). Advanced therapies rely on a multidisciplinary strategy

that involves expertise such as medical oncologists, gynaecologic oncologists, radiation oncologists, radiologists, and nursing specialists (FIGO, 2009).

2.12 Formalin-fixed paraffin-embedded (FFPE) tissues

Human tissue specimens are the most important material for translational clinical research, such as biomarker discovery and validation as well as the investigation of molecular disease pathways (Thompson et al., 2013). For more than a century, histopathology has been essential for diagnostics, prognostics, and therapeutics. Pathologists use histological or immunohistochemical stains to provide specific diagnoses on slides derived from fresh-frozen or formalin-fixed paraffin-embedded (FFPE) tissues (Hussen et al., 2022; Rosai & Ackerman, 1979). Fresh tissue samples are much easier to process compared to formalin-fixed paraffin-embedded (FFPE) tissue samples. FFPE is an economical archival option because tissue can be stored, maintained and sectioned in high density at room temperature (RT) for years or decades while maintaining integrity for pathology analysis (Grillo et al., 2017).

Although FFPE tissues provide invaluable resource material for proteomics studies and biomedical studies, sample preparation requires rather extensive protocols for adequate and good quality protein extraction before subjected to proteomic analysis. In addition, cross-links between proteins and chemical modifications occur as a result of formalin fixation, affecting the yield and quality of protein. However, this process can be reversed using heat-induced antigen retrieval (Buczak et al., 2020; Gustafsson et al., 2015). The establishment of heat-induced antigen retrieval by Shi et al. enables mass spectrometry-based proteomic studies of FFPE tissues with efficient protein extraction and peptide identification (Shi et al., 1991). Poorly soluble proteins, such as integral membrane proteins, necessitate the use of strong denaturants