ROLES OF SELECTED GENETIC VARIATIONS AND MOLECULAR ALTERATIONS IN HUMAN PAPILLOMAVIRUS-MEDIATED CANCER OF UTERINE CERVIX

TAN SHING CHENG

UNIVERSITI SAINS MALAYSIA

2016

ROLES OF SELECTED GENETIC VARIATIONS AND MOLECULAR ALTERATIONS IN HUMAN PAPILLOMAVIRUS -MEDIATED CANCER OF UTERINE CERVIX

by

TAN SHING CHENG

Thesis submitted in fulfillment of the requirements

for the degree of

Doctor of Philosophy

September 2016

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest appreciation and sincere gratitude to my Ph.D. supervisor, Prof. Dr. Ravindran Ankathil, for his continuous advice, support and encouragement all this while. It has been an honor for me to work under his supervision for my Ph.D. study. His selfless guidance and invaluable comments have been a great help for me during my study period. I am also indebted to my co-supervisors, Prof. Dr. Nor Hayati Othman and Assoc. Prof. Dr. Mohd. Pazudin Ismail, for their enlightenment, assistance and motivation during this period.

I wish to express my cordial appreciation to Dr. Daniel Roza Duski (Hospital Sultan Ismail, Johor Bahru), Dr. Biswa Mohan Biswal, Dr. Wan Fadhlina Wan Adnan and Dr. Norhayati Che Soh (Hospital Universiti Sains Malaysia, Kubang Kerian), and Dr. Nik Ahmad Nik Abdullah (Hospital Raja Perempuan Zainab II, Kota Bharu) for their help in subject recruitment and sample collection. I also wish to thank Dr. Norul Badriah Hassan (School of Medical Sciences, Universiti Sains Malaysia) for her advice and guidance on questionnaire construction and validation. In addition, I am also extremely grateful to Dr. Lim Poh Ying (School of Health Sciences, Universiti Sains Malaysia) for offering me invaluable consultations, in the capacity of a friend, on statistical analyses related to my research study. I would also like to thank Dr. Lee Kui Soon (Department of Molecular Biology, Universiti Malaysia Sarawak) for his advice and helpful discussion on molecular techniques and troubleshooting methods.

I wish to extend my gratitude to all members of Human Genome Centre, for creating a conducive environment for me to carry out the thesis work comfortably in the past five years. My special thanks go to the Head of Department (Dr. Sarina Sulong), lecturers (Prof. Dr. Gan Siew Hua, Dr. Tan Huay Lin, Dr. Teguh Haryo Sasongko and Dr. Nazihah Mohd. Yunus), and science officers (Mr. Mohamad Ros Sidek, Mdm. Nor Atifah Mohd Adam and Mr. Chia Boon Hock), for providing me all the necessary facilities and technical assistance for my research study. I would also like to thank Dr. Marjanu Hikmah Elias, Dr. Nizam Zahary, Mr. Ahmad Aizat Abdul Aziz, Ms. Yam Yan Yan, Mr. Au Zian Liang, Ms. Najlaa Maddin, Ms. Maziras Makhtar, Mr. Aminudin Mustapha, Ms. Fazreen Nasir and Ms. Sathiya Maran, for the stimulating discussion and knowledge sharing throughout my study period.

I would like to acknowledge research funding from Fundamental Research Grant Scheme of the Ministry of Higher Education (No. 203/PPSP/6171167) and Research University Grant of Universiti Sains Malaysia (No. 1001/PPSP/812109), without which it would have been impossible for me to conduct the research work. I am also grateful to the Ministry of Higher Education for awarding me the MyBrain15 MyPhD scholarship from February 2012 to January 2016.

Besides, I wish to thank all my beloved friends, especially Dr. Xiao Li, Dr. Teoh Wei Ming, Dr. Leong Kai Sing, Dr. Chang Kah Haw, Dr. Chen Seong Ting, Dr. Leow Chiuan Yee, Dr. Chuah Candy, Dr. Nicole Seng, Dr. Om Parkash, Mr. Kho Swee Liang, Ms. Cheung Tian Pei, Ms. Ong Mei Gee, Ms. Wong Jia Hui,

Ms. Guo Yiqiang, Ms. Tan Han Yann, Mr. Tee Jong Huat, Ms. Ng Sze Han and Mr. Nur Waliyuddin Hanis, for all the fun that we have had together in the past five years.

No words can express my appreciation to my family members for their constant love and support. I am indebted to my parents, Mr. Tan Tau Kai and Mdm. Chong Ah Moi, for their never-ending faith and inspiration. I am also grateful to my siblings, Mr. Tan Shin Uei, Ms. Tan Shing Tai, Mr. Tan Shing Long and Mr. Tan Shing Yong as well as their respective spouses, for giving me the strength and courage to carry on. Finally, I am thankful to my fiancée, Ms. Fang Jie, for her trust, understanding and endless encouragement.

Last but not least, I wish to convey my heartfelt gratitude to everyone who helped me towards the completion of my thesis, whom I failed to mention by name.

TAN SHING CHENG

September 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

%	Percent
®	Registered trademark
°C	Degree Celsius
AC	Amplification control
ADC	Adenocarcinoma
AGCC	Affymetrix GeneChip Command Console
ANOVA	Analysis of variance
aUPD	Acquired uniparental disomy
bp	Base pair
CCND1	Cyclin D1
CCR2	Chemokine (C-C motif) receptor 2
CDK	Cyclin-dependent kinase
CDKN1A	Cyclin-dependent kinase inhibitor 1A
cDNA	Complementary DNA
CDR	Complementarity-determining region
ChAS	Chromosome Analysis Suite
CI	Confidence interval
ChIP	Chromatin immunoprecipitation
CIN	Cervical intraepithelial neoplasia
CN	Copy number
CNA	Copy number alteration

CNLOH	Copy-neutral loss of heterozygosity
CNV	Copy number variation
C_q	Quantification cycle
CTLA4	Cytotoxic T-lymphocyte antigen 4
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
E2F1	E2F transcription factor 1
FAS	Fas receptor
FIGO	International Federation of Gynecology and Obstetrics
HBOC	Hereditary breast and ovarian cancer syndrome
HC	Hybridization control
HNPCC	Hereditary non-polyposis colorectal cancer
HPV	Human papillomavirus
HSD	Honest significant difference
HSIL	High-grade squamous intraepithelial lesion
IBM	International Business Machines Corporation
IL4R	Interleukin-4 receptor
kb	Kilobase
kDa	Kilo Dalton
LCR	Long control region
LOH	Loss of heterozygosity
LSIL	Low-grade squamous intraepithelial lesion
MCP-1	Monocyte chemoattractant protein-1

MDM2	Murine double minute 2 homolog
$MgCl_2$	Magnesium chloride
ml	Milliliter
mm	Millimeter
mM	Millimolar
mRNA	Messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
ng	Nanogram
OR	Odds ratio
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pRb	Retinoblastoma protein
r^2	Coefficient of determination
REST	Relative Expression Software Tool
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
RT-qPCR	Reverse transcription-quantitative real-time PCR
SCC	Squamous cell carcinoma
siRNA	Small interfering ribonucleic acid
SNP	Single nucleotide polymorphism
snRNA	Small nuclear ribonucleic acid

SPSS	Statistical Package for the Social Sciences
STAT-1	Signal transducer and activator of transcription 1
TACE	Tumor necrosis factor-alpha converting enzyme
TNF	Tumor necrosis factor
TP53	Tumor protein 53
tRNA	Transfer ribonucleic acid
™	Trademark
V	Voltage
µg	Microgram
µl	Microliter
µM	Micromolar

PERANAN VARIASI-VARIASI GENETIK DAN PERUBAHAN-PERUBAHAN MOLEKULAR TERPILIH DI DALAM KANSER PANGKAL RAHIM YANG DICETUSKAN OLEH VIRUS PAPILOMA MANUSIA

ABSTRAK

Walaupun virus papilloma manusia (HPV) adalah faktor yang wajib bagi mencetuskan pembentukan kanser pangkal rahim, ia secara bersendirian adalah tidak cukup. Maklumat berkenaan peranan variasi-variasi genetik serta perubahan-perubahan molekular terpilih (dalam bentuk ekspresi gen, perubahan nombor salinan DNA dan kehilangan heterozigositi) di dalam kanser pangkal Rahim adalah tidak banyak diketahui dalam kalangan wanita di Malaysia pada masa ini. Maka, kajian ini dijalankan untuk mengisi jurang kesusasteraan ini. Bagi tujuan itu, hubung kait antara 12 polimorfisme genetik dalam sembilan gen yang terlibat dalam pelbagai rangkaian karsinogenik dengan risiko kanser pangkal rahim telah dinilai dalam 185 orang pesakit kanser pangkal rahim yang disahkan secara histopatologi dan 209 orang wanita kawalan tanpa kanser. Tahap ekspresi mRNA untuk sembilan gen tersebut, serta perubahan nombor salinan dan ketidakseimbangan alel keseluruhan genom, juga telah dikaji dalam 72 orang daripada jumlah subjek yang terlibat. Seterusnya, semua data yang diperolehi telah dibahagikan mengikut kumpulan etnik and subjenis HPV yang terlibat. Selain itu, data berkenaan perubahan-perubahan molekular tumor juga telah dibahagikan mengikut peringkat dan histopatologi kanser. Tambahan pula, hubung kait antara polimorfisme-polimorfisme genetik dengan tahap ekspresi gen-gen yang berkenaan juga telah dinilai. Hasil kajian ini menunjukkan bahawa 154 (83.2%) pesakit kanser pangkal rahim dan tiada (0.0%)

wanita kawalan tanpa kanser adalah positif untuk jangkitan HPV. Antara subjek yang positif untuk jangkitan HPV, 141 (91.6%) dijangkiti oleh sejenis HPV sahaja (kebanyakannya HPV16, HPV18 dan HPV58) dan 13 (8.4%) yang selebihnya dijangkiti oleh dua jenis HPV pada masa yang sama (di mana HPV16 + HPV18 merupakan jangkitan yang paling biasa dijumpai). Hasil kajian ini juga menunjukkan bahawa selepas pelarasan terhadap faktor-faktor risiko kanser pangkal rahim yang diketahui, polimorfisme *TNF* c.-488G>A, *CTLA4* c.49A>G, *FAS* c.-671A>G, *MDM2* c.14+309T>G and *CCND1* c.723G>A menunjukkan hubungan yang signifikan secara statistik dengan risiko kanser pangkal rahim apabila dianalisis secara bersendirian. Selain itu, pemerhatian yang signifikan secara statistic juga telah dijumpai bagi beberapa gabungan antara polimorfisme *CTLA4* c.-319C>T dengan polimorfisme *TNF* c.-418G>A, *CDKN1A* c.93C>A dan *CCND1* c.*687C>G. Di samping itu, kajian ini juga menunjukkan bahawa *PUM1*, *YWHAZ* dan *RPLP0* merupakan gen-gen rujukan yang diekspres dengan paling stabil di dalam kanser pangkal rahim, dan pelarasan tahap ekspresi sembilan gen calon tersebut terhadap gen-gen rujukan ini menunjukkan ketidakwujudan perbezaan tahap ekspresi relatif gen-gen calon itu di antara tisu kanser dan tisu tanpa kanser pangkal rahim. Walaupun begitu, kajian ini menjumpai hubung kait antara tahap ekspresi *CTLA4* yang tinggi dengan alel varian polimorfisme *CTLA4* c.49A>G, dan tahap ekspresi *FAS* yang rendah dengan alel varian polimorfisme *FAS* c.-671A>G. Kajian ini juga telah berjaya menyifatkan landskap perubahan nombor salinan dan kehilangan heterozigositi keseluruhan genom di dalam tisu kanser pangkal rahim, dan mengenal pasti bahawa lokus-lokus yang paling biasa menunjukan perubahan genomik secara berulang termasuk 3q21.3, 10q26.3 dan 5p15.33 (untuk penambahan nombor salinan), 11p11.12, Xp11.1 dan 12q11 (untuk kehilangan nombor salinan), dan

Xq11.1, Xq22.3 dan 16p11.2 (untuk kehilangan heterozigositi). Kesimpulannya, kajian ini telah berjaya menentukan kelaziman dan pengagihan subjenis HPV dalam kalangan wanita di Malaysia, mengenal pasti beberapa polimorfisme genetik yang berhubung kait dengan risiko kanser pangkal rahim, dan menerangkan perubahan-perubahan molekular utama yang berlaku di dalam tisu kanser dan tisu tanpa kanser pesakit kanser pangkal rahim. Hasil kajian yang diperolehi ini dijangka akan mempunyai prospek yang penting dalam bidang perubatan perseorangan yang menjaminkan.

ROLES OF SELECTED GENETIC VARIATIONS AND MOLECULAR ALTERATIONS IN HUMAN PAPILLOMAVIRUS-MEDIATED CANCER OF UTERINE CERVIX

ABSTRACT

Although human papillomavirus (HPV) is essential for cervical carcinogenesis, it alone is insufficient to result in the malignant transformation of the cervix. Information on the roles of selected genetic variations as well as molecular alterations (in the forms of gene expression, DNA copy number alterations and loss of heterozygosity) in cervical cancer is currently lacking among Malaysian women. The present study was therefore undertaken to fill this gap of the literature. For this purpose, the associations of 12 genetic polymorphisms in nine candidate genes involved in various carcinogenic pathways with cervical cancer risk were evaluated on 185 histopathologically confirmed cervical cancer patients and 209 cancer-free female controls. The mRNA expression levels of these nine candidate genes, as well as the genome-wide copy number alterations and allelic imbalances were also analyzed in cancerous and non-cancerous tissues of 72 of the subjects. All data obtained were subsequently stratified by ethnicity and HPV subtypes involved. In addition, data on tumor molecular alterations was also stratified by stages and histopathology of the cancer. Besides, the association between the genetic polymorphisms and their respective gene expression levels was investigated. It was demonstrated that 154 (83.2%) of the cervical cancer patients and none (0.0%) of the cancer-free female controls were positive for HPV infection. Among the HPV-positive subjects, 141 (91.6%) had single-type HPV infections (of which HPV16, HPV18 and HPV58 predominated), while the remaining 13 (8.4%) were simultaneously

infected by two HPV types (with HPV16 + HPV18 being the most common multiple-type infection). It was also observed in the present study that after adjustment to known risk factors of cervical cancer, *TNF* c.-488G>A, *CTLA4* c.49A>G, *FAS* c.-671A>G, *MDM2* c.14+309T>G and *CCND1* c.723G>A polymorphisms showed statistically significant associations with the risk of the cancer when analyzed singly. In addition, statistically significant observations were noted for the several combinations of *CTLA4* c.-319C>T polymorphism with *TNF* c.-418G>A, *CDKN1A* c.93C>A and *CCND1* c.*687C>G polymorphisms. Apart from that, it was shown in the present study that *PUM1*, *YWHAZ* and *RPLP0* were the most stably expressed reference genes in cervical cancer, and normalization of expression levels of the nine candidate genes against these reference genes showed no significant relative expression difference between cancerous and non-cancerous tissues of the cervix. However, an association was found between a higher *CTLA4* expression and the variant allele of *CTLA4* c.49A>G polymorphism, and between a lower *FAS* gene expression and the variant allele of *FAS* c.-671A>G polymorphism. The present study also successfully characterized the landscape of genome-wide copy number alterations and loss of heterozygosity in cervical cancer tissues, and observed that the most common loci with recurrent genomic alterations were 3q21.3, 10q26.3 and 5p15.33 (for copy number gains), 11p11.12, Xp11.1 and 12q11 (for copy number losses), and Xq11.1, Xq22.3 and 16p11.2 (for losses of heterozygosity). In conclusion, the present study has successfully determined the prevalence and distribution of HPV types among Malaysian women, identified several host genetic polymorphisms associated with risk of cervical cancer, and identified some major molecular alterations which occurred in the cancerous and non-cancerous tissues of cervical cancer patients. These results have a potential prospect in the promising field of individualized preventive medicine.

CHAPTER 1

INTRODUCTION

1.1 Research background

Cervical cancer is the seventh most common cancer worldwide and the fourth most common type of cancer among the female population (Ferlay *et al.*, 2013). In Malaysia, the age-standardized rate of cervical cancer is 7.8 per 100,000 females, according to the most recent Malaysian National Cancer Registry Report (Omar & Tamin, 2011). The cancer develops slowly from a premalignant stage called cervical intraepithelial neoplasia (CIN), during which the cure rate is nearly 100% (Berget *et al.*, 1991). Nevertheless, once CIN progresses into cervical cancer and spreads to distant organs, the average five-year survival rate of the patients drops drastically to 16.5% (National Cancer Institute, 2015). This implies that early detection and identification of cervical cancer is critical for appropriate management of the patients.

Currently, the most widely employed test for cervical cancer screening is the Papanicolaou test, more popularly known as the Pap test or Pap smear. However, a downside of this conventional screening test is its low sensitivity of approximately 37-66% (Gyllensten *et al.*, 2010; McCrory *et al.*, 1999; Wright, 2007), which denotes that a considerable portion of patients is tested negative for cervical anomalies when premalignant or malignant dysplasia of the cervix is actually present. In addition, the lack of Pap screening resources has resulted in the high incidence of cervical cancer in developing countries (Kitchener *et al.*, 2006). Therefore, there is a clear need for the

identification of potential biomarkers for early detection of individuals at risk of cervical cancer.

Over the past decades, it has become established that human papillomavirus (HPV) is the central etiologic agent for cervical carcinogenesis (Schiffman & Wentzensen, 2013). The viral DNA has been shown to be present in virtually all cervical carcinoma specimens examined worldwide (Walboomers *et al.*, 1999), and persistent infection of HPV has been definitively linked to the development of the cancer (Schiffman & Wentzensen, 2013). However, the fact that most precancerous cervical lesions regress spontaneously and only a small proportion of HPV-infected women eventually develop cervical cancer suggests that HPV alone is not sufficient to contribute to cervical carcinogenesis. Other cofactors must therefore be present in conjunction with HPV infection for the progression of CIN into malignancy.

Several environmental and/or lifestyle factors have now been recognized as important cofactors for cervical carcinogenesis. Cigarette smoking, for example, has been shown to induce severe oxidative DNA damage especially in HPV-infected cells, thus increasing the susceptibility of smokers to cervical cancer (Moktar *et al.*, 2011). In addition, multiparity and use of oral contraceptives may modulate the endogenous level of steroid hormones, which in turn affects the efficiency of immune response against HPV, thus facilitating cervical cancer development (Chung *et al.*, 2010). Besides, in recent years, it has become more and more apparent that host genetic factors may affect the persistency of HPV infection, and could therefore serve as a non-modifiable risk factor for cervical cancer development and progression (Tan & Ankathil, 2015).

High penetrance genetic mutations can potentially confer a high lifetime risk of cancer among their carriers (Fletcher & Houlston, 2010). However, the search for high penetrance genetic mutations associated with cervical cancer has proven difficult and has been unsuccessful. Many researchers have thus shifted their attention to the identification of low penetrance single nucleotide polymorphisms (SNPs) associated with the cancer. It is thought that these low penetrance genetic polymorphisms could influence cervical cancer risk only modestly, but the magnitude of the risk association could be increased dramatically when several low penetrance genetic polymorphisms are combined (Fletcher & Houlston, 2010). In addition, since SNPs are highly prevalent in the general population, they may account for a greater population-attributable risk (PAR) of cervical cancer compared to high penetrance mutations (Tan & Ankathil, 2015).

Among the large number of SNPs in the human genome, polymorphisms in genes which have been implicated to the complex mechanisms of carcinogenesis represent potential candidate risk markers for cervical cancer. SNPs of genes involved in mediating inflammatory and immune responses (e.g. *TNF*, *IL4R*, *CTLA4* and *CCR2*), apoptosis and anti-apoptosis (e.g. *TP53*, *FAS* and *MDM2*), cell cycle regulation and DNA repair (e.g. *CDKN1A* and *CCND1*), among others, have been frequently investigated for their association with cervical cancer susceptibility (de Freitas *et al.*, 2012; Tan & Ankathil, 2015; Vineis *et al.*, 2009; Wang *et al.*, 2009b; Zhang *et al.*, 2014c), although inconsistent findings have been obtained in many studies. The inconsistency of the results obtained could be due to the different ethnic and geographical backgrounds of the study subjects included, or the different sample size and experimental approaches employed (Tan & Ankathil, 2015). There is currently a

paucity of data available regarding the association of candidate polymorphisms in the above genes with cervical cancer risk in Malaysian population.

Besides, as carcinogenesis is a multistep process involving dysregulation of proto-oncogenes, tumor suppressor genes and other key genes involved in cell proliferation, inflammatory and immune responses, apoptosis, cell cycle regulation and DNA repair, disruption of gene expression is commonly observed in many cancers (Sadikovic *et al.*, 2008; Sharma *et al.*, 2010). Alteration in gene expression is thought to arise as a result of point mutations, gene rearrangements, gene amplifications, epigenetic events or other carcinogenesis-related genetic events (Pierotti *et al.*, 2010; Sharma *et al.*, 2010). The resulting ectopic gene expression could in turn facilitate acquirement of cancer phenotypes by disrupting the downstream signaling pathways (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011). If the expression level of a gene exhibits a consistent pattern of distinct cut-off between normal and cancerous tissues across a large number of specimens, it may potentially be used as a genetic marker for predicting disease prognosis or for molecular classification of cancer. Recent years have witnessed the identification of several potential genetic markers which are either upregulated (Xiao *et al.*, 2014; Zhang *et al.*, 2014b) or downregulated (Zhang *et al.*, 2015) in cervical cancer. However, little is known about the expression levels of the aforementioned candidate genes (*TNF*, *IL4R*, *CTLA4*, *TP53*, *FAS*, *MDM2*, *CDKN1A*, *CCR2* and *CCND1*) in cancerous and non-cancerous tissues of uterine cervix. Given the important role of these genes in various mechanisms of oncogenesis, evaluation of the differential gene expression of these genes in cervical cancer remains a promising avenue of research.

Another molecular event which may arise during the process of carcinogenesis is somatic copy number alterations (CNAs), in which segments of DNA is present in non-diploid copies. Somatic CNAs are arguably the most abundant genetic alteration in the cancer genome (Beroukhim *et al.*, 2010; Zack *et al.*, 2013), although its prevalence in cervical cancer has not been thoroughly investigated. The most apparent functional impact of somatic CNAs is the alteration of gene dosage, thereby resulting in over- or under-production of the gene products (Handsaker *et al.*, 2015). In addition, somatic CNAs could result in the inactivation of tumor suppressor genes, hence facilitating malignant transformation. Considering the potential impact of somatic CNAs in oncogenesis, characterization of these molecular aberrations in cervical cancer could provide important insights into the etiology of cervical cancer.

Apart from that, loss of heterozygosity (LOH) represents another class of genetic alteration frequently observed in various cancers (Beroukhim *et al.*, 2006). Conventionally, LOH has been thought to arise due to hemizygous deletion of one allele, thus resulting in a state of DNA copy number loss (Beroukhim *et al.*, 2006). Recent studies, however, showed that LOH does not necessarily be associated with a net change in DNA copy number (Kumar *et al.*, 2015; Melcher *et al.*, 2011; O'Keefe *et al.*, 2010; Stirewalt *et al.*, 2014). In fact, this condition, termed copy-neutral LOH (CNLOH), is extremely ubiquitous in the human genome and may constitute as high as 70% of allelic imbalance events in human cancers. CNLOH is manifested when the remaining undeleted allele becomes duplicated. Since both resulting alleles are derived from a single parental chromosome, CNLOH is also sometimes referred to as acquired uniparental disomy (aUPD). CNLOH is potentially pathological if the wild type allele is

deleted and the mutant allele is duplicated. In addition, LOHs with copy number gains have also been reported, in which the remaining undeleted allele becomes amplified (Beroukhim *et al.*, 2006; Kuga *et al.*, 2008). So far, the contribution of LOH, CNLOH and LOH with copy number gains in modulating cervical cancer development is less well understood. Genome-wide screening for these genetic events in cervical cancer may serve as an important step for identification of key genomic regions harboring novel cancer-related genes.

1.2 Rationale of the study

HPV is by far the most important risk factor for cervical cancer. More than 100 HPV subtypes have been identified, but only several subtypes are classified as "high-risk" which are potentially carcinogenic to humans. Current HPV vaccines in Malaysia are available in bivalent (which targets on HPV16 and HPV18) and quadrivalent forms (which targets on HPV6, HPV11, HPV16 and HPV18) (Ministry of Health Malaysia, 2011), although the United States Food and Drug Administration (U.S. F.D.A.) has recently approved a nonavalent (9-valent) form of the vaccine, which protects against HPV6, HPV11, HPV16, HPV18, HPV31, HPV33, HPV45, HPV52 and HPV58 (Chatterjee 2014; U.S. F.D.A, 2014). Presently, the prevalence and type distribution of HPV in the Malaysian population are incompletely understood. Therefore, the present study assessed the prevalence and type distribution of HPV in cervical cancer, to help appraising the efficiency of the current vaccines in protecting against the disease.

Apart from that, it is known that genetic polymorphisms typically influence cervical cancer risk in a population-specific manner. A number of candidate genetic polymorphisms have been investigated with regard to their association with cervical cancer risk in various other populations. These include the *TNF* c.-418G>A (rs361525), *TNF* c.-488G>A (rs1800629), *IL4R* c.1727A>G (rs1801275), *CTLA4* c.49A>G (rs231775), *CTLA4* c.-319C>T (rs5742909), *TP53* c.215C>G (rs1042522), *FAS* c.-671A>G (rs1800682), *MDM2* c.14+309T>G (rs2279744), *CDKN1A* c.93C>A (rs1801270), *CCR2* c.190G>A (rs1799864), *CCND1* c.723G>A (rs9344) and *CCND1* c.*687C>G (rs678653) polymorphisms. However, information is lacking regarding the allelic and genotypic distributions of these polymorphisms in the Malaysian population, as well as the contribution of these polymorphisms in modulating the susceptibility of HPV-mediated cervical cancer in Malaysia. This study addressed this issue employing a case-control analysis.

Besides, it has now been widely recognized that disruption of gene expression represents an important feature of cancers. Among the many methods of gene expression profiling, reverse transcription-quantitative real-time PCR (RT-qPCR) is considered as the gold standard due to its simplicity, rapidity, sensitivity, specificity and reproducibility. However, reliability of RT-qPCR findings depends heavily on the appropriateness of its data normalization against reference genes whose expressions are constant across diverse biological conditions. There is no single 'universal' reference gene which can be used for normalization of RT-qPCR data in all types of tissue, and the search for the optimal reference genes for a particular tissue type is critically important (Jacob *et al.*, 2013; Kozera & Rapacz, 2013). There is currently a paucity of information

regarding the suitable reference genes to be used for normalization of RT-qPCR data in human cervical tissues. This study therefore screened a large panel of candidate reference genes to identify the optimal RT-qPCR reference genes for cancerous and non-cancerous tissues of human uterine cervix.

In addition, little is known about the expression changes of *TNF*, *IL4R*, *CTLA4*, *TP53*, *FAS*, *MDM2*, *CDKN1A*, *CCR2* and *CCND1* during the process of cervical carcinogenesis. Thus, this study investigated the mRNA expression of the above-mentioned genes in cancerous and non-cancerous tissues of the uterine cervix and evaluated the difference of their expression between the two groups of the specimens, in order to determine whether dysregulation of these genes occurs during cervical cancer development. The present study also evaluated the usability of these genes as biomarkers of cervical cancer staging, by investigating whether a trend exists in the expression levels of the genes in increasing stages of cervical cancer. Besides, the influence of genetic polymorphisms on the expression of the respective genes was also investigated.

Lastly, somatic copy number alterations (CNAs) and loss of heterozygosity (LOH) are known to be common events in many cancers. However, not much is known about the presence and genomic loci of CNAs and LOH in cervical cancer. Moreover, previous technologies which did not permit simultaneous screening of CNAs and LOH have failed to appreciate the significance of copy-neutral LOH (CNLOH) and LOH with accompanying copy number gain in oncogenesis. Hence, another aspect of this study was to employ the latest SNP array technology for concurrent profiling of CNAs and LOH in cancerous and non-cancerous tissues of the cervix.

1.3 Objectives

General objective:

To investigate the role of host genetic variations and tumor molecular alterations in HPV-mediated cancer of uterine cervix in Malaysian population

Specific objectives:

1. To determine the prevalence and type distribution of HPV in cervical cancer patients and cancer-free healthy female controls in Malaysia.
2. To investigate the allele and genotype frequencies of selected polymorphisms in genes involved in mediating inflammatory and immune responses, apoptosis and anti-apoptosis, cell cycle regulation and DNA repair in cervical cancer patients and cancer-free healthy female controls in Malaysia, and to determine the association of these polymorphisms, either singly or in combinations, with cervical cancer risk.
3. To determine the most stably expressed reference genes for normalization of reverse transcription quantitative real-time PCR (RT-qPCR) data in cancerous and non-cancerous tissues of human uterine cervix.
4. To investigate the mRNA expression levels of *TNF*, *IL4R*, *CTLA4*, *TP53*, *FAS*, *MDM2*, *CDKN1A*, *CCR2* and *CCND1*, and to assess the difference in gene expression of the above-mentioned genes, in cancerous and non-cancerous tissues of uterine cervix.

5. To determine the association between *TNF*, *IL4R*, *CTLA4*, *TP53*, *FAS*, *MDM2*, *CDKN1A*, *CCR2* and *CCND1* gene expression levels and their respective genetic polymorphisms.
6. To characterize the genome-wide copy number alterations (CNAs) and allelic imbalances (LOH, CNLOH and LOH with copy number gains) present in cancerous and non-cancerous tissues of human uterine cervix.

CHAPTER 2

LITERATURE REVIEW

2.1 Cancer

Cancer refers to a group of diseases developed as a result of uncontrolled cell proliferation. This occurs when the normal biomolecular circuits for cell division and cell death have become dysregulated, leading to the formation of a mass of cells called tumor or neoplasm. Tumors can, however, be classified into benign or malignant, and only malignant tumors contribute to the development of cancer. Since every cell types have the possibility to become cancerous, more than 200 types of cancer have been identified to date (Pelengaris & Khan, 2006). These can be classified into four main forms, namely sarcomas, carcinomas, leukemia or lymphoma (Weinstock, 1984), in which the tumor originates from mesenchymal tissue, epithelial tissue, blood-forming tissues and immune cells, respectively. Of these, carcinoma is the most commonly seen form of cancer, and it represents more than 80% of all cancers (Liu, 2007).

2.2 Hallmarks of cancer

Hanahan and Weinberg (2000) have proposed that most, if not all, of human cancers manifest six essential physiological changes at the cellular level, which are commonly referred to as the core hallmarks of cancer. These changes are acquired during the multistep process of carcinogenesis, and each indicates that an anticancer defense mechanism within the cells has successfully been overcome. Extensive research in the field of oncology during the past decade has added two “emerging hallmarks” to the

above list, namely reprogramming of energy metabolism and evasion of immune destruction (Hanahan & Weinberg, 2011). Central to the acquisition of all the eight cancer hallmarks are two "enabling characteristics", namely genome instability and mutation, as well as tumor-promoting inflammation (Hanahan & Weinberg, 2011). The core and emerging hallmarks of cancer are summarized in Table 2.1.

Table 2.1: Core and emerging hallmarks of cancer (adapted from Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011)

Hallmarks of cancer	**Example of mechanism**
Core hallmarks:	
Self-sufficiency in growth signals	Activation of H-Ras oncogene
Insensitivity to anti-growth signals	Loss of retinoblastoma suppressor
Evasion of apoptosis	Production of insulin-like growth factor (IGF) survival factors
Limitless replicative potential	Activation of telomerase
Sustained angiogenesis	Production of vascular endothelial growth factor (VEGF) inducer
Tissue invasion and metastasis	Inactivation of E-cadherin
Emerging hallmarks:	
Reprogramming of energy metabolism	Upregulation of glucose transporters
Avoidance of immune destruction	Secretion of TGF-β or other immunosuppressive factors

2.3 Cervical cancer

Cervical cancer is the malignant neoplasm occurring to the cells lining the cervix uteri. The cancer arises primarily from the squamocolumnar cell junction at an area termed the transformation zone. Cervical carcinoma typically develops slowly, beginning as a premalignant stage known histologically as cervical intraepithelial neoplasia (CIN). There are three grades of CIN, namely CIN 1, CIN 2 and CIN 3, which indicate mild, moderate and severe dysplasia, respectively (Richart, 1969). The premalignant cervical dysplasia can begin in any of the three grades. Cytologically, CIN 1 is known as low-grade squamous intraepithelial lesion (LSIL), while CIN 2 and CIN 3 are known as high-grade squamous intraepithelial lesion (HSIL). CIN usually regresses spontaneously, but can sometimes progress into cervical cancer in a small proportion of patients (Ostör, 1993; Schiffman & Wentzensen, 2013; Walboomers *et al.*, 1999).

Cervical cancers are histopathologically categorized according to the type of cells from which they are derived (International Agency for Research on Cancer, 2005). Cervical cancer that develops from squamous epithelial cells is called squamous cell carcinoma, while that which develops from columnar epithelial cells is known as adenocarcinoma. Less commonly, cervical cancers possess features of both squamous cell carcinoma and adenocarcinoma, in which case it is referred to as adenosquamous carcinoma. Squamous cell carcinoma represents the most common type of cervical cancer, with approximately 70-80% incidence rate (Vizcaino *et al.*, 2000), as compared to around 20% incidence of adenocarcinoma (Schorge *et al.*, 2004). Besides squamous cell carcinoma, adenocarcinoma and adenosquamous carcinoma, a rare subset of cervical cancer is manifested as neuroendocrine carcinoma, small cell carcinoma, clear

cell carcinoma, mixed Müllerian tumor, glassy cell carcinoma, among several other uncommon histopathological types (International Agency for Research on Cancer, 2005; Khalbuss *et al.*, 2013).

Globally, cervical cancer is the seventh most common type of cancer overall and the fourth most common cancer in women. Incidence data from GLOBOCAN database indicates that 527,624 new cervical cancer cases were diagnosed in year 2012 alone, and approximately 84.3% of these occur in developing countries (Ferlay *et al.*, 2013). The mortality rate of cervical cancer, relative to its incidence rate, is approximately 50%.

2.3.1 Staging of cervical cancer

A number of systems have been employed for cervical cancer staging. Instead of being surgically-based as in the case of many other cancers, staging of cervical cancer is clinically-based, since many early studies have established that surgical staging of cervical cancers can complicate the downstream therapy processes (Delgado *et al.*, 1978; Piver & Barlow, 1974; Wharton *et al.*, 1977).

The most widely used system for cervical cancer staging is the International Federation of Gynecology and Obstetrics (Fédération Internationale de Gynécologie et d'Obstétrique, or FIGO) system. Staging using the FIGO system relies principally on physical examinations, examination under anesthesia, colposcopy, cystoscopy, sigmoidoscopy, intravenous pyelogram and chest X-ray (Moore, 2006). The current FIGO staging system for cervical cancer (revised 2009) is shown in the Table 2.2 (Pecorelli, 2009). Although discernment of carcinoma *in situ* from later stages of

cervical cancer is helpful for clinical management of patients, the FIGO staging system does not include a stage for the former. Thus, a Stage 0 is often informally used to denote carcinoma *in situ*.

Table 2.2: FIGO staging system for cervical cancer (adapted from Pecorelli, 2009)

Stage	**Description**
I	The carcinoma is strictly confined to the cervix (extension to the corpus would be disregarded)
IA	Invasive carcinoma which can be diagnosed only by microscopy, with deepest invasion ≤5 mm and largest extension ≥7 mm
IA1	Measured stromal invasion of ≤3.0 mm in depth and extension of ≥7.0 mm
IA2	Measured stromal invasion of >3.0 mm and not >5.0 mm with an extension of not >7.0 mm
IB	Clinically visible lesions limited to the cervix uteri or pre-clinical cancers greater than stage IA
IB1	Clinically visible lesion ≤4.0 cm in greatest dimension
IB2	Clinically visible lesion >4.0 cm in greatest dimension
II	Cervical carcinoma invades beyond the uterus, but not to the pelvic wall or to the lower third of the vagina
IIA	Without parametrial invasion
IIA1	Clinically visible lesion ≤4.0 cm in greatest dimension
IIA2	Clinically visible lesion >4 cm in greatest dimension
IIB	With obvious parametrial invasion

Table 2.2, continued

Stage	Description
III	The tumor extends to the pelvic wall and/or involves lower third of the vagina and/or causes hydronephrosis or non-functioning kidney
IIIA	Tumor involves lower third of the vagina, with no extension to the pelvic wall
IIIB	Extension to the pelvic wall and/or hydronephrosis or non-functioning kidney
IV	The carcinoma has extended beyond the true pelvis or has involved (biopsy proven) the mucosa of the bladder or rectum. A bullous edema, as such, does not permit a case to be allotted to Stage IV
IVA	Spread of the growth to adjacent organs
IVB	Spread to distant organs

2.3.2 Risk factor of cervical cancer: human papillomavirus (HPV)

Human papillomavirus (HPV) (Figure 2.1) is the central etiologic agent in the development of cervical cancer. To date, more than 100 types of HPV have been identified, and about 40 of these can infect human genitalia (Bernard *et al.,* 2010; de Villiers, 2001). The virus can be categorized into two categories, namely cutaneous and mucosotropic, with the latter being relevant to the development of cervical carcinoma. Mucosotropic HPV can further be classified into high and low risk groups, depending on their ability to immortalize human keratinocytes (de Villiers *et al.*, 2004). The high risk group includes HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82 (Muñoz *et al.*, 2003), and infection of any of these can result in cervical intraepithelial neoplasia, which can potentially progress into invasive cervical cancer. A total of 99.7% of cervical carcinoma samples have been tested positive for some of the above-mentioned HPV DNA (Walboomers *et al.*, 1999). HPV16 and 18 predominate in cervical cancer, accounting for 50-60% and 10-20% of cervical cancer cases, respectively (Muñoz *et al.*, 2004). In fact, the International Agency for Research on Cancer (1995) has officially classified the two above-mentioned HPV types as Group 1, i.e. carcinogenic to humans.

HPV possesses a circular, double-stranded DNA genome of approximately 8 kb in size, which encodes for six early genes (E1, E2, E4, E5, E6 and E7) and two late genes (L1 and L2) that are respectively expressed early and late in the HPV life cycle. There is also a long control region (LCR) in the HPV genome, which lies between the early and late gene regions that plays important regulatory roles in the replication and transcription of HPV genome (Figure 2.2). Upon infection, HPV genome typically exists as an episome in the nucleus of infected cells, and hijacks the host cell machinery for the

viral gene expression. The normal expression of E2 gene represses the expression of E6 and E7, which are implicated in cervical neoplasia due to their abilities to inactivate p53 and pRb tumor suppressor proteins, respectively (Figure 2.2) (Dyson *et al.*, 1989; Scheffner *et al.*, 1990; Tan & Ankathil, 2015; Werness *et al.*, 1990). In cervical cancer and some cervical intraepithelial neoplasia, integration of HPV genome into host chromosome has been observed (Lehn *et al.*, 1988). The integration results in the disruption of HPV E2 open reading frame (ORF), which abolishes the function of the E2 protein (Tan & Ankathil, 2015). This causes the increased expression and stability of E6 and E7 gene products, which can eventually lead to the immortalization and subsequent malignant transformation of the infected cells (Figure 2.2) (Münger *et al.*, 1989; Tan & Ankathil, 2015; Villa & Schlegel, 1991).

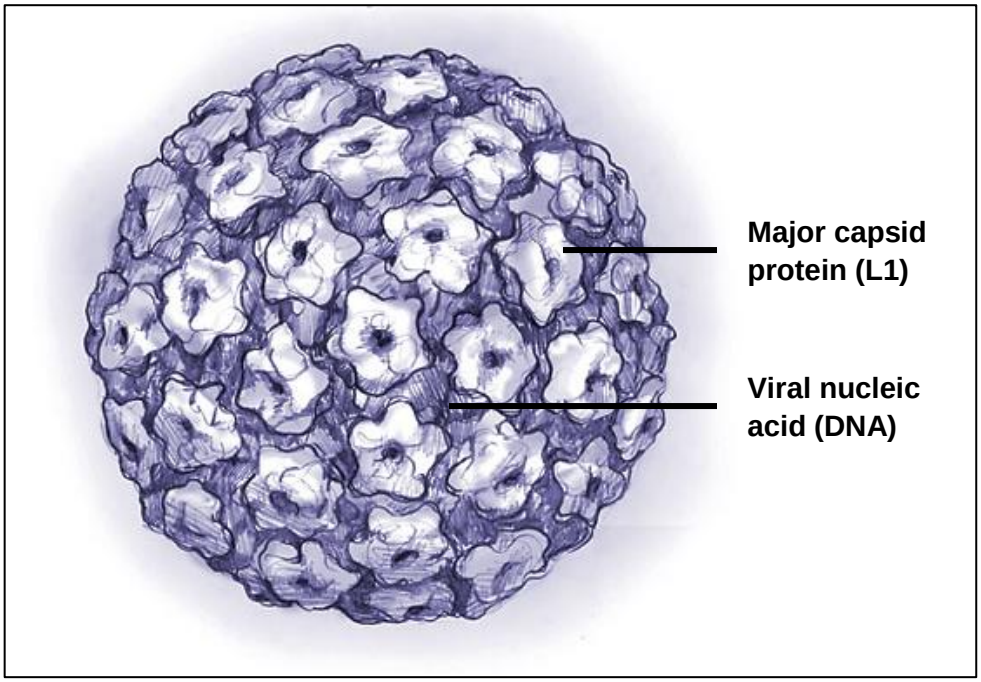

Figure 2.1: Structure of HPV (adapted from American Medical Association, 2012).

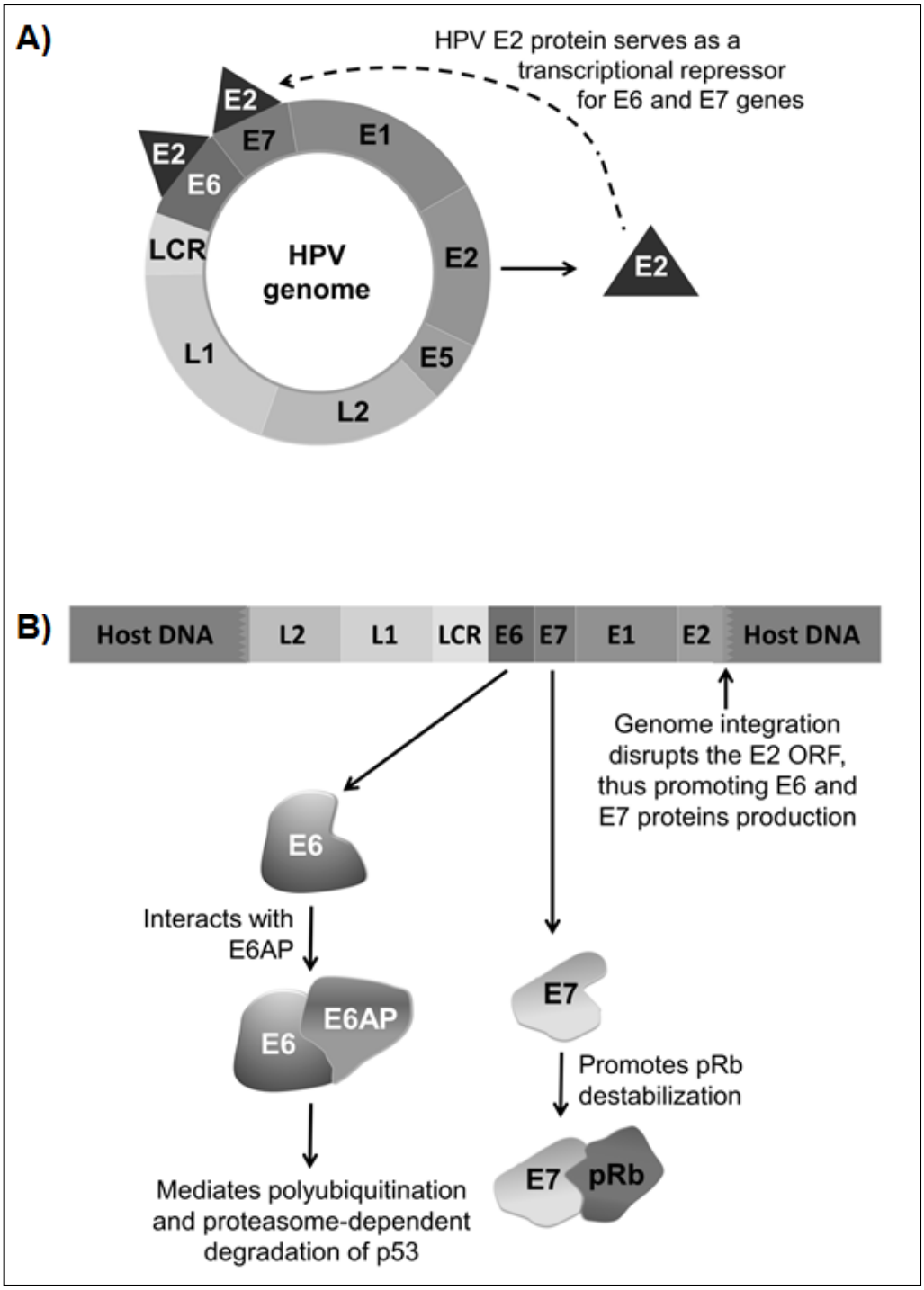

Figure 2.2: The HPV genome. (A) The HPV genome comprises six early genes and two late genes, and a long control region in between. The encoded E2 protein serves as a transcriptional repressor for E6 and E7 genes. (B) Integration of HPV DNA into the host genome disrupts the E2 gene, allowing production of E6 and E7 proteins which respectively inhibit p53 and pRb tumor suppressor proteins (adapted from Tan & Ankathil, 2015).

2.4 Genetic predisposition to cancer

The observations that cancers tend to cluster in families and segregate in a Mendelian fashion suggest that the predisposition of an individual to the development of cancer can be inherited. In cervical cancer, a familial aggregation of the disease has also been noted (Hemminki & Chen, 2006). Besides, it has been shown that the risk of developing cervical cancer was considerably higher in biologically-related first degree relatives of cervical carcinoma patients, compared to non-biologically-related relatives (Ahlbom *et al.*, 1997). Furthermore, a recent large-scale study showed that monozygotic twins shared significantly more similar results in cervical smear test compared to dizygotic twins and other first-degree relatives (Vink *et al.*, 2011). Based on these epidemiological studies, the genetic heritability of cervical smear abnormality and invasive cervical cancer has been proposed.

Genetic predisposition of cancer can occur through the inheritance of highly penetrant germline mutations in key cancer susceptibility genes (Fletcher & Houlston, 2010). The inheritance of such mutations is equivalent to the first hit of Knudson two-hit hypothesis of tumorigenesis (Knudson, 1971), thereby increasing an individual's lifetime risk of developing a cancer. Although numerous cancer susceptibility genes, such as *BRCA1* and *BRCA2* in hereditary breast and ovarian cancer (HBOC) syndrome, MMR genes in hereditary nonpolyposis colorectal cancer (HNPCC) and *RB1* in hereditary retinoblastoma, have been characterized in other types of human cancer, identification of high penetrance mutations in cervical cancer has proven difficult. Considering this, and the fact that high penetrance mutations typically accounts for less than 10% of all human cancers, many genetic susceptibility studies of cancers have

shifted their focus to identification of low penetrance genetic polymorphisms over the past decades.

In genetic epidemiological studies of cancers, low penetrance polymorphisms in genes involved in the complex mechanisms of oncogenesis have naturally become the potential candidates for investigation. These include the genes involved in inflammation, DNA repair, immune responses, detoxification of carcinogens, hormonal responses, among others. Although the cancer risk that can be conferred by these polymorphisms is relatively lower compared to the above-mentioned high penetrance mutations, the variant alleles of these genes frequently interact with environmental factors and predispose a significant portion of the population to the development of cancer. Many common SNPs in genes involved in various biological pathways, such as xenobiotic metabolism, DNA repair, inflammation, immune response, cell cycle regulation etcetera, have been shown to be associated with cancer. However, varying results have been obtained for the relationship between these SNPs and cancer risk, possibly due to the differences in experimental approaches and sample size used, as well as the ethnicity, genetic and geographical background of the subjects studied.

2.5 *TNF* polymorphisms and cervical cancer

Tumor necrosis factor (TNF) is a cytokine encoded by the *TNF* gene on chromosomal locus of 6p21.3 in human. The gene spans a region of approximately 2.7 kb, and comprises 4 exons, with the last exon encoding for the majority of the protein (Nedwin *et al.*, 1985). Together with two lymphotoxin genes (*LTA* and *LTB*; Browning *et al.*,

1993) and *Lst-1* gene (Holzinger *et al.* 1995), *TNF* occurs within the major histocompatibility complex (MHC) class III region in the *TNF* gene cluster.

TNF is initially produced as a 26 kDa transmembrane protein, which is composed of 212 amino acids arranged in stable homotrimers (Kriegler *et al.*, 1988; Tang *et al.*, 1996). The mature soluble cytokine, which is 17 kDa in size and 185 amino acids long, is produced following proteolytic cleavage of the initial transmembrane protein by TNF-alpha converting enzyme (TACE) (Black *et al.*, 1997). The secreted protein functions in homotrimeric structure and is unstable below the nanomolar range as it tends to dissociate into monomeric subunits at low concentrations (Corti *et al.*, 1992).

TNF can potentially play a role in the progression and development of cancer, as ectopic expression of the *TNF* has been observed in certain cancer cells (Locksley *et al.*, 2001). The role of TNF in cancer development is further envisaged by considering the fact that it is one of the most crucial cytokine involved in the induction of inflammation, a process which allows various bioactive molecules such as growth factors, survival factors, proangiogenic factors, extracellular matrix-modifying enzymes and inductive signals to be supplied to the tumor microenvironments (Hanahan & Weinberg, 2011). Moreover, inflammatory cells have been shown to release reactive oxygen species and other mutagenic chemicals which can promote the malignancy of their nearby cancer cells (Grivennikov *et al.*, 2010). In addition, TNF have been shown to be able to promote angiogenesis either directly by stimulating the proliferation of endothelial cells,

or indirectly by modulating the expression of various proangiogenic factors (Leek *et al.*, 1998).

Aberrant expression of *TNF* could result in an increased risk of cervical cancer. Promoter region of the *TNF* gene contains a number of regulatory sequences which can serve as the binding sites of transcriptional activators or repressors and lead to either an increase or a decrease expression level of the gene. Polymorphisms in the promoter region of the gene could disrupt these regulatory sequences and could thus influence the level of *TNF* expression. Certain *TNF* promoter polymorphisms, such as the well-studied c.-418G>A polymorphism (rs361525, previously known as the -238G>A polymorphism) and c.-488G>A polymorphism (rs1800629, previously known as the -308G>A polymorphism), have been associated with the risk of developing cancers, although the results obtained in many studies are inconsistent. The inconsistency appeared to occur in a population-specific manner. For example, a genetic susceptibility study on the Korean population revealed a significant association between the c.-418A allele and a lower predisposition risk of cervical cancer, and lack of association between the c.-488G>A polymorphism and the malignancy (Jang *et al.*, 2001). On the contrary, Zuo *et al.* (2011) reported no association between c.-418G>A polymorphism and cervical cancer risk among Southwest Chinese population, while Duarte *et al.* (2005) observed that the c.-488A genotype of *TNF* could confer a higher risk of developing cervical cancer among Portuguese population. Despite this, a meta-analysis of 13 studies which involved a total of 3,294 cases and 3,468 controls showed that the c.-418A allele was significantly associated with a decreased risk of cervical cancer, while the c.-488G>A polymorphism was associated with an increased cervical cancer risk only among

Caucasians and Africans, but not among Asians or subjects with mixed ethnicities (Pan *et al.*, 2012). No study has been performed to investigate the association between the polymorphisms and cervical cancer risk in the Malaysian population.

2.6 Interleukin-4 receptor (IL4R) and cervical cancer

Interleukin-4 receptor (IL4R) is a transmembrane cytokine receptor found in between 50 and 5,000 units on the surface of most cell types (Nelms *et al.*, 1999; Zamorano *et al.*, 2003). There are two major forms of IL4R, namely the type I and the type II receptors, which are predominantly expressed in immune and cancer cells, respectively (Murata *et al.*, 1998). Both types of receptors contain an alpha chain (IL4Rα), which is a polypeptide belonging to the hematopoietin receptor superfamily that is encoded by the *IL4R* gene located on chromosome 16p12.1 in humans.

Upon IL4 binding, IL4Rα in type I receptors heterodimerizes with the gamma chain of IL2 receptor (IL2Rγ), while that in type II receptors forms a heterodimer with the alpha-1 chain of IL13 receptor (IL13Rα1) (Orchansky *et al.*, 1999). IL2Rγ and IL13Rα1 are encoded by *IL2RG* on chromosome Xq13.1 and *IL13RA* on chromosome Xq24, respectively. In certain types of cells, type III IL4 receptors, which contain all the three above-mentioned chains, are present, although IL4 was found to be in association with only two of the chains at any given time following heteromerization (Murata *et al.*, 1998). Heterodimerization between IL4Rα and IL2Rγ and/or IL13Rα1 is necessary for the activation of the Janus family of tyrosine kinases (JAK), which brings about the downstream signaling cascades (Ihle, 1995).