COMPARISON OF PROTEIN EXPRESSION BETWEEN BREAST AND COLORECTAL CANCERS

By

LEE HUI HUA

Thesis submitted in fulfillment of the

requirements for the degree of Master of Science

March 2012

ACKNOWLEDGEMENT

First of all, I would like to deliver my deepest gratitude to my supervisor, Assof. Prof. Dr. Gam Lay Harn for her guidance and helps throughout this study. Her knowledge and motivation greatly assisted me in the completion of this project. I would also like to express my gratitude to my co-supervisors, Assof. Prof. Dr. Shaharum Shamsuddin and Dr. Khoo Boon Yin for assisting, teaching and sharing of their knowledges to me.

I would like to thank Institute of Postgraduate Studies, Universiti Sains Malaysia for providing me with Fellowship Scholarship Scheme. I would like to thank the staffs of School of Pharmaceutical Science, National Poison Centre, INFORMM, School of Biological Sciences and Malaysian Institute of Pharmaceuticals and Nutraceuticals for providing facilities and instruments throughout the study. I would like to express my gratitude to the surgeons and staffs of OT room and Surgery department of Penang General Hospital and Hospital Universiti Sains Malaysia, Kubang Kerian to provide me breast and colorectal tissues for this study. Special thanks to Dr. Saravanan Dharmaraj for helping me in the statistic analysis of my result.

Finally, sincere thanks to my family, friends and labmates for their helps and encouragement during my master's study life.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
ACN	Acetonitrile
AEBSF	4-(2-ammoethyl) benzenesulfonyl fluoride
AP	Alkaline phosphotase
APS	Ammonium persulfate
BSA	Bovine serum albumin
CHAPS	3-[3-(cholamidopropyl) dimethylammonio]-1-propanesulfonate
CID	Collison induced dissociation
DTT	1,4-dithiothreitol
ESI	Electrospray ionization
HPLC	High performance liquid chromatography
Hsp 60	Heat shock protein 60
Hsp 70	Heat shock protein 70 variant 8
Hsp 90	Heat shock protein 90/endoplasmin
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
kDa	Kilodalton
LDA	Linear discriminant analysis
LC-MS/MS	Liquid chromatography tandem mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass to charge ratio

NaCl	Sodium chloride
$Na_2HPO_4.2H_2O$	Disodium hydrogen phosphate dehydrate
$NaH_2HPO_4.2H_2O$	Sodium dihydrogen phosphate dehydrate
PBS	Phosphate buffered saline
PCA	Principal component analysis
RP-HPLC	Reverse phase high performance liquid chromatography
SDS	Sodium dodecyl sulfate
SPSS	Statistical package for social science
TCA	Tricloroacetic acid
TEMED	N, N, N'N'-tetramethylenediamine
TLB	Thiourea lysis buffer

LIST OF APPENDICES

(e) LC-MS/MS analysis result for alpha-1-antitrypsin in colorectal cancer

(f) LC-MS/MS analysis result for selenium binding protein in colorectal cancer

(g) LC-MS/MS analysis result for annexin A3 in colorectal cancer

(h) LC-MS/MS result for tumor rejection antigen (GP96) in colorectal cancer

PERBANGAN PENGESPRESAN PROTEIN ANTARA KANSER PAYUDARA DENGAN KANSER KOLOREKTAL

Abstrak

Sehingga hari ini, kanser kekal sebagai masalah kesihatan yang utama secara global. Kanser payudara merupakan kanser yang paling biasa dalam kalangan wanita manakala kanser kolorektal ialah kanser yang ketiga umum untuk kedua-dua jantina di seluruh dunia. Memandangkan kanser payudara dan kanser kolorektal adalah antara kanser yang paling umum di Malaysia, kami ingin mengenalpasti penanda bio yang boleh digunakan untuk tujuan diagnosis dan rawatan bagi kedua-dua jenis kanser ini. Keinginan ini dicapai melalui perbandingan pengekspresan protein antara tisu-tisu kanser payudara dan kanser kolorektal dengan tisu-tisu normalnya. Dua puluh lima pasang tisu-tisu kanser payudara dan kolorektal serta tisu normal dikutip dari Hospital Umum Pulau Pinang (GH) selepas memperolehi keizinan daripada pesakit atau saudara terdekat pesakit. Protein-protein diekstrak daripada tisu-tisu dengan keadah pengekstrakan bersambungan yang mengandungi dua jenis penyangga, iaitu penyangga TRIS dan penyangga TLB. Protein-protein tersebut dipisahkan oleh elektroforesi dwi-dimensi. Dengan menggunakan perisian analisis gambar gel, proteome bagi tisu-tisu daripada kedua-dua jenis kanser dibandingkan antara: i) tisu kanser payudara dengan tisu normal; ii) tisu kanser kolorektal dengan tisu normal dan iii) kanser payudara dengan

kanser kolorektal. Tompok-tompok protein yang diekpresi pada kadar yang berbeza dikeluarkan daripada gel untuk analisis pengenalpastian protein selanjutnya oleh LC-MS/MS. Bagi kanser payudara, tropomyosin isoform dikenalpasti sebagai protein kawalan meningkat pada semua peringkat N bagi semua bangsa. Calrecticulin (TLB), PDI A3 (TLB), Hsp 70 (TLB), Hsp 60 dan PDI dikenalpasti sebagai protein kawalan meningkat pada peringkat N0 dan N1 bagi semua bangsa, kecuali peringkat N2. Ujian statistik PCA dan LDA menunjukkan bahawa penggunaan PDI A3 (TRIS), 'activation-induced cytidine deaminase', 'alpha-1-antitrypsin precursor', 'alpha-1-antitrypsin, chain A', 'haptoglobin', PDI, PDI A3 (TLB), 'annexin A3', 'annexin V', 'heat shock protein 70 kDA 1A', 'calrecticulin' dan 'tubulin alpha chain' dapat digunakan dengan pasti untuk menunjukkan keadaan patologi kanser payudara. Bagi kanser kolorektal, annexin A3 dikenalpasti sebagai protein kawalan meningkat pada semua peringkat N (iaitu N0, N1 dan N2) serta bangsa yang dikaji. Ujian statistik PCA dan LDA menunjukkan bahawa 'alpha-1-antitrypsin precursor', 'alpha-1-antitrypsin, chain A', PDI A3 (TRIS), HSP 90 (TRIS), 'tubulin, beta polypeptide' (TRIS), 'tubulin alpha chain', PDI A3, 'tubulin, beta polypeptide' (TLB), 'ATP synthase', HSP 70 (TLB), 'annexin A3', dan 'tropomyosin alpha-3 chain' boleh digunakan secara kolektif untuk membezakan tisu-tisu dalam keadaan normal atau kanser untuk kanser kolorektal. Pada masa yang sama, 'tubulin, beta polypeptide' (TLB) dan PDI A3 (TRIS) dikenalpasti sebagai protein kawalan bagi kedua-dua jenis kanser. Ujian 'Western blot' terhadap calrecticulin dan PDI A3 dijalankan untuk mengesahkan keputusan

daripada analisis LC-MS/MS dan mesin pencarian MASCOT database. Penyelidikan yang selanjutya terhadap calon protein sebagai penanda bio dengan menggunakan bilangan sampel yang lebih besar adalah wajar sebelum meneruskan pengembangan diagnosis dan rawatan kanser yang selanjutnya.

COMPARISON OF PROTEIN EXPRESSION BETWEEN BREAST AND COLORECTAL CANCERS

Abstract

Up to date, cancer remains a major health problem across the globe. Breast cancer is the most common cancer among women while colorectal cancer is the third most common cancer in both genders worldwide. In view of breast and colorectal cancer being among the most common types of cancer in Malaysia, we aimed to identify common biomarkers that can be used in the diagnosis or treatment of the cancers. This aim is accomplished by studying the protein expression profiles of breast and colorectal cancers tissues in comparison with their respective normal tissues. Twenty-five pairs of normal and cancerous tissues, respectively for breast and colorectal cancer were collected from Penang General Hospital (GH) with written informed consent obtained from the patients or their close relatives. Proteins were extracted from tissues by using sequential extraction method comprised of two extraction buffers, namely TRIS buffer and TLB buffer. The proteins were separated by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). By using the image analysis software, the proteome of each tissues type was compared between: i) breast cancerous and normal tissues; ii) colorectal cancerous and normal tissues and iii) breast and colorectal cancerous tissues. The differentially expressed protein spots were excised and subjected to LC-MS/MS analysis. In breast cancer, tropomyosin

isoform was up-regulated in all studied N stages of all ethnics. Calrecticulin (TLB), PDI A3 (TLB), Hsp 70 (TLB), Hsp 60 and PDI were detected up-regulated in N0 and N1 stage in all the ethnic groups, but not in N2 stage. PCA and LDA statistic test revealed that the collectively used of PDI A3 (TRIS), activation-induced cytidine deaminase, alpha-1-antitrypsin precursor, alpha-1-antitrypsin, chain A, haptoglobin PDI, PDI A3 (TLB), annexin A3, annexin V, heat shock 70 kDa protein 1A, calrecticulin and tubulin alpha chain can reliably indicate pathological state of breast cancer. In colorectal cancer, annexin A3 was found consistently up-regulated in the entire N stages studied (namely N0, N1 and N2) and all ethnics. The PCA and LDA statistic test revealed that alpha-1-antitrypsin precursor, alpha-1-antitrypsin, chain A, PDI A3 (TRIS), Hsp 90 (TRIS), tubulin, beta polypeptide (TRIS), tubulin alpha chain A, PDI A3, tubulin, beta polypeptide (TLB), ATP synthase, Hsp 70 (TLB), annexin A3, and tropomyosin alpha 3 chain can be collectively used to differentiate normal and cancerous state of colorectal cancer. At the same time, tubulin, beta polypeptide (TLB) and PDI A3 (TRIS) were detected up-regulated in both cancer types. Western blot on calrecticulin and PDI A3 was carried to validate the result of LC-MS/MS analysis and MASCOT database search engine. The usefulness of this particular protein as biomarker for colorectal and breast tissues is worth to be further evaluated by using a larger sample set before proceeded to subsequent cancer diagnosis and treatment intervention.

CHAPTER 1

INTRODUCTION

1.1 Cancer

1.1.1 Definition of cancer

Cancer is a class of clonal malignant disease characterized by uncontrolled cell growth, cell invasion and sometimes cell metastasis in body *via* lymph or blood. It is caused by abnormalities in genetic materials involved in cell proliferation, differentiation and cell death regulation. The genetic materials involved include oncogene, tumor suppressor gene, suicide gene and DNA-repair gene. Mutation or damage to these genes make the cell loss or have less ability to correct DNA damage and undergo programmatic cell death, resulting in mass of transformed cells to form cancer (Hejmadi, 2010, Petrova and Toncheva, 2008).

1.1.2 Carcinogenesis

Carcinogenesis is a multistage development of normal cells turning into cancer cells which include tumor initiation, promotion and progression (Vincent and Gatenby, 2008, Petrova and Toncheva, 2008). The tumor initiation is a malignant transformation of normal cells after exposure to carcinogen, such as chemical, radiation or virus. Carcinogen distorts cells by inducing DNA mutation, gene rearrangement or gene amplification to produce a genotypically altered cell. This initiation step is irreversible, permanent and heritable in the transformed cells

(Ruddon, 1994, Vincent and Gatenby, 2008). Tumor promotion is the stage that involves cell proliferation and clonal expansion of transformed cells upon stimulation of mitotic stimuli such as chronic inflammation. The stimulus is non-mutagenic, it caused cell proliferation enhancement and finally produces localized tumor that is non-malignant and with self limited growth properties. The gradual genetic change is occurred in this preneoplastic lesion (Ruddon, 1994, Vincent and Gatenby, 2008). On the contrary, tumor progression refers to formation of limitless or invasive growth of tumor. At this stage, the tumor is characterised by its karyotypic instability, aneuploidy, and chromosomal abnormalities (Ruddon, 1994). Figure 1.1 showed the development of a single normal cell to malignant tumor (Funes, 2002).

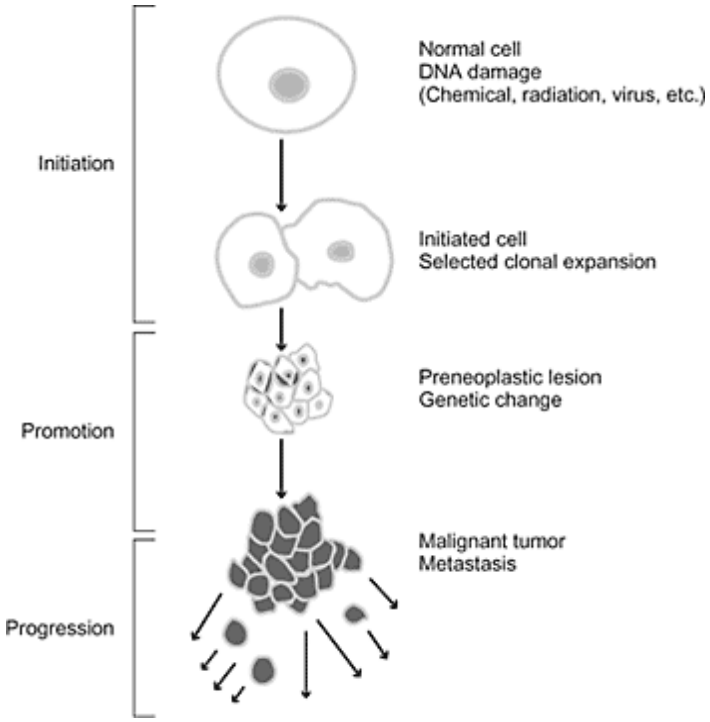

Figure 1.1 Development of cancer cells (Funes, 2002)

1.2 Breast cancer

1.2.1 Breast

Breasts are located in front of chest wall where extend from clavicle to the middle of sternum and continue into the axilla till sixth rib. The breasts are supported and hold in place by Cooper's ligaments. It is composed of fat that covered mammary glands that consisted of 15-20 glandular lobes. Each lobe possesses single lactiferous duct that further enlarge to form a small and spindle-shaped lactiferous sinus. The duct subdivides into smaller duct connect to a gland lobule. There is raised nipple surrounded by circular and pigmented areola at the external part of breast (Seeley *et al.*, 2006).

Axillary lymph nodes are classified into three levels according to their location relative to pectoralis minor muscle, which laid deep unto the pectoralis major muscle; level I lymph nodes are lateral to the border of muscle, level II is beneath to the muscle and level III is medial to the edge of the muscle. The intramammary nodes are located along the lateral edge of axillary tail and breast. The axillary, internal mammary, and intercostals veins are found parallel to the lymphatic pathway. Figure 1.2 showed the anatomy of female breast (Pass, 2008).

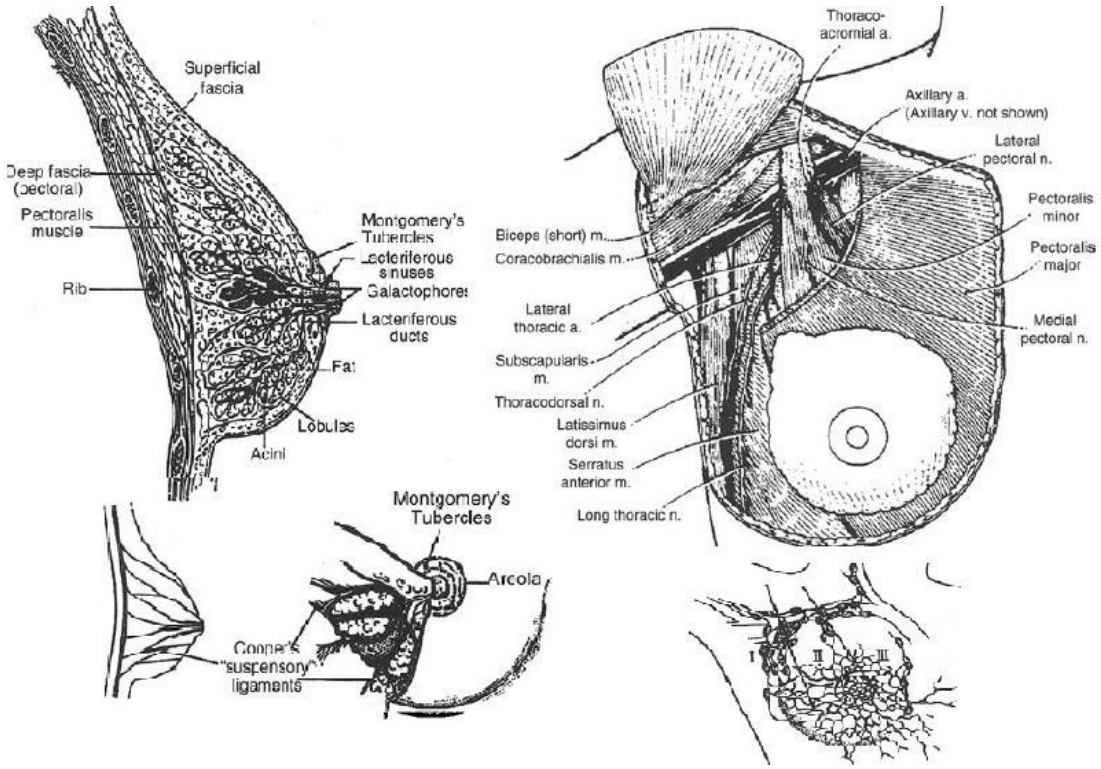

Figure 1.2 Anatomy of female breast (Pass, 2008)

1.2.2 Definition of breast cancer

Breast cancer is the cancer originated from breast tissues. It usually forms in lactiferous ducts or lobules that supply milk. There are several types of breast cancer that are categorized by their origin: ductal carcinoma in situ (DCIS), infiltrating ductal carcinoma (IDC), medullary carcinoma and infiltrating lobular carcinoma. Among these, infiltrating ductal carcinoma is the most common and aggressive form of breast cancer. It is originated and developed in the lactiferous duct, break through the duct tube and invade or infiltrate to surrounding tissue of breast. It is one of the invasive breast cancer where it can spread to other parts of the body through lymph and blood system (Kabbage *et al.*, 2008).

1.2.3 Incidence of breast cancer

Breast cancer is the most common type of cancer among women in Malaysia (Hisham and Yip, 2003) and worldwide (Agarwal *et al.*, 2007). The incidence was reported to be the lowest in Asia and Africa, intermediate in Mediterrance countries and South America, and the highest in North Europe and North America. In year 2000, about 3825 new cases and 1707 death due to breast cancer were reported in Malaysia. On the contrary, about 1,050,346 new cases were reported and 372,969 death occurred worldwide (Hisham and Yip, 2003). In Peninsular Malaysia, the incidence was highest among Chinese, followed by the Indian and Malay (Omar, 2006). The risk for breast cancer is increased with age, especially those women who used estrogen and progestin MHT after postmenopausal, overweight or obese after 18 years old, long menstrual history, physical inactivity, personal or family history of breast cancer, and inherited genetic mutation (Society, 2010)

1.2.4 Tumor, node, metastasis (TNM) staging of breast cancer

TNM staging system is one of the commonly used cancer staging systems. It is maintained by American Joint Committee on Cancer (AJCC) and International Union against Cancer (UICC). TNM staging is determined according to the result of physical examination, biopsy, and imaging tests. It is based on the size or extent of tumor (T), the extent of spreading to lymph nodes (N) and the presence of metastasis (M). It is useful to guide doctor in treatment option and prognosis (Society, 2010).

The classification of stage grouping and TNM staging system is showed in Table 1.1. The description of T, N and M of fifth edition TNM staging system is showed in Table 1.2.

Table 1.1 Classification of stage grouping and TNM staging system in breast cancer

AJCC Stage	TNM stage
0	Tis, N0, M0
IA	T1, N0, M0
IIA	T0 -1, N1,M0 T2, N0, M0
IIB	T2, N1, M0 T3, N0, M0
IIIA	T0 -2, N2, M0 T3, N1 -2, M0
IIIB	T4, N0-2, M0 Any T, N3, M0
IV	Any T, any N, M1

Table 1.2 Description of T, N and M of TNM staging system in breast cancer

Components		Description
T	Tis	Cancer cells within duct/lobules/not found
	T1	Tumor size ≤ 2cm
	T2	Tumor size ≥ 2cm and ≤ 5cm
	T3	Tumor size ≥ 5cm
	T4	Tumor grown into chest wall/skin
N	N0	No spread to lymph nodes
	N1	Spread to 1 to 9 axillary nodes
	N2	Spread to internal mammary nodes
	N3	Spread to 4 or more axillary lymph nodes, tiny amounts found in internal mammary lymph nodes
M	Mx	No description. Incomplete information
	M0	No spread to distant site
	M1	Spread to distant site

1.3 Colorectal cancer

1.3.1 Colon, rectum and appendix

Colon, rectum and appendix are parts of large intestine in digestive tract system. In digestive system, the water and mineral nutrients are absorbed from food in colon whereas the waste is then passed through rectum and expelled from anus.

Colon, a muscular tube about 5 feet long in large intestine, is located either in the abdominal cavity or behind the abdominal cavity in the retroperitoneum. It consists of four sections: ascending colon, transverse colon, descending colon and sigmoid colon (Society, 2008, Seeley *et al.*, 2006). The inner surface is aligned from serosa, muscularis, submucosa to mucosa with numerous straight tubular glands called crypts. The lymphatic nodule is located within a loose connective tissue called lamina propria in mucosa. The blood vessels and nerve is found between connective tissue layer and peritoneum in serosa. The rectum is the ~12 cm straight muscular tube begins at the end part of sigmoid colon and extended until anal canal. Appendix is a ~9 cm small blind tube attached to the caecum (Seeley *et al.*, 2006). Figure 1.3 showed the anatomy of colon, rectum and appendix (Seeley *et al.*, 2006).

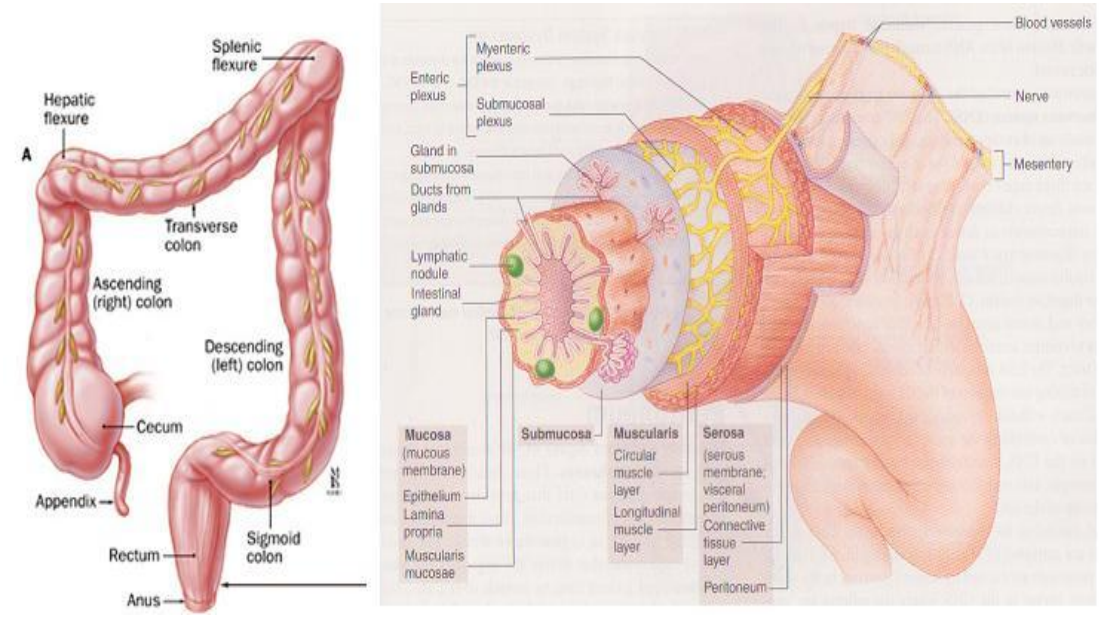

Figure 1.3 Anatomy of colon, rectum and appendix (Seeley *et al.*, 2006)

1.3.2 Definition of colorectal cancer

Colorectal cancer is cancer that grows in the colon, rectum or appendix. It is also called colon cancer or large bowel cancer. There are many forms of colorectal cancer which include adenocarcinoma, leimysarcoma, melanoma and neuroendocrine tumors. Adenocarcinoma is the most common type of colorectal cancer and represents 70% of all malignant disease developed in gastrointestinal tract system (Society, 2008, Kumar *et al.*, 1997). It refers to tumor originates from epithelial cell with gland pattern or derived from glandular tissue but not necessarily with gland pattern (Kumar *et al.*, 1997).

1.3.3 Incidence of colorectal cancer

In year 2006, 2866 colorectal cancer cases were registered with National Cancer Registry (NCR) and represent 13.2% of all cancer cases registered (Omar, 2006). The second report of NCR noted that colon cancer is ranked as the third most frequent cancer with rectal cancer being ranked as the fifth most common cancer for both genders in Malaysia (Balraj and Ruhana, 2007). The occurrence and mortality rate of colorectal cancer were higher in men than women (Society, 2008). In Peninsular Malaysia, the incidence was the highest among Chinese and the lower in Indian and Malay (Omar, 2006). The risk is increased with age (Balraj and Ruhana, 2007). Besides, personal or family history of colorectal cancer, familial adenomatous polyposis, chronic ulceration colitis, obese, diabetes and physical inactivity are also the risk factors for development of this malignant disease (Ruddon, 1994).

1.3.4 Tumor, node, metastasis (TNM) staging of colorectal cancer

TNM staging system is the most commonly used staging system standardized by American Joint Committee on Cancer (AJCC) and International Union against Cancer (UICC) for colorectal cancer. The T in TNM staging for colorectal cancer is different from breast cancer where it describes how far the primary tumor has grown into the intestine wall and nearby areas instead of tumor size in breast cancer (Society, 2010). The classification of stage grouping and TNM staging system for colorectal cancer is shown in Table 1.3. The description of T, N and M of fifth edition TNM staging system is shown in Table 1.4.

Table 1.3 Classification of stage grouping and TNM staging system in colorectal cancer

AJCC Stage	TNM stage
0	Tis, N0, M0
I	T1-2, N0, M0
IIA	T3, N0, M0
IIB	T4, N0, M0
IIIA	T1-2, N1, M0
IIIB	T3-4, N1, M0
IIIC	Any T, N2, M0
IV	Any T, any N, M1

Table 1.4 Description of T, N, and M of TNM staging system in colorectal cancer

Components		Description
T	Tis	Grown on mucosa
	T1	Grown through mucosa, submucosa and extends to muscularis
	T2	Grown through submucosa and extends to muscularis
	T3	Grown through muscularis and outermost layers of colon/rectum but not through it.
	T4a	Grown through serosa
	T4b	Grown through wall of colon/rectum, attach or invades into nearby tissues or organs
N	N0	No spread to lymph nodes
	N1a	Spread to 1 nearby lymph node
	N1b	Spread to 2-3 nearby lymph nodes
	N1c	Small deposits of cancer cells at adipose tissue near lymph nodes, but not lymph node.
	N2a	Spread to 4-6 nearby lymph nodes
	N2b	Spread to 7 or more nearby lymph nodes
M	Mx	No description, incomplete information
	M0	No spread to distant site
	M1a	Spread to 1 distand organ or set of distant lymph nodes
	M1b	Spread to more than 1 distant organ or set of distant lymph nodes/distant parts of peritonuem

1.4 Comparison between breast and colorectal cancer

Extensive works on biomarkers for cancer have been carried out by other researchers for many years. Nevertheless, this study is still worth to be carried out. The geographical variation in cancer incidence predominantly reflected the environmental influences on carcinogenesis, which include dietary, social and culture environment factor. This fact indicating that there is a need to identify the most beneficial biomarkers specific for Malaysian, making the treatment decision with more predictive evidence and diagnosis, and hence reduce the mortality due to cancer in Malaysia.

Currently, most cancer proteomic studies are restricted to individual cancer type. In general, this caused the biomarkers to be less useful for cancer staging, monitoring and treatment, where the identified biomarkers might not be related to the specific organ where tumor spread, growth and metastasis. In this study, we compare the proteins expression pattern between breast and colorectal cancer. Both are among the most common cancers in Malaysia (Hisham and Yip, 2003, Omar, 2006). There are differences in anatomic site, structure, cancer progression and histopathology type of tumor between both cancers. Breast cancer is classified based on Bloom-Richardson grading system (Grades) and TNM staging system(Bloom and Richardson, 1957, Cancer, 1997). Meanwhile, colorectal cancer is classified according to Dukes and TNM staging system (Dukes, 1980, Cancer, 1997). In order to compare both cancers, a common cancer classification system is employed. The

TNM staging system is chosen instead of Grades or Dukes classification. Unfortunately, the TNM staging system for breast cancer and colorectal cancer has certain classification difference. In general, T describes the tumor size while N the region of lymph nodes involved. Precisely, T refers to the size of tumor in breast cancer and the level of invasion into nearby structures in colorectal cancer; N denotes the lymph nodes location involved in breast cancer and the number of lymph nodes involved in colorectal cancer (Cancer, 1997).

Hence, the first stage of comparison was conducted regardless of the cancer stages. Secondly, the comparison was performed between the N0 and N1 stages for both cancers. This is because the N0 and N1 classification for both the cancer types are the same. Therefore, allowing a common ground to understand the protein expression changes within the stages. The study of protein expression pattern among ethnicity of patients was also included.

1.5 Cancer proteomic

1.5.1 Proteomic

Proteomic is a study of all proteins encoded by genome in a cell or tissue, which include structure and function of the proteins (Dwek and Rawlings, 2002, Karp and Lilley, 2007). It is the science that study disease biology with qualitative and quantitative comparison of proteomes under two or several different conditions (normal versus cancer)(Carpenter and Melath, 2003). In comparison to genomics,

proteomics have several advantages in that proteins are more reflective of the existing condition of the cell's microenvironment, where the levels of gene activity do not exactly correlate to the corresponding protein expression levels (DaoHai *et al.*, 2006, Kennedy, 2001, Laronga and Drake, 2007). The proteins can also picture a vast cadre of posttranslational modifications which affecting the protein stability, localization, interaction and function (Laronga and Drake, 2007). An organism is largely governed through the function of proteins rather than deciphered by gene alone. Thus, in combination with genomics, proteomics can provide a holistic understanding of the biology of cells, organisms, disease processes and response to treatment (Kennedy, 2001, Srivastava, 2005).

1.5.2 Protein structure and function

Proteins are polymer that made up of 20 or more different amino acids linked by peptide bond. The sequence of amino acids is determined by sequence of nucleotide encoded in genetic code. Amino acids are molecules consisted of amine group, carboxylic group and side chain that vary between amino acids. The key elements of the amino acid are carbon, hydrogen, oxygen, and nitrogen. A short polymer that made up of less than 20 amino acids is termed peptide instead of protein. Peptide bond is also known as amide bond, where a chemical bond formed between carboxylic acid and an amino group with the expulsion of water molecule (Petsko and Ringe, 2004).

There are four levels of protein structure. The primary structure is referred to the amino acid sequence in linear form. The secondary structure is the repeating amino acid sequences in either alpha helix, beta sheet or turns form. The tertiary structure is the overall shape of a single protein molecule, and finally, several protein molecules fold to form quaternary structure. There are mainly three group of protein characterized according to their function, which include globular protein (enzyme), fibrous protein (structural), and membrane protein (receptor) (Petsko and Ringe, 2004).

1.5.3 Biomarker

Cancer biomarkers discovery is one of the major interests in cancer study. Biomarkers are substances that function as indicator in a biological process. It can be genes, proteins, small molecules or metabolites (Hale *et al.*, 2003, Issaq and Blonder, 2009). During the transformation of normal cell into a cancer cell, changes occur at protein level due to distorted expression, differential protein modification, changes in metabolite activity, and aberrant localization. Proteome analysis yields information of these changes and facilitates finding of biomarkers (Issaq and Blonder, 2009, Somiari *et al.*, 2005). More specifically, these cancer biomarkers refer to proteins that indicate changes in expression due to the abnormal condition or progression of cancer (Issaq and Blonder, 2009).

The biomarkers are critical for early detection, diagnosis, treatment, monitoring, and prognosis of disease (Srinivas *et al.*, 2002). A biomarker that can accurately detect the onset of disease can facilitate diagnosis, permit choice of therapy in its pre-invasive stage and consequently reduces the suffering and cost associated with the disease (David and Evelyn, 2011, Alfonso *et al.*, 2005). The effectiveness and progress of treatment can be monitored with it capability to indicate the abnormal condition or progression of cancer. It also can be a potential therapeutic target for chemotherapy (Dwek and Rawlings, 2002). Furthermore, biomarkers aid drug development, as it indicates the effect of drug during treatment of disease (Hale *et al.*, 2003).

The drawback of early detection of cancer using protein biomarker is the reliability of the expression of the biomarker itself. There were instances where false-positive results that were due to differential carcinogenesis development pattern among population subgroups, and thus resulting in over diagnosis or overtreatment on patients. Hence, the study on protein expression pattern within N stages and ethnics was carried out in order to identify potential biomarkers that more specific for each cancer stages and ethnic of the patients, consequently facilitate and optimized personal care on each patient.

1.5.4 Proteome analysis of cancer and adjacent normal tissue

The most common and effective approach for biomarkers discovery is the differential analysis of protein profiles between cancer and adjacent normal tissues (Bai *et al.*, 2010, Kim *et al.*, 2009, Luo *et al.*, 2005, Wulfkuhle *et al.*, 2002, Somiari *et al.*, 2005). Cell lines are common samples used in cancer proteomics due to its known genetic background and consistent cellular composition. However, the growth and development of the cancer cells is not only depending on their malignant potential. There are interactions within cells, among cells and between cells and agents in the microenvironment (Deng *et al.*, 2006). In order to obtain better information of protein expression, tissues specimens were used instead of cell lines in this study. By comparing the normal and cancerous tissues, the changes of proteins expression related to cancer cells development can be recognized (Kennedy, 2001).

1.5.5 Proteomic Research in Cancer

Numerous studies have been performed by researchers on different types of cancer using proteomic approach. The 2D-PAGE coupled with mass spectrometry, isotope coded affinity tags (ICAT), multidimensional protein identification technology (MudPIT), protein array technology and surface enhanced laser-desorption ionization-time of flight (SELDI-TOF) are among the common technologies applied in proteomic study (Dwek and Rawlings, 2002, Somiari *et al.*, 2005, Srivastava, 2005).

The proteomic analysis of breast normal and cancerous tissue (DCIS) was first reported by Wulfuhle *et al*., (2002). They identified 57 differentially expressed proteins between normal and DCIS, these proteins included VDAC, transgelin, Hsp 27(45), GRP78 (46), and cathepsin. In year 2005, Somiari *et al.*, reported that annexin V and HSP 90 were differentially expressed in DCIS, whereas carbonic dehyrase, protein disulfide isomerase, gelsolin and fibrinogen beta chain were differentially expressed in IDC. Meanwhile, Luo *et al*., (2005) discovered 10 up-regulated proteins which included manganese SOD, biliverdin reductase B, carbonic anhydrase I and annexin I in IDC. In addition, they also reported down-regulated proteins which included fatty acid binding protein 4, cofilin 1, profiling 1 and uracil DNA glycosylase in the same type of cancer. They proposed the potential used of these markers for diagnostic and also to determine the effectiveness of therapy against cancer. Deng *et al*., (2006) reported that alpha-1-antitrypsin, EF-1-beta, cathepsin D, TCTP, SMT3A, and PSMA1 as candidate biomarkers for patients with breast cancer.

In year 2006, Roth *et al*., (2006) reported a number of differential expression proteins between normal and colorectal cancerous tissue, comprised of 18 up-regulated proteins and 13 down-regulated proteins. Among the up-regulated proteins were NADH-ubiquinone oxireductase B_{18}, β-subunit of F_0F_1-ATPase, liver fatty acid-binding protein, soluble carrier family 5 Na^+-glucose cotransporter member 1, insulin-like peptide 5 and gluthathione transferase M_4. Among the

down-regulated proteins reported were mitochondrial transcription factor 1, lactate dehydrogenase, p28 BAB31, glucosamine-6-sulfatasse, caspase-9(Roth *et al.*, 2006). Kim *et al*., (2009) reported 3 novel WD-repeat-motif-bearing proteins that involved in actin remodeling, which included Arp2/3 complex subunit 2(p34-Arc), coronin-1A and WD-repeat protein 1(Wdr1). These proteins shown significant differ expression between normal and cancerous tissue when analyzed using Western blotting approach. Another study that evaluated the potential biomarkers or targets for novel therapeutic intervention of colorectal cancer was conducted by Dao Hai *et al*., (2006). Their study showed that the oncoprotein hnRNP A1 was highly expressed (80%) in cytoplasm of tumors. It could be used as a potential biomarker for treatment. Followed by the proteome analysis of adenocarcinoma and adjacent normal tissue, TAF9 RNA polymerase II (TAF9) and cytokine inducible SH-2 containing protein (CISH) were discovered as biomarker in colorectal cancer (Krasnov *et al.*, 2009). In year 2010, Bai *et al*. reported that carbonic anhydrase II and protein disulfide isomerase were down-regulated in colorectal cancer. On the other hand, APC-stimulated guanince protein 2B, glutathione S-transferase A3, Arginase and zinc finger protein 64 homolog were up-regulated in the same type of cancer.

1.6 Proteomic approach

1.6.1 Protein extraction

Sequential extraction method is an extraction method comprised of two extraction buffers to extract proteins in sequential order. The first buffer, namely Tris

buffer (TRIS), is used to extract highly soluble (hydrophilic) cytoplasmic proteins whereas the poorly soluble (hydrophobic) membrane-associated proteins are extracted using thiourea lysis buffer (TLB), the second extraction buffer in sequential extraction method. The TLB contains more potent chaotropes and detergents (Molloy *et al.*, 1998) when compared with TRIS. Sequential extraction method limits the number of protein extracted in each fraction. Therefore, it allows better separation and visualization of proteins in gel (Cordwell, 2008).

1.6.2 Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)

The 2D-PAGE is the most widespread tool for proteomic study due to its ability to map proteome and detect protein expression and post-translational modifications (Weiss and Görg, 2009). In first dimension separation, the proteins with difference isoelectric point (pI) are rehydrated into an immobilized pH gradient (IPG) strip and separated by isoelectric focusing (IEF). The pI is pH at which a protein carried net (zero) electrical charges in an electric field (Carpenter *et al.*, 2004). It is based on the principle that when a protein is placed in a medium with pH gradient and applied with electric forces, it moves toward the electrode with opposite charge. The proteins either pick up or lose protons until reach their pI and stop migrating (Garfin, 2001, Angelika *et al.*, 2000). In second dimension separation, the proteins are separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, according to molecular weight (Mw) (Weiss and Görg, 2009). The percentage of resolving gel used is 10% in order to resolve protein mass range from 30-150 kDa.

The proteins on gel are visualized as protein spots by Bio-SafeTM Colloidal Coomasie Blue G-250 stain (Garfin, 2001).

1.6.3 In-gel digestion

The protein spots of interest are excised from the gel and subjected to in-gel digestion. In-gel digestion is a sample preparation step for tandem mass spectrometric identification of protein identity after gel electrophoresis (Gam and Aishah, 2005, Rosenfeld *et al.*, 1992). It involves four steps which include desalting and destaining of gel, reduction and alkylation of cysteines in protein, proteolytic cleavage of protein using TPCK-treated trypsin and passive elution of the generated peptides from gel pieces. The eluted peptides are then analyzed by mass spectrometry (Resing and Ahn, 2004). Subsequently, the characteristic information of protein can be obtained. This method was first introduced by Rosenfeld *et al*., (1992).

1.6.4 Reverse phase high performance liquid chromatography tandem mass spectrometry (LC-MS/MS)

Reverse-phase (RP) mode HPLC is the most common approach for the separation of biological compounds. In HPLC analysis, the sample is applied onto a column filled with stationary phase. Then, the components in sample are partitioned onto the sorbent material and separated by stream of liquid mobile phase. For RV-HPLC, the stationary phase is made up of nonpolar matrix and mobile phase is

polar solvent (Dass, 2001). Tandem mass spectrometry is a branching application of MS, which capable of selecting an ion formed from a molecules or mixture of molecule through collision-induced dissociation and subsequently generate characteristic secondary fragment ions (Futrell, 2000). The peptide MW and partial sequences of amino acid obtained from MS data can be utilized in search of protein Swiss-Prot and NCBRnr database through Mascot Protein Database (MSDB) search engine (www.matrixscience.com) and Expasy proteomic server (expasy.org) for protein identification (Pyrek, 1997).

1.6.5 Western blot

The identity of protein identified by mass spectrometry analysis can be verified by Western blotting experiment. It is an immunoblot technique for detection of specific proteins in biological sample (Heidebrecht *et al.*, 2009). It involved electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to a nitrocellulose membrane, incubation with primary antibody specific to targeted proteins, and followed by enzyme linked secondary antibodies against primary antibody. When exposed to the substrate, colorimetric reaction occurred and the interaction between protein and specific antibody is visualized as color on the membrane (Burnette, 1981, Towbin *et al.*, 1979). The use of nitrocellulose membrane was started by Towbin *et al*. (1979). The name Western blot was given by Burnette (1981). There are two apparatus to transfer proteins from gel to nitrocellulose membrane, which are buffer-filled tanks (Towbin *et al.*, 1979) and semi-dry

(Lauriere, 1993) transfer tool. The apparatus used in this study was semi-dry transfer, where the gel and membrane were sandwiched horizontally between 2 stacks of buffer-wetted blotting paper within 2 closely spaced solid plate electrode in semi-dry transfer unit (Walker and Gravel, 2002).

1.6.6 Statistical test

1.6.6.1 Wilcoxon paired-sample test

Wilcoxon paired-sample is a univariate statistical method introduced by Wilcoxon (1945) to test the significant of means differences of paired samples. It is a non-parametric test with no assumption about the underlying statistical distribution, with at least six paired-samples are needed per class. The null hypothesis is that the population median of the paired different of the two samples is zero (Wilcoxon, 1945). In our case, the paired sample was referred to cancerous and normal tissue from the same patient. If the calculated p-value was below 0.05, then the changes of protein expression levels between normal and cancerous tissues were statistically significant. This test method is available in PDQuest software, which is utilized for 2D gel images analysis.

1.6.6.2 Principal component analysis (PCA)

Followed by the sample processing, proteins extraction and separation, 2D gel image and MS analysis, a high dimension of data is created. Therefore, the PCA, a multivariate statistical method is utilized to facilitate the data processing (Muir *et*

al., 2007). The theory of PCA was first introduced by Pearson (1901) and then further developed by Hotelling (1933). It allows the unsupervised data dimension reduction and representation of the original dataset with original variables into a new reference system with new variables called Principal Components (PCs) (Hotelling, 1933, Kim *et al.*, 2009, Marengo *et al.*, 2007, Pearson, 1901, Rodriguez-Pineiro *et al.*, 2007). The PCs are clustered and measured in hierarchical pattern, where the first PCs represent the maximum data variance, the second PCs accounted for the maximum residual data variance and so on (Marengo *et al.*, 2007). In our study, PCs were referred to the identified protein, where the first PCs represented the most differentially expressed protein between normal and cancerous tissue, the second and third PCs referred to other differentially expressed proteins in decreasing order, and so on. Subsequently, further supervised data reduction can be performed by selecting first PCs and preceded to LDA statistical analysis.

1.6.6.3 Linear discriminant analysis (LDA)

LDA, which is also known as Bayesian classification method, was first introduced by Fisher (1936). It is another multivariate statistical method that characterizes the groups, assigning or classifying the cases and measuring the degree of success of the classification model (Fisher, 1936, Marengo *et al.*, 2007, Rodriguez-Pineiro *et al.*, 2007). In LDA, original variables from PCA are transformed into new independent variables called discriminant functions (DF). The projection of the data in relation to the DF gives discriminant score. The DF accounts

for the maximum amount of separation of the different groups. Finally, LDA classifies observations into their respective groups based on their discriminant scores, where the groups are mutually exclusive and collectively exhaustive (Sharma and Mukherjee, 1996). Wilcoxon paired-sample test, PCA and LDA statistical test results are taken into consideration to determine whether the protein is a reliable marker and allow the differentiation of normal and cancer state

1.7 Research objectives

In view of breast and colorectal cancers are the most common types of cancer in Malaysia, we aimed to explore the common differential expressed proteins for breast and colorectal cancers using cancerous tissues obtained from human patients. These proteins will be valuable for treatment and diagnosis of both cancers. Furthermore, comparing the protein expression profile will lead to better understanding of the similarity and differences between these two cancers. The objectives of this study are:

a) To map the proteomes of breast and colorectal cancer tissues and their respective adjacent normal tissues by using 2D-PAGE.

b) To compare the differential protein expression on 2D gels.

   i. Breast normal and cancerous tissues according to categories of;

      - All patients
      - Lymph node metastasis
      - Ethnics group of patients

ii. Colorectal normal and cancerous tissues according to categories of;

- All patients
- Lymph node metastasis
- Ethnics group of patients

iii. To identify common and differentially expresses proteins in breast and colorectal cancers according to categories of;

- All patients
- Lymph node metastasis
- Ethnics group of patients

c) To identify the targeted protein using In-gel digestion and LC-MS/MS.