IDENTIFICATION OF BIOMARKERS FOR COLORECTAL CANCER FROM MALAYSIAN PATIENTS

by

YEOH LAY CHIN

Thesis submitted in fulfillment of the

requirements for the degree of

Master of Science

June 2010

ACKNOWLEDGEMENT

Firstly, I would like to extend my gratitude to my supervisor, Assof. Prof. Dr Gam Lay Harn for her help and guidance throughout my study. I would also like to my co-supervisors Assoc. Prof. Dr Shaharum Shamsuddin and Dr Khoo Boon Yin who offered their assistance and experience during the course of my study.

I would like to thank the Institute of Postgraduate Studies, Universiti Sains Malaysia for the Fellowship Scholarship Scheme which helped to finance my studies.

I would like to extend my gratitude to the staffs of the School of Pharmaceutical Sciences and the National Poison Centre for assisting me by providing facilities, equipment and assistance to my work. I would like to thank En. Hamid, En. Basri Jaafar, En. Fizal Jamaludin and En. Tan Siao Peng.

I would also like to thank the staff and surgeon of Penang General Hospital for providing the colonic tissue samples.

Finally, I would like to record my special thanks to Dr Saravanan Dharmaraj for helping me to analyze the statistics part of my study. I would also like to thank my labmates for their assistance, advice and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D-PAGE	: Two-dimensional polyacrylamide gel electrophoresis
ACN	: Acetonitrile
AEBSF	: 4-(2-Aminoethyl) benzenesulfonyl fluoride
AP	: Alkaline phosphatase
APS	: Ammonium persulfate
BPC	: Base peak chromatogram
BSA	: Bovine serum albumin
CHAPS	: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CID	: Collision induced dissociation
4-CN	: 4-Chloronaphthol
CRC	: Colorectal cancer
DAB	: Diaminobenzidine
DTT	: 1,4 – Dithiothreitol
ECM	: Extracellular matrix
ESI	: Electrospray ionization
FAP	: Familial adenomatous polyposis
HNPCC	: Hereditary non polyposis colorectal cancer
HPLC	: High performance liquid chromatography
HRP	: Horseradish peroxidase
IEF	: Isoelectric focusing
IHC	: Immunohistochemistry

IPG	: Immobilized pH gradient
kDa	: kilo Dalton
LDA	: Linear discriminant analysis
LC-MS/MS	: Liquid chromatography tandem mass spectrometry
MS	: Mass spectrometry
m/z	: Mass to charge ratio
NaCl	: Sodium chloride
$Na_2HPO_4.2H_20$	: Disodium hydrogen phosphate dihydrate
$NaH_2PO4.2H_20$	: Sodium dihydrogen phosphate dihydrate
NH_4HCO_3	: Ammonium bicarbonate
PAGE	: Polyacrylamide gel electrophoresis
PBS	: Phosphate buffered saline
PC	: Principal component
PCA	: Principal component analysis
pI	: Isoelectric point
RP-HPLC	: Reverse-phase high performance liquid chromatography
SDS	: Sodium dodecyl sulfate
SPSS	: Statistical Package for Social Science
TCA	: Trichloroacetic acid
TEMED	: N, N, N' N' - tetramethylethylenediamine
TLB	: Thiourea lysis buffer
TRIS	: Tris buffer
Tris	: Tris(hydroxymethyl)aminomethane

LIST OF APPENDICES

PENGENALPASTIAN PENANDA-PENANDA BIO BAGI KANSER KOLOREKTAL DARIPADA PESAKIT-PESAKIT MALAYSIA

ABSTRAK

Kes-kes kanser kolorektal (CRC) semakin meningkat. Akan tetapi, penanda-penanda bio yang dapat menggambarkan perkembangan CRC masih kekurangan walaupun kemajuan dalam alat-alatan proteomik dan "microarray" telah tercapai. Tujuan kajian ini adalah untuk mengenalpasti penanda-penanda bio berpotensi bagi diagnosis CRC dalam kalangan pesakit-pesakit Malaysia. Dalam kajian ini, pengekstrakan turutan dijalankan dengan menggunakan larutan tampan Tris dan diikuti dengan larutan lisis tiourea (TLB), hal ini adalah untuk mengurangkan kerumitan proteom CRC. Tisu-tisu mukosa normal dan kanser diperolehi daripada spesimen pembedahan 26 pesakit CRC. Protein-protein yang diekstrak dipisahkan oleh elektroforesis gel dwi-dimensi. Tompok protein-protein yang diekspres pada kadar yang berbeza dipencilkan untuk analisis pengenalpastian protein oleh kromatografi cecair tandam spektrometri jisim (LC-MS/MS). Kami telah mengenalpasti 37 protein hidrofilik dalam ekstrak Tris dan 24 protein hidrofobik dalam ekstrak TLB daripada sampel-sampel yang dikumpul. Leukocyte elastase inhibitor (LEI) dan 60 kDa heat shock protein (HSP60)

merupakan penanda bio yang berpotensi untuk CRC dalam ekstrak Tris manakala triosephosphate isomarase (TPI) dan annexin A3 merupakan penanda bio yang berpotensi bagi CRC dalam ekstrak TLB. Di samping itu, LEI ialah penanda bio berpotensi bagi pesakit CRC pada peringkat III yang berkombinasi dengan adenokarsinoma pembezaan sederhana (MDA), adenokarsinoma pembezaan baik (WDA) ataupun metastasis nodus limfa. Fibrinogen beta chain (FIBB) merupakan penanda bio yang baik bagi pesakit-pesakit CRC peringkat II dan MDA. Kami juga menemui corak ekspresi protein yang khusus dalam kalangan ketiga-tiga bangsa. TPI dikenalpasti sebagai penanda bio yang berpotensi bagi pesakit CRC pada peringkat III yang berkombinasi dengan WDA, MDA ataupun metastasis nodus limfa. Tisu-tisu CRC peringkat II dan MDA menunjukkan bahawa annexin A3 ialah penanda bio yang berguna. TPI dan annexin A3 diekspres secara umum dalam ketiga-tiga bangsa (Cina, Melayu dan India), hal ini mencerminkan kepentingan penanda-penanda bio ini dalam kalangan pesakit Malaysia. Keputusan pewarnaan-imun pada calreticulin dan heat shock 70 kDa protein (HSP70) pada sel HT29 dan HCT116 menunjukkan kedua-dua protein tersebut terletak pada permukaan luar membran sel. Analisis komponen utama (PCA) dan analisis pembezalayan linear (LDA) mencerminkan kenyataan protein-protein yang diekspres dalam kadar berbeza untuk menggambarkan keadaan kanser dalam tisu-tisu.

IDENTIFICATION OF BIOMARKERS FOR COLORECTAL CANCER FROM MALAYSIAN PATIENTS

ABSTRACT

Colorectal cancer (CRC) is on the rise. However, biomarkers for indication of CRC progression is still lacking despite several advances in proteomic and microarray tools. This study aims to identify potential biomarkers for the diagnosis of CRC among Malaysian patients. In this study, sequential extraction was first performed using Tris buffer followed by TLB buffer in order to reduce the complexity of the CRC proteome. The tissues used were surgical resections of normal mucosa and cancerous tissues from 26 CRC patients. The extracted proteins were separated by 2D gel electrophoresis. The differentially expressed protein spots were then mapped for further protein identification by the use of liquid chromatography tandem mass spectrometry (LC-MS/MS). We have identified 37 hydrophilic proteins in the Tris extracts and 24 hydrophobic proteins in the thiourea lysis buffer (TLB) extracts from the samples collected. Leukocyte elastase inhibitor (LEI) and 60 kDa heat shock protein (HSP60) were the potential biomarkers for CRC analyzed from the Tris extracts whereas triosephosphate isomerase (TPI) and annexin A3 were the potential biomarkers

for CRC analyzed from the TLB extracts. Moreover, LEI was the potential biomarker for CRC patients at stage III in combination with either moderately differentiated adenocarcinoma (MDA), well differentiated adenocarcinoma (WDA) or lymph node metastasis CRC. Fibrinogen beta chain (FIBB) was a good biomarker in CRC patients at stage II and MDA. We also noted distinctive protein expression patterns among the three races. TPI was the potential biomarker in CRC patients at stage III in combination either with MDA, WDA or lymph node metastasis CRC. Stage II and MDA CRC tissues revealed annexin A3 as a useful biomarker. TPI and annexin A3 were found commonly expressed in the three races (Chinese, Malay and Indian), reflecting the importance of these biomarkers in CRC amongst Malaysian patients. Immunostaining results of calreticulin and heat shock 70 kDa protein (HSP70) on HT29 and HCT116 colon cell lines revealed the cellular location of these proteins on the external surface of cell membranes. Principal component analysis (PCA) and linear discriminant analysis (LDA) reflected the significance of these differentially expressed proteins in indicating the state of cancer in tissues.

CHAPTER 1

INTRODUCTION

1.1 Cancer

Cancer is a complex and unpredictable genetic disorder (Cornelisse, 2003) which causes abnormal growth of cells. It is the result of the accumulation of genetic mutations that allows tumour cells to grow. Cancer is characterized by uncontrolled cell proliferation, division and disruption between apoptosis by loss of cell cycle control, sustained angiogenesis and increase of cell ability to invade other tissues and metastasize (Hanahan and Weinberg, 2000; Søreide *et al.* 2006). Today, cancer is one of the main causes of global mortality.

1.1.1 Carcinogenesis

Carcinogenesis is a multi-step progression of cancer development. It is divided into three main stages; initiation, promotion and progression (Chu *et al.*, 2007). In the initiation stage, normal cells are exposed to carcinogens for a period of time, which causes damage to their genetic information and predisposes them to become cancerous. In colorectal cancer, the carcinogenic agents include tobacco, alcohol and diet. However, the initiated cells usually do not immediately

become cancerous. In the promotion stage, the initiated cells are induced to grow and divide rapidly due to exposure to promoting agents that disrupt cells' normal growth cycle. In the progression stage, the cells will undergo further transformations which will result in autonomous growth and division, and become more aggressive. These cancerous cells will be able to grow and divide rapidly and spread to other parts of the body.

1.1.2 Metastasis

Metastasis is the process when cancer cells travel from the primary site (organ or tissue) to other parts of the body, where they influence the secondary site to divide abnormally. Cancer cells that metastasize exhibit several characteristics; they have the capability to invade, enter and exit the vasculature, deregulated adhesion, avoid being killed by immune cells and escape anoikis, which is the programmed cell death that is caused by the loss of cell matrix interaction (Stein and Schlag, 2007). Metastasis is the major cause of death in cancer.

The metastatic cascade occurs in several steps. Firstly, the cancer cell detaches itself from the primary mass by loss of adhesion and proteolysis of the extra-cellular matrix (ECM). The cancer cell then migrates through the intracellular space to reach the blood or lymphatic vessels. They invade the vessel

walls (intravasation) and enter the blood circulation. When the cell arrives at a distant site, they arrest in the capillary beds, breach the vessel wall again and adhere to secondary sites (extravasation). There, they will proliferate at the new site (Stein and Schlag, 2007). The metastatic cascade model is shown in Figure 1.1.

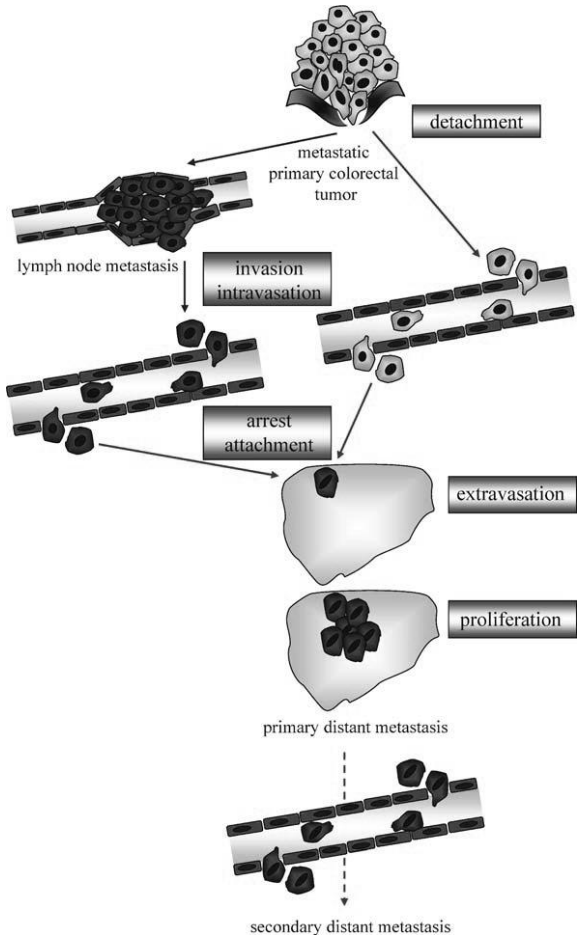

Figure 1.1 Model for the metastatic cascade of colorectal cancer metastasis (Stein and Schlag, 2007).

1.2 Colon and Rectum

The colon and rectum form the lower part of the human gastrointestinal tract. They consist of the caecum, ascending, transverse, descending colon, sigmoid colon, rectum and appendix. Colorectum, also known as the large intestine or large bowel, is a 5 feet long muscular segment. The colon serves to absorb water and nutrients from partially digested foods that enter into the colon from the small intestine. The caecum is located in the beginning of the large intestine and joins to the ascending colon on the right side of the abdomen. At one point, the hepatic flexure turns towards the midline to form the transverse colon. The transverse colon across the spleen turns downward at the splenix flexure. At the left side of the abdomen, the descending colon turns into S-shape sigmoid colon. The rectum is the last section of large intestine, passing the waste (stool) out from the body. It connects the end of the sigmoid colon to the anus. There are four layers of tissues that make up the colon; mucosa, submucosa, muscularis, and serosa. The muscularis is further divided into muscularis mucosa that is located between the mucosal and submucosal layers, while the muscularis propria is located in the middle of submucosal and serosa (John Hopkins Medicine, 2010). Colon cells reproduce in the crypts of Lieberkuhn found in the mucosal layer. When new cells are generated, the old matured colon cells tend to migrate out from the crypts and die off. Any damages to the crypt could bring harm to the cells and cause the formation of adenomas (Griffin-Sobel, 2001).

During a lifetime, colon cells may transform, leading to colorectal carcinoma. The anatomy of the colon is shown in Figure 1.2 (John Hopkins Medicine, 2010).

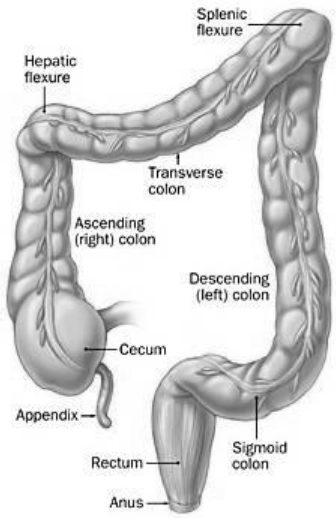

Figure 1.2 Anatomy of the colon and rectum (John Hopkins Medicine, 2010).

1.3 Colorectal Cancer

Colorectal cancer (CRC) includes the cancer of the colon, rectum and appendix. The mechanism of how CRC developed has yet to be understood, however, it is possible that an adenomatous polyp grows on the mucosal epithelium, the polyp slowly develops into colorectal carcinoma (Griffin-Sobel, 2001). Individuals with a family history of colon cancer or have inflammatory bowel disease, or cancer in the uterine, ovarian, ureter or bladder have a higher risk of getting colorectal cancer (Sweed and Meropol, 2001). Inherited syndrome, environmental factors, sedentary lifestyle (alcohol consumption and smoking)

and diet (low folate and fibre intake, high fat, red meat consumption) also increase the chance of developing colorectal cancer (Willett, 2001). Patients suffering from inflammatory bowel disease have a tenfold risk of developing CRC. Those with a family history of CRC, including their first degree relatives have an increased risk of contracting CRC. Patients that carry hereditary conditions such as Familial Adenomatous Polyposis or Hereditary Non Polyposis Colorectal Cancer have a higher risk of developing colorectal cancer.

Early colorectal cancer is usually asymptomic. In early stages, an unusual or continuing fatigue, weight loss will be experienced by the patient. When the tumours are in an advanced state, symptoms such as abdominal pain, diarrhea, and changes in bowel movement and blood in stool can be clearly observed in CRC (Sweed and Meropol, 2001).

Screening and surveillance of CRC substantially reduce both the incidence and mortality of CRC (Harford, 2006). CRC screening for individuals is recommended at the age of 50. Screening of CRC in normal-risk individuals can be done by several methods; faecal occult blood test (FOBT), colonoscopy, sigmoidoscopy, double contrast barium enema and digital rectal exam. Newer technologies for screening CRC include virtual colonoscopy, faecal immunochemical testing for blood and faecal DNA testing.

The treatment of CRC depends on the location of the tumour and stage of the disease. Colon and rectal cancer diagnosed at an early stage is treated by

surgical removal of the involved colon. Radiotherapy, chemotherapy and targeted therapy may also apply to the metastatic colorectal cancer.

Adenocarcinoma is the most common type of colorectal cancer (American Cancer Society, 2010a). Other histological types include adenoma *in situ*, mucinous carcinoma, signet ring carcinoma, squamous cell (epidermoid) carcinoma, adenosquamous carcinoma, small cell carcinoma, undifferentiated carcinoma and carcinoma. Tumours in the ascending colon resemble cauliflower-like fungating masses that become ulcerative and necrotic. These tumours are usually well-differentiated. However, tumours in the descending and sigmoid colon are ulcerative and are likely to penetrate the bowel wall. Rectosigmoid tumours occur as villous or frondlike lesions (Griffin-Sobel, 2001).

1.3.1 Incidence of Colorectal Cancer

CRC is the third most common cancer in the world, accounting for 8.6% of deaths worldwide. CRC is the fourth most frequent type of cancer affecting both males and females in the world (World Health Organization, 2009). The estimated incidence and estimated deaths each year from CRC is approximately 1.16 million and 630,000, respectively (Garcia *et al.*, 2007). In Malaysia, CRC was responsible for 14.2% of male cancers and 10.1% of female cancers, making it the most common cancer among men and the third most common cancer

among women (Lim and Halimah, 2004). In 2003, 1335 cases of CRC in males and 1217 cases of CRC in females were reported. CRC is the third most common cause of cancer-related death in Malaysia and the Chinese have the highest incidence of CRC (Malaysian Oncological Society, 1999; Lim and Halimah, 2004).

1.3.2 Colorectal Carcinogenesis

Colorectal cancer (CRC) is a well studied cancer model (Schulmann *et al.*, 2002), a small lesion slowly develops into premalignant lesions and to invasive cancer over years. This progression is caused by genetic changes, such as activation of protooncogenes and inactivation of tumour suppressor genes, which result in tumorigenesis in CRC. These molecular events are described in the colon cancer adenoma-carcinoma sequence model by Fearon and Vogelstein (1990). Two distinct pathways have been identified; the microsatellite instability (MSI) and the chromosomal instability pathways (CIN).

Adenoma-carcinoma sequence represents the stepwise progression from early lesions to carcinomas in a linear way (Stein and Schlag, 2007). This sequence is displayed in Figure 1.3. CRC is initiated by the growth of benign polyps. Polyps refer to a mass of cells that grow above the surface of surrounding normal mucosa. Inactivation of both alleles of the APC gene located on

chromosome 5 and activation of mutations in the β-catenin gene cause the onset of colorectal tumorigenesis. This is followed by the activation of the K-ras protooncogene through mutation of codon 12 or 13. The progression from adenoma to intermediate adenoma and then to carcinoma involves the loss of heterozygosity (LOH) of the long arm of chromosome 18q, near the SMAD4 (DPC4) locus. In addition, mutations in DCC and SMAD2 tumour suppressor genes may also be involved. At the late stage of tumorigenesis, mutation of p53 on chromosome 17p occurs, which is thought to allow tumour growth to evade cell cycle arrest and apoptosis. Mutation of DNA mismatch repair genes, including hMSH2, hMLH1, hPMS1 or hPMS2 also plays a role in CRC formation (Tejpar and van Cutsem, 2002; Stein and Schlag, 2007).

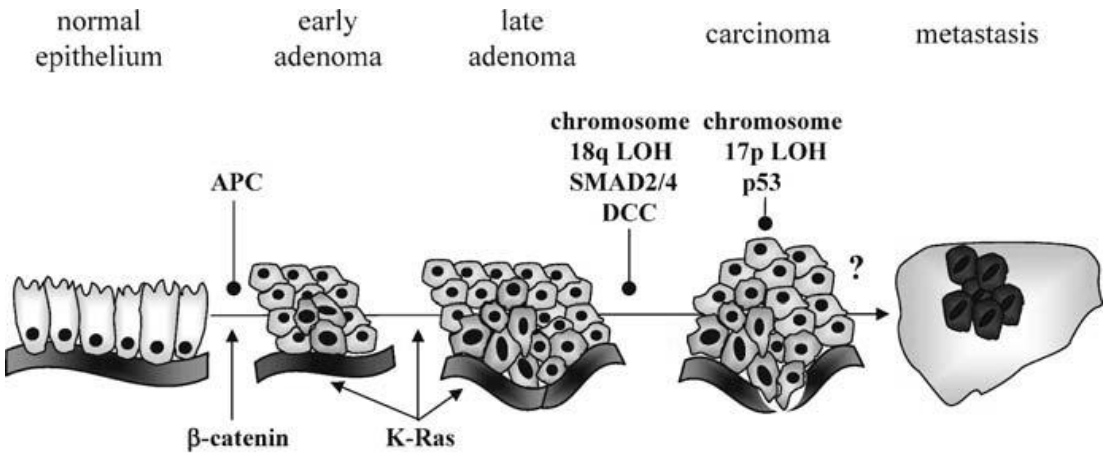

Figure 1.3 Adenoma-carcinoma sequence model of colorectal cancer (Stein and Schlag, 2007).

Colorectal polyps can be either neoplastic (adenomatous) or non-neoplastic (hyperplastic and hamartomous) (Schulmann *et al.*, 2002). In most cases, adenoma polyp remain benign and asymptomic lesions under endoscopy. However, these small lesions may evolve into malignancy and there are evidences indicating that the majority of colorectal carcinomas develop from adenomatous polyps (Konishi and Morson, 1982). Hamartomatous polyps have lower tendencies to malignant transformation whereas hyperplastic polyps are free from the risk. Adenomatous polyps are precursors of colorectal cancer (Muto *et al.*, 1975). The evidence of adenoma progress to adenocarcinomas supports the practice of removing adenomatous polyps to reduce and prevent colorectal cancer incidence (Winawer, *et al.*, 1993).

1.3.3 Colorectal Cancer Classification

Colorectal cancer is classified into three main groups; sporadic, hereditary and familial cancer, with each interposing specific clinical features.

1.3.4 Sporadic Colorectal Cancer

Sporadic colorectal cancer is the common type of CRC, accounting for 75% of all CRC cases. Sporadic CRC is a cancer with no apparent hereditary symptoms, where the patient has no prior family history of CRC (John Hopkins

Medicine, 2010). Sporadic CRC arises from the expansion of a subset of cells that have acquired a growth advantage through randomly acquired somatic mutations (Søreide *et al.*, 2009). The chances of developing sporadic CRC increases with age, the average lifetime risk of getting sporadic CRC after 50 years old is 5% (Schulmann *et al.*, 2002). This is consistent with the theory that cells must accumulate genetic defects, including gene mutations to undergo full transformation from normal to cancerous. Individuals that have one or more adenomas have a 45% chance of getting sporadic CRC. The size of these adenomas is 4-5 mm, and the chances for any adenoma to develop into cancer are 1 in 400 per year. However, if the size of adenomas is larger, the risk increases to 1 in 4 per year. Those with affected first-degree relatives have a higher risk of developing CRC (Chu *et al.*, 2007). Symptoms of sporadic CRC include abdominal pain, change in bowel habits and rectal bleeding.

1.3.5 Hereditary Colorectal Cancer

Hereditary colorectal cancer is a form of CRC where the gene that causes CRC in the family is known. Hereditary CRC makes up 10% of all CRC cases. Hereditary CRC is associated with a germline mutation in a known cancer susceptibility gene. If a person inherited a cancer susceptibility gene from a parent who has a family history of CRC, the person is predisposed to develop CRC. There are two groups of hereditary colorectal cancers; the polyposis

syndromes with multiple colorectal polyps and the non-polyposis syndromes with only one or a few polyps. The polyposis syndromes include familial adenomatous polyposis (FAP), Peutz-Jeghers syndromes and juvenile polyposis coli. The non-polyposis syndrome is then divided into hereditary non-polyposis colorectal cancer (HNPCC) and familial colorectal cancers with unknown genes (Tonelli, 2002).

1.3.6 Familial Adenomatous Polyposis

Familial adenomatous polyposis (FAP) is a disorder that leads to the formation of hundreds or thousands of benign adomatous polyps in the colon and rectum at a young age, usually as a teenager or young adult. Some of these polyps will progress into invasiveness and ultimately metastasize. FAP was the first polyposis syndrome to be recognized and therefore remains the best characterized. FAP is an autosomal dominant disease caused by the mutation of the APC gene, accounting for 1% of all CRC cases. It has been shown that the APC gene is not only responsible for sporadic CRC but also FAP. FAP is inherited and primarily affects the colon and less often the stomach and small intestine.

1.3.7 Hereditary Nonpolyposis Colorectal Carcinoma

Hereditary nonpolyposis colorectal carcinoma (HNPCC) or known as Lynch syndrome is the most common hereditary CRC. HNPCC is an autosomal dominant disease and is responsible for 1-6% of all CRC cases (Søreide *et al.* 2009). HNPCC is characterized by the development of CRC and endometrial cancer at an early age. HNPCC is caused by the mutation in the hMLH1 and hMSH2 genes, which when function normally would protect against CRC. HNPCC commonly develops in the right side of the colon and is undifferentiated, with the presence of mucus and signet ring cells, infilitration of peritumoral lymphocyte, Crohn-like reaction and lymphocyte infiltration (Coura *et al*., 2005). Incidence of HNPCC among the population is about 1:1000 in the United States (Umar *et al.*, 2004). The diagnosis of HNPCC is based on Amsterdam criteria and Bethesda guidelines as HNPCC does not have any apparent phenotypic signs. Figure 1.4 shows the different types of colorectal cancers and their distribution.

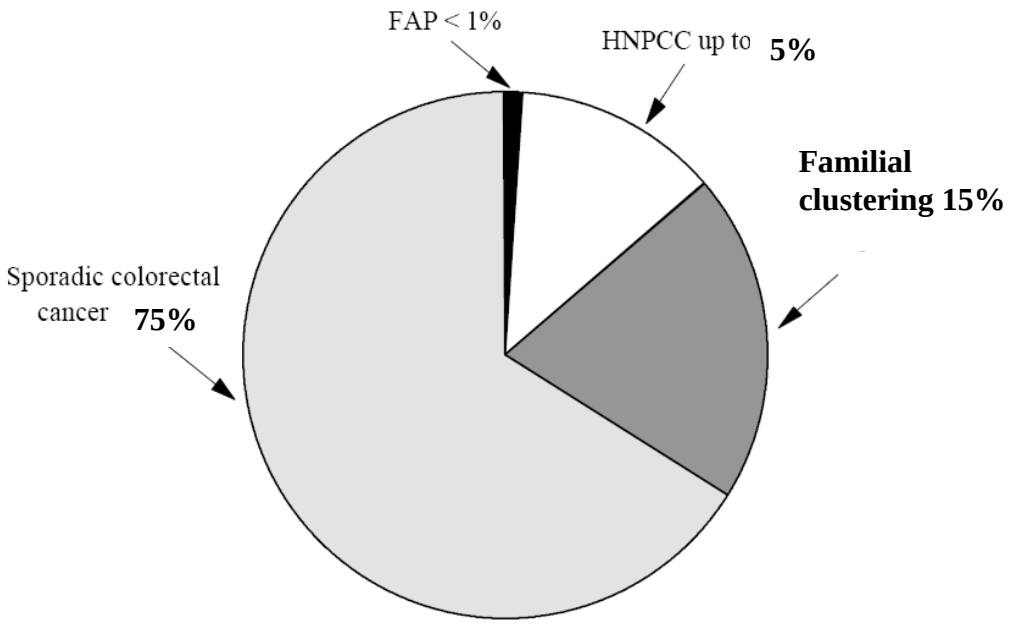

Figure 1.4 Distribution of different types of colorectal cancers (Card and Logan, 2003).

1.3.8 Familial Clustering

Familial clustering of CRC accounts for about 15% of all CRC cases is suspected when "familial accumulation" of CRC cases occurs (Lynch *et al.*, 2006). Firstly, there is an excess number of CRC cases in a family that would be statistically expected. The occurrence of CRC in more than one family member may be due to chance alone. But it could also mean the potential for developing CRC has been inherited from one generation of the family to the next. Relatives of a person with CRC in the same family have a higher chance of getting CRC themselves. Secondly, the family does not meet the criteria for either sporadic or hereditary CRC. The transmission pattern of familial clustering of CRC is not

understood and does not show any specific inheritance pattern (Susser and Susser, 1989). The exact gene that causes this type of cancer is still unknown. Relatives of a person with colorectal cancer are more likely to develop CRC (Sork, 2008).

1.3.9 Staging of Cancer

The cancer stage is the extent of cancer spread in the body (American Cancer Society, 2010b). The stage of colorectal cancer is based on how far the cancer cells have grown into the intestinal wall and whether or not they have spread to nearby structures, such as lymph nodes or distant organs (American Cancer Society, 2010b). The staging process is to determine how extensive a cancer has spread and is based on results of the physical examination, biopsies, and imaging tests.

The Dukes' classification system was used for early stages of colorectal cancer, where the invasion of tumour is limited to mucosa or submucosa. This Dukes' category includes all tumour confined to the bowel wall using an A, B, C and D system (Dukes, 1932). This is shown in Table 1.1.

Table 1.1 Dukes' classification of colorectal cancer (Dukes, 1932).

Duke System	Description
A	Penetration into but not through the bowel wall
B	Penetration through the bowel wall
C	Involvement of regional lymph nodes
D	Distant metastasis

Later, the American Cancer Society defined the stage of colorectal cancer by the depth of tumour invasion into the intestinal wall, involvement of lymph nodes or distant metastasis (American Cancer Society, 2010b). More recently, many modifications of the Dukes' scheme have been adopted to provide better information on prognosis and pattern of the disease. The tumour-node-metastasis (TNM) classification is used for colorectal tumour staging. In 2002, the American Joint Cancer Committee on Cancer (AJCC) has further grouped the stages II and III of colorectal cancer according to tumour size. The higher the stage, the bigger the tumour size is. The new staging system for colon cancer stratifies stage II and stage III further by use of the T stage (tumour depth of penetration), and the N stage (number of positive lymph nodes), resulting in a total of seven stages (O'Connell *et al.*, 2004) as shown in Table 1.2.

Table 1.2 Guidelines for Dukes' and TNM Staging System (American Joint Committee on Cancer, 2002; O'Connell *et al.*, 2004).

Dukes' Classification	**Stage system**	**T stage**	**N stage**	**M stage**
-	0	Tis	N0	M0
A	I	T1 or T2	N0	M0
B	IIA	T3	N0	M0
B	IIB	T4	N0	M0
C	IIIA	T1-T2	N1	M0
C	IIIB	T3-T4	N1	M0
C	IIIC	Any T	N2	M0
D	lV	Any T	Any N	M1

The layers of the colon and the tumour penetration in colorectal cancer are shown in Figure 1.5.

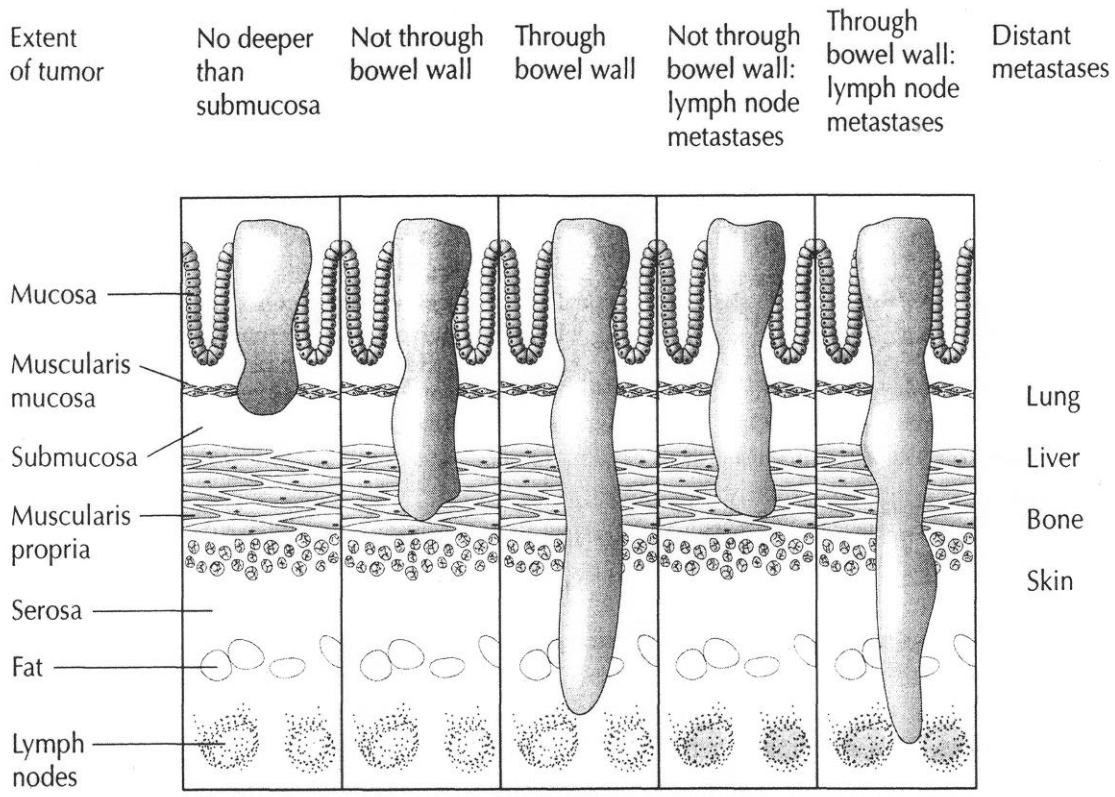

Figure 1.5 Layers of the colon noting tumour penetration (Griffin-Sobel, 2001).

The TNM staging system describes the extent of colorectal cancer into several terms;

Primary Tumor (T)

Tis Carcinoma in situ, intraepithelial or invasion of the lamina propria

T1 Tumor invades submucosa

T2 Tumor invades muscalaris propria

T3 Tumor invades muscalaris propria into subserosa, or into nonperitonealized or perirectal tissues

T4 Tumor directly invades other organs or structures, and/or perforates visceral peritoneum

Regional Nodes (N)

N0 No regional lymph node metastasis

N1 Metastasis to one to three regional lymph nodes

N2 Metastasis to four or more regional lymph nodes

Metastasis (M)

M0 No distant metastasis

M1 Distant metastasis

1.3.10 Grading of Cancer

Grading determines how similar a cancerous colorectal tissue is to a normal colorectal tissue when viewed under a microscope. It also determines how aggressive it is for the tumour to grow and spread. The grading scale for colorectal cancer ranges from G1 to G4 (National Cancer Institute, 2004; American Cancer Society, 2010b).

G1 Well differentiated tissue. Cells have the appearance of normal cells, and tend grow slowly. The least aggressive type of tumour.

G2 Moderately differentiated tissue. Cells have an appearance that is less like normal cells and often grow faster than G1 cancer cells. More aggressive than G1.

G3 Poorly differentiated tissue. Cells are the least like normal cells. They grow more rapidly and are highly aggressive, often spreading into the lymph nodes and bone.

G4 Undifferentiated tissue.

1.4 Biomarker

Biomarkers are substances or molecules in a biological system that responds to the change of the system. They can be measured and evaluated to determine the physiological and pathological changes in a normal condition or disease. (Atkinson *et al.*, 2001) The ability of biomarker to be used as an indicator helps to monitor disease conditions and assists in its diagnosis and prognosis.

1.4.1 Cancer Biomarkers

The progress of identification of protein biomarkers involves separation and analysis of protein samples from individual with and without a particular disease. Historical studies revealed the presence of these biomarkers especially in the malignancies of gastrointestinal system. One of the examples is carcinoembryonic antigen (CEA), discovered in human colon cancer tissue extracts that was found to be elevated in the serum of patients with gastrointestinal tumors (Gold and Freedman, 1965). Another glycoprotein, carbohydrate antigen 19-9 (CA 19-9), originated from patients with colon and pancreatic cancer, it was found highly expressed in the serum of patients with pancreaticobiliary tumours (Morris-Stiff *et al.*, 2009). Finally, alpha fetoprotein is commonly expressed in the serum of patients with hepatocellular cancer (Zhou *et al.*, 2006).

Although the clinical practice of CRC screening using fecal occult blood test (FOBT) and colonoscopy reduces mortality; however, to the best currently available blood test, carcinoembryonic antigen exhibits low sensitivity and specificity particularly in the setting of early disease. Most of the carcinoma (70%) is often diagnosed at late stage and showed poor prognosis (Habermann *et al.*, 2008). With the recently advancement of proteomic tools, it is now possible to identify CRC biomarkers from the biological matrices, such as serum, tissue or cell lines (Jimenez *et al.*, 2010).

1.5 Proteomics

Proteomics is a study of total protein expression in a living system or cell. Proteomics has wide applications in the field of biomedicine for the identification of new markers, therapeutic targets and markers of therapeutic response (Cole *et al.*, 2000; Minowa *et al.*, 2000). Analysis of the genome and transcriptome may provide information on important proteins in the development of disease. However, proteome analysis is the most accurate reflection of protein expression. The expression of a certain gene caused by alternative splicing of the mRNA, posttranslational modifications, temporal and functional regulation may not correlate with its corresponding protein (Greenbaum *et al.*, 2003). Nevertheless, proteomics is an unbiased approach to examine proteins in complex biological fluids, serum as well as tissues from patients. Body fluids, for example nipple fluid aspirates, cerebrospinal fluid (CSF) and synovial fluid have been applied in the search of biomarkers for various diseases. Blood and urine are other specimens that were used to locate candidate biomarkers for clinical diagnosis and treatment of diseases (Reichelt *et al.*, 2006; Li, *et al.*, 2007).

Basically, proteomics can be grouped into four applications; Global proteome mining, protein expression profiling, protein-protein interactions identification and posttranslational modification profiling. Global proteome mining is the process of identifying all proteins in a sample. Protein expression profiling detects and measures differences in protein expression that are

characteristic of a particular state of the organism. Protein-protein interactions identification is used to determine the condition and manner in which proteins interact with each other in a living organism. Posttranslational modification profiling establishes how, when and where proteins are modified after their synthesis (Liebler, 2002; Simpson, 2003a; Somiari, 2005). Conventional proteomic analyses utilized tools based on two-dimensional polyacrylamide electrophoresis (2D-PAGE) and mass spectrometry for protein separation, analysis and identification.

1.5.1 Electrophoresis

Electrophoresis refers to the movement of charged particles or molecules under the influence of an electric field. Separation of proteins by means of electrophoresis depends on size, charge, shape, temperature and gel medium viscosity. Protein's molecular weight (MW) and isoelectric point (pI) can be determined by using electrophoresis separation (Nelson and Cox, 2005). Polyacrylamide gel electrophoresis (PAGE) is a widely used matrix for protein separation. PAGE with its small pore size acts as molecular sieve and is suitable to separate proteins according to shape, size and charge density. Gel matrix formation is through free vinyl polymerization of acrylamide monomers. Ammonium persulfate (APS) was used to initiate the polymerization process by producing free persulfate radicals that convert the acrylamide monomer to a free

radical state. This free radical acrylamide then reacts with other acrylamide monomers to form polymers. N,N,N',N'-tetramethylenediamine (TEMED) is added to accelerate the polymerization process The elongating polymers are crosslinked with bisacrylamide to form a matrix with a pore size that is influenced by the acrylamide/bisacrylamide monomer concentration (Hames, 1981). PAGE serves as the support matrix in isoelectric focusing (IEF) and sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

1.5.2 2D Gel Electrophoresis

Two-dimensional (2D) gel electrophoresis is used to separate complex protein mixtures. It separates proteins according to the MW and pI. The first dimension of protein separation by 2D electrophoresis is by isoelectric focusing (IEF). IEF separates proteins based on their isoelectric point (pI). Proteins in solution exist as amphoteric molecules. They have a net positive, negative or zero charge depending on the surrounding pH. The specific pH where the protein has zero net charge is known as its pI (Stochaj *et al.*, 2003). In a pH gradient under an electric field, protein moves to a position where its net charge is zero. When a protein with positive charge migrates to the cathode, it becomes less charged until it reaches its pI where it receives no net charge and therefore stops migrating (Garfin and Heerdt, 2000; Berkelman and Stenstedt, 2002; Stochaj *et al.*, 2003). The resolution of IEF depends on the pH gradient slope and voltage applied

(Berkelman and Stenstedt, 2002). Originally, the IEF was based on carrier ampholyte-generated pH gradients in polyacrylamide gel rods in tubes (Klose, 1975; O'Farrell, 1975). Since then, many improvements focus on reproducibility of 2D pattern, because of limitation of carrier ampholytes to generate stable pH gradient. These issues were overcome when Gasparic *et al.* (1975) developed immobilized pH gradients (IPG) to form a stable pH gradient within the gel matrix. The concept was further expanded until IPG strips were designed by Bjellqvist *et al.* (1982). IPG strip was rehydrated with protein in rehydration buffer until swollen before IEF was carried out.

1.5.3 Equilibration

The equilibration step on IPG strips was used to transfer proteins from first to second dimension. It was first introduced by Görg *et al.* in 1987. The step is essential to let IPG strips be equilibrated after IEF in equilibration buffer. Equilibration buffer contains urea (6M), glycerol (30%), SDS (2%), Tris-HCl buffer and dithiothreltol (DTT). Urea and glycerol reduce the effects of electroendosmosis by increasing the viscosity of the buffer. They also improve the transfer of protein from the first to second dimension. Tris-HCl buffer maintains the IPG strip and the correct pH for electrophoresis. SDS denatures and unfolds proteins to their linear form and confers uniform negative charge to them. DTT is a reducing agent to break disulfide bonds to keep the proteins in their