IDENTIFICATION OF POTENTIAL BIOMARKERS IN BREAST CANCER TISSUES AMONG MALAYSIAN PATIENTS

by

LIM CHU AI

Thesis submitted in fulfillment of the
requirements for the degree of
Master of Science

July 2011

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following people who have contributed to the success of this project. First of all, I would like to extend my appreciation to my main supervisor Assc. Prof. Dr. Gam Lay Harn for her invaluable help and guidance in completing this project. Sincere thanks to my co-supervisors, Assc. Prof. Dr. Shaharum Shamsuddin and Dr. Khoo Boon Yin, for their guidance and assistance throughout the course of my study.

A special word of appreciation to the staffs of School of Pharmaceutical Sciences, Institute for Research in Molecular Medicine (INFORMM) and National Poison Centre for providing me the facilities, equipments and their assistance in my work. I would like to express my gratitude as well to the surgeons and staff in Penang General Hospital for helping me to collect the breast tissue samples.

To my lab mates and friends, thank you for your concerns, encouragement, helpful suggestions and comments.

Lastly, I would like to extend my deepest gratitude to my parents and brothers. Thank you for your support, understanding and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

2D-PAGE	: Two-dimensional polyacrylamide gel electrophoresis
ACN	: Acetonitrile
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
DAB	: Diaminobenzidine
DTT	: 1,4 – Dithiothreitol
ESI	: Electrospray ionization
FBS	: Fetal bovine serum
HPLC	: High performance liquid chromatography
HRP	: Horseradish peroxidase
HSP	: Heat Shock Protein
IDC	: Infiltrating Ductal Carcinoma
IEF	: Isoelectric focusing
IPG	: Immobilized pH gradient
kDa	: kilo Dalton
LC/MS/MS	: Liquid chromatography tandem mass spectrometry
MS	: Mass spectrometry
MS/MS	: Tandem mass spectrometry
m/z	: Mass to charge ratio
NaCl	: Sodium chloride

$Na_2HPO_4.2H_20$: Disodium hydrogen phosphate dihydrate

$NaH_2PO_4.2H_20$: Sodium dihydrogen phosphate dihydrate

NH_4HCO_3 : Ammonium bicarbonate

PAGE : Polyacrylamide gel electrophoresis

PBS : Phosphate buffered saline

PDI : Protein Disulfide Isomerase

pI : Isoelectric point

RT : Room temperature

SDS : Sodium dodecyl sulfate

TCA : Trichloroacetic acid

TEMED : N, N, N' N' - tetramethylethylenediamine

TLB : Thiourea lysis buffer

Tris : Tris(hydroxymethyl)aminomethane

PENGENALPASTIAN PENANDA-PENANDA BIO BERPOTENSI BAGI KANSER PAYUDARA DI KALANGAN PESAKIT-PESAKIT MALAYSIA

ABSTRAK

Kanser payudara merupakan penyakit barah yang ketiga paling umum di dunia dan paling meluas dalam kalangan wanita Malaysia. Pada tahun 2006, sebanyak 3525 kes baru kanser payudara telah didaftar dengan Pendaftaran Kanser Kebangsaan Malaysia. Protein merupakan komponen yang mengawal aktiviti sel. Melalui perbandingan antara protein yang diekspres dalam tisu normal dan kanser, protein yang berpotensi untuk mengesan, mengubati, mengawal penyakit payudara boleh dikenalpasti. Dalam kajian ini, protein-protein telah diekstrak daripada 25 pasangan tisu normal dan kanser daripada pesakit-pesakit payudara jenis infiltrating ductal carcinoma(IDC). Protein-protein diasing berdasarkan titik isoelektrik dan berat molekul. Tompok-tompok protein dianalisis dan dipencir daripada gel. Selepas itu, protein dicerna oleh tripsin untuk menghasilkan tangkaian peptide-peptida yang dapat dianalisis oleh LC/MS/MS. Data LC/MS/MS digunakan untuk mengenalpasti identiti protein melalui carian database. Dua puluh enam protein telah dikenalpasti dalam ekstrak Tris dan 18 protein dalam ekstrak TLB yang mempunyai pengekspressan yang berbeza antara tisu-tisu normal and kanser. Fibrinogen beta chain merupakan protein kawalaturan-turun (>70%) dalam ekstrak Tris, manakala 80K protein H precursor, tubulin, beta polypeptide, calreticulin variant and protein disulfide-isomerase (PDI) ialah protein-protein kawalaturan-naik (>70%) dalam ekstrak TLB. Ujian Western blot dan perwarnaan immunosito bagi PDI dan HSP 60 dijalankan untuk mengesah dan mengenalpasti lokasi protein pada sel. Keputusan

menunjukan bahawa PDI berlokasi pada permukaan membran sel MCF-7, T47D dan MDA-MB-231. Manakala, HSP 60 hanya berlokasi pada permukaan membran sel MCF-7 dan T47D. Protein-protein ekstra selular ini berpotensi untuk dijadikan penanda bio bagi rawatan terapi sasaran. Selain itu, kami menyusulkan PDI, heat shock protein gp96, HSP60, 80K protein H precursor, calreticulin variant and tubulin, beta polypeptide berpotensi sebagai penanda bio bagi kanser payudara. Memandangkan kanser ialah sejenis penyakit heterogeneous, penggunaan bersama penanda-penanda bio akan member diagnosis yang lebih tepat. Kegunaan penanda-penanda bio ini boleh dikaji secara lebih dalam pada masa hadapan.

IDENTIFICATION OF POTENTIAL BIOMARKERS IN BREAST CANCER TISSUES AMONG MALAYSIAN PATIENTS

ABSTRACT

Breast cancer is the third most common cancer worldwide and is the most common cancer among Malaysian women. In year 2006, there were 3525 new cases of female breast cancer registered with Malaysia National Cancer Registry. Protein is the expression component of cells that regulates cellular activity. By comparing healthy and cancerous tissues at the protein level, it is possible to identify the potential protein targets that can serve to detect, treat, monitor, and for prognosis of breast cancer. In this study, proteins were extracted from 25 pairs of breast normal and cancerous tissues from patients who suffered from infiltrating ductal carcinoma (IDC). These proteins were separated according to their pI's and molecular weights. The protein spots were analyzed and the protein spots of interest were excised from the gel. After that, they were digested with trypsin to produce fragment of peptides that can be analyzed using LC/MS/MS. The data obtained from LC/MS/MS analysis were subjected to protein database search engine for protein identity determination. Twenty-six protein spots from Tris extract and 18 protein spots from TLB extract were identified due to their differential expression between cancerous and normal tissues. Fibrinogen beta chain was the down-regulated protein (>70%) in Tris extract, while 80K protein H precursor, tubulin, beta polypeptide, calreticulin variant and protein disulfide-isomerase (PDI) were the up-regulated proteins (>70%) in TLB extract. PDI and heat shock protein 60 (HSP60) were selected for Western blot and immunocytostaining studies to validate the presence of these proteins and their

extracellular location. The results showed that PDI was located on the external membrane surface of MCF-7, T47D and MDA-MB-231 cell lines. Meanwhile HSP60 was located on the external membrane surface of MCF-7 and T47D cell lines. These extracellular surface proteins may serve as potential biomarkers for drug target therapy. In addition, PDI, heat shock protein gp96, HSP60, 80K protein H precursor, calreticulin variant and tubulin, beta polypeptide were the potential biomarkers in breast cancer. As cancer is a heterogeneous disease, the collective use of these markers may provide more accurate diagnosis of the disease. The usefulness of these proteins as biomarkers for IDC is worth further investigation.

CHAPTER 1

INTRODUCTION

1.1 Cancer

Cancer is a class of diseases with common feature of uncontrolled cell growth (King & Robins, 2006). The growth of cancer cell is different from normal cell where the cancer cells grow uncontrollably. The cancer cells are ultimately spread throughout the body and interfering with the functions of normal tissues and organs in which the cells are life threatening (Cooper & Hausman, 2004). The term "cancer" tends to be used differently in experimental and clinical settings. The term "neoplasm" refers to new growth and "tumour" refers to abnormal growth. Cancer is more precisely defined as invasive and able to metastasise (malignant). In contrast, benign growth refers to a localized growth of cells. The benign cells are not always life-threatening and therefore treatments of these two types of growth are different (King & Robins, 2006). However, the benign growth may sometimes lead to malignant growth.

The ability of cancer cells to overcome the normal containment mechanisms that inhibit cell growth reflects the modification of its membrane that results membrane modification in diminishing of cell-cell interactions and the excessive production of proteases that facilitate movement through the extracellular matrix. Besides that, cancer cells can use chemical messengers to signal normal cells to aid in their growth by promoting the development of new blood vessels to ensure an adequate supply of essential nutrients, this invasive property is what distinguishes cancer cells from normal cells (King & Robins, 2006). Metastasis is the process

when cancer cells get into the bloodstream or lymph vessels, and travel to other parts of the body. There, the cells begin to grow and form new tumours that replace normal tissues (Cooper & Hausman, 2004).

There are more than a hundred distinct types of cancer. Most cancers fall into one of three main groups: carcinomas, sarcomas, and leukemias or lymphomas. Carcinoma, which made up approximately 90% of human cancers are malignancies of epithelial cells. Sarcomas are solid tumors of connective tissues, such as muscle, bone, cartilage, and fibrous tissue. Sarcoma is rare in human. Leukemias and lymphomas account for approximately 7% of human malignancies, they arise from the blood-forming cells and from cells of the immune system, respectively (Cooper & Hausman, 2004).

Cancer is one of the major problems in Malaysia (Lim, 2002). In 2006, a total of 21,773 cancer cases were diagnosed and registered in the National Cancer Registry (NCR) in Peninsular Malaysia. The cases comprises of 9,974 males and 11,799 females. The Age standardised Incidence Rate (ASR) for all cancers in 2006, regardless of gender, was 131.3 per 100,000 population. The ASR among males in Peninsular Malaysia was 128.6 per 100,000 population while there was 135.7 per 100,000 population among female (Omar, 2006). Cancer is more predominant among Chinese as compared to Malay and Indian in Peninsular Malaysia. The ASR for Chinese male was 148.0 per 100,000 population and for Chinese female was 151.5 per 100,000 population. For Malay males, the ASR was 94.5 per 100,000 population and Malay female was 96.4 per 100,000 population. The ASR of cancer among Indian males was 97.4 per 100,000 population and 134.7 per 100,000 population for

Indian females (Omar, 2006). The five most common cancers in Peninsular Malaysia were breast, colorectal, lung, cervix and nasopharynx (Omar, 2006). Figure 1.1 shows ten most frequent cancers of Peninsular Malaysia in 2006.

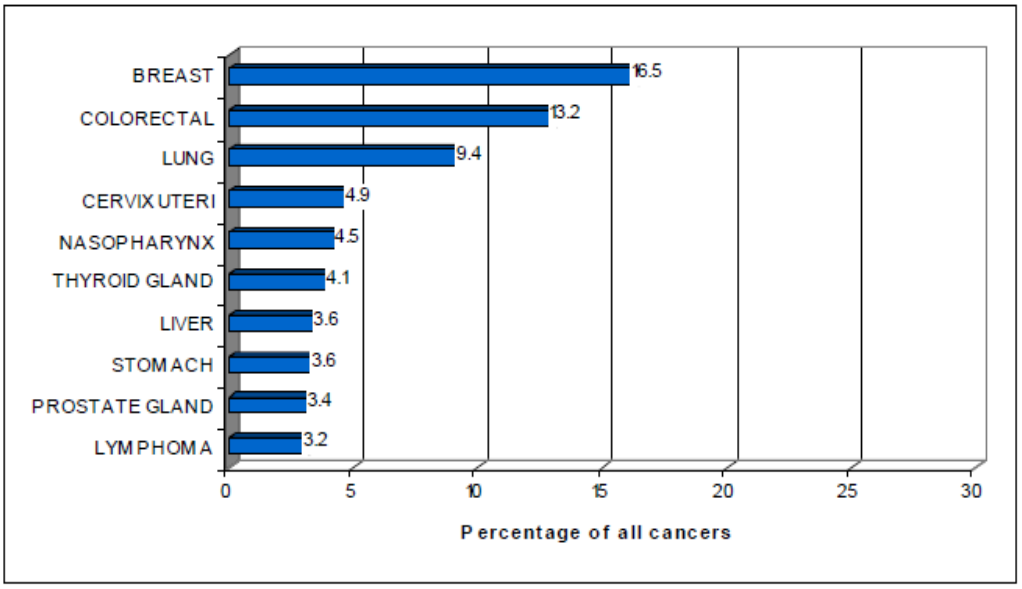

Figure 1.1**:** Ten most frequent cancers in Peninsular Malaysia in 2006 (Omar, 2006)

Many factors may increase the likelihood of developing cancer which includes radiation (atomic, ultraviolet and X-ray), chemicals (arsenic and chromium, carcinogenic in tobacco smoke, aflatoxin from fungus), bacteria and viruses such as *Helicobacter pylori* and Human papillomaviruses (HPVs). The radiation and chemicals act by damaging DNA and inducing mutations (Cooper & Hausman, 2004; King & Robins, 2006) whereas viruses initiate cancer under conditions not completely understood (Cooper & Hausman, 2004). Differences in the incidence of cancer around the world usually can be traced to environmental influences (Kumar *et al*, 1992). Geographical distributions and influences of lifestyle, such as diet, personal practices, workplace and ambient environment, are the factors that may contribute to cancer development (King & Robins, 2006). Figure 1.2 illustrates the prevalence of certain cancers in different countries. The most common cancer in

United States of America (USA), China and Malaysia were breast and prostate cancer, stomach and esophagus, colorectal and breast cancer, respectively (King & Robins, 2006; Omar, 2006).

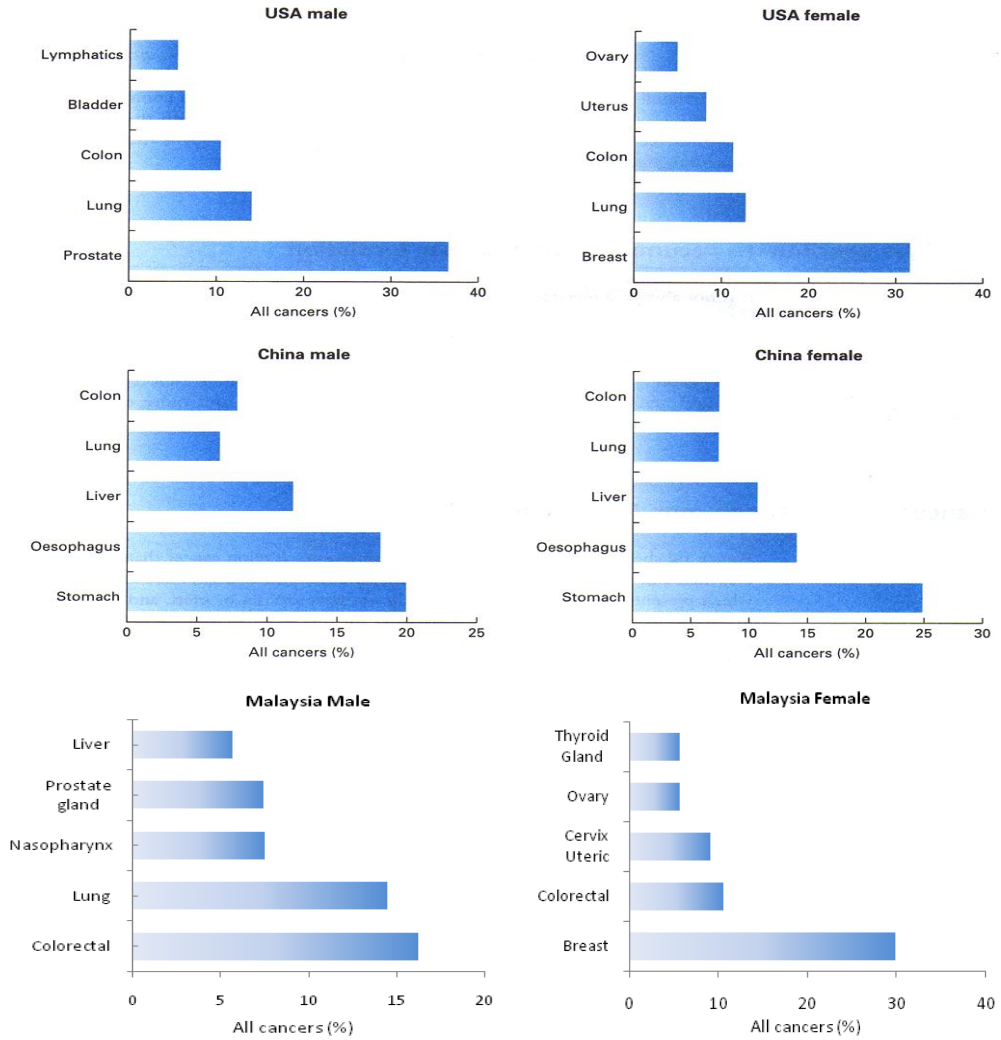

Figure 1.2: Comparison of the five most common cancers in the USA, China and Malaysia (King & Robins, 2006; Omar, 2006)

Diet is a major determinant of human carcinogenesis, which reflects the effects of complex chemicals mixture. There is increasing debate about the carcinogenicity of food additives, artificial sweeteners, and containing pesticides (King & Robins, 2006; Kumar *et al.*, 1992). Gastric carcinomas may relate to endogenous synthesis of carcinogens or promoters from diet. Nitrosamines and nitrosamides are believed to be carcinogens because they have been clearly shown to induce gastric cancer in animals. These compounds can be formed in the body from

nitrates and amines or amides derived from digested proteins. The sources of nitrites include sodium nitrate, a food preservative (Kumar *et al*., 1992).

1.2 Breast Cancer

Breast cancer is a malignant tumour that derives from the cells of breast. It is commonly found in women, but men can also get breast cancer. The ratio of male carcinoma to female carcinoma is 1:125 (Kumar *et al*., 1992). Breast cancer is the third most common cancer worldwide and is the most common cancer among Malaysian women (McPherson *et al.*, 2000b);(Hisham and Yip, 2004). In 2006, breast cancer was reported to be the most common cancer among female and also the most prevalence cancer among population in Peninsular Malaysia regardless of gender. There were 3525 female breast cancer cases registered in NCR, accounted for 16.5% of all cancer cases registered. The incidence of breast cancer varies among the ethnics in Peninsular Malaysia. According to the record in NCR for 2006, Chinese have the highest ASR among the three major ethnics, for example, 46.4 per 100,000 population followed by Indian and Malay, respectively with the ASR of 38.1 per 100,000 population and 30.4 per 100,000 respectively (Omar, 2006).

1.2.1 Types of Breast Cancer

There are many types of breast cancer, but some **of** the types **are** very rare. The common types of breast cancer include ductal carcinoma *in situ* (DCIS), lobular carcinoma *in situ* (LCIS), invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC), while the less common types of breast cancer are inflammatory breast cancer (IBC), medullary carcinoma, Paget's disease of the breast, tubular carcinoma, phylloides tumour, metaplastic carcinoma, sarcoma, micropapillary

carcinoma and adenoid cystic carcinoma. Sometimes, a breast tumour can be a mix of more than one types or a mixture of invasive and *in situ* cancer.

Below are the elaborations of some common types of breast cancer.

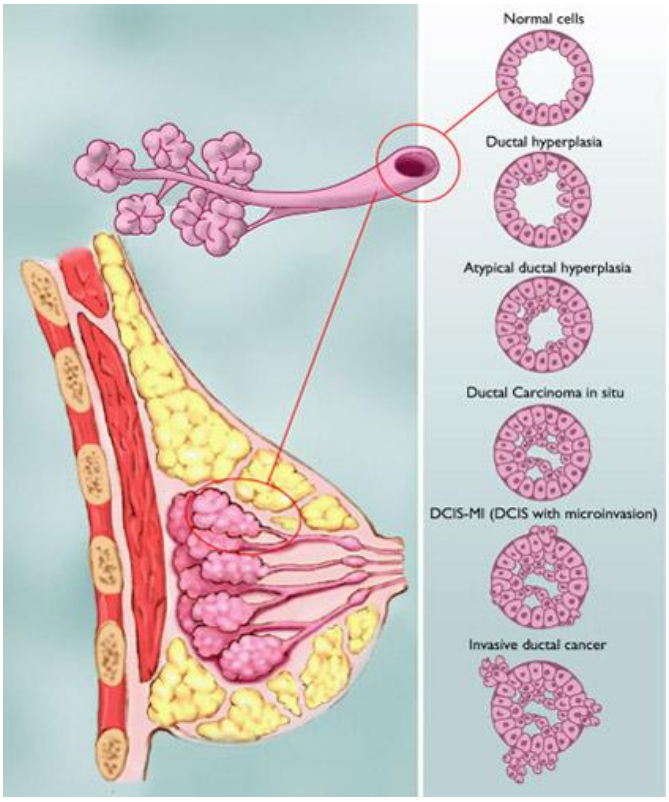

Figure 1.3 Range of ductal carcinoma in situ
(http://www.breastcancer.org/pictures/types/)

Ductal carcinoma *in situ* (DCIS) refers to abnormal cell growth in the lining milk duct. It is confined to the walls of the ducts and has not spread into the tissue of the breast. Figure 1.3 shows range of ductal carcinoma *in situ*. This is an early-stage breast cancer. Almost all women with DCIS can be successfully treated. However, it may eventually develop into invasive breast cancer if left untreated.

Lobular carcinoma *in situ* (LCIS) begins in the milk-making glands (lobules) without invading to the surrounding breast tissue. It is not a true cancer, but having LCIS increases a woman's risk of getting cancer later. Women who have LCIS are more likely to develop invasive lobular breast.

Invasive (or infiltrating) ductal carcinoma (IDC) is the most common breast cancer that accounts for 75% of breast carcinoma. It starts in a milk passage (a duct), breaks through the wall of the duct, and invades the tissue of the breast. The cancer cells may remain localized or it may be able to spread (metastasize) to other parts of the body through bloodstream or lymphatic system.

Invasive lobular carcinoma (ILC) starts in the milk-producing lobule, invades to the surrounding breast tissue and can metastasize. ILC is harder to be detected as it is less likely to appear on a mammogram and patients might not be able to detect a breast lump. Patients may perceive only a general thickening or feel some different in the breast tissue.

Inflammatory breast cancer (IBC) is uncommon but aggressive type of invasive breast cancer that accounts for about 1% to 3% of all breast cancers. IBC makes the skin of the breast becomes red, swollen and feels warm. The skin looks thick and pitted something like an orange peel. The breast may get bigger, hard, tender, or itchy. Medullary carcinoma is a specific type of invasive breast cancer in which the tumour's borders are clearly defined, the cancer cells are large, and immune system cells are present around the border of the tumour. Mucinous (colloid) carcinoma is a type of invasive breast cancer, the cancer cells produce mucus and grow into a jelly-like tumour. The tissues are extremely soft, bulky, gray-blue masses consist of gelatin. Paget's disease of the breast is rare type of breast cancer that affects nipple, and skin surrounding the nipple (areola). Tubular carcinoma gets its name from the appearance of the cancer cells under a microscope. Though it is an invasive breast cancer, the prognosis is more favourable than invasive ductal

carcinoma or invasive lobular carcinoma. Phylloides tumour is a large, bulky tumour. phylloides tumours develop in the connective tissue of the breast. The prognosis for a phylloides tumour is uncertain. If the tumour cannot be removed, it will be difficult to treat. Metaplastic carcinoma tends to remain localized and contains several different types of cells that are not typically seen in other forms of breast cancer. Sarcoma is a tumour that develops in the connective tissue of the breast, which is usually cancerous (malignant). Micropapillary carcinoma is an aggressive invasive breast cancer, often spreading to the lymph nodes even when they were very small. Adenoid cystic carcinoma is characterized by a large, local tumour. It is an invasive but slow-growing type of breast cancer and is unlikely to spread.

1.2.2 Breast Cancer Risk Factor

Several factors have been identified as potentially responsible for increasing breast cancer risk, although their mechanisms of action are unclear (McPherson *et al.*, 2000b). These factors are:

a) Previous history of breast cancer.

b) Increasing age-estimated 64% of women diagnosed with cancer are over 55 of age.

c) Early menarche (age 12 and younger) starting menarche at age 11 or earlier is considered an established breast cancer risk factor, and starting menarche at age 15 or older is considered an established protective in risk (Kelsey & Bernstein, 1996).

d) Late menopause (age 55 and older). Each one-year delay in the onset of menopause is associated with a 3% increase risk (Collaborative Group on Hormonal Factors in Breast Cancer, 1997). This may be due to the proliferative response of

breast epithelial cells to estrogen through the estrogens' receptor. (Roses & Giuliano, 2005).

e) Nulliparity at age 40 years. Women who have their first birth before age of 20 have 30% lower risk compare to women with their first birth after the age of 35 (Ewertz *et al.*, 1990).

f) Obesity in postmenopausal women.

g) Hormone replacement therapy. The Women's Health Initiative study found an increased risk of breast (cancer ratio 1.26) in women who are on hormone replacement therapy combined with estrogen and progestin given in a continuous manner (Women's Health Initiative Investigators, 2002).

h) Family history. Women with family history of breast cancer are at increased risk, especially those that having first-degree relative with breast cancer increase risk by 80%.

i) Evidence of specific genetic susceptibility (such as carriage of BRCA1, BRCA2 or BRCA3). BRCA1 and BRCA2 are tumor suppressor genes. When they are mutated, cancer is more likely to develop because they no longer cause cells to die at the right time.

Some studies have shown that the public are not aware of the risk factor for breast cancer by being exposed to some incorrect information. The knowledge of risk factors and perception of personal risk are important for motivating people to prevent, detect and manage disease as early as possible (Washbrook, 2006). Therefore, it is important that accurate information is communicated by physicians and the media.

1.2.3 Breast Cancer Treatment

1.2.3.1 Surgery

The primary aim of surgery is to remove the entire tumour and metastases in regional lymphatic (King & Robins, 2006). Surgery can be conservative (removing the lump or a segment of the breast only) or may be in the form of a mastectomy (removing the whole breast) (McPherson *et al.*, 2000a). Breast conservation surgery followed by radiotherapy is an appropriate method of primary therapy for the majority of women with stage I or II breast cancer. It is preferable because it provides survival equipment to total mastectomy and axillary clearance while preserving the breast (McPherson *et al.*, 2000b)

1.2.3.2 Radiotherapy

Radiation therapy plays an important role in the curative treatment of breast cancer. After conservative surgery, radiation of breast reduces the risk of local recurrence rate and possibly prolongs survivals (Stevens, 2005). Radiotherapy uses high doses of radiation to kill cancer cells by causing irreparable double-stranded DNA breaks (King and Robins, 2006, Hall and Giaccia, 1995) The cell cannot divide when it enters mitosis, and cell death occurs (Hall and Giaccia, 1995). Radiation therapy will cause side effects, the most common acute effects are fatigue, erythema, edema and hyperpigmentation (Stevens, 2005).

1.2.3.3 Chemotherapy

Chemotherapy uses cytotoxic drugs that disrupt the cell cycle, thus causing the death of proliferating cells (King & Robins, 2006). For breast cancer, chemotherapy drugs are given intravenously (directly into a vein) or orally (by

mouth). Once the drugs enter the bloodstream, they travel to all parts of the body in order to reach cancer cells that may have spread beyond the breast. Hence, chemotherapy is considered a "systemic" form of breast cancer treatment. For late stage disease, chemotherapy is used to control the growth rate of the tumour and to downsize the bulk of the disease. Therefore, chemotherapy may also help to control symptoms, prolong survival and improve the quality of life.

1.2.3.4 Hormone Therapy

Hormone or anti-hormone treatment is only given to patients whose cancer cells are positive to oestrogen or progesterone receptors. Determination of hormone receptor status [estrogen (ER) and progesterone receptor (PR)] has become standard practice in the management of invasive breast cancers. It is useful as a prognostic and predictive factor (Bauer *et al.*, 2007). By blocking these receptors, the growth of the cancer cells can be curtailed. Drugs such as Tamoxifen and the new generation of aromatase inhibitors may help to control cancer for many years after primary treatment.

Tamoxifen is the standard drug for endocrine therapy in patients with estradiol and progesterone receptor positive breast cancer. It is a drug taken orally in tablet, interferes with the activity of estrogen. Estrogen can promote the development of breast cancer. Tamoxifen works against the effects of estrogen on the cells. It is often called an antiestrogen or a SERM (Selective Estrogen Receptor Modulator). Tamoxifen has been used to treat patients (men and women) with early-stage breast cancer, as well as those with metastatic breast cancer for more than 30 years. Tamoxifen has serious side effects, such as blood clots, strokes, uterine cancer, and

cataracts. The most common side effects are hot flashes and vaginal discharge. The clinical studies of tamoxifen therapy in postmenopausal breast cancer patients showed that tamoxifen reduced the incidence of tumours in the second breast but increased the risk of endometrial carcinomas (Fornander *et al.*, 1991)

1.2.3.5 Drug Targeted Therapy

Approximately 25% to 30% of human metastatic breast cancers over-express the human epidermal growth factor receptor 2 (HER-2). Overexpression of HER-2 is a significant predictor of reduced survival and shorter time to relapse (Slamon *et al.*, 1987, Huston and Osborne, 2005). This is because these tumours tend to grow faster and are more likely to metastase than tumours that do not overexpress HER-2. Thus, HER-2 has become an important therapeutic target for this subtype of breast cancer. When breast cancer was diagnosed, HER-2 status is routinely assessed by either immunohistochemical (IHC) analysis of HER-2 protein expression or fluorescent *in situ* hybridization (FISH) analysis of HER-2 gene copy number in breast tumour biopsy (Sauter *et al.*, 2009). The presence of HER-2 overexpression is a predictive factor that may indicate success with the use of trastuzumab (Herceptin) (Huston & Osborne, 2005)

Trastuzumab (Herceptin) is a recombinant humanized monoclonal antibody targeted against an extracellular region of the HER-2 receptor, it was approved by the FDA in 1998. The initial clinical trials of trastuzumab as a single agent in HER2-over-expressing metastatic breast cancer demonstrated response rates ranging from 12% to 34% for a median duration of 9 months (Vogel *et al.*, 2002).

1.3 Cancer Biomarker

Biomarkers are biochemical substrates that are indicators of physiologic state and also development during a disease process (Srinivas *et al*., 2002). Cancer biomarker can be classified into cellular and humoral markers. Cellular markers are associated with cancer cells; they can provide prognostic information and information are for the selection of the optimal drugs or a specific therapeutic plan for systemic treatment. (Hunt *et al.*, 2001) D'Arcy *et al*., 2006). Humoral cancer markers are characterized by their appearance in body fluids, for examples urine, blood and body fluid. These markers can be released on tumour disintegration or may be secreted by tumours (D'Arcy *et al*., 2006).They are also useful in the early detection of cancer, especially in asymptomatic people who are at high risk for cancer development (Hunt *et al.*, 2001).

Biomarkers are important tools for detection, diagnosis, treatment, monitoring, and prognosis of diseases (Srinivas *et al*., 2002). Cancer biomarkers can be divided into three categories: diagnostic (screening) biomarkers, prognostic biomarkers and stratification biomarkers. Diagnostic markers are used to detect a given type of cancer in an individual. Prognostic biomarkers are commonly used when the disease status has been established. They can be used to predict the course of disease and its recurrence. Therefore, they indicate the aggressiveness of the therapy. Stratification biomarkers serve to predict the likely response to drug (based on DNA) before starting a treatment. They classify an individual as "responders" or "nonresponder". Such prediction is important in designing clinical drug trials in order to define an intended use for the drug under investigation (Hamdan, 2007).

1.3.1 Currently Available Breast Cancer Biomarkers

There are a few markers for breast cancer. But only two serum markers were approved by the Food and Drug Administration (FDA) for monitoring treatment of advanced breast cancer or recurrence, these markers are MUC-1 (CA27.29 and CA15-3) and carcinoembryonic antigen (CEA) (Kirmiz *et al*., 2007).

MUC1 mucins are expressed physiologically at the luminal surface of glandular epithelia. The expression of MUC1 mucin is usually up-regulated in breast cancer and high amount of the protein is released into the blood stream of breast cancer patients (Gender *et al.*, 1991); Graves *et al*., 1998). Family members of MUC-1-gene include, MCA, BRMA CA549, CA27.29 and CA15-3. CA15-3 is effective in evaluating the clinical course of patients being treated for metastatic disease. Serum level of CA15-3 is related to tumour size. Several reports suggested that CA27.29 is more sensitive but less specific than CA15-3 (Diamandis *et al.*, 2002)

Carcinoembryonic antigen (CEA) is a glycoprotein present in the serum of cancer patients that can be detected using radioimmunoassay or enzyme-linked immunosorbent assay (Levenson, 2007). Several studies have reported that the probability of finding rising level of CEA in patients who were clinically detectable with metastatic breast cancer is between 15% to 68% (Hayes *et al*., 1991). The clinical value of CEA detection is limited due to high false-positive rate detected in normal populations (Levenson, 2007), the false positive rates is range from 10% to 27% (Daniel, 2005). CEA main application is in gastrointestinal cancers, especially in colorectal malignancy. It is a very useful marker for the early detection of liver metastasis in patients with diagnosed with colorectal cancer (Duffy, 2006).

Early detection of breast cancer increases the survival rate of the patient (Levenson, 2007). However, early symptoms of breast cancer are sometimes absent or not recognized. It is often detected in an advanced stage of progression and untreatable by the time the cancer is finally diagnosed (Kirmiz *et al*., 2007). Thus, a reliable biomarker is needed to rule out breast cancer in the early state. Unfortunately, current available tumour markers are lack of the specificity and sensitivity to be use in early detection (Kirmiz *et al*., 2007; D'Arcy *et al*., 2006).

HER-2/neu (c-erbB-2) protein is over expressed in about 20% to 30% of invasive breast carcinomas (Lipton, 2005). HER-2 belongs to type I transmembrane tyrosine kinase receptor family. HER-2 is an important regulator of cell growth and differentiation during embryogenesis and for mammary development during puberty. Deregulation of HER-2 signalling in mammary cells promotes breast tumourigenesis (Nahta *et al*., 2009). Two HER2-targeted therapies are approved by the United States Food and Drug Administration (FDA) for treatment of HER2-overexpressing metastatic breast cancer, namely Trastuzumab (Herceptin) and Lapatinib (Tykerb).

CA-125 is a useful serum marker for monitoring ovarian cancer patients and possibly for prediction of the patient's response to therapy, but it has insufficient sensitivity for the diagnosis of the disease (Bast *et al.*, 2005). Serum CA-125 is lack of specificity, it has been shown to be elevated in various form of cancer including, ovarian, pancreatic, breast, colon, lung and endometrial carcinoma (Bast, 2003). It is commonly agreed that CA-125 is lack both sensitivity and specificity as a marker for

early stage disease (stage I and II). However, the specificity can be improved by combining CA-125 with various forms of sonography (Hamdan, 2007).

1.4 Proteomics

Proteomics is a study of the proteome, the protein complement of the genome. The terms "proteome" and "proteomics" were introduced by Marc Wilkins and colleagues in early 1990s (Liebler, 2002). Proteomics is a branch of functional genomics (Palzkill, 2002), the large-scale study of proteins, particularly their structure and functions, including detection, identification, measurement of their concentration, characterization of modification, characterization of protein-protein interaction and regulation (Xiao *et al.*, 2008). The proteome is the end product of genome. The proteome is a highly dynamic entity and differs from cell to cell (Xiao *et al.*, 2008), while genome is comparatively static (Rabilloud & Humphery-Smith, 2000). The proteome in cells is constantly changing through its biochemical interactions with the genome and the surrounding environment, physiological state of the cell (e.g. position in the cell cycle), stress, drug administration, health and disease (Rabilloud & Humphery-Smith, 2000). In addition, some proteins may be expressed during very short periods of time in the life of an individual, while others may be continually expressed (Xiao *et al.*, 2008)

Proteomic technologies allow the identification of the protein changes caused by the disease process in a relatively accurate manner (Srinivas *et al.*, 2002).The proteome contains all of the gene products that represent the functional output of a cell rather than nucleic acids that are derived from an individual's full genetic code (Palzkill, 2002). This makes proteomics a promising tool for characterizing cells and

tissues of interest and for biomarker discovery (Xiao *et al.*, 2008). Protein analysis is considerably more challenging than DNA/RNA-based analyses. DNA is composed of only four different nucleotides, while proteins are composed of at least 20 unmodified and a lot of modified amino acids. Thus the physiochemical characteristics of proteins vary considerably. Besides that, expression levels of protein cover an extremely large range, and many proteins undergo a myriad of co- and postranslational modifications, including phosphorylation, acetylation, sulfation, glycosylation and etc. These posttranslational modifications are adding to the complexity of protein analysis (Jolles & Jornvall, 2000).

Proteomics directly analyze protein expression at the post-translational level, it permits the qualitative and quantitative assessment of a broad-spectrum of proteins that can be related to specific cellular responses (Alaiya *et al*., 2005; Xiao *et al*., 2003; Xiao *et al*., 2005). Qualitative proteomics experiments or protein expression mapping aims to study changes in protein expression under different physiologically relevant conditions (Ong *et al*., 2003; Xiao *et al*., 2008). For clinical applications, this may help to discover differences between healthy and diseased patients (Xiao *et al*., 2008). On the other hand, quantitative proteomics not only provides lists of identified proteins but also the quantitative information for all proteins in a sample (Mann, 1999; Ong & Monn, 2005). Mass spectrometry (MS)-based quantitative proteomics has become an increasingly popular approach to study changes in protein abundances and diversity in biological samples (Xiao *et al*., 2008).

1.4.1 Proteomics in the Development of Cancer Biomarkers

It has been estimated that only 2% of human diseases are resulted from single gene defects. For remaining 98% of human diseases, epigenetic and environmental factors are taken into accounts that affect both the etiology and the severity of the affliction (Montreuil *et al*., 1996). Therefore, cancer biomarker discovery strategies that target on protein expressions are becoming increasingly popular because the proteomic approaches characterize both the modified or unmodified proteins that are involved in cancer progression (Srinivas *et al.*, 2002). In recent years, the emerging technologies in the field of genomics and proteomics have led to the discovery of a great number of cancer biomarker candidates (Pritzker, 2002). Unfortunately, very few of these biomarkers have been validated and approved for prognostic and diagnostic purpose (Polanksi & Anderson, 2006).

1.5 Proteomics Tools

A proteomic analysis from tissues involved a few steps which include extraction of proteins from tissue matrix, separation of the protein mixture and qualitative and quantitative analysis of the protein by mass spectrometric analysis spectrometric analysis, each of these steps are explain as below.

1.5.1 Sequential Protein Extraction

Protein extraction from tissues and cells is perhaps the most critical step in proteomic studies because this step influences protein yield, biological activity and the structural integrity of the specific target protein (Simpson, 2003). The sequential extraction procedure is based on protein solubility in Tris buffer for the initial removal of highly soluble protein, whereas proteins from the insoluble pellet are then

extracted with thiourea lysis buffer (TLB), a stringent buffer which contains 8M Urea, Thiourea and CHAPS. The sequential extraction buffer can aid in the visualization of a greater percentage of the proteome for a complex cell or tissue (Cordwell *et al*., 2000; Molloy *et al*., 1998).

1.5.1.1 Tris Extraction Buffer

Tris [tris(hydroxymethyl)aminomethane] is widely used in biochemistry and molecular biology as a buffer (Gomori, 1955). Tris is slightly basic with a pKa of 8.08 (Bates & Pinching, 1949) which makes it suitable for studies in biological systems as well as pH control in vitro (Bates *et al*., 1978; Nahas, 1961). It is used at a low concentration of 40 mM to solubilize aqueous soluble proteins (Molloy *et al*., 1998; Santoni *et al*., 2000).

1.5.1.2 Thiourea Lysis Buffer (TLB)

Thiourea lysis buffer (TLB) modified from Rabilloud (1998) consis of urea, thiourea, CHAPS, DTT and carrier ampholyte. Urea and thiourea are chaotropic agents which alter all the solvent parameters, exert profound effects on all types of interactions function (Herskovits *et al.*, 1970). For example, by changing the hydrogen bond structure of the solvent, chaotropes disrupt hydrogen bonds and hydrophobic interactions both between and within proteins (Herskovits *et al.*, 1970); Rabilloud *et al*., 1997). When chaotropic agents are used at high concentrations, it disrupts protein secondary structure and bring into solution, proteins that are not otherwise soluble (Rabilloud *et al*., 1997).

CHAPS, a zwitterionic detergent also disrupt hydrophobic interactions between and within proteins and are particularly useful in solubilizing the insoluble proteins. Normally neutral or zwitterionic detergents are used in 2-D electrophoresis sample preparation, because they are most compatible with subsequent IEF separation. Carrier ampholytes are low molecular mass components with both amino and carboxyl groups (Monribot & Boucherie, 2000). The presence of carrier ampholytes is required to prevent proteins from interacting with each other ionically, but without disturbing the pH gradient or interfering with electrophoresis.

1.5.2 2D Gel Electrophoresis

Two-dimensional electrophoresis is a powerful proteomic tool which combines two dimensions of protein analysis, isoelectric focusing (IEF) and polyacrylamide gel electrophoresis (PAGE) for separation of complex protein mixtures. It utilizes the principal that all proteins in an electric field migrate at a speed that is dependent on their conformation, size, and electric charge (Simpson, 2003). In the first dimension, the proteins are separated according to net charge or isoelectric point (pI), isoelectric focusing (IEF) is done in a polyacrylamide gel with a pH gradient.

1.5.2.1 Isoelectric Focusing (IEF)

Proteins are amphoteric molecules; they are either a positive, negative, or zero net charge, depending on the pH of their surroundings. The isoelectric point (pI) of a protein is the pH value of the protein's surroundings at which the protein has a zero net charge. A protein carries a net negative charge when pH values is above its pI, while the protein carries a net positive charge at the pH values below the pI. IEF

manipulate advantage of this phenomenon (Simpson, 2003). When introduced into an electric field, a protein molecule will migrate according to its surface charge. A positively charged molecule will gradually loose positive charge and gain negative charge through deprotonation of carboxyl or amino functional groups. Eventually, it will stops when it reaches its pI (Righetti, 1983).

1.5.2.2 SDS-PAGE

SDS-PAGE is a method of separating proteins on the basis of their molecular weight. The separation is done on SDS polyacrylamide gels (Fichmann & Westermeier, 1998). SDS is anionic surfactant that binds to proteins according to a constant weight ratio of 1.4g SDS per gram of protein, independent of the nature of the protein (Reynolds & Tanford, 1970). The bound of SDS gives the proteins a net negative charge per unit mass. SDS disrupts hydrogen bonds, blocks hydrophobic interactions, and unfolds the protein, thus eliminating secondary and tertiary structures. Reducing agents such as DTT or 2-mercaptoethanol can be added in the presence of SDS to break the disulfide bridges and therefore unfold the protein completely. There is a linear relationship between the logarithm of the molecular mass of the protein and relative migration distance of SDS-polypeptide micelle (Laemmli, 1970; Weber & Osborne, 1969)

1.5.3 Liquid Chromatography Mass Spectrometry

1.5.3.1 Reversed-phased high-performance Liquid Chromatography

Reversed-phase chromatography separates molecules based on their reversible interaction with the hydrophobic surface of a chromatographic medium. It is very useful for the separation of protein and peptides. Since the mid 1980s,

reversed-phase chromatography has been widely used for the separation of proteins on both analytical and preperative scales. However, reversed-phase chromatography is not recommended for protein purification if recovery and return of correct tertiary structure are required, because many proteins are irreversibly denatured in the presence of organic solvents (Simpson, 2003)

The mobile phase in reversed-phase chromatography is an aqueous solution, which dissolves the sample and equilibrates the column. Gradient elution allows the separation of a complex mixture component that exhibits a broad range of receptivity in a single run. The more hydrophobic the solute, the higher is the affinity of the solute to the stationary phase and therefore the greater is its retention time. As a result, molecules with different equilibrium constants elute at different times and are separated (Kastner, 2000). The order of the protein desorption is based on their relative hydrophobicity, least hydrophobic proteins elute first, followed by more hydrophobic proteins. Because the functional groups on the column support are hydrophobic in nature, the elution of protein and peptide from the column requires organic solvents and other additives for elution (Simpson, 2003). Due to the retention mechanism, a small change in concentration of the organic eluent component can dramatically affect retention of peptides and proteins (Kastner, 2000).

1.5.3.2 Mass Spectrometry

Mass spectrometry is an analytical technique that measures the masses of individual molecules and atoms. It consists of three major elements: an ion source, mass analyzer and a detector. Mass spectrometers are capable of forming, separating, and detecting molecular ion based on their mass-to-charge ratio (m/z) (Simpson,

2003). The first essential step in mass spectrometry analysis is to convert the analyte molecules into gas-phase ions because the motion of ions can be manipulated and detected. The excess energy transferred to the molecule during the ionization leads to fragmentation. Then, a mass analyzer separates these molecular ions and their charged fragments according to their mass-to-charge ratio (m/z). After that, a detector will detect the ion current due to these mass-separated ions and display a mass spectrum. Ions from colliding or interacting with other ions, each of these steps is carried out under high vacuum (Dass, 2001).

There are only two ionization modes relevant to proteomics, namely MALDI and ESI (Roepstorff, 1997). ESI is an atmospheric pressure ionization technique applicable to a wide range of liquid matrix compounds (Dass, 2001). Charged ions from molecules are formed by electrospray when a solution containing the sample enters the ESI source through a flow stream. In the presence of a strong electrical field, a fine spray of charged droplets is created from the solution through a small-diameter needle. A flow of hot-bath gas, usually nitrogen is added to assist in the evaporation of the solvent from those charged droplets. The ions are then transferred from the atmospheric pressure region to the high-vacumm region of mass analyzer for further analysis. All sample ions are stripped off the solvent molecules before enter into a mass analyzer (Dass, 2001; Simpson, 2003). ESI ionization forms multiple charge states of ions from a single precursor. Composition and pH of the electrospray solvent, as well as the chemical nature of the analyte will influence the charging of the peptides. ESI is widely used, due to its continuous-flow operation, tolerance to different types of solvent, acceptance of wide solvent flow rates, and

ability to generate intact multiply charged ions from electrically charged liquid droplets (Dole *et al*., 1968).

After the ionization, the peptide ions are ejected to the gas phase. These ions are then drawn into MS which operates under vacuum, ions are trapped by a radio frequency trapping field. A small amount of helium gas is used to collisionally cools ions and forces them into the center of the ion trap. Ions above a certain minimum m/z remain trapped, cycling in a sinusoidal motion. The magnitude of the RF voltage determines the frequency and motion of the ions in the trap. The RF voltage is ramped up linearly to generate a mass spectrum. When a small voltage is applied across the two endcap electrodes, it causes ions of successive m/z values to become unstable and to be ejected axially from the trap. This process is termed resonance ejection. As ions are ejected from the trap they are detected by an off-axis conversion dynode with an electron multiplier detector (Corthals *et al*., 2000).

1.6 Western Blotting

Western Blotting combines the resolution of gel electrophoresis with the specificity of immunochemical detection (Harlow, 1999). It is a powerful tool to detect one or more antigens in a mixture (Hames *et al*. 1990) and characterize a variety of proteins, especially those of low abundance. Western blotting is particularly useful when dealing with antigens that are insoluble, difficult to label or easily degraded (Harlow, 1999). Protein blotting or western blotting involves transfer of proteins to an immobilizing membrane. Electrophoretic transfer of resolved proteins from a polyacrylamide gel to a nitrocellulose or polyvinylidene difluoride (PVDF) sheet is the most widely used blotting method. It uses the driving force of an

electric field to elute proteins from gels to immobilize them on a matrix (Gravel, 2002). This method is fast, efficient and maintains the high resolution of the protein pattern (Garfin and Bers, 1982). There are two types of electrophoretic transfer methods: wet transfer or semi-dry transfer.

For semidry blotting, the gel and membrane are sandwiched horizontally between two stacks of filter papers that were buffered-wetted and directly contact with two closely spaced solid-plate electrodes. The name semidry refers to the limited amount of buffer that is confined to the stacks of filter paper. Semidry blotting requires considerably less buffer than the tank method (Gravel, 2002). Wet transfer devices are typically composed of a tank that holds the transfer buffer and one or more cassettes that slightly compress the fragile polyacrylamide gel against the transfer membrane and hold it perpendicular to the electrical field generated by the unit (Lee and Nilsen-Hamilton, 2001).

Once transferred to a membrane, proteins are more readily and equally accessible to various ligands than they were in the gel (Gravel, 2002). A blocking buffer containing bovine serum albumin (BSA) or non-fat skimmed milk is added to saturate unoccupied protein-binding sites in order to prevent non-specific binding of the antibody to the membrane. This is done before adding the primary antibody. After that, the location of the specific antigens is determined using a labeled primary antibody or an unlabeled primary antibody, followed by a labeled secondary antibody. The secondary antibody that is specific to the primary antibody can be radiolabeled or covalently conjugated to an enzyme. Finally, a substrate is added to