CHAPTER II

LITERATURE REVIEWS

2.1 Cholangiocarcinoma

Cholangiocarcinomas are tumors originated from cholangiocytes. Mostly, tumors occur at the biliary confluence (Klatskin or hilar cholangiocarcinoma), in some cases tumors occur within the liver or distal to the hilum. It is a rare malignancy but the incidence of cholangiocarcinoma is increasing globally (Patel, 2001; Taylor-Robinson et al., 2001). However, it remains the second most common primary hepatobiliary malignancy. Some advances have been made in the understanding and treatment of this withering disease since the description of Klatskin (Klatskin, 1965). Most cholangiocarcinomas arise in the absence of any known previous inclination (Chapman, 1999). Many investigators have focused on the molecular and cellular perturbations that characterize this malignant phenotype.

2.1.1 Diagnosis

Bile-duct obstruction and jaundice, clay-colored stools, cola-colored urine, and pruritus are the most common presenting symptoms of biliary tract cancer. These symptoms tend to occur early if the common hepatic duct, the common bile duct, or the ampulla of Vater is the tumor location site. They develop later in perihilar disease and, when present, are often markers of advanced disease in cancer of the gallbladder and intrahepatic cholangiocarcinoma. Pain in the right upper quadrant is the most frequent presenting symptom in gallbladder cancer but not in cholangiocarcinoma. Usually, pain, fatigue, malaise, and weight loss occur in advanced disease. The combination of acute right upper-quadrant pain, fever, and chills, in association with cholestasis, strongly suggests cholangitis. Cholecystitis in elderly patients is sometimes the first manifestation of gallbladder cancer. Sometimes, pancreatitis is the first manifestation of a periampullary tumor. The physical

examination may reveal jaundice, right-upper-quadrant pain, hepatomegaly, and a palpable gallbladder or mass, depending on the location and stage of the tumor (de Groen et al., 1999).

Currently, the central figure of diagnosis is imaging modality, i.e., Magnetic resonance imaging (MRI) with concurrent magnetic resonance cholangiopancreatography (MRCP) is the radiologic modality of choice (Angulo et al., 2000; Oberholzer et al., 1998; Textor et al., 2002). It allows visualization of the location and extent of biliary disease as well as hepatic parenchyma. Cholangiocarcinomas appear hypointense on T1-weighted images and hyperintense on T2-weighted images. Image enhancement can be observed using superparamagnetic iron and delayed gadolinium images (Braga et al., 2001; Peterson et al., 1998). MR angiography can be performed to assess vascular encasement (Lee et al., 2003). Computed tomography (CT) can assess hepatic parenchyma, intrahepatic tumors, biliary dilatation, and lymph nodes. CT angiography allows excellent visualization of the vasculature (Teefey et al., 1988). Ultrasound is non-specific; it may identify intrahepatic mass lesions, and bile duct dilatation proximal to the obstructing lesion. To assess respectability or eligibility for transplantation, the endoscopic ultrasound guided regional lymph node sampling can be performed in early disease (Fritscher-Ravens et al., 2000). However, endoscopic aspiration of hilar masses is not recommended because of the potential for tumor seeding.

In the setting of biliary obstruction, biliary instrumentation is necessary. It should also be performed to sample suspicious lesions for histologic and cytologic analyses. In the study, the sensitivity of routine cytology varied from 9 to 24% and specificity varied from 61 to 100% (Harewood et al., 2004). Besides to routine pathology and cytology, advanced cytologic techniques can now be recommended for the evaluation of aneuploidy, a hallmark of cancer. These techniques are fluorescence in situ hybridization (Chertov et al.) and digital image analysis (DIA) (Baron et al., 2004; Kipp et al., 2004). FISH utilizes fluorescent probes to identify chromosomal amplification (i.e., the actual number of a given chromosome in a cell) and DIA quantitates nuclear DNA as a ratio of normal ploidy (2N). The addition of either DIA or FISH to routine cytology increases the sensitivity without

compromising the specificity for diagnosis of cholangiocarcinoma. The combination of DIA and FISH fulfill the highest sensitivity for the diagnosis of malignant biliary stricture in patients with and without primary sclerosing cholangitis (PSC). Either DIA or FISH positivity had 67% sensitivity in the diagnosis of cholangiocarcinoma in PSC patients with normal cytology; however the specificity was 75%. When combined, DIA and FISH had a sensitivity of 14% in patients with PSC with normal cytology for the diagnosis of cholangiocarcinoma while retaining a specificity of 98% (Malhi and Gores, 2006).

2.1.2 Classification

Three different types of cholangiocarcinoma can be classified by anatomic location including intrahepatic cholangiocarcinoma, hilar cholangiocarcinoma and distal extrahepatic bile duct cancers (Figure 2-1). Hilar cancers or Klatskin tumors are the most frequent (50–60%) that occur at the confluence of the right and left hepatic ducts, and can involve the liver by direct extension (Khan et al., 2002). Because of the closeness of hilar and left and right branch tumors to the liver, they are included in intrahepatic cholangiocarcinomas in the SEER (Surveillance Epidemiology and End Results) database frequently used for epidemiologic studies. Intrahepatic cholangiocarcinoma refers only to tumors that originate within the hepatic parenchyma. Ten percentages of non-hilar tumors are intrahepatic and 20–30% are extrahepatic distal bile duct tumors.

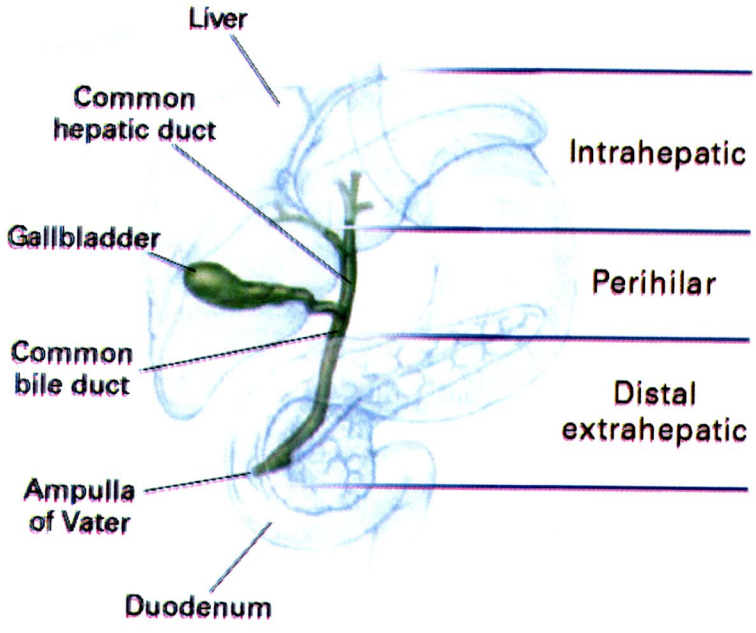

Figure 2-1 Classification of cancers of the human biliary tract according to the anatomic locations (de Groen et al., 1999).

Gross tumor morphology is the basement of further classification of both intrahepatic and extrahepatic tumors. Include of growth pattern: mass-forming, periductal-infiltrating, or intraductal-growing cholangiocarcinoma that use to identify 3 categories of cholangiocarcinoma (Lim and Park, 2004). In an alternative system, extrahepatic tumors alone are classified into nodular, sclerosing, or papillary (Figure 2-2), a description that matches respectively to the broader morphologic classification (Weinbren and Mutum, 1983). As the suggestion by Bismuth-Corlette, hilar lesions are further classified, a pragmatic classification which corresponds to surgical decision making rather than tumor biology. Histologically, most tumors (>90%) are well to moderately differentiated tubular adenocarcinomas. The other outstanding histologic feature is the presence of a desmoplastic reaction. The fibrosis is variable, but may be profound in some cases, leading to a low diagnostic yield of random biopsies. Papillary adenocarcinoma, signet-ring carcinoma, squamous cell or muco-epidermoid carcinoma and a lymphoepithelioma-like form are rare histologic variants.

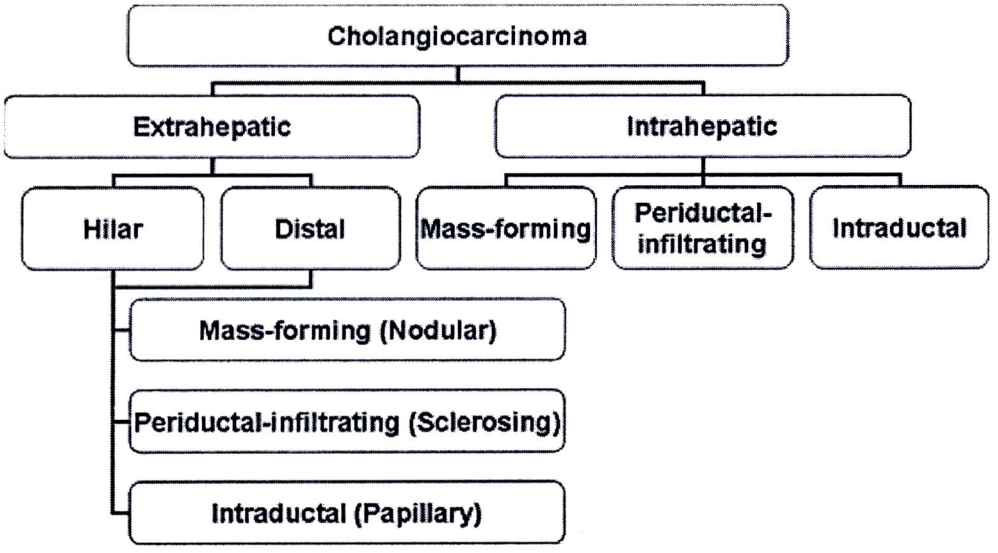

Figure 2-2 Classification of cholangiocarcinoma; cholangiocarcinomas are broadly classified into intrahepatic (also known as peripheral) or extrahepatic tumors. Each is morphologically classified further into mass-forming, periductal-infiltrating or intraductal- growing. This classification also corresponds to the depiction of extrahepatic tumors as nodular, sclerosing or papillary (Malhi and Gores, 2006).

2.1.3 Biomarkers

The important tools for cancer detection and monitoring are biomarkers. They provide as hallmarks for the physiological status of a cell at a given time and change during the disease process. Gene mutations, alterations in gene transcription and translation, and alterations in their protein products can all potentially serve as specific biomarkers for disease.

Serologic testing is supportive and in most instances done prior to or concurrent with imaging studies, despite not necessary for the diagnosis of cholangiocarcinoma. The focus of this discussion is on tumor associated markers and not routine liver test abnormalities. Actually, a large number of potential markers of biliary tract cancers have been identified. Many markers are not specific and may also be present under nonmalignant conditions (Table 2-1)

Table 2-1 Potential tumor markers in gallbladder cancer and cholangiocarcinoma.

Marker	Reference
In bile	
Tumor antigens or products	
Carcinoembryonic antigen	Ker et al, 1989
CA 19-9	Ker et al, 1991
CA 125	Ker et al, 1991
Sialyl-Tn antigen	Sasaki et al, 1996
Fibronectin	Korner et al, 1994
Oncogene	
K-*ra s*	Rijken et al, 1998, Voravud et al, 1989
Tumor-suppressor gene	
p53	Suto et al, 1998
Metabolic product	
Lactate	Nishijima et al, 1997
In serum	
Tumor antigens or products	
Carcinoembryonic antigen	Kuusela et al, 1991
CA 19-9	Kuusela et al, 1991, Su et al, 1996
CA 50	Kuusela et al,1991
CA 125	Su et al, 1996
CA 195	Bhargava et al, 1989
CA 242	Kuusela et al, 1991
DU-PAN-2	Maeda et al, 1996

Table 2-1 Potential tumor markers in gallbladder cancer and cholangiocarcinoma. (Cont.)

Marker	Reference
Cytokeratin 19 fragment	Kashihara et al, 1998
Protein induced by the absence of vitamin K or antagonist II (PIVKA-II)	Nakao et al, 1997
Cytokine	
Interleukin-6	Goydos et al, 1998
Proteases	
Trypsinogen-2	Hedstrom et al, 1996
Trypsin-2–a1-antitrypsin complex	Hedstrom et al, 1996
Peptide	
Pancreatic polypeptide	Bruckner et al, 1993

(de Groen et al., 1999)

From all possible biomarkers, the most studied tumor associated markers with cholangiocarcinomas are CA 19-9, CA 125 and CEA. Among them, CA 19-9 is the most utilized marker. In patients with PSC a value of >100 U/mL has a sensitivity of 89% and specificity of 86%, and in patients without PSC the sensitivity is 53%, for the diagnosis of cholangiocarcinoma (Lim and Park, 2004; Nichols et al., 1993). When increase the cutoff to >129 U/mL in patients PSC provided a sensitivity of 78.6%, and improved the specificity to 98.5%, and a positive predictive value of 56.6% (Levy et al., 2005). In patients with advanced disease, the level may be conspicuously elevated, despite not universally. CEA and CA-125 are not specific as they are elevated in patients with cholangiocarcinoma and other gut-derived malignancies, and they have a low sensitivity as well (Chen et al., 2002). The warning to the interpretation of an elevated CA 19-9 is the coexistence of bacterial cholangitis, which may not always be clear. Hepatolithiasis also commonly leads to increased levels of CA 19-9, CA 125, and also CEA. Hence, cholangiocarcinoma should not be diagnosed on the basis of these tests alone except in patients with known risk factors or virtually diagnostic imaging studies. Another study in unrelated at-risk population demonstrated the utility of testing serum interleukin-6 levels in conjunction with CA

19-9 in patients with chronic biliary parasitosis. A CA 19-9 value >100 U/mL combined with an IL-6 value >50 pg/mL had a sensitivity of 80% and accuracy of 76% (Tangkijvanich et al., 2004).

In conclusion, cancer in doubting cases can confirm by using multiple parallel diagnostic algorithms. Cancer should be carefully obtained and confirmed or excluded in patients with high grade strictures, elevated CA 19-9 and suspicious masses. Histological diagnosis is the gold standard, percutaneous or trans-luminal approach is not recommended because of the risk of tumor seeding. In an ideal manner, ERCP can obtain biliary tissue. However, the diagnosis of cholangiocarcinoma could not exclude the non-diagnostic biopsy or cytology. Screening biliary instrumentation in stable, asymptomatic PSC patients is not recommended due to the concurrent risk of pancreatitis, when clinically indicated advanced testing should be utilized to diagnose early cancers in PSC patients. Asymptomatic high risk patients may be surveyed non-invasively, such as with a serum CA 19-9 value and MRCP annually to achieve early cancer detection, despite the fact that there are no outcome studies or cost-effectiveness information to support this approach (Malhi and Gores, 2006).

2.2 Plasma and Serum Protein

Plasma is the liquid element of blood (Figure 2-3), in which the blood cells are suspended. Plasma is the largest single component of blood, containing about 55% of total blood volume. For any type of hematological test, plasma is obtained from whole blood. To preclude clotting, an anticoagulant, such as citrate or heparin can be added to the blood instantly after it is gained. To separate plasma from blood cells, the sample is then centrifuged. For many biochemical laboratory tests, plasma and blood serum can be used exchangeably. Serum looks like plasma in composition but lacks the coagulation factors. It is obtained by allowing a blood specimen to clot before centrifugation. For this aim, a serum-separating tube can be used which contains an inert catalyst (such as glass beads or powder) to help clotting as well as a distribution of gel with a density designed to sit between the liquid and cellular layers in the tube

after centrifugation, making their separation more suitable.

Even though serum is flavored for many tests because the anticoagulants in plasma can sometimes interfere the tests. Many authors shown that protein profiles derived from plasma and serum are very different. While serum is used for most studies, further research on this topic is needed. The use of both of them was ideal resolution, although that would complicate data analysis and need longer processing times (Luque-Garcia and Neubert, 2007).

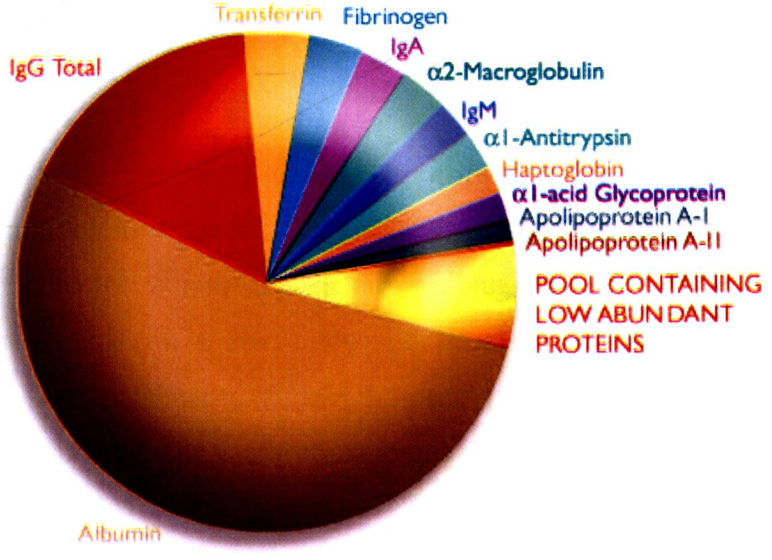

Figure 2-3 Twelve proteins comprise up to 96% of the protein mass in plasma. (Rodriguez, 2009)

The plasma proteome includes 1) proteins secreted by solid tissues, which play in the plasma, 2) immunoglobulins, 3) ''long distance'' receptor ligands include the classical peptide- or proteohormones, 4) ''local'' receptor ligands, such as cytokines and short distance mediators, 5) temporary passengers such as non-hormone proteins on their way to the structural site of primary function, 6) aberrant secretions, that released from tumor cells and diseased tissue, and finally 7) foreign proteins resulted from infectious organisms. Interestingly, only a small amount of proteins in the plasma are in fact extracellular proteins. Plasma or serum protein profiling is promising to detect any of these proteins with their variation, often tissue-specific origins and disease-mediated metabolic modifications.

2.2.1 Variation in plasma/serum protein

There are many influence factors that affect the concentration of proteins in plasma of patient with disease, and if these are not taken into consideration, the detection of medically significant changes will be more difficult. Influence factors can be considered as 1) genetic, 2) non-genetic other than medical treatment, 3) related to medical treatment, such as drug therapy, and 4) related to sample handling. Since methods of venipuncture typically used in studies of plasma proteins are well standardized, the last factor is typically a small effect and a recent report presents that extended storage at -70 °C preserves the structure and activity of plasma proteins tested.

The effects of differing genetic constitutions are evident both at the population level (*e.g.* race differences in plasma protein abundances) and the individual level. In twin studies, which quantitative protein measurements are compared within and between monozygotic twin pairs, have been performed for a small number of individually assayed plasma proteins as well as by simple one-dimensional electrophoresis. In combination these studies show that 12–95% of the quantitative variation in specific plasma protein levels is genetic in origin with an average of 62% for the proteins shown. From the result of the one-dimensional or two-dimensional approaches, the gel patterns of plasma proteins from monozygotic twins are quantitatively almost indistinguishable.

This level of genetic control is consistent with the observation made frequently in epidemiological studies and reference value investigations that an individual's level of a specific protein can be quite stable over time, showing less variation than the population variance. From a series of published studies (Table 2-2), the average intraindividual coefficient of variation (CV) (an individual sampled over time) for a series of proteins is ~23%, and the average interindividual CV is ~45%, mention that on average the individual's CV may be half that of the population. The C-reactive protein data in Table 2 provides a clear example of this relative individual steadiness where the average CV for an individual's repeated samples is 37%, while the population overall CV is 94%. In the published studies examined, both genetic

Table 2-2 Intraindividual and interindividual coefficients of variation for selected plasma proteins TPA, tissue plasminogen activator; CEA, carcinoembryonic antigen.

Protein	**CVi %** **(individual)**	**CVg %** **(group)**
Soluble interleukin 2 receptor (sil2r)	5.84	38.81
Prothrombin	7.3	14.5
Lipoprotein(a)	8.6	85.8
Myoglobin	11	24.3
Sex hormone-binding globulin	12.1	42.7
CEA	13.9	58.3
Follitropin	17.3	36.0
Haptoglobin	20.0	27.9
Interleukin-8	24.0	31.0
Lutropin	24.0	37.0
Interleukin-1B	30.0	36
TPA	31.4	62.5
CA 125	35.5	70.6
C-reactive protein (CRP)	42.2	92.5
Tumor necrosis factor-a	43.0	29.0
Interleukin 6 (IL-6)	48.5	39.4

(Anderson and Anderson, 2002)

and intra-CV *versus* inter-CV methods examined the only one protein that is lipoprotein A, and it showed strong evidence of genetic control of plasma abundance in both cases.

Non-genetic or non-treatment factor effects on plasma protein concentrations are extremely varied. The concentration of total protein in plasma varies by up to 10% depending onto the posture of the subject. The higher arterial pressure forces liquid out of the capillaries when one stands, and the blood volume is about 10% less than when one lies down (the change occurring over 10–30 min). Since this effect relates to a change in blood volume, it is likely to affect only absolute rather than relative protein concentrations. Correspondingly, there is a significant and progressive increase in the concentration of proteins with time after venous closure by the tourniquet, which is always applied before venipuncture, as well as differences between the first and last tube of blood drawn (particularly in proteins related to hemostasis). Other influences factors that effect protein composition of plasma are bed rest (short term increase, long term decrease in plasma protein), exercise, time of day, smoking, age (with four major periods distinguished: newborn, childhood to puberty, adult, and elderly adult), gender (differences generally appearing postpuberty), season, diet (effects of both protein and fat content), sleep, and weightlessness during space flight. The difference between venous blood (the usual source) and lancet-derived ("finger-stick") blood (which is a mixture of extracellular fluid and plasma) has received little attention but may develop into an important issue if serial samples from an individual are to be obtained comfortably. For convenient blood sampling by diabetics, the disposable lancet technology (and the laser puncture equivalent) have been extensively developed, and the volumes obtained (from <1 up to 100 ul) may be sufficient for some proteomic approaches.

In other study, Young (Young, 1997; Young et al., 1972) examined extensively and tabulated periodically for the effects of drugs on protein levels. The effects observed can be more or less direct as when an intramuscular drug injection causes an increase in muscle enzymes in plasma through leakage or localized tissue toxicity or when a diuretic alters total protein concentration by changing blood volume. In exchange, they can result from regulatory changes in target organs as

when liver enzymes are elevated (a side effect associated with a substantial proportion of commonly prescribed drugs, albeit in only a subset of patients). This effects of drugs should not be thought as a noise because they represent real biological effects and because these effects may (indeed should) be related (ideally inversely) to the disease effects they are meant to counteract.

From all factors, genetic effects on protein abundance in plasma are usually the largest and in many ways the least useful from the individual's attitude. Both effects of environment or drug administration are either small or else useful as signals of medically related events. All considerations suggest that the most productive approach to avoid the difficulties of the classical reference range problem is to use the individual as his or her own reference, comparing measurements taken over time to see changes associated with disease or treatment. Like an approach would eliminate the genetic effects on protein levels and go some distance toward minimizing many of the non-genetic influences as well since most lifestyle changes are made slowly. Especially, it appears clear that individual reference intervals could be decreased on average about 50% if a self-standardized approach were taken, potentially improving the detection of many disease states at an earlier stage.

2.2.2 Intrinsic property of plasma/serum: problems and solution

2.2.2.1 Highly abundant proteins

At present, scientists are still at the stage of cataloguing all the plasma proteins, and the duty is frightening enough. The proteins in plasma have a huge dynamic range. There is a difference in the abundance of the most abundant protein, such albumin, and the lowest abundance proteins that have now been measured clinically for medical use by immunoassays, approximately a 10–100 billion-fold, (Anderson and Anderson, 2002) (Figure 2-4). Most proteomic researchers have shied away from exploration plasma because the high-abundance proteins make it so difficult to detect the lower abundance ones. Because the distribution of protein abundances is much more even in tissue and cell culture samples, most of the history of proteomics has been in those ones.

Until currently, 2-DE was the first method for the separation of complex protein mixtures. Recently, there have been important developments in Mass spectrometry (MS), supplemented by better separation methods. Inspirited by these developments, the economic proteomics programmes are turning their attack ships to put serious attempt into discovering the plasma proteome (Habeck, 2003).

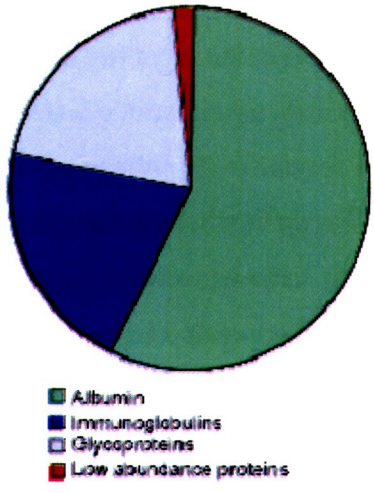

Figure 2-4 Distribution of plasma proteins. Albumin contibutes about 55% of plasma protein; immunoglobulins G, M and A include another 20%; approximately 20 intermediate-abundance proteins (such as glycoproteins, macroglobulins, transferrin) are singly 0.2 – 4.0% of plasma proteins and in total contain another 24% of plasma protein. The low abundance proteins in total include about 1% of the plasma protein (Habeck, 2003).

In fact, as shown in Figure 2-5, even the study of Righetti in 2005, might not be enough for obtaining the deep proteome, in that by this process, they take off 90% of the serum proteins. Recently, HUPO International meeting in Beijing (October 24–27, 2004), it has been advised that one should more treat this drained serum with an additional set of 12 anti-sera able to deplete the next 12 most abundant proteins: this would take away an additional 9% serum proteins, finally revealing to view the truly hidden proteome, the staying 1% buried under this hill of major species.

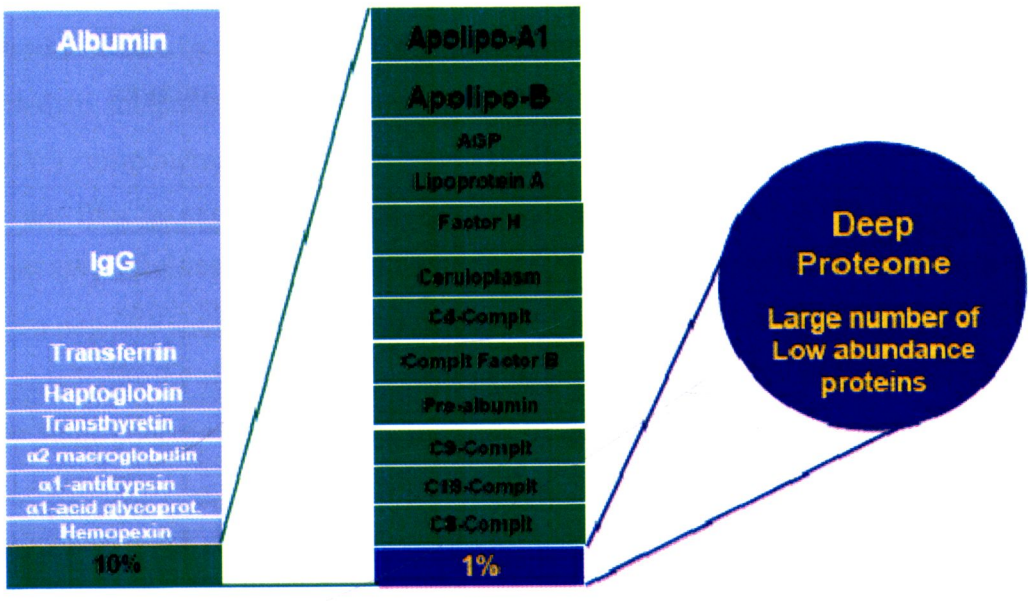

Figure 2-5 The blood proteome; dynamic range, disparity protein concentration and large number of species. The immunosubtraction of the first set of nine most abundant proteins deletes 90% of serum proteins, while deplete of the next set of twelve species takes out another 9%, so leaving the staying 1%, called the deep proteome (Righetti et al., 2005).

2.2.2.2 Depletion of highly abundant proteins

As commented before, broad analyses of serum and plasma have tested to be difficult, particularly for low molecular weight and low abundance proteins, because of the wide range of concentration with the 10 most abundant proteins founding almost 90% of the serum proteome by mass. The different procedures and techniques for deplete of high abundance proteins (see Table 2-3) are reviewed here.

Table 2-3 Procedures used for depletion of highly abundant proteins.

Strategy	Advantages	Disadvantages
Centrifugal ultrafiltration	Fast Easy to operate Inexpensive	Potential loss of components binding to HMW proteins
Solid phase extraction		
- Columns	High selectivity High reproducibility High sensitivity when using a series of different columns	High cost (especially antibody-based depletion columns) Low sample capacity
- Disk plates	Highly suitable for automation	
Organic solvent extraction	Fast Easy to operate	Requires organic solvents Dilution of the sample

(Luque-Garcia and Neubert, 2007)

1) Centrifugal ultrafiltration

Centrifugal ultrafiltration is a category of membrane filtration in which centrifugation forces a liquid against a semi-permeable membrane. While the liquid and low molecular weight solutes pass through the membrane depending on the molecular weight cut off (MWCO) of the membrane used, the suspended solids and solutes of high molecular weight are held. This simple technique has been widely used for the depletion of high molecular weight species for serum/plasma based biomarker exposure for clinical diagnosis of several diseases, such as lung tumors, ovarian cancer, hepatocellular carcinoma, etc. Different MWCO

membranes ranging from 10 kDa to 50 kDa and different centrifugation speeds between 3000 x *g* and 4000 x *g* have been used. Special solvent conditions are needed to interrupt protein–protein/peptide interactions, thus low molecular weight (LMW) components that may be bound to albumin or other larger species are released and are free to pass through the membrane. The addition of acetonitrile to diluted serum or plasma to a final concentration of 20% or 25% (v/v) has been presented to have a positive cause on the enhancement of LMW protein/ peptides without affecting the capabality of the process to remove larger proteins. After adding acetonitrile to 25% final concentration a 15% increase of protein content in the ultrafiltrate was observed. However, a study reported that ultrafiltration unsuccessed to remove albumin and other high molecular weight proteins from human plasma, the ultrafiltration was directed at 12,000 x *g* (Georgiou et al., 2001) and, as it was suggested by Tirumalai et al, 2003, it is possible that at this high centrifugal force the completeness of the membrane may be accommodation to allow high molecular weight components, such as albumin, to pass through. Additionally, non-diluted plasma was used and the ultrafiltration was directed under non-denaturing solvent conditions. Depend on these results, it seems that low-speed centrifugation, using the diluted serum or plasma and use of denaturing conditions are key factors for favorably showing plasma/serum centrifugal ultrafiltration for deplete of highly abundant proteins together with the enrichment of LMW protein/peptides (Tirumalai et al., 2003).

2) Solid phase extraction (SPE)

Solid phase extraction (Echan et al.) is a separation technique that uses a solid phase to separate one, or one type, of analyte from a solution. It is generally used in serum/plasma samples as a clean-up step to deplete high abundance proteins.

(Solid phase extraction columns)

Solid phase extraction columns are seemingly the method most widely used for deletion of high abundance proteins in serum/plasma and they have been widely used as a first step in biomarker-related proteomic studies.

Different types of SPE columns based on ion-exchange, metal-chelating, affinity ligands, dye-ligands, bacterial proteins, antibodies or combinations of these have been used. Any high abundance protein depletion kits have been marketed by companies, such as Agilent, GenWay Biotech, Bio-Rad, Sigma–Aldrich, Amersham Biosciences, Pierce and others. Different chromatography designs are used that include columns or cartridges, microcolumns and spin columns.

Deletion procedures based on dye-ligands or ion-exchange are not protein specific. By comparative relation, biological affinity separation based on antibodies, proteins, peptides, nucleotides, lectins, etc. is much more careful for specific target proteins. Bacterial Protein A and Protein G, which specifically bind to the Fc region of immunoglobulin G (IgG), have been favorably used for specific separation of IgGs from serum and plasma, and the antibody-based columns are the first choice as they give more competent, selective and reproducible depletion of high abundance proteins. The major disadvantages of antibody-based depletion columns are those features included to working with antibodies, that is to say, relatively high cost and low sample capacity.

An attractive method by Guerrier et al, 2005 uses a unimodal multidimensional concept to fast attain an efficient fractionation of human serum. The method is based on the use of a column constituted of an overlaid sequence of sorbents. As the sample passes through the different adsorbent layers, proteins within are consequently trapped in agreement with the complementary properties of the sorbent. Instantly the loading and capturing is achieved, the sequence of columns is not collected and each column contains a different supplement of proteins, is eluted separately in a single step and underneath optimal elution conditions. When correlated to classical single-chemistry fractionation based on anion-exchange and pH stepwise elution, the new advanced method showed much lower protein overlap between fractions, and as a result, greater resolution. Importantly higher sensitivity for low abundance species, so found as evidenced by spiking trials (Guerrier et al., 2005).

In accordance with a study by Björhall et al., 2005, they tested the ProteoExtract™ albumin/IgG removal kit was the most choosy and reproducible among the 5 common disposable (single-use) depletion columns (Bjorhall et al., 2005). Currently a method based on a mix of six polyclonal antibodies in a column has been developed to quickly and effectively deplete the six most abundant proteins from serum and plasma in a single purification step. This method is able to deplete 90–95% of the total serum proteins in serum/plasma, while, for example, the combination of a dye-ligand affinity column to deplete albumin and Protein A column to deplete Igs provided a depletion of only 70% of total serum proteins. This multiple affinity removal system (MARS), also called multiple affinity removal columns (MARC) has been commercialized by Agilent Technologies. The major advantage of the MARS antibody column is that it can effectively deplete the six most highly abundant proteins that include different molecular forms and many proteolytic products of these proteins with low nonspecific losses of other proteins, so giving a larger number of detectable species (Figure 2-6) in complex samples, such as serum and plasma.

For example, Echan et al., 2005, comparing different SPE columns, performed that while the most extensive losses happened with all dye-based affinity columns, the MARS columns had the lowest losses of non-targeted proteins. Further, antibodies are relatively stable proteins, so the columns last for many purification cycles (up to 200) if suitable care is taken to minimize proteolysis and column clogging. It has been shown that the MARS system also works efficiently in the spin column format, which allows parallel processing of multiple samples and does not need complex arrangement. Two MARS columns have also been used in tandem to increase the loading capacity to 75 μl of plasma (Echan et al., 2005).

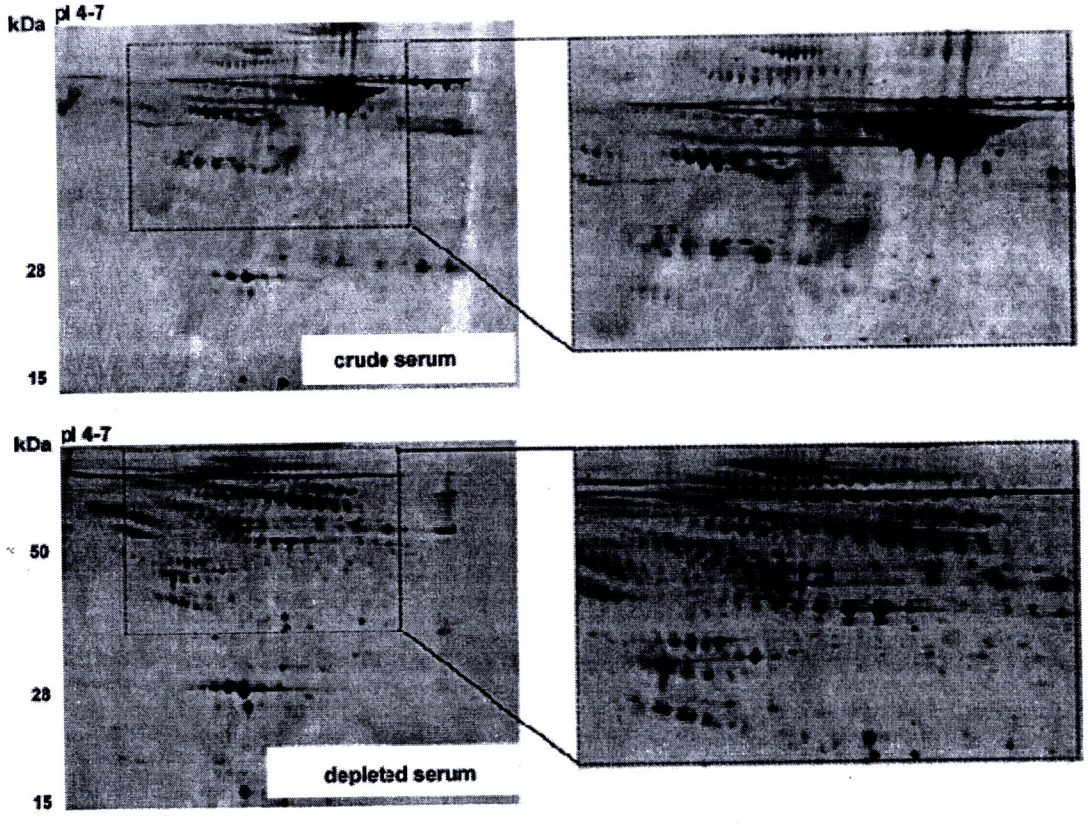

Figure 2-6 Two-DE protein profiles of crude (70 μg protein) and depleted (100 μg protein) serum samples. The multiple affinity removal column was used to delete a total of six highly abundant proteins, such as albumin, IgG, IgA, a1-antitrypsin, transferrin, and haptoglobin. Depletion of these proteins clearly improved the resolution in the area of the gel containing albumin and increased the intensity of low abundance proteins (Luque-Garcia and Neubert, 2007).

Although, the MARS column, has been shown to be the best SPE column to deplete the most abundant proteins from serum/plasma. However, after removing these six abundant proteins, the next most abundant ones rapidly become a problem for identifying lowabundance proteins. Consequently, ideally a highly choosy column that could deplete at least 18–22 of the most abundant proteins, which contain 98–99% of total serum protein component, would be likable. By similar approach, a new column containing 12 polyclonal immunoglobulin yolk (IgY)

antibodies, the Seppro^{TM} mixed 12 spin column (Genway Biotech), may be a hopeful step. The 12 IgY antibodies against the 12 most abundant proteins are covalently bonded to microbeads used to pack the column. High reproducibility and preservation of the separation capacity over multiple cycles has been observed. Recycling of a spin column up to 135 times did not cause clear loss of specificity or capacity.

(Solid phase extraction disk plates)

The high abundance protein is depleted from serum/plasma samples by using SPE disk plate base on the same underlying chemical principles as SPE columns. Different functionalities, such as ion-exchange, dye-ligand and reverse phase disks have been used. The main advantage connected with the use of SPE disk plates is the increased capability for mechanization. The disk plates are commonly used in a 96-well plate format allowing contemporary processing of an advanced number of samples by using a robot. It is also general to use different elution/washing buffers that are applied respectively to the disks in order to guarantee that all the adsorbed protein/peptides are eluted, or for fractionation.

3) Organic solvent extraction

The potential of selectively deleting large abundant proteins from serum has been tested by precipitating them with contemporary extraction of peptides and low molecular weight proteins using organic solvents. The precipitation with organic solvents in the presence of ion-pairing agents separates peptides and smaller proteins from large abundant proteins, by which simplified their extraction. It has been shown that two volumes of acetonitrile added to serum/plasma samples effectively precipitate large abundant proteins, such as albumin, while smaller proteins and peptides still in solution and consequently can be analyzed by MS. Chertov *et al.*, applied this process for preparing of mouse serum samples before analysis by SELDI MS, showing a importance development of the mass spectra. Some of the polypeptide signals shows in the SELDI MS spectra of the extracted material were not present in the spectra of total serum. The review of these peptides in the serum extract was significant for the detection of the two markers that were of lower abundance in the serum extracts from the tumor bearing mice. The extraction process was also very reproducible. Analysis of 40 aliquots of human serum by

MALDI-MS gave actually identical spectra after extraction with acetonitrile containing 0.1% trifluoroacetic acid (TFA). In addition, the extracted material had much less total protein, making it easier to purify and identify the potential markers (Chertov et al., 2004).

2.3 Proteomics

2.3.1 Proteomics: the definition

Currently, the biologic research in early stage tends to study for the role of protein in living organism. Protein is derived from the Greek term "*proteios*" that mean the first rank. It was used to explain these important molecules at the first time by Berzelius in 1838. From movement to mitosis, many functions of specific cell are regulated by 9000 specialized protein types. The posttranslational modifications (PTMs), such as sulfation, phosphorylation, farnesylation, hydroxylation, methylation and glycosylation are the factors that made microheterogeneity, which adds the complexity to specific protein population.

The cell interactions and modifications are the organization of proteins that found within the cell, that make us to understanding biologic systems. The term "proteome" is involved, that can be described as the protein population of the cell in any given time included of localization, PTM, interaction and turnover. The proteome is basically dynamic and has an intrinsic complexity that overcomes the genome or mRNA supplement that found within the cell (transcriptome; Figure 2-7). The analysis of gene expression at the mRNA level is allowed by the development of DNA microarray technology for the reason that transcriptomics provides information about the degree of gene activity in individual tissues and the association to cell function, response to external stimuli, development stage and disease. Even though, transcriptome data will show the genome's objective for the protein synthesis but they don't give information for the finalization of those objectives. Proteome analysis offeres an understanding of the pathologic and physiologic states of an organism by provides a view of biologic processes at their level of apperance and becomes the

important step in the development and validation of diagnostics and therapeutics. The relationship between transcriptome and proteome data have been studied to explain in both useful and poor association between the mRNA and protein concentration and turnover too. The application of proteomic approaches for protein characterization included relative quantification and study of protein-protein interaction can improve the processes of crucial hematopoietic cell development, leukemogenesis and the functional activity of mature cell.

2.3.2 Proteomic technologies development

It is estimated that 45,000 human gene generated to 250,000 spliced variants of RNA then translated to over 1.5 million proteins due to posttranslational processing and modifications. This complexity lead to the study for associating the protein circuits that operates in cancer cells and their surrounding microenvironment to achieve essentially clinical prediction of the disease. Presuming that the proteome is the global example of all biological processes that happen in cancer cell, It would seem difficult to discover the specific biomarker in biological complexity if without of ultra-high resolving analytical techniques for quantitative measurement of thousands of components, strong data acquirement and analysis techniques to productively and reliably process these large datasets.

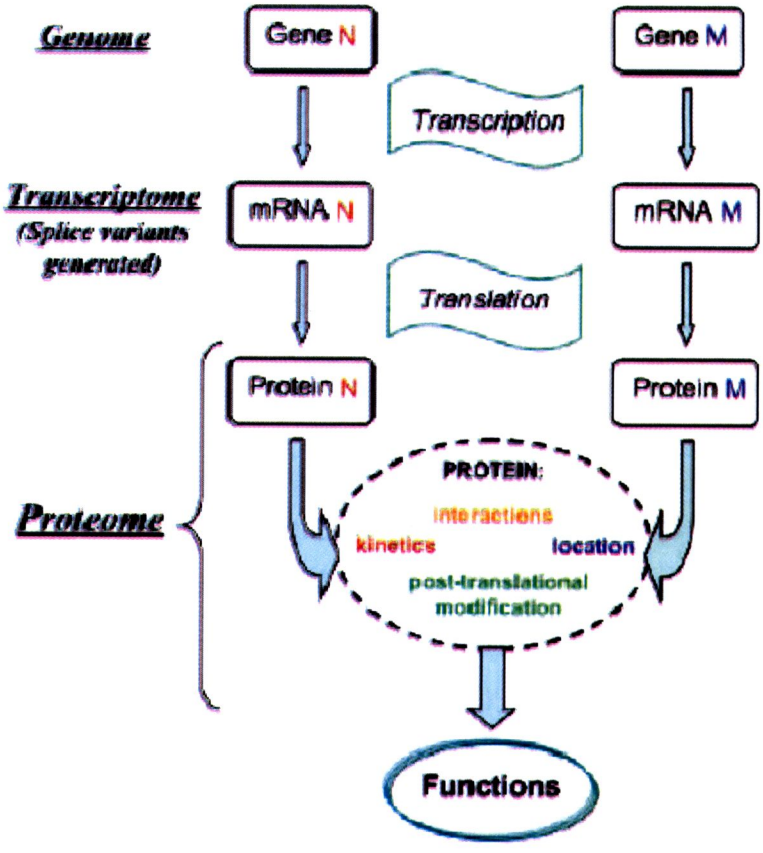

Figure 2-7 Advancement from genome to proteome. The coding DNA strand is transcribed into an mRNA, which is edited by intron excision and the joining of exons when a gene is expressed. At the transcriptome level, the study of mRNA expression by a genome at a given time is usually achieved using microarray analysis. Proteins are synthesized and may experience co-translational and posttranslational modification processes that are often involved in the formation of the functionally active structure of the protein. A given mRNA sequence can give rise to more than one protein. This figure illustrates aspects of this process. Because proteome analysis gives a view of the biologic processes at their level of appearance, proteomics offers a better understanding than genomics of cell cycle, cell death, development stage, cell function, and cellular responses to external stimuli and disease. Proteomics has become an important step in the development and validation of diagnostics and therapeutics (Steiner and Witzmann, 2000).

Current improvement in proteomics has been largely caused recent developments in mass spectrometry (Kipp et al.) based technologies. Especially, new techniques for the ionization of proteins and peptides such as matrix-assisted laser desorption-ionization (MALDI) and electrospray ionization (ESI) combined with time-of-flight (TOF), as well as new hybrid mass spectrometers, are now suitable for the tools of choice for protein characterization. These advances have been highly accepted by the scientific community to include two mass spectrometrists, Drs. John B. Fenn and Koichi Tanaka as co-recipients (with the developer of NMR Dr. Kurt Wüthrich) of the 2002 Nobel Prize for chemistry. These techniques have also been supplemented, although with a significant lag, by dramatic improvements in bioinformatics' tools for analysis of complex datasets. Moreover, powerful multi-dimensional chromatographic and sample labeling techniques have been developed to further benefit from the improvements in mass spectrometry.

The standard proteomic approach for biomarker research include of (i) isolation of cell proteins from clinical specimens (tissue or biological fluids such as serum, ascites, saliva, etc.), (ii) digestion with proteases such as trypsin, and (iii) separation of the resulting mixture by two-dimensional polyacrylamide gel electrophoresis (2-DE) or liquid chromatography (LC). The desired spots (2-DE) or protein fractions (LC) are isolated, digested, and peptides are separated by LC and depending on the sample complexity, the low-molecular weight fractions may be further fractionated by ion-exchange chromatography. For qualitative and quantitative comparisons, the peptides are then put through electrospray or MALDI mass spectrometry (Kipp et al.) or MS/MS analysis. The mass spectrometer measures the mass to charge ratio (m/z) of the components and then can achieve more detailed experiments to identify the components. For peptides, one gains a partial sequence that is usually enough for identification of the peptide or parent protein from which the peptide was derived. As a result, for each sample, one can get a three dimensional dataset (retention time, m/z and intensity).

For a number of years, 2-DE followed by protein identification using Mass Spectrometry has been the primary technique for biomarker discovery in conventional proteomic analyses. This technique is uniquely suited for direct

comparisons of protein expression and has been used to identify proteins that are differentially expressed between normal and tumor tissues in various cancers, such as liver, bladder, lung, oesophageal, prostate and breast.

Despite its utility, there are several inherent disadvantages to 2D-PAGE. It requires a large amount of protein as starting material, and the technique cannot be reliably used to detect and identify low-abundance proteins (Table 2-4).

2.3.3 Serum proteome in cholangiocarcinoma

Cholangiocarcinoma (CCA) is a rare but disastrous neoplasm that counting for 3% of all gastrointestinal cancers and 15% of all primary liver cancers the entire world. The new cases are diagnosed for 5,000 on average per year in the United States, and the trend is increasing. Northeast Thailand sees the highest rate, with 96 cases per 100,000. Surgical reapportion is the only recent opportunity of heal, with no tested for help therapy; yet, this program has a death rate of 5% to 10%. Furthermore, margin-free reapportion, 5-year survival numerals only get 20% to 40%. Non regardable disease is normally fatal within 6 months to 1 year, and over one third of patients show at a stage too late for reapportion. This confirms the need for primary and more certain diagnostic procedures.

Currently, diagnosis of CCA depends on imaging of the biliary tree with computed tomography or ultrasonography in the residence of high clinical doubt. For intrahepatic tumors, tissue diagnostic proving is possible but more distal and more common (2/3 of cases) extrahepatic cases are very difficult. These often show as a limit of the common bile duct, which can be further characterized by indicates of endoscopic backward cholangio-pancreatography. Using this technique, tissue diagnosis can be made through brush cytology and intraductal biopsies, but these methods have poor sensitivities (50%) cause the high desmoplastic nature of the neoplasm. For this reason, clinicians often look for other methods of certain diagnoses, and serum tumor markers give important further keys to the malignant nature of such restrictions.

Table 2-4 Comparison of proteomic technologies and their contributions to biomarker discovery and early detection.

	ELISA	2D-PAGE	Multidimensional protein identification technology (MudPIT)	Proteomic pattern diagnostics	Protein microarrays
Sensitivity	Highest	Overall low, particularly for less-abundant proteins; sensitivity limited by detection method; LCM can improve specificity via enrichment of selected cell populations	High	Medium sensitivity with diminishing yield at higher molecular weights; will improve with new MS instrumentation	Medium/high
Direct identification of markers	N/A	Yes	Yes	No, newer MS technologies might make this possible	Possible when coupled with MS technologies
Use	Detection of single, specific well-characterized analyte in body fluid or tissue; gold standard of clinical assays	Means for discovery and identification of biomarkers, not a direct means of early detection in itself	Detection and identification of potential biomarkers	Diagnostic pattern analysis in body fluids and tissues; potential biomarker identification	Multiparametric analysis of many analytes simultaneously
Throughput	Moderate	Low	Very low	Highest	High

Table 2-4 Comparison of proteomic technologies and their contributions to biomarker discovery and early detection. (Cont.)

	ELISA	2D-PAGE	Multidimensional protein identification technology (MudPIT)	Proteomic pattern diagnostics	Protein microarrays
Advantages/drawbacks	Very robust; well-established use in clinical assays; requires well-characterized antibody for detection and extensive validation; not amenable to direct discovery (strictly measurement based)	All IDs require validation and testing before clinical use; tried and true methodology, reproducible and more quantitative combined with fluorescent dyes	Significantly higher sensitivity than 2D-PAGE (much larger coverage of the proteome for biomarker discovery)	Protein IDs not necessary for diagnostic pattern analysis; reproducibility issues need to be addressed; need for validation; coupling to adaptive informatics tools might revolutionize the field of clinical chemistry	Format is flexible: can be used to assay for multiple analytes in a single specimen or a single analyte in a large number of specimens; requires prior knowledge of analyte being measured; limited by antibody sensitivity and specificity; requires use of an amplified tag detection system

(Wulfkuhle et al., 2003)

Verified serum tumor markers connected with CCA include carbohydrate antigen 19.9 (CA19.9) and carcinoembryonic antigen (CEA) but these markers are not always helpful, with sensitivities of approximately 70% and 50%, severally. Therefore, there is a need for new markers of the disease. Proteomic profiling of serum has been used with a number of other malignancies to find potential biomarkers.

The serum levels of CA19.9 and CEA inside the study group, used diagnostically, provided sensitivities and specificities in good agreement with the literature. Their use in CCA was currently measures by Nehls *et al.*, 2004 who reviewed 11 studies with over 1,200 patients resulting in an average sensitivity of 71% and 51% and an average specificity of 78% and 88% for CA19-9 and CEA, severally. The higher identifying power of CEA is reported both in this study and in the literature. An important finding of this study is that the single protein peak *m/z*, 462 was more identifying than CEA or CA19-9. More characterization of this peak is recently underway (Nehls et al., 2004).

Patrakitkomjorn et al., 2000 used 2-DE approach to compare between plasma proteins from CCA and normal subjects. They found some spots of plasma protein from CCA patients were significantly different from those of normal subjects in both quality and quantity (Patrakitkomjorn, 2000).

In 2006, Scarlett et al. used proteomic techniques to improve the diagnosis of CCA in both tissue and serum by the using of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) to identify potential protein biomarkers of CCA. They found that the training models developed panels of peaks that distinguished CCA from bile duct tissue (92.5% sensitivity, 92.3% specificity), CCA from benign serum (65.0% sensitivity, 70.0% specificity), and CC from sera of healthy volunteers (75.0% sensitivity, 100% specificity) (Scarlett et al., 2006).

2.3.4 Proteomics in other cancers

Early diagnosis of cancer is difficult because of the lack of specific symptoms in early disease and the limited understanding of etiology and oncogenesis. For example, for hepatocellular carcinoma (HCC), the common method of screening high risk patients by alpha-fetoprotein (AFP) and ultrasonography has been shown to result in earlier detection and consequently more easily treatable tumors and longer survival. Of the other tumor markers, the newer high sensitive des-gamma-carboxy-prothrombin has been found to be useful. In addition, the AFP fractions L3, P4/5, and the +II band are highly specific for HCC. Among routinely assayed tumor markers in the laboratory, CA-125 is more sensitive for HCC than AFP but far less specific.

Table 2-5 Common serum cancer markers used in primary care.

Biomarker	Cancer type	Specificity	Example of non-cancer pathology	Primary clinical use
α-fetoprotein	Hepatocellular, non-seminomatous testicular	Moderate	Prostatitis	Staging
Human chrionic gonadotropin-β	Testicular, ovarian	Low	Pregnancy	Staging
CA15-3	Breast	Poor	Cirrhosis, benign diseases of ovaries and breast	Disease monitoring
CA19-9	Gastro, pancreatic, stomach	Poor	Gastritis	Disease monitoring
CA125	Ovarian, cervical, uterine, fallopian tube	Moderate	Pancreatitis, kidney or liver disease	Disease monitoring
CA27-29	Breast			Disease monitoring
CEA	Colorectal, pancreas, lung, breast, medullary thyroid	Low	Non-malignant disorders	Disease monitoring
Epidermal growth factor receptor	Colon, non-small cell lung cancer	Low	Non malignant disorders, such as benign prostatic hyperplasia	Selection of therapy
Her2/Neu	Breast, ovarian	Moderate	Benign breast disease	Disease monitoring; selection of therapy
PSA	Prostate	High	Benign prostatic hyperplasia	Screening; disease monitoring
Thyroglobulin	Thyroid	Poor	Grave's disease thyroiditis	Disease monitoring

CA: Cancer antigen; CEA: Carcinoembryonic antigen; PSA: Prostate-specific antigen

(Alaoui-Jamali and Xu, 2006)

However, most currently available screening tests for cancers lack high sensitivity and specificity (Table 2-5) to be useful in screening the general population, so the differentiation between some benign and malignant tumors is still a clinical challenge. The advent of oncoproteomics has provided the hope of discovering novel

biomarkers for use in the screening, early diagnosis, and prediction of response to therapy (Table 2-6).

Table 2-6 Comparison of proteomic biomarkers and current tumor markers.

Cancer	Proteomic biomarkers			Current tumor markers		
	Sensitivity	Specificity	Reference	Markers	Sensitivity	Specificity
Bladder	80%	90–97%	Mueller,2005	NMP22	31%	95%
Breast	93%	91%	Li, 2002	CA 15-3	63%	80–88%
Colorectal	91%	93%	Chen, 2004	CEA	43%	****
Gastric	83%	95%	Poon, 2006	CEA	49%	****
Liver	94%	86%	Ward, 2006	AFP	50%	90%
Lung	87%	80%	Yang, 2005	Cyfra21-1	63%	94%
Ovarian	83%	94%	Zhang, 2004	CA-125	57%	****
Pancreatic	78%	97%	Koopmann, 2004	CA 19-9	72%	****
Prostate	83%	97%	Adam, 2002	PSA	86%	20–34%

(Cho, 2007)

In other study, Celis and co-workers have utilized 2-DE and MS analysis to identify differential protein expression between bladder cancer and healthy tissue including squamous cell carcinomas versus normal urothelium, which has defined some of the steps involved in the squamous differentiation of the bladder transitional epithelium. On the other hand, making use of 2-DE and MS/MS, Sheng *et al* recently discovered that fatty acid binding proteins, annexin V, heat shock protein (Hsp) 27, and lactate dehydrogenase were associated with bladder cancer. They also found altered expression of a group of proteins in bladder cancer that have not been documented previously, including annexin I, 15-hydroxyprostaglandin dehydrogenase, galectin-1, lysophospholipase, and mitochondrial short chain enoyl-coenzyme A hydratase 1 precursor.

Six proteins (ANXA3, BMP4, LCN2, SPARC, MMP7, and MMP11) were found to be over-expressed in colorectal tumoral tissues by using immunoblotting and tissue microarray analysis. Two of them (LCN2 and MMP11) were clearly over-expressed in late Dukes stages. To identify proteins with colorectal cancer (CRC) specific regulation, comparative 2-DE of individual matched normal and neoplastic colorectal tissue specimens was performed. Endocrine cell-expressed protein secretagogin exhibited a marked down-regulation in CRC tissues. This finding may represent the basis for the clinical application of secretagonin as a biomarker for a distinct subgroup of CRCs.