Chapter - 2

*Literature Review*
2.1. WHAT IS A BIOMARKER

Cancer remains a major cause of mortality in men and women and is the second leading cause of death in the United States. About half a million people died from this disease in 2006 and it accounts for approximately 25% of all deaths each year in the United States (Jemal et al., 2010). One of the most important characteristic of the tumor cells is to spread from the primary tumor to form secondary tumors at the same site or a different location in the body in a process called “metastasis” (Poste and Fidler, 1980). The course of cancer metastasis is shown in Figure 2.1 (from Ref. (Poste and Fidler, 1980)). There are different types of cancers which are named after the location they emerge from.

![Diagram of tumor progression and metastasis](image)

**Figure 2.1: Basics of tumor progression and metastasis.** One or few cells in normal tissue attain the ability to proliferate uncontrollably. These cells get nutrients from the newly developed vasculature and rapidly proliferate and form local tumors. Eventually the cells enter the blood/lymphatic stream and migrate to other tissues (from Ref. (Poste and Fidler, 1980)).
Cancerous cells release various substances into the surrounding tissue produced by the cancer cells themselves, endothelial cells, fibroblasts, cells of the immune system and proteins secreted from the tumor cells. This results in the formation of what is known as the Extra Cellular Matrix (ECM) of the tumor. The molecular signature of these cells is different from that of the surrounding normal cells (Sund and Kalluri, 2009). In many cases, these substances are released into the blood stream. The tumor cells, the substances released by them along with the molecular and physiological changes that occur in these cells during the course of cancer progression or carcinogenesis may serve as good indicators of the disease. They can reflect the type, stage and seriousness of the disease. Identifying these “indicators” or biomarkers at an early stage enables the physicians to detect the disease and choose the best therapeutic measures to treat the patients. Secreted proteins such as DKK1, CHI3L1, LAMB3 and LAD1, membrane proteins such as ITGB6 and ECM proteins such as JUP in the current study have been reported to be associated with various types of cancers and therefore can serve as potential tumor markers (see Table 1.1 in Chapter 1).

One of the major challenges in “the war on cancer” is to identify the disease at an early stage and prevent it before it spreads and takes over the resistance mechanisms of the human body making it incurable (Etzioni et al., 2003). Early detection of cancer has been one of the biggest challenges for clinicians and it is the ultimate goal of the cancer biomarker study. Biomarkers are cellular indicators of the physiological state reflecting changes during the disease process. Active genes and their corresponding protein products and other chemical substances made by the tumor cell can all serve as
these “cellular indicators” (Srinivas et al., 2001). Biological marker or biomarker has been defined in different ways by various researchers. According to Atkinson et al., “biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological process, pathogenic process or pharmacological responses to a therapeutic intervention” (Atkinson et al., 2001). Hulka and colleagues defined biomarkers as “cellular, biochemical or molecular alterations that are measurable in biological media such as human tissues, cells or fluids” (Griffith et al., 1989; Hulka BS, 1990).

2.2 MILESTONES IN THE CANER BIOMARKER FIELD

The first cancer biomarker ever discovered was the light chain of immunoglobulin (popularly known as Bence-Jones protein) in the urine of multiple myeloma patients in a 1848 study and is still in use by clinicians (Jones, 1848) to diagnose this disease. The modern era of tumor markers began with the discovery of alpha-fetoprotein (AFP) (Abelev, 1963) and carcinoembryonic antigen (CEA) (Gold and Freedman, 1965) in hepatomas and gastro-intestinal cancers respectively in the early 1960s. Prostate Specific Antigen or PSA, which is considered as one of the best cancer biomarkers, was discovered in 1980 (Papsidero et al., 1980). The development of Radioimmunoassay (RIA) in 1960s revolutionized the field of cancer biomarkers. It is a very sensitive assay used to measure the concentration of various antigens such as hormones and other proteins in the serum of patients. Yalow and Berson first introduced this technology in 1960 by detecting endogenous plasma insulin in humans. They received Nobel Prize for their discovery (Yalow and Berson, 1960). The major disadvantages with this
technology are that it needs expensive instruments and the use of radioactive materials which are highly regulated and difficult to dispose of. Hence this technology has been replaced in the later years by ELISA. It is a sensitive biochemical technique and one of the most important diagnostic tools in medicine used to detect antigen or antibody in a given sample. This technique was developed by Eva Engvall and Van Weemen independently in the early 1970s (Engvall and Perlmann, 1971; Lequin, 2005; Nakane PK, 1967; Van Weemen and Schuurs, 1971). The principle of ELISA is based on the chemistry of antigen-antibody binding, in a similar way to RIA.

Another major breakthrough in the field of cancer biomarkers was the discovery of hybridoma technology in 1980s. Hybridomas are the cells that have been engineered, by fusing an antibody producing B-cell with a myeloma tumor cell, to produce a hybrid cell that secretes the desired antibody in large amounts. This technology was first discovered by Cesar Milstein, Georges J. F. Köhler and Niels Kaj Jerne in 1975 (KÖHLER and MILSTEIN, 1975) They received Nobel prize for their contribution to the field of medicine. The first antibody employed as a diagnostic marker using hybridoma technology was the ovarian epithelial cancer marker, carbohydrate antigen (CA) 125 (Robert C. Bast et al., 1981), discovered in 1981.

Researchers and clinicians used only a single biomarker to predict the disease progression. But the heterogeneity of cancer development, lack of sensitivity and/or specificity of certain biomarkers has made the diagnosis with just a single biomarker very difficult. To overcome this limitation and improve testing characteristics,
researchers started combining or multiplexing more than one biomarker (Laxman et al., 2008). Highly advanced proteomic and genomic technologies have made the multiplexing of biomarkers more effective in cancer detection (Fredriksson S, 2007; Fredriksson et al., 2008; Vo-Dinh, 2009). Labeling the antibodies with fluorescent probes has changed the way scientists looked at ELISA before. This technology, known as Multiplex ELISA, facilitated the detection of more than one protein in the sample. Various high throughput screening (HTS) ELISA techniques have been introduced in the early 2000s which reduced, to a greater extent, the time, amount of the screening material and expenses without compromising the sensitivity and specificity of the assay. Proteomic and genomic advancements in these areas of cancer biology have opened the gates to a new era in the tumor marker studies. Completion of the human genome project, expression array techniques for nucleic acids and proteins, advanced bioinformatics, MS based profiling and high throughput sequencing has generated optimism among researchers and clinicians for novel biomarker discovery. Some of the milestones in the cancer biomarker field are mentioned in Table 2.1 (modified from (Diamandis et al., 2002)).
Table 2.1: Milestones in cancer biomarker discovery

<table>
<thead>
<tr>
<th>Year</th>
<th>Biomarker discovered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1846</td>
<td>Bence-Jones protein</td>
</tr>
<tr>
<td>1940</td>
<td>Acid phosphatase</td>
</tr>
<tr>
<td>1960</td>
<td>Immunoassay</td>
</tr>
<tr>
<td>1963</td>
<td>Alpha-fetoprotein (AFP)</td>
</tr>
<tr>
<td>1965</td>
<td>Carcinoembryonic antigen (CEA)</td>
</tr>
<tr>
<td>1971</td>
<td>ELISA</td>
</tr>
<tr>
<td>1975</td>
<td>Hybridoma technology and Monoclonal antibodies</td>
</tr>
<tr>
<td>1980</td>
<td>CA125, PSA, carbohydrate antigens</td>
</tr>
<tr>
<td>1970-1990</td>
<td>Oncogenes and tumor suppressor genes</td>
</tr>
<tr>
<td>2001</td>
<td>Microarrays, Mass Spectrometry, Neural networks, Bioinformatics</td>
</tr>
<tr>
<td>2005-present</td>
<td>Proteomics, Peptidomics, SNP arrays, high throughput sequencing</td>
</tr>
</tbody>
</table>

2.3 GENOMICS AND PROTEOMICS IN CANCER BIOMARKER STUDY

In the recent years, two major and rapidly emerging fields of biology are playing key role in the study of biomarkers: genomics and proteomics. Genomics is the study of the complete genome of a cell or organism. Changes in the genetic make up, including down-regulation or up-regulation of gene expression, mutations and deletions, are some of the most important events in the carcinogenesis and these events can serve as valuable biomarkers (Sidransky, 2002). The major advantage with these types of markers is that they facilitate early detection by high throughput technologies since they rely on genetic instability which is an integral part of carcinogenesis and the molecular signature changes occur well before the development of clinical symptoms (Chen and Hunter, 2005; Srivastava, 2001; Verma and Manne, 2006; Vineis and Perera, 2007).
Even though not many genomic-based biomarkers have been validated so far, there are a few that are worth mentioning. Signatures of interest include microsatellite instability (MSI), DNA hypermethylation and single nucleotide polymorphisms (SNPs) (Srinivas et al., 2001).

Microsatellites are polymorphic tandem repeats of short nucleotide sequences of less than six bases and are distributed throughout the genome. Frequent association of inherent instability of microsatellite loci, known as MSI, has been first reported in colon cancer (Thibodeau et al., 1993) and later identified in many other cancers which include tumors of lung (Yi-Ching et al., 2006), breast (Zhu et al., 2003) and bladder (Schneider et al., 2000). Another important genetic event that is well associated with various types of cancers is hypermethylation. DNA methylation, which is thought to be one of the epigenetic mechanisms used by mammalian cells to modify gene functions, has been shown to be associated with cancer (BaylIn et al., 1997; Peter A Jones, 1999). Various studies show that hypermethylation can be used as biomarkers in different types of cancers including endometrial (Jiang et al., 2008), ovarian (Barton et al., 2008), prostate (Masashi et al., 2004), breast (Ross et al., 2003) and colorectal (Giovannucci and Ogino, 2005; Ogino S, 2006). Pathologic features of certain cancers such as colorectal are well correlated with the hypermethylation of CpG islands (Ogino et al., 2006b; Ogino et al., 2006c). Single Nucleotide Polymorphisms or SNPs comprise about 90% of the DNA polymorphisms in human genome (Collins, 1998). According to Brookes, SNPs can be defined as “single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some population(s),
wherein the least frequent allele has an abundance of 1% or greater” (Brookes, 1999). Point mutations, a type of SNP, in KRAS gene have been associated with colorectal and pancreatic cancers. Interestingly, these mutations were observed in the DNA extracted from plasma and urine of the patients (Anker et al., 1999; Botezatu et al., 2000). SNPs have been used as tumor markers in many cancers (Jianfeng et al., 2009; Meng et al., 2009; Pepe and Janes, 2008; Tejpar, 2007). Signet ring cell carcinoma and mucinous carcinoma are two subtypes of colorectal cancers. These two pathologically distinct subtypes can be distinguished based on their molecular signatures (Ogino et al., 2005a). It is also reported that combined expression analysis of certain targets such as COX-2 and p53 may be useful for molecular classification of colorectal cancers (Ogino et al., 2006a). High throughput SNP arrays, quantitative real-time PCR, gene expression analysis by Microarrays, serial analysis of gene expression (SAGE) and array CGH (complementary genomic hybridization) are just some of the important and highly sensitive techniques that have been recently developed for genomic studies.

Proteomics is the large scale study of the structure and function of proteins. Proteome is the entire complement of proteins expressed by a genome, cell or organism (Wilkins MR, 1996). Proteomics can be broadly classified into three categories (Graves PR, 2002). 1. Expression proteomics: study of expression levels of various proteins in a given tissue, body fluid or cell extract. This facilitates the comparison of protein profiles between a health or control sample and a diseased or drug-treated sample; 2. Functional proteomics: study of specific biological function, protein-protein and protein-nucleic acid interactions and post-translational modifications of various
proteins; 3. Structural proteomics: study of structures of various proteins and protein complexes in a given cell or tissue. Change in the expression levels of various proteins is one of the key events during cancer development and the altered proteins can be detected in tissue, blood, urine or other body fluids. Even though this has been the basis for biomarker discovery, sensitive assays to detect and quantitatively measure the altered proteins at early stages of cancer have always been a challenge. But recent advancements in quantitative proteomics have changed this view (Qing-Yu and Jen-Fu, 2003; Srinivas et al., 2002). A number of potential cancer biomarkers have been discovered using sensitive and high throughput proteomic technologies (Wong et al., 2009). An outline of strategies for biomarker discovery using emerging cutting edge technologies is shown in Figure 2.2.
Figure 2.2: Strategies for biomarker discovery using emerging technologies. Various genomic and MS based proteomic techniques are available today that can be used to effectively dig through for potential tumor markers in various clinical specimens.

Investigation of the plasma/serum proteome is gaining much importance in recent years because serum is an ideal specimen for diagnosis. Pathological and cancerous changes in the body can be well reflected as protein changes in the serum (Chen et al., 2005). Proteomic profiling of serum/plasma to identify novel cancer biomarkers utilizing current proteomic technologies has changed the face of cancer biomarker field. Genomic and proteomic profiling of tumor cells or cultured cell lines of a given cancer type has also provided valuable information about the potential markers in a particular cancer. Tumor-derived cells or cell lines can be cultured and subcellular compartments including cell surface proteins, secreted proteins and intracellular proteins can be analyzed by various MS methods and the data obtained can be
subsequently mined to identify potential biomarkers (Faca et al., 2008; Veenstra et al., 2005). MS-based techniques are the most widely used proteomic assays for biomarker discovery. Two dimensional electrophoresis (2-DE) followed by mass spectrometric identification of selected or all detected proteins has been the traditional approach for quantitative protein profiling. But the inability of 2-DE to distinguish small proteins, very basic, acidic or hydrophobic proteins combined with the low-throughput scale of research has limited its use (Aebersold and Goodlett, 2001).

In the recent years, extensive separation of peptides and the incorporation of signature tags into protein or peptides for quantitative analysis is being done by multidimensional chromatography techniques. Proteomic-pattern is a recent approach with the rationale that proteins produced by cancer cells or their microenvironment may eventually enter the circulation and their expression could be assessed by MS. Various MS-based techniques that have been developed recently to analyze the serum proteins with greater sensitivity include SELDI-TOF-MS (surface-enhanced laser-desorption-ionization time-of-flight MS) (Hutchens and Tai-Tung, 1993), MALDI (matrix-assisted laser desorption/ionization) (Karas et al., 1985) and ICAT (isotope-coded affinity tags) (Steven P. Gygi, 1999). The potential of proteomic pattern analysis was first demonstrated in the diagnosis of ovarian cancer (Petricoin III et al., 2002). Later this technology has been successfully extended to various other types of cancer as well (Alex J. Rai and Lori J. Sokoll, 2002; Honda et al., 2005; Huang et al., 2009; Li et al., 2002; Mian et al., 2005; Ou et al., 2008; Wu et al., 2009).
2.4 CANCER BIOMARKERS IN DIAGNOSIS AND TREATMENT

Biomarkers serve a wide range of purposes in drug development, clinical trials, and therapeutic assessment strategies. They also contribute to the understanding pharmacology of drug candidates and help with characterization of the disease subtypes. They may have the greatest value in early efficacy and safety evaluations such as in vitro studies in tissue samples, in vivo studies in animal models and early-phase clinical trials to establish “proof of concept.” They can be used for the accurate evaluation and management of the disease in different stages. Diagnostic tests to detect a particular type of cancer greatly facilitate and accelerate the development of targeted cancer therapies. There are various forms of cancer biomarkers that include physiologic (patient performance status), images including mammograms, specific molecules such as PSA, genetic changes, cell based markers and gene and protein expression profiles (Dalton and Friend, 2006).

According to Kulasingam (Kulasingam and Diamandis, 2008), biomarkers can be broadly classified into three groups: 1. Diagnostic or screening biomarkers which are used to detect and identify a given type of cancer. 2. Prognostic markers, which predict probable course of the disease, are used once the disease status has been established and 3. Stratification or predictive markers which serve to predict the response of a drug before treatment is started. For example, p21 expression in combination with p53 mutation is a predictor of resistance to the combination chemotherapy with gefinitib in colorectal cancer (Ogino et al., 2005b). Elevation of certain cell types such as plasmocytoid dendritic cells in Multiple myeloma can serve as an indicator of the
aggressiveness of the disease which in turn can serve as a valuable tool in designing the therapeutic strategy (Chauhan et al., 2009). Malfunction of apoptotic pathway and proteasome pathway is a common feature of cancer cells. Alterations in protein expression of certain targets such as Bcl-2, Bcl X\(_L\), caspases, USP2a etc. that are involved in these pathways are good indicators of drug response. Therefore they serve as good therapeutic targets and are very useful in the development of novel drugs (Chauhan et al., 2008; Chauhan et al., 2006; Priolo et al., 2006). The usefulness of a biomarker lies in its ability to provide early indication of disease or the progression of the disease. They serve as valuable tools in cancer disease staging or grading (Ludwig and Weinstein, 2005). Molecules or substances that serve as tumor markers include nucleic acids, DNA and mRNA in particular, metabolic products such as proteins, carbohydrates, lipids and various biological processes including blood vessel growth and cellular responses (Hayes et al., 1996). Substances that are produced by the tumor itself or by other tissues in response to the disease are the most widely used tumor markers because they directly reflect the severity of the disease.

An ideal tumor marker should (1) be produced by the tumor and enter the circulation, (2) be present at low level in the body fluids of healthy individuals and at significantly higher levels in cancer patients, (3) be present at detectable amounts even at pre-clinical stages, (4) be easily quantifiable with an inexpensive assay and reflect the tumor the tumor burden and most important of all (5) be highly sensitive with least possible false negatives and highly specific with least possible false positives.
Easy accessibility of the clinical specimen is another important factor in the biomarker study. Some of the most commonly used clinical specimens include various types of body fluids including serum, plasma, cerebrospinal fluid (CSF), urine, ascites fluid, pancreatic juice, ductal lavage, sputum, saliva and stool depending on the location of the tumor. Among all these, serum is the most widely used clinical specimen because of its easy and inexpensive accessibility. Moreover, many biological and pathological changes that occur even before the onset of the disease are reflected as protein changes in the serum, making it an ideal specimen for the early detection of cancer.

For pancreatic cancer, in particular, Imaging techniques such as Dual-phase helical computed tomography (CT) (Diehl et al., 1998; Fuhrman et al., 1994; Midwinter et al., 1999) and Endoscopic ultrasound (EUS) (Hawes, 2010) are widely employed for diagnosis and staging of suspected PDAC. Fine needle aspiration (FNA) combined with EUS has also offered the possibility to reach a definite diagnosis (Galasso D, 2010). Some other imaging techniques for PDAC diagnosis include MRCP (Magnetic Resonance Cholangiopancreatography), ERCP (Endoscopic Retrograde Cholangiopancreatography) and MRI (Magnetic Resonance Imaging). Early detection of PDAC is unfortunately an infrequent situation at the present time. Despite its low sensitivity and specificity, CA19-9 is the only serum biomarker available to date used for the detection of PDAC. Needle biopsy followed by histopathological evaluation of pancreatic tissue can only confirm the disease but cannot be used to detect the cancer at an early stage. Other serum biomarkers that are reported for PDAC diagnosis include CEA, Haptoglobin, CA242 etc (Bünger S, 2010).