1

OVERVIEW

1.1. INTRODUCTION

Cancer exacts a tremendous price on society through devastating effects on patients and their families, tremendous economic costs in terms of direct medical care for its treatment, and the loss of capital because of early mortality. The idea of the early detection of various forms of cancer before they spread and become incurable has tantalized both physicians and research scientists for decades. Although such an objective is still too far, it is encouraging to note that our increasing understanding of the biology of cancer, including genetic, molecular, and cellular mechanisms, is now providing clear objectives for the early detection, prevention, and therapy of a number of cancer forms. Understandably, the question of finding specific and reliable biomarkers for the early detection of various forms of cancer is attracting both enthusiasm and scepticism. The enthusiasm is driven by the completion of genome sequencing for a number of species including humans and by the availability of a spectrum of high-throughput technical platforms for both proteomic and genomic analyses. The scepticism on the contrary is partially derived from some inflated expectations, which are frequently followed by disappointment when the original results of certain investigations could not be reproduced. This scepticism, however, is not directed toward the final objective of defeating these devastating diseases; instead, it can be looked upon as some form of cautious assessment of current achievements and an attempt to dampen likely overenthusiasm generated by recent successes in this area of research. In other words, there seems to be a general

Cancer Biomarkers: Analytical Techniques for Discovery, by Mahmoud H. Hamdan

agreement within the research community that impressive steps have been made in the direction of discovering new and more specific biomarkers, yet there is a continuous and legitimate debate within the same community on what is needed to be done to translate laboratory successes into concrete clinical applications. Regardless of which side one would take, it is encouraging to note that the search for cancer biomarkers is one of the areas that bring together the scientist's quest to understand the biology and the molecular basis of these devastating malignancies with the physician's dedication to relief suffering and improve the quality of life of his patients.

Although spectacular advances in molecular medicine, genomics, and proteomics have been made, current efforts to combat cancer remain extremely disappointing. One main reason for the lack of such desired success is that in many cases, cancer is diagnosed and treated too late, when the cancer cells have already invaded adjacent tissues and established new colonies. The capability for invasion and metastasis enables cancer cells to escape the primary tumor mass and colonize new terrain in the body where, at least initially, nutrients and space are not limiting (Hanahan and Weinberg, 2000). These distant settlements of tumor cells are the cause of 90% of human cancer deaths (Sporn,1996). Currently, there are a number of platforms leading to the search for new biomarkers in cancer research. On the proteomic side, we have a number of emerging technologies that are applied in the area of biomarkers discovery, including surface enhanced laser desorption ionization (SELDI) (Hutchens and Yip, 1993; Tang et al., 2004), mass spectrometry combined with two-dimensional liquid chromatography (Link et al., 1999; Washburn et al., 2001; Wang and Hanash, 2003) or two-dimensional gel electrophoresis (O'Farrel, 1975; Klose and Kobalz, 1995; Aebersold and Goodlett, 2001; Abersold and Mann, 2003; Hamdan and Righetti, 2003), protein microarrays (MacBeath, 2002; Espina et al., 2003; Liotta et al., 2003), and imaging mass spectrometry (Caprioli et al., 1997; Chaurand et al., 1999; Stoeckli et al., 2001). On the genomic side, there are equally powerful platforms for biomarkers discovery, which use polymerase chain reaction (PCR) (Datta et al., 1994; Krismann et al., 1995), serial analysis of gene expression (SAGE) (Velculescu et al., 1995), and DNA microarrays (Young, 1995; Ramaswamy and Golub, 2002). These techniques and relevant references are fully covered in Chapter 2. Brief description of certain aspects of these technologies together with other elements relevant to biomarkers discovery is described in the following sections of this overview.

1.2. CANCER BIOMARKERS

Broadly speaking, cancer biomarkers can be divided into three categories: (a) Diagnostic (screening) biomarkers are used to detect and identify a given type of cancer in an individual. This type of biomarkers is expected to have high levels of diagnostic sensitivity and specificity, especially if it is used in large-screening trials. (b) Prognostic biomarkers are commonly used, once the disease status has been established. They are expected to predict the likely course of the disease, its recurrence, and thus they have an important influence on the aggressiveness of the

therapy. (c) Stratification (predictive) biomarkers are often DNA based and serve to predict the likely response to a drug before starting treatment, classifying individuals as "responders" or "nonresponders." This type of biomarkers is the result of recent advances in genetic research, which made it possible to predict clinical outcome from the molecular characteristics of the patient's tumor (Van de Vijver et al., 2002). Such predictive classification is of a major importance in designing clinical drug trials to define an intended use for the drug under investigation. In my opinion, the dividing line between screening and prognostic markers is rather flexible. In other words, there is no valid reason as to why a screening marker cannot be used as prognostic marker and vice versa. The role of a chosen marker does not end once the target cancer has been diagnosed. For example, the expression levels of a protein can be exploited as a biomarker and at the same time for assessment of the therapeutic response and recurrence following the diagnosis of the disease. Regardless of which definition is used, cancer biomarkers can be DNA, mRNA, metabolites, or processes such as apoptosis, angiogenesis, or proliferation that can be associated with a given type of cancer and can be measured quantitatively or qualitatively by an appropriate assay or technique. These markers can be found in a wide range of specimen, including body fluids (plasma, serum, urine, saliva, etc.), tissues, and cell lines. If the source of the biomarker is not the tumor itself then the term remote media is used. Such term refers to body fluids, lavages, detached cells, biopsies of nonmalignant tissues, and so on.

Tremendous amount of work in the area of biomarkers made it abundantly clear that the efficacy of a given biomarker assay is determined by its sensitivity and specificity. Both the terms take on precise meanings in the development of biomarker tests for population-based screening or for clinic-based surveillance of high-risk population. The clinical sensitivity of a biomarker can be simply defined as the proportion of individuals with confirmed disease who test positive for the biomarker assay, whereas the specificity refers to the proportion of control subjects (individuals without the disease) who test negative for the biomarker assay (Sullivan Pepe et al., 2001). A receiver-operating curve is commonly used to evaluate the efficacy of an assay. This is a graphical representation of the relationship between sensitivity and specificity. The ideal graph is the one giving the maximum area under the curve.

At present, there are many clinical situations in which tumor biomarkers are already being used; these existing markers are still the focus of further research efforts to increase their specificity, optimize them, and gain further information relevant to future generation of these or other new classes of markers. A commonly cited marker is the prostate-specific antigen (PSA), which is commonly used to screen male patients for prostate cancer (Stamey et al., 1987; Hudson et al., 1989; Thomson et al., 2004). Despite the tremendous impact of this marker on many aspects of the management of prostate carcinoma, the fact remains that this marker lacks specificity resulting in false-positives as high as 30%. In other words, almost a third of the patients with an elevated level of PSA do not necessarily suffer from this form of cancer. The most thoroughly assessed ovarian cancer biomarker is Carcinoma-associated glycoprotein antigen (CA-125), which was first identified by Bast et al., (1981; 1983). This biomarker and PSA are discussed in more details in Chapter 3.

Another serum-based marker is carcinoembryonic antigen (CEA), which was first reported by Gold and Freedman (1975). Levels of this protein are normally used to monitor disease progression and response to therapy in patients with colorectal cancer. One of the main limitations of this marker is that only a proportion of colorectal cancers tend to express elevated CEA levels at the time of diagnosis (Benson et al., 2000). This biomarker is a representative case of how initial findings regarding a promising biomarker are not always reproducible. CEA was initially purported to be nearly 100% sensitive and specific for colorectal cancer screening (Thomson et al., 1969). Subsequent studies have demonstrated that such levels of sensitivity and specificity were rather too optimistic (Read et al., 1995). The failure to reproduce the initial results was, in large part, due to the fact that individuals who were initially studied had an advanced stage of this disease, whereas individuals who were later studied had less extensive asymptomatic cancer in which CEA levels did not experience the expected increase. It has to be pointed out that this negative experience with CEA had some positive influence on methods development and rules of validation by which diagnostic tests are judged today.

Various forms of cancer are currently the target of major interdisciplinary efforts aiming at elucidating the molecular mechanisms governing disease pathogenesis, discovering new biomarkers for diagnosis, prognosis, and response to therapy. Within these interdisciplinary efforts, protein- and DNA-based technologies are expected to play a key role in the understanding and treatment of various human disorders including cancer.

1.3. PHASES OF BIOMARKERS DEVELOPMENT

The surge in research to develop cancer-screening biomarkers prompted the establishment of the Early Detection Research Network (EDRN) by the National Cancer Institute (Srivastava and Kramer, 2000). The aim of the EDRN is to coordinate research among biomarker-development laboratories, biomarker validation laboratories, clinical repositories, and population-screening programs. With the goals of EDRN in mind, Sullivan Pepe et al. (2001) proposed five phases that a biomarker needs to pass through to become a useful population-screening tool. These five phases can be summarized as follows: (i) The first phase is based on preclinical exploratory studies, comparing tumor with nontumor specimen. The aim of this phase is to identify unique characteristics of tumor that might lead to ideas for clinical tests capable of detecting the cancer. In this phase, various techniques can be employed, including immunochemistry, western blots, gene-expression profiles, protein-expression profiles, and levels of circulating antibodies against thousands of cancer antigens. (ii) Phase two involves the development of clinical assays preferably using specimen, which can be obtained noninvasively. A protein uniquely expressed by tumor and measured with serum antibody can be considered as an example of such assays. The main aim of this phase is to establish the true-positive rate (the proportion of case subjects who are biomarker positives), false-positive rate (the proportion of control who are biomarker positive), and the receiver operating

characteristics. The authors of these guidelines noted that since the case subjects in this phase have established disease, such phase does not determine if the same disease can be detected early with the same biomarker. (iii) Comparison of clinical specimens collected from subjects with cancer before their clinical diagnosis and compared with those from control subjects (subjects who have not developed the disease) can provide initial evidence on the capability of the biomarker to detect disease in the preclinical phase. In this phase, retrospective longitudinal repository studies are conducted. The aim of this phase is to evaluate, as a function of time before clinical diagnosis, the capability of the biomarker to detect preclinical disease, and to define criteria for a positive screening test in preparation for phase 4. In other words, if the levels of the biomarker in case subjects measured at a time close to clinical diagnosis show little deviation from those in control subjects, the biomarker has little promise for screening. On the contrary, if the levels in case subjects demonstrate distinct differences from those in control subjects months or years before the appearance of clinical symptoms, then the biomarker's potential for screening is enhanced. (iv) This final phase estimates the reduction in mortality of a given type of cancer as a result of screening tests employing a selected biomarker for that type of cancer. This phase has to address a number of difficulties before its findings can be truly related to the benefits of screening. Some of these difficulties have been pointed out within the guidelines by Sullivan Pepe et al. (2001) and include the following: (a) ineffective treatment for screen-detected tumors, (b) poor compliance with the screening program, (c) prohibitive economic costs of screening itself and of the diagnostic work-up of subjects who falsely screen positive for the disease, and (d) the overdiagnosis.

Before considering the applicability of such guidelines, it is relevant to take into account the indications by The World Health Organization (WHO) regarding early detection and disease control. These indications can be summarized as follows: first, the disease must be common and associated with serious morbidity and mortality. Second, screening tests must be able to accurately detect early-stage disease. Third, treatment after detection through screening must have been shown to improve prognosis relative to usual diagnosis. Fourth, evidence must exist that the potential benefits outweigh the potential harms and costs of screening (Winawer et al., 1995).

To appreciate the difficulties in constructing practical and reliable screening tests, it is sufficient to consider existing screening tests, which are in use for a number of cancers. For example, in the case of colorectal cancer screening, many guidelines recommend sigmoidoscopy (hollow tube inserted into the rectum for imaging the lower part of the colon and rectum) or colonscopy (similar to sigmoidoscopy, but examines the entire length of the colon), which are both expensive and above all are not well accepted in terms of time required, discomfort involved, and the risk of adverse outcome. Biomarkers that are in use for prostate-cancer screening, the PSA carries a substantial risk of overdiagnosis due to its poor specificity. A similar situation is found in ovarian cancer, where the use of CA-125 as a biomarker can result in false-positive rates that lead to unacceptably high number of surgeries to confirm the disease; the same biomarker fails to detect many early-stage cancers.

1.4. NEW APPROACH TO BIOMARKERS DISCOVERY

The early approaches to discover and identify cancer biomarkers were mainly based on preliminary clinical or pathological observations. A representative example of such approaches is the overexpression of the CEA, which was first reported by Gold and Freedman (1975). The isolation and purification of PSA (Wang et al., 1979), which is currently the only biomarker for prostate cancer is another example. A simple comparison between the methods used to discover these well-established cancer biomarkers and those currently employed in the search for new biomarkers reveals an unmistakable new approach for such discovery. Such comparison underlines two apparent differences: first, there is a clear shift in investigative strategy from an orderly inquiry into biological mechanisms toward a "brute force" approach that can be described as "collect the set, generate, and mine data." Furthermore, present methods attempt to identify distinguishing pattern (s) and/or multiple markers rather than a single one. Second, the discovery is conducted by different techniques even within the same laboratory. It is reasonable to suggest that such changing approach to biomarkers discovery is the direct result of technical advancement and newly acquired knowledge of the biology and molecular basis of various forms of cancer. Human genome sequencing, the discovery of oncogenes, tumor-suppressor genes, and tremendous advances in DNA-based and proteomic-based analyses have started to have a tangible impact on the landscape of biomarkers discovery. So, how such emerging technologies together with the newly acquired knowledge of cancer biology have influenced the discovery of cancer biomarkers? A partial answer to this question can be postulated through the following considerations:

1.4.1. New and Powerful Technologies

The past 10 years have witnessed an impressive growth in the field of large scale and high-throughput biology, resulting in a new era of technology development and the accumulation of new knowledge, which highlighted a number of challenges, including the need to elucidate the function of almost every encoded gene and protein in an organism and to understand the basic cellular events mediating a host of complex processes and their possible role in various diseases. Such newly acquired knowledge made it clear that a comprehensive analysis of the molecular basis of cancer and other disease states requires the integration of the distinct, but complementary information gained from genomics and proteomics. A number of emerging approaches have been used to tackle this prohibitive task, including large-scale analysis of genes and proteins. Over the past few years, miniaturized and parallel assay systems have already demonstrated a part of their potential in large-scale and high-throughput biological analysis. Today, the expression of thousands of genes can be simultaneously assessed under different conditions, including disease state and treatment. Powerful technologies including PCR, SAGE, single nucleotide polymorphism analyses, and microarrays can target almost any DNA, RNA, or protein sequence. These microarrays have been used for the detection of sequence variations and for mapping the targets of transcription factors (Lyer et al., 2001; Heller, 2002; Horak et al., 2002). A drawback

of DNA microarrays is their unsuitability for protein analysis. There are two experimentally demonstrated reasons behind such limitation: first, there is little correlation between mRNA and protein expression levels (Anderson and Seihamer, 1997; Gygi et al., 1999). Second, proteins are often derived from different alternative spliced RNAs, and/or contain posttranslational modifications, which result in distinct functions and activities (Harada et al., 2004; Rammensee, 2004).

Although it is still too early to compare the success of protein microarrays with that already achieved by their DNA counterparts, there is no doubt that the first type of microarrays has made substantial progress in terms of construction and applications, including the area of biomarkers discovery. In recent years, there have been considerable achievements in preparing microarrays containing over 100 proteins and even an entire proteome (Madoz-Gúrpide et al., 2001; Cahill and Nordhoff, 2003; Michaud et al., 2003; Haab, 2005). Different array formats have been developed, including tissue, living cells, peptides/small molecules, antibody/antigen (s), protein, and carbohydrate arrays, which are described in more details in Chapter 2. The capability of these formats to provide simultaneous assessment of expression/interaction of 100s and even 1000s of proteins can be considered one of the emerging developments, which is paving the way to new and more powerful strategies in biomarkers discovery.

Mass-spectrometry-based methods for proteomic analysis have been improved on various fronts; new generation of mass spectrometers allows higher mass accuracy, higher detection capability, and shorter cycling times, allowing higher throughput and more reliable data. Two-dimensional chromatography coupled to MS/MS is getting more acceptance as a powerful tool for the analysis of complex protein mixtures. Recent improvements on the chromatography side included high-pressure LC systems and smaller diameter packing material allowing shorter analysis times and higher detection limits. With regard to data analysis, there are now several data mining tools for analyzing global protein expression data generated by this approach. Several publicly available software packages are currently in use to map proteomic data sets generated by searching peptide collision induced dissociation spectra against one or more major protein databases such as Swiss-Prot/TrEMBL, International Protein Index (IPI), and the National Center for Biotechnology Information (NCBI). Other relevant Web sites are given in Table 1.1.

In the emerging field of systems biology, accurate quantification of proteins and their changing patterns represent an important component. Recently, MS-based quantitative proteomics has become an important component in biological and clinical research. Over a number of years, multidimensional chromatography coupled to tandem MS has demonstrated its capability to identify hundreds to thousands of proteins within complex mixtures. However, the same platform fails short of routinely providing accurate quantitative analysis of proteins in complex media such as serum or cell lysate. A number of strategies have been devised to enhance the potential of this approach for protein quantification, including some posttranslational modifications. Many of these modifications, such as phosphorylation and glycolysation, have well-documented roles in signal transduction, regulation of cellular processes, clinical biomarkers, and therapeutic targets. A limited number of recent

TABLE 1.1. Some links relevant to DNA- and protein-based analysis available for public use.

ArrayExpress: http://www.ebi.ac.uk/arrayexpress
Biocarta: http://www.biocarta.com
Biomolecular Interaction Database:
http://www.blueprint.org/bind/bind.php
CaCORE: http://ncicb.nci.nih.gov/core
Cancer Biomedical Informatics Grid:
http://cabig.nci.nih.gov
Cancer Genome Anatomy Project: http://cgap.nci.nih.gov/
Cytoscape: http://www.cytoscape.org
Database of Interacting Proteins:
http://dip.doe-mbi.ucla.edu/
Cancer Genome Anatomy Project: http://cgap.nci.nih.gov
Cancer Genome Project: http://www.sanger.ac.uk/CGP
dbEST: http://www.ncbi.nlm.nih.gov/dbEST
Gene Expression Omnibus: http://www.ncbi.nlm.nih.gov/geo
Human Cancer Genome Project:
http://www.ludwig.org.br/ORESTES
IMAGE Consortium: http://image.llnl.gov
Mitelman Database of Chromosome Aberrations in Cancer:
http://cgap.nci.nih.gov/Chromosomes/Mitelman
SAGE Genie: http://cgap.nci.nih.gov/SAGE
SAGEmap: http://www.ncbi.nlm.nih.gov/SAGE
Spectral Karyotyping/Comparative Genomic Hybridization
Database: http://www.ncbi.nlm.nih.gov/sky
Access to this interactive links box is free online.
ExPASy: http://ca.expasy.org
Gene Ontology Consortium: www.geneontology.org
Swiss-Prot/TrEMBL: http://ca.expasy.org/sprot
International Protein Index(IPI): www.ebi.ac.uk/IPI/IPIhelp.html
National Center for Biotechnology Information (NCBI): www.ncbi.nlm.nih.gov
MouseSpec: http://tap.med.utoronto.ca/~posman/mousespec
Protein families data base: www.sanger.ac.uk/Software/Pfam
InterPro: www.ebi.ac.uk/interpro
PSORT II: http://psort.ims.u-tokyo.ac.jp
TreeView: http://jtreeview.sourceforg.net
GenMAPP; www.GenMapp.org

strategies have demonstrated the potential for large-scale analysis of phosphorylated and glycosylated proteins. MacCoss et al. (2002) have described what they termed "shotgun" approach for the identification of various forms of protein modifications (including phosphorylation) in complexes and in lens tissue. To digest the investigated protein mixtures, the authors used three different enzymes, one that cleaves at a specific site, whereas the other two cleave at nonspecific sites. The mixture of the resulting peptides was separated by multidimensional liquid chromatography and

analyzed by tandem mass spectrometry. This approach has been applied to simple protein mixture, Cdc2p protein complexes isolated by affinity tag, and to lens tissue from a patient with congenital cataracts. These results yielded various sites of phosphorylation, acetylation, methylation, and oxidation.

In two relatively recent articles, which appeared in the same issue, two independent groups (Kaji et al., 2003; Zhang et al., 2003) described similar strategies for the identification and quantification of N-linked glycoproteins. The first group used a strategy that combines hydrozyde chemistry, stable isotope labeling, and mass spectrometry, whereas the second group used lactin affinity capture in combination with isotope-coded tagging and mass spectrometry. The approach by Kaji et al. (2003), termed as isotope-coded glycosylation-site-specific-tagging (IGOT), is based on the lactin column-mediated affinity capture of glycopeptides generated by tryptic digestion of protein mixtures, followed by peptide-N-glycosidase-mediated incorporation of stable isotope tag ^{18}O specifically into the N-glycosylation site. The tagged peptides are then identified by multidimensional LC coupled to mass spectrometry. This approach was tested on N-linked high-mannose and and/or hybrid-type glycoproteins derived from an extract of *Caenorhabditis e/egans.* The authors reported the identification of 250 glycoproteins, including 83 putative transmembrane proteins, with the simultaneous determination of 400 unique N-glycosylation sites. To demonstrate the potential of IGOT strategy for protein quantification, the authors processed two peptide aliquots differentially labeled with ^{18}O and ^{16}O and the mixed preparation was examined by LC/MS. Although the isotope distribution of the two-tagged peptides partly overlapped owing to the natural isotopic abundance, both spectra were good enough to permit relative quantification of ^{16}O- and ^{18}O-tagged peptides.

Visible-coded affinity tag (VICAT) is a tagging reagent which allows absolute quantification of protein(s) in a complex biological sample (Lu et al., 2004). This tagging procedure can be considered a variant of the well-established ICAT procedure. VICAT reagents target thiol groups of Cys or thioacetylated amino groups and introduce into the tryptic peptide a biotin affinity handle, a visible moiety for tracking the chromatographic location of the target peptide by a detection method other than mass spectrometry. Initial capability of this reagent was demonstrated by the absolute determination of human group V phospholipase A_2, in eukaryotic cell lysates.

Another approach for high-throughput quantitative analysis has been recently reported by Zhang et al. (2005). This approach is designed to simplify the analysis of serum and allow targeted quantification of proteins, which happen to have relatively low concentration. This method is based on the selective isolation of those peptides from serum proteins that are N-linked glycosylated in the native protein and the use of LC/MS and LC/MS–MS to analyze the peptide mixture of the deglycosylated forms of these peptides. This method has two apparent advantages: first, a dramatic reduction in the total number of peptides, and second, a reduction in the complexity of the acquired spectra due to the removal of oligosaccharides that contribute significantly to the peptide pattern heterogeneity. The potential of this method was demonstrated by generating peptide patterns, which could distinguish the serum proteome of cancer bearing mice from genetically identical normal mice.

A method, termed stable isotopelabeling with amino acids in cell culture (SILAC), has recently gained popularity for its ability to compare the expression levels of hundreds of proteins in a single experiment (Everley et al., 2004). SILAC is based on the use of ^{12}C- and ^{13}C-labeled amino acids added to the growth media of separately cultured cell lines, giving rise to cells containing either light or heavy proteins. Lysates collected from these cells are then mixed, separated on SDS-Page, separated bands are excised and digested, and can be injected into a tandem mass spectrometer for protein identification/quantification.

1.4.2. Promising Sources for Biomarkers

1.4.2.1. DNA Methylation. The past few years have seen a substantial advance in our understanding of the functional consequences of DNA methylation and its interaction with chromatin structure and the transcriptional machinery (Laird, 2003). First insights into what causes DNA methylation patterns to undergo changes in cancer cells have also been acquired (Di Croce et al., 2002; Song et al., 2002). From a clinical perspective, DNA methylation changes in cancer represent a highly attractive therapeutic target, as epigenetic alterations, including DNA methylation are, in principle, more readily reversible than genetic events (Karpf and Jones, 2002). However, the great strength of DNA methylation in clinical applications promises to be in the areas of molecular diagnostics and early detection. The introduction of a highly sensitive methylation specific PCR (MSP) procedure by Herman et al. (1996) rendered DNA methylation a fertile ground for biomarkers research. The main advantage of the MSP assay is its sensitivity and capability to detect methylation in the presence of contaminating normal tissue or cells. The same assay can be conducted directly on tissue sections (in situ MSP) to identify clonality of the gene silencing in tumors and premalignant lesions (Nuovo et al., 1999). Recent improvements in the sensitivity of this assay are: A more sensitive assay called methylLight capable of detecting methylated alleles in the presence of 10^4-fold excess of unmethylated alleles has been described by Eads et al. (2000). This is a high-throughput assay capable of quantitative determination of a particular pattern of DNA methylation. A further improvement in MSP assays has been introduced by Palmisano et al. (2000). This improved assay was designated nested MSP and had the capability to detect a single methylated allel in ~50,000 unmethylated alleles. The list of cancer-associated methylated genes detected by this type of assays is expanding (see Chapter 4, Tables 4.4 and 4.5).

1.4.2.2. Mitochondrial DNA Mutations. This is another area targeted as a possible source for the identification of cancer biomarkers. Mitochondria dysfunction was proposed to be involved in cancer over 50 years ago (Warburg et al., 1967). Mitochondria are believed to be more susceptible to exogenous mutagens and also have less efficient DNA-repair mechanisms. There is accumulating evidence suggesting that mitochondria regulate several cellular processes that are linked to apoptosis, which include electron transport and energy metabolism. They are also the storage site for a number of soluble proteins that mediate apoptosis, including

cytochrome c. Many of the signals that elicit apoptosis converge on the mitochondria, which respond to apoptotic signals by releasing cytochrome c (Verma et al., 2003). Information gained recently on the connection between mitochondrial dysfunction, deregulation of apoptosis, and tumorigenesis together with an increasing knowledge of proteins that are involved in cancer progression may lead to a new class of markers for the early detection and possible prevention of certain types of cancer.

1.4.2.3. Phosphatidylinositol-3 Kinases (PI3Ks). PI3Ks constitute a lipid kinase family characterized by their ability to phosphorylate inositol ring 3′-OH group in inositol phospholipids to generate the second messenger phosphatidylinositol-3,4,5-triphosphate (PIP3) at the inner side of the cell membrane (Cantley and Neel, 1999; Cantley, 2002). PIP3 in turn contributes to the recruitment and activation of a wide range of downstream targets, including the serine–threonine protein kinase Akt (also known as protein kinase B). The PI3K-Akt signaling pathway regulates many normal cellular processes including cell proliferation, survival, growth, and motility, processes that are critical for tumorigenesis. In the last decade, much of the cancer research has focused on the central role of *RAS*, the first identified oncogene, in neoplastic transformation. Extensive biochemical and genetic studies of the signaling components upstream and downstream of this small GTPase in model organisms led to the model of mitogenic signaling by receptor tyrosine kinases (RTKs) through RAS and motigen-activated protein kinases (MAPKs). The central importance of this pathway in neoplastic cell proliferation in humans has been strongly supported by the clinical success of therapeutics that target tyrosine kinases.

In recent years, a second pathway downstream of RTKs that involves phosphatidylinositol-3 kinase and Akt has come onto the scene and is proving to be an important regulator of mammalian cell proliferation and survival. Indeed, the role of this pathway in oncogenesis has been extensively investigated and altered expression or mutation of many of its components has been implicated in various forms of human cancer (Vivanco and Sawyers, 2002).

Currently, there are a number of therapeutic strategies in development, which target this pathway. Quantification of signaling throughput in the PI3K to Akt pathways has the potential of providing prognostic information to distinguish clinically important subsets of cancer. For example, a number of findings have specifically linked this pathway to prostate cancer. *PTEN* inactivation or loss of heteozygosity is common in prostate cancers, especially metastatic carcinoma (Suzuki et al., 1998; Sansal and Sellers, 2004), and targeted deletion of *PTEN* in mouse prostate activates Akt and induces prostate carcinoma (Wang et al., 2003). In a xenograft model for progression of the androgen-dependent (or androgen-sensitive) LNCaP cell line to androgen independence, Akt activity (but not expression) was elevated and correlated with Ser^{473} phosphorylation (Graff et al., 2000). Introduction of constitutively activated Akt into these cells permitted androgen-independent growth. Progression from normal prostate epithelium to prostatic intraepithelial neoplasia or carcinoma is associated with elevated Akt phosphorylation (Paweletz et al., 2001; Malik et al., 2002). These studies reported that mitogen-activated protein kinase activation

monitored with phospho-extracellular signal-regulated kinase (ERK) antibodies was enhanced in prostatic intraepithelial neoplasia but reduced in carcinoma.

1.4.2.4. Profiling Tyrosine Phosphorylation. Over the last two decades, it has become clear that tyrosine phosphorylation plays a central role in a variety of important signaling pathways in multicellular organisms. Such role has been recently enforced by the success of specific tyrosine kinase inhibitors in cancer treatment (Druker, 2002). Functional profiling of the tyrosine phosphoproteome is likely to lead to the identification of novel targets for drug discovery and provide exciting and novel molecular diagnostic approaches. A major challenge in this direction is to develop the means to rationally control and manipulate the cellular tyrosine phosphorylation state. It is reasonable to state that the detection, identification, and quantification of phosphoproteins, and mapping of their phosphorylated sites, are the main objectives of phosphoproteomics. Over the past few years, a number of approaches have shown the potential to achieve some of these objectives. These include MS-based phosphoproteomic approaches (Conrads et al., 2002; Mann et al., 2002), two-dimensional gel electrophoresis with and without ^{32}P labeling (Immler et al., 1998; Larsen et al., 2001; Yoshimura et al., 2002), immunoaffinity-based methods (Pandey et al., 2002; Steen et al., 2002), and western blotting (Nollau et al., 2001; Machida et al., 2003).

1.4.2.5. Proteins Expression. Extensive activities dealing with protein profiling and analyses have generated a tremendous amount of data on the expression/modification of proteins under disease conditions, including various forms of cancer. Subsequent interpretation and assignment of biological roles of a part of such data allowed the identification of families which have the potential to be translated into biomarkers capable of an early detection of some forms of cancer. Three of these families are considered in the present text, two of which are briefly introduced below. These two families together with kallikerins are described in more details in Chapter 4.

Extensive research activities over the last 30 years have shown that HSPs and their close constitutively expressed relatives are in effect molecular chaperons. Following exposure to proteotoxic stressors, the cells in most tissues dramatically increase the production of this group of proteins. Various studies have proposed diverse roles for chaperones as sensors and regulators of stress-induced apoptosis. The molecular pathways that mediate apoptosis are tightly regulated by a series of positive and negative signals, the balance of which determines whether or not cells commit suicide. Increasing evidence suggests that HSPs can influence this process through direct interaction with key components of the apoptotic machinery. In other words, these proteins serve as cellular safeguards to protect the network of protein–protein interactions that sense stress signals and relay them to the apoptotic machinery (Mosser and Morimoto, 2004). These authors suggested that the ability of HSPs proteins to influence a cell's fate through modulation of numerous control points endows these proteins with the unusual capacity to contribute in a decisive way and at multiple points in the process of tumorigenesis.

Molecular cloning and biochemical characterization of 14-3-3 proteins have revealed seven homologous isoforms in mammalian cells, which were designated with

the Greek letters β, γ, ε, η, σ, τ (sometimes referred to as θ), and ζ (Ichimura et al., 1988; Fu et al., 2000). Most of these isoforms are expressed in all human tissues, although the σ form expression is restricted to epithelial cells (Leffers et al., 1993). This family of proteins is implicated in the regulation of numerous cellular signaling circuits that are involved in the development of various forms of cancer. These proteins have attracted interest because they are involved in important cellular processes such as signal transduction, cell-cycle control, apoptosis, stress response, and malignant transformation. These different roles are in part due to their capability to bind more than 100 different binding partners. Of all the 14-3-3 genes, 14-3-3σ has been most directly linked to cancer. Inactivation of this gene occurs at many levels, and the high frequency of its inactivation suggests that it has a crucial role in tumor formation. This role has been consolidated by a number of recent investigations. Osada et al. (2002) have demonstrated frequent and histological-specific inactivation of 14-3-3σ in human lung cancer. The loss of 14-3-3σ expression in breast carcinoma was attributed to methylation silencing (Umbricht et al., 2001). Another report along these lines was given earlier by Ferguson et al. (2000). By using SAGE analysis the authors reported that the expression of this gene was 7-fold lower in breast carcinoma cells compared with normal breast epithelium.

1.5. INITIATIVES RELEVANT TO BIOMARKERS DISCOVERY

The complexity of the process of biomarkers discovery and validation together with the tremendous research activities involved in such process have underlined the urgent need for various initiatives, both at the national and international levels, to facilitate scientific collaboration and access to data generated by various research groups working in the field of biomarkers discovery. Some of these initiatives are briefly discussed in the sections below, and a more detailed description of these and other initiatives will be given in the latter part of this book.

1.5.1. Initiatives of the Human Proteome Organization (HUPO)

The HUPO formed in 2001 has launched several major initiatives. These are focused on the plasma proteome, the liver proteome, brain proteome, protein standards/bioinformatics, and certain technologies, including large scale antibody production. The aim of these initiatives is to foster organized international efforts in the field of proteomics, including more effective strategies for early disease detection (Hanash, 2004). The initial planning meetings around this initiative have drawn some sort of a checklist to be followed by various research groups. Regarding the plasma proteome project, interdisciplinary groups of experts have proposed a pilot phase to address the following issues: assessment of the sensitivity of various techniques; guideline on all aspects of specimen collection and handling; methods of depleting or prefractionation of the most abundant proteins; comparison of advantages and limitations associated with plasma versus serum; enumeration and categorization of visualized and identified proteins,

with particular attention to their posttranslational modifications and tissue of origin; separation of intact proteins versus separation of their digested peptides, comparing gel-based methods with multidimensional liquid chromatography, and assessment and advancement of specific labeling chemistry. Looking at this list of issues, it is not difficult to realize that such pilot stage is an attempt to address some of the drawbacks and limitations, which have been highlighted by previous and more recent proteomic analyses generated by various techniques. Whether such list of issues can be rigorously implemented is difficult to predict. Regardless of such prediction, such initiative will no doubt contribute to more reproducible data and the identification of the most suitable platform(s) to handle the complexity of plasma or serum proteome.

1.5.2. Data Mining in Cancer Research

The complexity of tumors biology renders the use of interdisciplinary approaches a necessity rather than a choice. Furthermore, the tremendous amount of data generated by a wide and diverse proteomic and genomic approaches underlined the need for the creation of easily accessible repositories to allow the interrogation of databases and other tools. Currently, there are a number of data repositories containing enormous amount of data on gene expression in normal and cancer cells. These data are the result of initiatives such as the Cancer Anatomy Project and the Director's Challenge Initiative, funded by the National Cancer Institute (NCI). There are other resources available for data mining. One of these resources, GoMiner, was developed by Zeeberg et al. (2003), a program package that organizes lists of genes, such as down- and overexpressed genes from a range of microarray experiment(s). This program package provides quantitative and statistical output files and two different visualizations. Genes displayed in GoMiner are linked to major public bioinformatics resources. The NCBI at the National Institute of Health was created almost 20 years ago to develop information systems for molecular biology. In addition to maintaining the GenBank nucleic acid sequence database, to which data are submitted by the scientific community, NCBI provides data retrival systems and computational resources for the analysis of GenBank data and a variety of other biological data. Wheeler et al. (2006) described the major resources of NCBI, which are available on a home page at the Web site, http://www.ncbi.nlm.nih.gov, and download of bulk data underlying these resources is also available through a link ftp.ncbi.nih.gov from the NCBI home page.

There are resources relevant to data mining in cancer, including the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2002). This publicly available web database contains more than 150 pathways with emphasis on well-defined metabolic pathways. Gene Microarray Pathway Profiler (GenMAPP) is a freely available program for viewing and analyzing expression data on microarray pathway profiles representing biological pathways or other functional grouping of genes (Doniger et al., 2003). Other useful links are listed in Table 1.1.

1.6. CONCLUDING REMARKS

The material in this book is an attempt to underline tremendous advances in the technologies and current knowledge of the biology and molecular basis of cancer. The same material, however, leaves no doubt that despite such advances there are still a number of challenges before many forms of cancer can be defeated. The correlation between what has been already achieved and what is remained to be done has been elegantly described by the following phrase; "One day, we imagine that cancer biology and treatment at present, a patchwork quilt of cell biology, genetics, histopathology, biochemistry, immunology, and pharmacology will become a science with a conceptional structure and logical coherence that rivals that of chemistry or physics" (Hanahan and Weinberg, 2000). Before moving to the next chapter it is useful to have in mind a number of general considerations, which in part illustrate some of the difficulties still facing research scientists in their efforts to discover biomarker(s) for the various forms of cancer:

- Despite enormous proteomic and genomic efforts, in only a few tumor diseases have relevant markers been established that can be used for early diagnosis or improved therapy in cancer. We are still facing the dilemma where in many cases cancer is not diagnosed and treated until cancer cells have already invaded surrounding tissues and metastasized throughout the body. No one disputes the fact that serum-based markers such as CA125 and PSA have saved many lives, yet both markers suffer two well-recognized limitations: The first is their low specificity, which in turns results in a high rate of false-positives, whereas the second limitation is associated with what can be considered an unacceptable time-lag between the detection and the in situ state of the disease. In other words, an elevated level of these markers seems to manifest at an advanced stage of the malignancy. To address certain limitations there have been some recent attempts to emphasize the utility of multiple markers rather than relying on "one-at-a-time" approach. On the proteomic side, for example, the use of serum-based proteomic patterns analysis started to gain more momentum. The SELDI analysis is a representative example of such emerging approach, where patterns containing a number of different molecular ions are used to distinguish between healthy and diseased samples. The basic principle of such analysis is not substantially different from that applied in two-diminsional gel analysis, where alteration in protein expression is monitored for multiple rather than for single proteins. The same principle is also applied in some DNA-based analysis, where, for example, patterns of DNA methylation are sought rather than the methylation of a single entity.
- The inefficacy of some existing serum-based cancer markers particularly regarding their capability for the early detection of the disease cast some shadow on the search strategies used to discover them. These strategies have been criticized on the ground that the media in which the target markers are detected (e.g., serum) do not necessarily reflect the in situ situation of the tumor. In other words, data delivered by these strategies are not easily traced back to the

biological properties or the heterogeneity of the tumor itself. Such criticism can be partially justified if we consider the in situ complexity of many forms of cancer. Such complexity has been underlined by Hanahan and Weinberg (2000) in an article entitled "The hallmarks of cancer." The authors suggested that the vast catalog of cancer genotypes is a manifestation of six essential alterations in cell physiology that collectively dictate malignant growth: self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, evasion of programmed death (apoptosis), limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis. The same authors went on to propose that all six capabilities are common to most and perhaps to all types of human tumors. The complexity of many forms of cancer is further aggravated by the absence of a comprehensive knowledge on the signaling pathways within a cell, which are more and more mimicking complex electronic integrated circuits, where transistors are replaced by proteins and electrons by phosphates and lipids. If we apply the same principle to the signaling pathways in cancerous cells, then we can appreciate that an elevated level of a single protein in serum may represent a useful marker for a given type of cancer, yet at the same time we have to accept that such marker is going to have a number of limitations.

- Over the last 20 years, both proteomic and genomic activities have begun to make extensive use of products of human origin. This new trend has raised many ethical and social issues, particularly those involving the individual rights, including issues of consent. This means that researchers should carefully consider several aspects when designing studies in which samples of human origin are required. These aspects will surely include the extent of risk for human volunteers, biosafety in particular when international collaboration is needed, rules on data acquisition and storage that also has to be carefully assessed particularly when using computerized databases. Presently, the ethical and regulatory framework for using human tissues in various areas of research is still vague and lacks precise guidelines. Besides these ethical and social problems, research scientists looking for new disease markers are still facing hurdles related to approvals by Food and Drug Administration (FDA) and European Agency for the Evaluation of Medicinal Products (EMEA). As far as proteomic-based tests are concerned, neither of the two regulatory bodies has an official guideline on how such tests should be submitted. However, it has to be said that the FDA has taken an important step toward defining a policy designated "Multiplex Tests for Heritable DNA Markers, Mutations, and Expression Patterns which mainly focuses on DNA tests, including DNA microarrays." Although such draft guideline could also cover some proteomic tests, it is hoped that a similar guideline, which may cover a wider range of proteomic tests, is something that can materialize in the near future. The situation is more complicated in the case of EMEA, which is not in charge of evaluating diagnostic tests, yet it does monitor developments in the field, which may impact on pharmaceuticals.

- Continuing the discussion on the theme of serum-based markers, it is worth considering a current point of contention regarding the relevance of establishing the identity of such markers. For example, if we use mass spectrometry-based method to analyze healthy and diseased samples, would differences in the pattern of unidentified MS peaks be sufficient for use as a diagnostic tool? The answer to this question strongly depends on the person who gives it. A research scientist would argue that the identification of each peak is important for current and future attempts to decipher the complex biology and signaling pathways associated with cancer. Another line of thought advocates that a pattern of unidentified MS peaks, which has been tested on extremely large number of samples, might be more than sufficient to satisfy doctor–patient perspective. In other words, if we have a reliable and selective marker for a given type of cancer, then its identity is not the top priority of either the patient or his physician. Leaving aside the difference between the two opinions, it is not difficult to spot a common objective, which is called discovering new and reliable markers for a class of devastating diseases.
- High-throughput analysis techniques raise the question of overfitting of data generated in discovery-based research. Such danger can be encountered when large amount of data are generated and analyzed for discriminatory patterns to use in diagnosis or prognosis (Stears et al., 2003). For example, RNA expression levels of thousands of genes from a cancer specimen can be analyzed for patterns that predict a patient's prognosis or response to therapy. Similarly, thousands of peaks generated by mass spectrometry of serum sample can be analyzed for protein/peptide patterns that discriminate between a healthy person and a patient. Ransohoff (2004) underlined the problem of overfitting by a simple yet efficacious example, which is worth considering. According to this author, overfitting can occur when large numbers of potential predictors are used to discriminate among a small number of outcome events. This scenario has been exemplified by imagining 10 people with cancer and 10 without who are screened using 20,000 features with no relation to cancer, such as the type of films they watch or the number of times they chew their food. The author commented that if enough predictors are examined, even if nonsensical and random, a pattern could be found to discriminate among the group of individuals derived from a training set, but it would not discriminate in an independent validation set. Biostatistics and empirical assessment has also demonstrated how overfitting can occur in RNA expression analysis (Ambroise and McLachlan, 2002; Simon et al., 2003). Simon and collaborators constructed a group of imaginary individuals, 10 with and 10 without cancer, along with expression data for 6000 genes. They then applied different methods of cross-validation, in a manner highly representative of real experiments, to discover discriminatory patterns. The authors reported that using one common method, 98% of the models fit perfectly in the training set, indicating how frequently overfitting can occur. Such overfitting has to be carefully assessed in approaches that use multivariable analysis such as artificial neural networks (Selaru et al., 2002), genetic algorithms (Petricoin et al.,

2002), boosted decision-tree analysis (Qu et al., 2002), and metagenes (Huang et al., 2003) are commonly used in discovery–based research.

REFERENCES

Aebersold, R., Goodlett, D. R. (2001) Chem. Rev. **101**, 269.
Aebersold, R., Mann, M. (2003) Nature **422**, 198.
Ambroise, C., McLachlan, G. J. (2002) PNAS. USA **99**, 6562.
Anderson, L., Seihamer, J. (1997) Electrophoresis **18**, 533.
Bast, R. C., Jr, Freeney, M., Lazarus, H., et al. (1981) J. Clin. Invst. **681**, 1331.
Bast, R. C., Jr, Klug, T. L., St John, E., et al. (1983) N. Engl. J. Med. **309**, 883.
Cahill, D. J., Nordhoff, E. (2003) Adv. Biochem. Engin. Biotechnol. **83**, 177.
Cantley, L. C. (2002) Science **296**, 1655.
Cantley, L. C., Neel, B. G. (1999) PNAS. USA **96**, 4240.
Caprioli, R. M., Farmer, T. B., Gile, J. (1997) Anal. Chem. **69**, 4751.
Chaurand, P., Stoeckli, M., Caprioli, R. M. (1999) Anal. Chem. **71**, 5263.
Conrads, T. P., Issaq, H. J., Veenstra, T. D. (2002) Biochem. Biophys. Res. Commun. **290**, 885.
Datta, Y. H., Adams, P. T., Drobyski, W. R., et al. (1994) J. Clin. Oncol. **12**, 475.
Di Croce, L., Raker, V. A., Corsaro, M., et al. (2002) Science **295**, 1079.
Doniger, S., Salomonis, N., Dahlquist, K., et al. (2003) Genome Biol. **4**, R7.
Druker, B. J. (2002) Cancer Cell **1**, 31.
Eads, C. A., Danenberg, K. D., Kawakami, K., et al. (2000) Nucleic Acids Res. **28**, e32.
Espina, V., Mehta, A., Winters, M. E., et al. (2003) Proteomics **3**, 2091.
Everley, P. A., Krijgsveld, J., Zetter, B. R., et al. (2004) Mol. Cell. Proteomics **3.7** 729.
Ferguson, A. T., Evron, E., Umbricht, C. B., et al. (2000) Proc. Natl. Acad. Sci. USA **97**, 6049.
Fu, H., Subramanian, R. R., Masters, S. C. (2000) Annu. Rev. Pharmacol. Toxicol. **40**, 617.
Gold, P., Freedman, S. O. (1975) JAMA. **234**, 190.
Graff, J. R., Konicek, B. W., McNulty, A. M., et al. (2000) J. Biol. Chem. **275**, 24500.
Gygi, S. P., Rochon, Y., Franza, B. R., et al. (1999) Mol. Cell. Biol. **19**, 1720.
Haab, B. B. (2003) Proteomics **3**, 2116.
Hamdan, M., Righetti, P. G. (2003) Mass Spectrom. Rev. **22**, 272.
Hanahan, D., Weinberg, R. A. (2000) Cell. **100**, 57.
Hanash, S. (2004) Mol. Cell. Proteomics **3.4**, 298.
Harada, K., Yewdell, J. W., Young, J. C. (2004) Nature **427**, 252.
Heller, M. J. (2002) Ann. Rev. Biomed. Eng. **4**, 129.
Herman, J. G., Graff, J. R., Nelkin, P. D., et al. (1996) PNAS. USA **93**, 9821.
Horak, C. E., Luscombe, N. M., Qian, J., et al. (2002) Genes Dev. **16**, 3017.
Huang, E., Cheng, S. H., Dressman, H., et al. (2003) Lancet **361**, 1590.
Hudson, M., Bahnson, R., Catalona, W. (1989) J. Urol. **142**, 1011.
Hutchens, T. W., Yip, T. T. (1993) Rapid Commun. Mass Spectrom. **7**, 576.

Ichimura, T., Isobe, T., Okuyama, T., et al. (1988) PNAS. USA **85**, 7084.
Immler, D., Gremm, D., Kirsch, D., et al. (1998) Electrophoresis **19**, 1015.
Kaji, H., Saito, Y., Yamauchi, T., et al. (2003) Nat. Biotecnol. **21**, 667.
Kanehisa, M., Goto, S., Kawashima, S., et al. (2002) Nucleic Acids Res. **30**, 42.
Klose, J., Kobalz, U. (1995) Electrophoresis **16**, 1034.
Krismann, M., Todt, B., Schroder, J., et al. (1995) J. Clin. Oncol. **13**, 2769.
Laird, P. W. (2003) Nat. Rev. Cancer **3**, 253.
Larsen, M. R., Sorensen, G. L., Fey, S. J., et al. (2001) Proteomics **1**, 223.
Lu, Y., Bottari, J., Stemman, O., et al. (2004) Anal. Chem. **76**, 4104.
Karpf, A. R., Jones, D. A. (2002) Oncogene **21**, 5496.
Leffers, H. P., Madsen, H. H., Rasmussen, B., et al. (1993) J. Mol. Biol. **231**, 982.
Link, A. J., Eng, J., Scvhieltz, D. M., et al. (1999) Nat. Biotechnol. **17**, 676.
Liotta, L. A., Espina, V., Mehta, A., et al. (2003) Cancer Cell **3**, 317.
Lyer, V. R. (2001) Nature **409**, 533.
MacCoss, M. J., McDonald, W. H., Saraf, A., et al. (2002) PNAS. USA **99**, 7900.
MacBeath, G. (2002). Nat. Genet. Supl. **32**, 526.
Machida, K., Mayer, B. J., Nollau, P. (2003) Mol. Cell. Proteomics **2**, 215.
Madoz-Gúrpide, J., Wang, H., Misek, D. E., et al. (2001) Proteomics 1, 1279.
Malik, S. N., Brattain, M., Ghosh, P. M., et al. (2002) Clin. Cancer Res. **8**, 1168.
Mann, M., Ong, S. E., Gronborg, M., et al. (2002) Trends Biotechnol. **20**, 261.
Michaud, G. A., Salcius, M., Zhou, F., et al. (2003) Nat. Biotechnol. **21**, 1509.
Mosser, D. D., Morimoto, R. I. (2004) Oncogene **23**, 2907.
Nollau, P., Mayer, B. J. (2001) PNAS. USA **98**, 13531.
Nuovo, G. L., Plaia, T. W., Belinsky, S. A., et al. (1999) PNAS. USA **96**, 12754.
O'Farrell, P. H. (1975) J. Biol. Chem. **250**, 4007.
Osada, H., Tatematsu, Y., Yatabe, Y., et al. (2002) Oncogene **21**, 2418.
Palmisano, W. A., Divine, K. K., Saccomanno, G., et al. (2000) Cancer Res. **60**, 5954.
Pandey, A., Blagoev, B., Krachmarova, I., et al. (2002) Oncogene **81**, 8029.
Paweletz, C. P., Charboneau, L., Bichsel, V. E., et al. (2001) Oncogene **20**, 981.
Petricoin, E. F., Ardekani, A. M., Hitt, B. A., et al. (2002) Lancet **359**, 572.
Qu, Y., Adam, B-L., Yasui, Y., et al. Clin. Chem. **48**, 1835.
Ramaswamy, S., Golub, T. R. (2002) J. Clin. Oncol. **20**, 1932.
Rammensee, H-G. (2004) Nature **427**, 203.
Ransohoff, D. F. (2004) Nat. Rev. Cancer **4**, 309.
Read, M. C., Lachs, M. S., Feinstein, A. R. (1995) JAMA 274, 645.
Sansal, I., Sellers, W. R. (2004) J. Clin. Oncol. **22**, 2954.
Selaru, F. M., Yan, X., Yin, J., et al. (2002) Gastroenterology **122**, 606.
Simon, R., Radmacher, M. D., Dobbin, K., et al. (2003) J. Natl. Cancer Inst. **95**, 14.
Song, J. Z., Stirzaker, C., Harrison, J., et al. (2002) Oncogene **21**, 1048.
Sporn, M. B. (1996) Lancet. **347**, 1377.
Srivastava, S., Kramer, B. S. (2000) Lab. Invest. **80**, 1147.

Stamey, T. A., Yang, N., Hay, A. R., et al. (1987) N. Engl. J. Med. **317**, 909.
Stears, R. L., Martinsky, T., Schena, M. (2003) Nature Med. **9**, 140.
Steen, H., Kuster, B., Fernandez, M., et al. (2002) J. Biol. Chem. **277**, 1031.
Stoeckli, M., Chaurad, P., Hallahan, D. E., et al. (2001) Nat. Med. **7**, 493.
Sullivan Pepe, M., Etzioni, R., Feng, Z., et al. (2001) J. Natl. Cancer Inst. **93**, 1054.
Suzuki, H., Freije, D., Nusskern, D. R., et al. (1998) Cancer Res. **58**, 204.
Tang, N., Tornatore, P., Weinberger, S. R. (2004) Mass Spectrom. Rev. **23**, 34.
Thomson, I. M., Pauler, D. K., Tangen, C. M., et al. (2004) N. Engl. J. Med. **350**, 2239.
Thomson, D. W., Krupey, J., Freedman, S. O., et al. (1969) PNAS USA **64**, 161.
Umbricht, C. B., Gabrielson, E., Ferguson, A., et al. (2001) Oncogene **20**, 3348.
Van de Vijver, M. J., He, Y. D., van 't Ver, L. J., et al. (2002) N. Engl. J. Med. **347**, 1999.
Velculescu, V. E., Zhang, L., Vogelstein, B., et al. (1995) Science **270**, 484.
Verma, M., Kagan, J., Sidransky, D., et al. (2003) Nature Rev. **3**, 789.
Vivanco, I., Sawyers, C. L. (2002) Nat. Rev. Cancer **2**, 489.
Wang, S., Gao, J., Lei, Q., et al. (2003) Cancer Cell **4**, 209.
Wang, H., Hanash, S. M. (2003) J. Cromatogr. B. **787**, 11.
Wang, M. C., Valenzuela, L. A., Murphy, G. P., et al. (1979) Investig. Urol. **17**, 159.
Warburg, A., Geissler, A. W., Lorenz, S. (1967) Physiol. Chem. **348**, 1686.
Washburn, M. P., Wolters, D., Yates, J. R., 3rd. (2001). Nat. Biotechnol. **19**, 242.
Wheeler, D. L., Barrett, T., Benson, D. A., et al. (2006) **34**, D173.
Winawer, S. J., St. John, D. J., Bond, J. H., et al. (1995) Bull. World Health Organization. **73**, 7.
Young, R. A. (1995) Cell **102**, 9.
Yoshimura, Y., Shinkawa, T., Taoka, M., et al. (2002) Biochem. Biophys. Res. Commun. **290**, 948.
Zeeberg, B. R., Feng, W., Wang, G., et al. (2003) Genome Biology **4**, R28.
Zhang, H., Li, X-J., Martin, D. B., et al. (2003) Nat. Biotechnol. **21**, 660.
Zhang, H., Yi, E. C., Li, X-J., et al. (2005) Mol. Cell. Proteomics **4**, 144.