

## **Chapter 2**

# **Technologies for Discovery of Biomarkers**

### **Introduction**

Biomarkers are present in all parts of the body including body fluids and tissues. Most of the clinical laboratory examinations are done on body fluids such as blood and urine. Biomarkers can be detected on imaging studies or examination of body tissues. Even exhaled breath contains biomarkers. A wide range of technologies is utilized for detection of biomarkers and a number of assays are already available.

### **Detection of Biomarkers in Tissues and Body Fluids**

Biomarkers occur in all parts of the body. For practical purposes, the focus of investigation is on detection of biomarkers in body fluids as most of such specimens can be obtained by non-invasive or minimally invasive techniques. Fluids include blood, urine, and cerebrospinal fluid (CSF). Tissue biomarkers can be examined in biopsy specimens of diseases organs. Molecular imaging enables non-invasive in vivo study of biomarkers of disease in various internal organs including the brain. Since most of body fluid biomarkers are proteins, they are described under the section on proteomics. Beyond these conventional sources, biomarkers can also be found in the exhaled breath.

### **Disease Biomarkers in Breath**

Since ancient times, physicians have known that human breath can provide clues to diagnosis of various diseases, e.g., the sweet, fruity odor of acetone indicates uncontrolled diabetes. Researchers have identified over 1,000 different compounds contained in human breath. These molecules have both endogenous and exogenous origins and provide information about physiological processes occurring in the body as well as environment-related ingestion or absorption of contaminants.

Among commonly used exhaled biomarkers, nitric oxide (NO) on exhaled air and some constituents of exhaled breath condensate in volatile or non-volatile form

may represent suitable biomarkers (Maniscalco et al. 2009). Nasal, bronchial, and alveolar NO could be analyzed separately, with implications in the assessment of systemic disease and endothelial dysfunction. Moreover, the profiles of several exhaled gases have a place in phenotyping diabetic patients and their risk of complications. Accordingly, metabolomics of the airway fluid using exhaled breath condensate is useful for the evaluation of both volatile and non-volatile biomarkers. Reference values are, however, lacking, and the influence of preanalytical variables on the methods requires further studies.

Several methods for trace molecular detection have been applied to the problem of breath analysis. These include optical detection, mass spectrometry, and electronic noses.

### ***Portable Breath Test for Volatile Organic Compounds***

Volatile organic compounds (VOCs) are present in normal human breath and can be detected by gas chromatography (GC). Some of these breath VOCs may be biomarkers of disease, but there are technical problem in collecting breath, analyzing low concentrations of VOCs, and distinguishing normal from abnormal findings. Menssana Research Inc. has overcome these technical problems and developed a portable breath collection apparatus which can collect breath samples virtually anywhere. It captures breath VOCs onto a small sorbent trap, which is sent to the laboratory for analysis by GC and MS. Each analysis usually identifies more than 200 different VOCs. This breath test has identified a new and comprehensive set of biomarkers of oxidative stress known as the breath methylated alkane contour (BMAC). Changes in the BMAC have revealed distinctive patterns in a number of different diseases each can be identified. This breath test is now being evaluated in several clinical studies, including

- Lung cancer.
- Breast cancer.
- Heart transplant rejection.
- Kidney disease.
- Ischemic heart disease.
- Diabetes mellitus.
- Pulmonary tuberculosis.

### ***Detection of Breath Biomarkers by Sensation Technology***

Analysis of breath for various respiratory biomarkers is an emerging field for Sensation<sup>TM</sup> technology (Nanomix Inc.), which is a nanoelectronic detection platform based on carbon nanotube networks. It consists of a tiny detector chip with one or more individually addressable detection elements, each capable of being independently functionalized to detect a specific target analyte. Sensation<sup>TM</sup> technology is

enabling a new non-invasive device for asthma monitoring by measuring the level of nitric oxide in exhaled breath.

### ***Detection of Breath Biomarkers Optical Frequency Comb Spectroscopy***

Broad-bandwidth, high-spectral-resolution optical detection of human breath has identified multiple important biomarkers correlated with specific diseases and metabolic processes (Thorpe et al. 2008). This optical-frequency-comb-based breath analysis system has shown excellent performance in high detection sensitivity, ability to identify, and distinguish a large number of analytes, and simultaneous, real-time information processing. A series of breath measurements have been done with this method including stable isotope ratios of CO<sub>2</sub>, breath concentrations of CO, and the presence of trace concentrations of NH<sub>3</sub> in high concentrations of H<sub>2</sub>O. It will enable detection of molecules that may be biomarkers for diseases like asthma or cancer by blasting a person's breath with laser light. New designs of prism cavities promise to expand the spectral coverage of optical cavities from 200 nm to several microns, greatly increasing the number of biomarkers that can be measured by a single system. Even the currently available system can be used for clinical trials to gather statistics for the feasibility and cost effectiveness of breath measurements for non-invasive health screening tests.

## **Genomic Technologies**

A major impact of the Human Genome Project (HGP), which identifies the 3 billion base pairs that comprise the human genome, is on molecular diagnostics. Many technologies developed during sequencing the human genome are being applied in molecular diagnostics and have led to discovery of biomarkers. Several genomic technologies used for biomarker discovery are also listed in molecular diagnostics. The presence of a gene mutation or specific gene alleles tested at the level of the patient's genome is not the same as gene expression and merely indicates that the patient is at risk for a certain disease. It does not tell us if the patient has the disease or when it may develop. Gene expression is important for this information and will be discussed here.

### ***Gene Expression***

Gene expression is used for studying gene function. Knowledge of which genes are expressed in healthy and diseased tissues would allow us to identify both the protein required for normal function and the abnormalities causing disease. Gene expression is a biomarker that will help in the development of new diagnostic tests

for various illnesses. Gene expression analysis provides a means for researchers to survey and measure gene activity in various disease populations, stages of disease development, and other defined research conditions. It may also be used for safety assessment of therapeutics and chemical substances in toxicology studies, to validate drug targets, and in preclinical drug-development research. Comprehensive, high information content, gene expression profiles obtained at early stages in drug discovery will ensure selection of the optimal path from target and lead to commercial product and minimize the risk of failure at later stages in development. A classification of methods of gene expression analysis is shown in Table 2.1.

**Table 2.1** Classification of methods of gene expression analysis

---

Genome-wide methods
Microarrays: whole-genome expression array
Serial analysis of gene expression (SAGE)
Expressed sequence tags (ESTs) analysis
Gene expression profiling based on alternative RNA splicing
Tangerine expression profiling
Individual sequences
Real-time RT-PCR
Competitive RT-PCR
RNase protection assay
T cell receptor expression analysis
Analysis of single-cell gene expression
RNA amplification
Monitoring in vivo gene expression
Magnetic resonance imaging (MRI)

---

©Jain PharmaBiotech

### Whole-Genome Expression Array

The Expression Array System (Life Technologies Corporation) is based on highly sensitive, chemiluminescent technology for gene expression detection and incorporates DNA probes that are each approximately 60 base pairs long. Chemiluminescent detection uses an enzymatic-initiated chemical reaction to generate light without the need for a laser. The combination of chemiluminescence and long oligonucleotides is expected to provide higher sensitivity than other microarray products that incorporate shorter oligonucleotide lengths or use fluorescence detection methods. The long oligonucleotides enable tighter binding to the target, which leads to the detection of more genes with greater selectivity and specificity. A single microarray has probes to detect an annotated and fully curated set of more than 20,000 human genes. To facilitate analysis of results, the system is combined with an industry standard database that includes gene acronyms, gene names, cross-references for gene identification, gene ontologies, and protein characterization. This system is also integrated with the sequence detection system-based products for gene expression analysis. Together, these products provide researchers with a seamless solution for gene expression analysis – from whole-genome analysis

to further quantitation and single gene analysis. TaqMan<sup>®</sup> Array Gene Signature Plates enable cost-effective measurement of the activity of multiple gene targets known to be expressed in biological pathways, processes, or disease states. These tools enable researchers to adopt a systems biology approach to investigate the role gene networks play in complex biological processes and molecular pathologies by performing real-time PCR assays to develop biomarkers based on patterns of gene expression.

### **Gene Expression Profiling on Whole Blood Samples**

Whole blood is one of the most common samples used in clinical research and the most readily available tissue for biomarker analysis. However, expression profiling on whole blood samples presents considerable technical challenges. Different whole blood preparation methods and the high relative concentration of globin RNA found in red blood cells can lead to changes in expression profiles *ex vivo*. Globin RNA can mask the RNA found in the transcriptionally active cells of interest in the white blood cell fraction, including the lymphocytes and monocytes. The PAXgene Blood RNA System (PreAnalytiX GmbH) has set a new standard for stabilizing whole blood cellular RNA profiles at the time of blood collection in an evacuated blood collection tube. This enables researchers and clinicians to perform more accurate analysis of gene expression profiles without the variations caused by sample collection, storage, transport, or fractionation while relying on highly standardized and proven sample collection principles. Microarray analyses from stabilized whole blood samples can be standardized to reduce variability in the preparation processes and decrease the signal contributed by globin RNA. The original information in the whole blood sample is preserved to enable the accurate application of molecular test methods, such as gene expression profiling.

### **Profiling Gene Expression Patterns of White Blood Cells**

White blood cells (WBCs) express tens of thousands of genes, whose expression levels are modified by genetic and external factors. Blood genomic profiles, created from distinct gene expression patterns of WBCs obtained by microarray examination of a minimally invasive blood sample, can provide biomarkers of several different disease states. These profiles may be used for diagnostic, prognostic, and therapeutic evaluations and also provide a method for the evaluation of the safety and efficacy of various therapeutics. Gene expression fingerprints are useful tools for monitoring exercise and training loads and thereby help to avoid training-associated health risks (Büttner et al. 2007).

There is marked alteration in WBC gene expression in animal models of injury and inflammation; the majority of the differentially expressed genes appear to be uniquely associated with the type of injury and/or the inflammatory stimulus (Brownstein et al. 2006). Although some pathological states such as hypoxia may have direct impact on white blood cells that is manifested by specific expression profiles, seemingly unrelated events affecting various organs can markedly alter white

blood cell gene expression in a predictable, characteristic way that provides a novel approach to diagnosis of diseases such as those involving the nervous system.

### ***Tissue Microarrays for Study of Biomarkers***

Microarray experiments produce a large volume of data that require validation. The validation of these experiments can be carried out in many fashions. In the reduction to clinical utility, the use of tissue microarrays (TMAs) has become a common tool to both validate and generalize the results of microarray experiments. A TMA is a collection of tissue specimens presented on a glass slide in a grid layout. TMAs contain between tens and hundreds of samples allowing the generalization of microarray findings to a large number of samples. TMAs can be used for in situ hybridization (ISH) and immunohistochemical analysis (IHC), confirming the results of microarray experiments at both the transcriptional and the proteomic levels (Hewitt 2006).

The traditional IHC methods for the study of biomarkers by TMAs are subjective approaches using manual analysis of tissues according to observable morphological patterns. Disparate results by these methods regarding the relationship between biomarker expression and patient outcome decrease the credibility of TMA studies. Some of these disparities result from subjective optimization of antibody concentrations. AQUA<sup>TM</sup> (HistoRx Inc), an automated quantitative analysis platform, dramatically improves the accuracy and speed of tissue biomarker analysis while reducing error through objective, rigorous measurement of protein expression (McCabe et al. 2005). TMAs are discussed further in Chapter 6.

### **Epigenomic Technologies**

Epigenomics is one of the many “omics” that have developed in the wake of the Human Genome Project (HGP). The Human Epigenome Project (HEP) map DNA methylation sites throughout the human genome. Whereas HGP provides the blueprint for life, the HEP tell us how this whole thing gets executed, what determines when and where genes are switched on and off to produce a person. And knowing more about the human epigenome may provide clues to what goes wrong in cancer and other diseases. Latest information can be obtained at the HEP web site <http://www.epigenome.org/>

As a prelude to the full-scale HEP, a pilot study of the methylation patterns within the major histocompatibility complex (MHC) has been completed. This region of chromosome 6 is associated with more diseases than any other region in the human genome. Methylation variable positions (MVPs) were identified in the vicinity of the promoter and other relevant regions of approximately 150 loci within the MHC in tissues from a range of individuals. This will provide an unprecedented insight into the complex relationship between genetics and epigenetics that underlies both normal cellular homeostasis and disease states, in particular autoimmune diseases.

For the pilot project, an integrated genomics-based technology platform was developed. The pipeline involves the automated bisulfite treatment of DNA from minute tissue biopsies, gene-specific bisulfite PCR, and large-scale sequencing of PCR amplicons. Analysis and quantification of methylation patterns are achieved by mass spectrometric and microarray assays.

### ***Discovery of Methylation Biomarkers***

Methylation is the only flexible genomic parameter that can change genome function under exogenous influence. Hence it constitutes the main and so far missing link between genetics, disease, and the environment that is widely thought to play a decisive role in the etiology of virtually all human diseases. Methylation occurs naturally on cytosine bases at CpG sequences and is involved in controlling the correct expression of genes. Differentially methylated cytosines give rise to distinct patterns specific for tissue type and disease state. Such MVPs are common epigenetic markers. SNPs promise to significantly advance our ability to understand and diagnose human disease. DNA methylation is an important cellular mechanism modulating gene expression associated with aging, inflammation, and atherosclerotic processes. Global DNA hypermethylation is associated with inflammation and increased mortality in cardiovascular disease (Stenvinkel et al. 2007).

In the last few years, DNA methylation has become one of the most studied gene regulation mechanisms in carcinogenesis. Advances in the technologies that enable detection of DNA methylation in a variety of analytes have opened the possibility of developing methylation-based tests. A number of studies have provided evidence that specific methylation changes can alter the response to different therapeutic agents in cancer and, therefore, be useful biomarkers. Application of technologies for methylation biomarkers in cancer is described in Chapter 6. Several companies are involved in detection of methylation biomarkers.

Pyro Q-CpG<sup>TM</sup> (Biotage AB) is a quantitative, high-resolution analysis solution for CpG methylation (epigenetic biomarkers), which can harmonize methylation data with observed biological phenomena. Quantitative methylSNP analysis by SNaPmeth or PyroMeth is a favorable alternative to existing high-throughput methylation assays. It combines single CpG analysis with accurate quantitation and is amenable to high throughput.

Orion's proprietary biomarker discovery platform, MethylScope<sup>®</sup> technology, is used for the discovery of methylation biomarkers. A single MethylScope<sup>®</sup> microarray is capable of quantitatively detecting the methylation status of each and every human gene. By comparing methylation profiles of two or more samples, Orion and its partners discover biomarkers associated with specific diseases. MethylScope<sup>®</sup> technology is the only platform capable of detecting inappropriate DNA methylation for all human genes on a single array, providing a fast, cost-effective, and comprehensive biomarker discovery tool.

A method using a split GFP tethered to specific DNA recognition elements, called mCpG-Sequence Enabled Reassembly (mCpG-SEER) of proteins, employs

a zinc finger attached to one-half of GFP to target a specific sequence of dsDNA, while a methyl-CpG-binding domain protein that is attached to the complementary half of GFP targets an adjacent methylated CpG dinucleotide site (Stains et al. 2006). It was shown that the presence of both DNA sites is necessary for the reassembly and concomitant fluorescence of the reassembled GFP. The GFP-dependent fluorescence reaches a maximum when the methyl-CpG and zinc-finger sites are separated by two base pairs and the fluorescence signal is linear to 5 pmol of methylated target DNA. The specificity of mCpG-SEER was found to be >40-fold between a methylated and a non-methylated CpG target site. This sequence-specific detection of DNA methylation at CpG dinucleotides will provide a possible diagnostic biomarker for cancer.

Although previous studies described methylation of isolated DNA extracted from cells and tissues using a combination of appropriate restriction endonucleases, no application to tissue cell level has been reported. A new method, called histo endonuclease-linked detection of methylation sites of DNA (HELMET), was designed to detect methylation sites of DNA with a specific sequences in a tissue section (Koji et al. 2008). In this study, the authors examined changes in the methylation level of CCGG sites during spermatogenesis in paraffin-embedded sections of mouse testis. They found hypermethylation of CCGG sites in most of the germ cells although non-methylated CCGG were colocalized in elongated spermatids. Some TUNEL-positive germ cells, which are frequent in mammalian spermatogenesis, became markedly Hpa II reactive, indicating that the CCGG sites may be demethylated during apoptosis.

## Proteomic Technologies

Although analysis of proteins has been an integral part of the field of clinical chemistry for decades, recent advances in technology and complete identification of the human genome sequence have opened up new opportunities for analysis of proteins for clinical diagnostic purposes. New analytical methods allow the simultaneous analysis of a large number of proteins in biological fluids such as serum and plasma, offering partial views of the complete set of proteins or proteome. Several proteomics approaches have been used to identify novel biomarkers. These are described in detail in a special report on this topic (Jain 2010b).

Not only has the number of proteins that can be detected in plasma expanded dramatically from hundreds to thousands, there is an increased capability for detecting structural variations of proteins. Recent studies also identified the presence of complex sets of small protein fragments in plasma. This set of protein fragments, the fragmentome or peptidome, is potentially a rich source of information about physiologic and disease processes. Advances in proteomics, therefore, offer great promise for the discovery of biomarkers that might serve as the basis for new clinical laboratory tests. There are many challenges, however, in the translation of newly discovered biomarkers into clinical laboratory tests (Hortin et al. 2006). Only 10%

of the proteins in human serum can be detected with currently available approaches, indicating the potential for further discovery of biomarkers. Protein variation is an untapped resource in the biomarker space, but only a selected few forms of proteomics applications are suitable for their analysis, and such variation could have a significant impact in disease diagnostics and therapeutic intervention (Kiernan 2008).

Proteomics is a key technology for the discovery of biomarkers for pharmaceutical and diagnostic research. Although gene expression provides the level of proteins that is the key to the effect of the gene, it can be due to other factors in addition to the concentration of mRNA that codes for it. These factors include protein post-translational modifications, turnover, transport, and excretion. Therefore quantitative proteomics is essential for monitoring different pathways in blood samples of patients. Such biomarkers help in differential diagnosis as well as provide an understanding of pathomechanism of the disease and assessment of response to treatment. Non-invasive measurement (e.g., in serum) is the key feature of a biomarker that can be identified in diseased tissue. Multidimensional protein fractionation schemes are used to achieve appropriate sensitivity. Proteomics is a key technology for the discovery of biomarkers for pharmaceutical and diagnostic research.

## ***2D GE***

Two-dimensional (2D) gel electrophoresis (GE) offers the highest resolution separations available for protein components of cells when gels of sufficient size are used. Proteins are separated in the first dimension on the basis of their charge and in the second dimension on the basis of their molecular mass. 2D GE is still the workhorse for obtaining protein expression patterns in cells. In high-format mode, it can produce gels containing up to 10,000 distinct proteins and peptide spots. The major problem with this technique is that most of the spots cannot be sequenced as they are beyond the capacity of current high-sensitivity sequencers. Standard format 2D GE yield up to 2,000 spots and are easy to sequence. During 2D PAGE (polyacrylamide GE), the proteins are separated in two dimensions (by isoelectric focusing and mass) and a pattern is achieved that places each of the 2,000 proteins of the cell at a grid reference point. By reference to the databases, individual proteins on the map can be identified as the product of genes that have been sequenced.

A second generation 2D GE called differential in gel electrophoresis (DIGE) has been commercialized by GE Healthcare and involves labeling two distinct protein mixtures with two different cyanine dyes, each of which fluoresces at a distinct wavelength. The labeled protein samples are then separated on a single 2D gel. The size- and charge-matched proteins enable co-migration of identical proteins.

While comparing different samples, controlling the position of the protein spots can be critical and is completely dependent upon the fidelity of the isoelectric focusing first dimension and the molecular weight separating gel slab of the second dimension. Differences between the test samples are determined by quantifying the ratios of spot intensities in independent 2D gels and techniques such as mass

spectrometry (MS) can then be used to help identify the proteins through peptide mass fingerprinting or direct sequencing.

### *Isotope-Coded Affinity Tags*

Isotope-coded affinity tag (ICAT) peptide labeling is an approach that combines accurate quantification and concurrent sequence identification of individual proteins in complex mixtures. This method is based on a newly synthesized class of chemical reagents used in combination with tandem mass spectrometry. The method consists of four steps:

- The reduced protein mixtures representing two-cell state are treated with two different versions of ICAT reagent – one light and one heavy.
- The labeled samples are combined and proteolytically digested to produce peptide fragments.
- The tagged cysteine-containing fragments are isolated by avidin affinity chromatography.
- The isolated tagged peptides are separated and analyzed by microcapillary tandem MS which provides both identification of peptides by fragmentation in MS-mode and relative quantitation of labeled pairs by comparing signal intensities in MS mode.

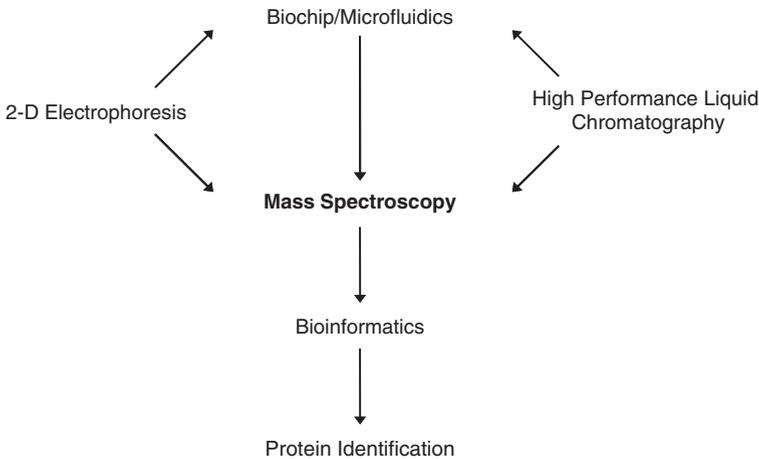
The advantages of ICAT over 2D GE have the potential for full automation, and thus for high-throughput proteomic experiments. There is no need to run time-consuming experiments and because it is based on stable isotope labeling of the protein, there is no need for metabolic labeling or no radioactivity is involved. ICAT can be used for the analysis of several classes of proteins such as membrane proteins and low copy number proteins that are poorly tractable by 2D gels. Most importantly, it provides accurate relative quantification of each peptide identified. The limitations of this technique are that the proteins must contain cysteine and the large size of the tag compared to some small peptides and may interfere with peptide ionization. These, however, can be overcome by designing different reagents with specificities for other peptide chains and using a smaller tag group.

ICAT is an emerging technique for differential expression proteomics, and its full potential remains to be fully evaluated. Advances in sample fractionation at the protein level, sample fractionation at the peptide level, and improved data acquisition schemes will all be required for the full potential of ICAT to be realized. New separation systems, such as ultra-high pressure nanoscale capillary LC, will improve the peak capacity for ICAT experiments, leading to improved proteome coverage. New MS technologies, such as the high sensitivity, high-throughput MALDI-TOF instrument, can be expected to have a very significant impact in ICAT proteomics.

## Mass Spectrometry

Mass spectrometry (MS) is the measurement of molecular mass. A mass spectrometer consists of three essential parts: (1) an ionization source with conversion of molecules into gas-phase ions; (2) a mass analyzer to separate individual mass to charge ratios ( $m/z$ ); and (3) an ion detector. Several variants of mass spectrometry are described in the following sections. Proteomics methods based on MS hold special promise for the discovery of novel biomarkers that might form the foundation for new clinical blood tests, but to date their contribution to the diagnostics has been disappointing, partly due to the lack of a coherent pipeline connecting marker discovery with well-established methods for validation. Advances in methods and technology now enable construction of a comprehensive biomarker pipeline from six essential process components: candidate discovery, qualification, verification, research assay optimization, biomarker validation, and commercialization (Rifai et al. 2006).

Biomarker discovery using MS techniques requires sensitivity, mass accuracy, and reproducibility. Various MS-based techniques used for biomarker discovery are described in the following text. The central role of mass spectrometry in proteomics is shown in Fig. 2.1.



**Fig. 2.1** The central role of spectrometry in proteomics

### MALDI Mass Spectrometry for Biomarker Discovery

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) has become a widely used method for determination of biomolecules including peptides, proteins, carbohydrates, glycolipids, and oligonucleotides. MALDI-MS has emerged as an effective bioanalytical tool having unique capabilities in handling

complex mixtures (such as proteolytic digests) and in high-sensitivity (femtomole or even subfemtomole) measurements.

The direct analysis of tissue sections by MALDI mass spectrometry holds tremendous potential for biomarker discovery. This technology routinely enables many hundreds of proteins to be detected over a mass range of approximately 2,000–70,000 Da while maintaining the spatial localization of the proteins detected (Reyzer and Caprioli 2005). This technology has been applied to a wide range of tissue samples, including human glioma tissue and human lung tumor tissue. In many cases, biostatistical analyses of the resulting protein profiles revealed patterns that correlated with disease state and/or clinical endpoints.

Mass spectrophotometry platforms using MALDI MS and MS/MS or LTQ ion trap MS are capable of delivering sensitive and accurate identifications of hundreds of proteins contained in individual samples including individual forms of processing intermediates such as phosphopeptides. The systems biology approach of integrating protein expression data with clinical data such as histopathology, clinical functional measurements, medical imaging scores, patient demographics, and clinical outcome provides a powerful tool for linking biomarker expression with biological processes that can be segmented and linked to disease presentation.

MALDI-TOF MS is a particle-counting method that responds to molar abundance, and ranking of plasma proteins by molar abundance increases the rank of small proteins relative to traditional ranking by mass abundance. Detectors for MALDI-TOF MS augment the bias for detecting smaller components by yielding stronger signals for an equivalent number of small versus large ions. Consequently, MALDI-TOF MS is a powerful tool for surveying small proteins and peptides comprising the peptidome or fragmentome, opening this new realm for analysis. It is complementary to techniques such as electrophoresis and HPLC, which have a bias for detecting larger molecules. Virtually all of the potential markers identified by MALDI-TOF MS to date represent forms of the most abundant plasma proteins. Analyses of serum or plasma by MALDI-TOF MS thus provide new information mainly about small proteins and peptides with high molar abundance. The spectrum of observed proteins and peptides suggests value for applications such as assessment of cardiovascular risk, nutritional status, liver injury, kidney failure, and systemic immune responses rather than early detection of cancer. Extending analysis by MALDI-TOF MS to lower-abundance components, such as biomarkers for early-stage cancers, probably will require more extensive specimen fractionation before analysis (Hortin 2006a).

## **2D PAGE and Mass Spectrometry**

The essence of this approach is to separate proteins from a specific cell or tissue type, record the pattern, and then produce a Western Blot. Proteins in the blot are digested with a proteolytic enzyme, which has well-defined cleavage specificity. Peptide fragments can be analyzed by matrix-assisted laser desorption mass spectrometry (MALDI-MS). The resulting peptide masses are then compared with theoretical masses calculated from amino-acid sequence databases. This technique

has been used successfully to identify yeast proteins. For completely sequenced genomes, 90% of the proteins can be identified rapidly and automatically by searching databases with lists of peptide masses obtained by 2D gel technique and matrix-assisted laser desorption/ionization. This study established that mass spectrometry provides the required throughput, the certainty of identification, and the general applicability to serve as the method of choice to connect genome and proteome. Nano-electrospray tandem mass spectrometry is then used to characterize novel proteins either by searching EST databases with peptide sequence tags or by generating sufficient sequence for cloning. This approach can be automated.

### **Quantitative Tandem MS**

Mass spectrometers are making large inroads into the modern day life science industry. Their high-throughput capability, resolution, and precision have ensured their place in pharmaceutical industry and life science research institutes as pivotal tools to screen hundreds and thousands of DNA and protein samples per day. The advantages are obvious when proteins or nucleotides can be rapidly analyzed in a few seconds by MALDI.

Modern tandem mass spectrometers have many benefits. Among these is the ability to accurately measure down to femtomolar amounts in complex biological mixtures without chromatography. This enables samples to be prepared and analyzed using high-throughput technologies such as robotic liquid handlers and microtiter solid phase extraction systems. The operator of tandem mass spectrometry (MS/MS) experiments can choose metabolite specific fragmentations for definitive analytic identification. Thereby, arrays of metabolites in multiple pathways can be accurately identified and quantified per sample run akin to DNA and protein multiplexing technologies.

### **Single-Molecule Mass Spectrometry Using a Nanopore**

A 2D method for mass spectrometry (MS) in solution that is based on the interaction between a nanometer-scale pore and analytes (Robertson et al. 2007). As an example, poly(ethylene glycol) molecules that enter a single  $\alpha$ -hemolysin pore cause distinct mass-dependent conductance states with characteristic mean residence times. The conductance-based mass spectrum clearly resolves the repeat unit of ethylene glycol, and the mean residence time increases monotonically with the poly(ethylene glycol) mass. This single-molecule analysis technique could prove useful for the real-time characterization of biomarkers (i.e., nucleic acids, proteins, or other biopolymers). With automated, unsupervised analytical, and statistical methods, this technique should prove viable as a generalized analytical technique with nanopore arrays containing nanopores both with specific affinities for single biomarkers and with non-specific transducers. In situ monitoring of cellular metabolism with such arrays should provide the sensitivity to monitor subtle changes observed through the release of biomarkers.

## Imaging Mass Spectrometry

Imaging MS (IMS) is an emerging technology that enables the direct analysis and determination of the distribution of molecules in tissue sections. Pinpointing the location of the cells producing different levels of specific proteins enables the identification of differences in normal and diseased tissues that could lead to a better understanding of the onset and treatment of disease. IMS is a combination of MS with tissue imaging. Molecular imaging is achieved through secondary ion mass spectrometry (SIMS), which has unique analytical capabilities for mapping a variety of biological samples at the tissue level. SIMS provides information on the spatial distribution of the elements and low molecular mass compounds as well as molecular structures on these compounds, while MALDI yields spatial information about higher molecular mass compounds, including their distributions in tissues at very low levels, as well as the molecular structures of these compounds. To take molecular photographs, a chunk of tissue is first frozen so that it can be cut into super-thin slices. A tissue slice to be analyzed is then coated with a matrix material and introduced into the mass spectrometer, where a laser beam blasts successive spots on the tissue to release molecules for analysis. Each spot becomes a pixel in the final image, with each pixel containing a record of the molecules located in that tiny spot. Computer processing can then be used to display the locations of selected proteins, based on their size.

Biological molecules such as proteins, peptides, lipids, xenobiotics, and metabolites can be analyzed in a high-throughput manner with molecular specificity not readily achievable through other means. Tissues are analyzed intact and thus spatial localization of molecules within a tissue is preserved. IMS has the potential to deliver highly parallel, multiplexed data on the specific localization of molecular ions in tissue samples directly, and to measure and map the variations of these ions during development and disease progression or treatment. There is an intrinsic potential to be able to identify the biomarkers in the same experiment, or by relatively simple extension of the technique (Goodwin et al. 2008). Unlike many other imaging techniques, no a priori knowledge of the biomarkers being sought is necessary.

Several studies that focus on the unique types of information obtainable by IMS, such as A $\beta$  isoform distributions in Alzheimer plaques, protein maps in mouse brain, and spatial protein distributions in human breast carcinoma (Seeley and Caprioli 2008). The analysis of a biopsy taken 100 years ago from a patient with amyloidosis illustrates the use of IMS with formalin-fixed tissues. The technique can also assist in tracking the location of drugs and their metabolites (see chapter on pharmaceutical applications of IMS) in the treatment of various disorders, including cancer, thus potentially providing a powerful new research tool.

Separation of shotgun-produced peptides by the use of immobilized pH gradient-isoelectric focusing (IPG-IEF) has been combined with IMS (Vaezzadeh et al. 2008). The peptides are then transferred by capillarity to a capture membrane, which is then scanned by the mass spectrometer to generate MS images. This high-throughput method enables a preview of the sample to be obtained in a single

day and has been used for differential comparison of the membrane proteome of two different strains of *Staphylococcus aureus* bacteria in a proof-of-principle experiment.

### **Requirements for MS-Based Proteomic Biomarker Development**

The success of the MS process is critically dependent on the ability to identify as many putative active or informative molecules as possible and having efficient methods to weed out weak candidates in subsequent steps. In biomarker development this requires definition of performance goals, the ability to test these goals in relevant populations, and the ability to evaluate the performance of multianalyte panels and discover covariance in addition to single markers (Heegaard 2008). Some requirements for successful MS-based proteomic biomarker development are

*Samples:* These should be collected, preprocessed, and stored in a uniform way, but should be split into two sets. The validation sample sets are identical to training samples but used only for validation.

*Separation:* Discriminatory peaks must be identified to assess relevance for intended use and to develop immunoassays.

*Statistics:* Tools used include support vector machines, discriminant analysis, classification trees, random forest, and neural networks.

Found or not found, performance with and without.

### ***Liquid Chromatography–MS Combination***

Classical LC (liquid chromatography) has been combined with MS. However, some emerging strategies are shown to be more suitable for protein characterization and identification. Microanalytical protein characterization with Ettan multidimensional liquid chromatography (GE Healthcare) achieves reproducible separation of proteins based on more than one physical property. MDLC/mass spectrometry (MDLC/MS) improves the throughput and reliability of peptide mapping for the following reasons:

- Faster protein digestion with immobilized enzymes.
- Automated generation of peptide maps.
- Online detection of peptide maps by electrospray interface and MALDI-TOF (matrix-assisted laser desorption ionization mass spectrometry).

### ***Protein Tomography***

Light microscopy techniques can also provide images of cellular events. However, these techniques do not show structure and function on a protein level. NMR, X-ray, and single-particle electron microscopy provide higher resolution structural

information but the proteins are studied in an artificial environment. In addition, the protein image is compiled of an average of thousands to millions of objects. As a result, the latter methods are unable to give information about flexibility and shape of individual proteins or protein complexes as they are not in situ.

Protein Tomography<sup>TM</sup> (Sidec Technologies) can be used for imaging of molecular events in situ or in vitro, combining cryo-EM, electron tomography, and advanced data processing algorithms. It can accomplish the following:

- It can take images of individual proteins and protein complexes.
- It is the only method that picture membrane proteins in their cellular context.
- The results are directly comparable between in vitro and in situ analysis.
- It enables visualization of the different states of the same molecule in a mixture: unbound receptors, receptors attached to ligands, recruitment of signaling proteins, etc.
- It is an ideal method for study of proteins as biomarkers.

### ***Protein Biochips/Microarrays and Biomarkers***

Protein microarray technology is a powerful tool for biomarker discovery and validation. Protein microarrays have this potential and it will, in future, be possible to use them to determine simultaneously a variety of parameters from a minute sample volume, as well as in the discovery and validation of biomarkers. However, before protein microarrays find their way into routine and high-throughput applications, their robustness, sensitivity, and automation must be tested thoroughly and they must be available at an affordable price. Nevertheless, for focused protein-profiling approaches searching for only a few parameters in parallel, current technologies are already mature enough to deliver reliable data sets.

ProteinChip (Vermillion) has been used for discovery of biomarkers. The advantage of ProteinChip over 2D GE is that the chip platform used to identify the biomarker can also be used to develop a high-throughput assay. Laser capture microdissection has been used in conjunction with ProteinChip to study protein expression profiles in cancer.

### **Detection of Biomarkers Using Peptide Array Technology**

Peptide arrays can be used to diagnose almost every disease providing their biomarkers are available. Peptides recognizing those biomarkers can be synthesized on a single array and one test can simultaneously screen for all biomarkers. High throughput and high resolution make it possible to distinguish the slightest difference between two disease biomarkers. In addition, unlike DNA arrays, the detection of disease biomarkers on a peptide array is technically very straightforward and easy to operate. Therefore, the peptide array can provide a powerful tool for clinical diagnosis in the future.

### **Protein Nanobiochip**

Biomarker proteins as early warning signs for diseases such as cancer can be identified for diagnostic purposes by finding their isoelectric points and their molecular weights. Isoelectric points are chemical features that refer to the electrical state of a molecule when it has no net charge. Conventional protein chips use a gel across which an electric current is applied to find the targeted protein's isoelectric points. In the nanobiochip, instead of being filtered through a block of gel, the protein molecules are separated by their isoelectric points by a capillary action as the proteins flow in a solution along channels in the chip. The protein molecules are then dried and irradiated by a laser. Their molecular weights are then measured by a mass spectrometer. The laser helps the proteins leave the chip, and the mass spectrometer is used to judge the molecular weights of the protein molecules in the samples by measuring how early they reach a detector. In the mass spectrometer, light molecules fly faster than heavy ones in an electric field. The mass spectrometer judges the weight of the molecules by monitoring the timing of when each molecule reaches a detector. In addition to being faster than techniques that use gel blocks, the nanobiochip method requires blood samples of about 1  $\mu\text{L}$  compared to about 20  $\mu\text{L}$  or more that are needed using gel-based techniques.

### **Antibody-Based Biomarker Discovery**

Antibody arrays (also called protein expression microarrays) can be used for screening of biomarkers. As in the case of gene expression microarrays, it is expected that protein expression microarrays will help in biomarker discovery, prediction of disease outcomes and response to treatments, and detection of molecular mechanisms and/or pathways associated with a particular disease state. However, accurately achieving these aims is dependent upon suitable experimental designs, normalization procedures that eliminate systematic bias, and appropriate statistical analyses to assess differential expression or expose expression patterns. A considerable amount of research has been devoted to two-color cDNA arrays to improve experimental design, normalization, and statistical analyses to assess differential expression and classification. These methods are directly applicable to two-color antibody arrays. Statistical methods have been developed for cDNA arrays and their application to antibody arrays has been described (Eckel-Passow et al. 2005).

### ***Real-Time PCR for Quantification of Protein Biomarkers***

The TaqMan Protein Expression assays (Life Technologies) are real-time PCR assays that enable researchers to rapidly detect and quantify proteins in human cell samples as well as correlate relative levels of specific proteins with cell functions in different disease conditions. These assays detect and quantify proteins by an innovative technology that combines an antibody–oligonucleotide-tagged immunoassay with a TaqMan assay to generate real-time PCR data for specific proteins present in

as little as 10–250 cells. These tools offer researchers a more quantitative, simpler, and standardized approach to protein analysis of various cell types, especially stem cells, compared to other more complex methods that also require large amounts of cell sample. When combined with TaqMan assays for microRNA and mRNA, they form the only quantitative protein analysis method that enables researchers to make comparisons of protein and RNA molecular biomarkers identified on the same platform with the same starting samples. The assays for targets in stem cells have already been used to quantify protein biomarkers for pluripotency (OCT 3/4, NANOG, SOX2, and LIN28) using small quantities of human stem cell and testicular germ cell samples. These types of biomarkers can potentially be used to identify and characterize malignant cells.

### ***Magnetic Beads for Protein Biomarker Discovery***

Magnetic beads are providing ways to identify protein biomarkers faster, more efficiently, and at lower cost than other methods. ProMag™ (Bangs Laboratories) is a new magnetic polymer-based 3 μ sphere, which has a hydrophilic surface to reduce non-specific binding in protein-based systems. The 3 μ size provides an ample surface for capturing/purifying targets. For biomarker discovery, investigators can coat beads with ligand of interest.

### ***CellCarta® Proteomics Platform***

CellCarta® (Caprion Proteomics) is an excellent method for profiling proteins in solid tissues and plasma and serves as an engine for both disease target identification and predictive medicine. The platform has been successfully employed in oncology target identification and clinical biomarker discovery. It has four steps:

1. Sample preparation and purification. Novel methods have been developed for isolating organelles such as the plasma membrane, phagosomes, and golgi to a high degree of purity. It enables more comprehensive proteomic comparisons of normal and diseased cells and provides the biological and functional context for the identification of high-value, disease-relevant protein targets.
2. Mass spectrometry analysis of samples in a given project consists of two phases, an initial phase of analysis in LC-MS mode for the detection of differentially expressed peptide ions contained in the samples, and a subsequent re-injection of the same samples in LC-MS-MS mode in order to obtain sequence information for those differentially expressed peptides.
3. Quantitative expression profiling of peptides. Proteomics data analysis software is capable of routinely detecting over 25,000 peptides per sample, reproducibly tracking these peptides across large sample sets and measuring their relative expression levels with high accuracy.
4. Bioinformatics analysis (protein identification).

## ***MASStermind™***

MASStermind™ (Pronota NV), a biomarker discovery engine, is based on novel proteomics technologies that enable fast, automated, and highly flexible and sensitive qualitative as well as quantitative proteomics without using gels and without the use of an affinity tag. The key difference from currently available profiling technology is MASStermind's high information content; it delivers both a profile and the identity of each of the underlying protein biomarkers. Analysis of one blood sample allows the quantitative assessment of close to 3,000 different proteins and their processed isoforms in a high-throughput fashion. The basic strategy comprises the following steps:

- Isolation of proteins from a biological sample.
- Cleavage of proteins to peptides.
- Chromatographic fractionation of the complex peptide mixture.
- Modification of a target subset of peptides.
- Specific isolation of the modified peptides by a second automated chromatography.
- Analysis by mass spectrometry of the isolated peptides and,
- Protein identification via intelligent interrogation of databases with novel software.

The procedure, in which fractions of the first chromatographic step are combined, modified, and run in a diagonal chromatographic manner, is called COmbined FRActional DIagonal Chromatography (COFRADIC™).

## ***Search for Biomarkers in Body Fluids***

The first decision to be made in a search for a biomarker is whether to look in a body fluid or a tissue. Body fluids have the advantage of being more easily accessible and are more likely to be of clinical use because serum or urine can be obtained by non-invasive methods as a routine (Jain 2007). However, plasma (together with serum) is the largest and deepest version of the human proteome and presents challenges for the investigators.

## ***Challenges and Strategies for Discovery of Protein Biomarkers in Plasma***

Plasma is the most difficult sample to work with in proteomics, despite the relatively good behavior (i.e., solubility) of its protein components. The daunting size of the plasma proteome is a reflection of the sheer number of different proteins to be tested –  $>10^6$  different molecules representing products of all 25,000–30,000

genes. Approximately half of the total protein mass in plasma is accounted for by only one protein (albumin, present at  $\sim 55,000,000,000$  pg mL<sup>-1</sup>), while roughly 22 proteins, which also include transferrin, fibrinogen, and other, make up 99% of the total. The remaining 1% contains secreted proteins, some of which are promising biomarkers. At the other end of the concentration histogram are the cytokines, such as interleukin-6 (IL-6), which is normally present at 1–5 pg mL<sup>-1</sup>. The difference in concentration between albumin and IL-6 is thus  $\sim 10^{10}$ . This range covers the proteins that are known and considered to be useful as biomarkers but ignores those to be discovered in the future at even lower concentrations. Characterization of the plasma proteome, therefore, poses many challenges. The aim is to create a reliable human blood plasma reference base as a basis for future biomarker discoveries. Many of the established biomarkers were detected by immunoassays and not proteomics, where currently technology is limited to a dynamic range of  $10^3$ – $10^4$ . A number of strategies have been used to overcome the problems and some research projects are underway.

Focus of research during the past decade has been on complete analysis of plasma samples to see all differences rather than targeted analysis to measure one or more hypothesis-generated candidates. Complete analysis has the advantage that it enables the direct selection of optimal biomarker proteins at the outset. The number of proteins detectable in plasma has risen to over 1,000 reported in various recent studies. Targeted analysis, however, is more meaningful for disease–biomarker association and hybrid approaches have been proposed that combine the multiprotein view of proteomics and the advantages of targeted-specific assays and are termed targeted proteomics.

Identification of rare proteins in blood is often hindered by highly abundant proteins, such as albumin and immunoglobulin, which obscure less plentiful molecules. A solution to this problem is an immunoaffinity column, Multiple Affinity Removal System (Agilent Technologies), which comprises antibodies to the six most abundant proteins found in human blood. By merely running a sample over the matrix, one can specifically remove all six proteins at once, unveiling lower-abundance species that may represent new biomarkers for disease diagnosis and therapy. The process removes about 85% of the total protein mass. The multiple affinity removal system works with blood, cerebrospinal fluid, and urine, all of which contain the same major proteins. Blood serum is the favored source for investigators interested in large-scale proteomics because it has the most proteins. However, so far only about 500 of the 30,000 proteins in serum have been identified. By removing albumin and the other five major proteins, the scientists will be able to dig further into the proteome.

VIVAPURE Anti-HSA/IgG Kit (Sartorius AG) provides both a highly specific, antibody fragment-based system for albumin removal and a protein G-coupled system for IgG removal. These affinity systems are well established for their excellent sensitivity and specificity. 2D separation of proteins from depleted serum showed better resolution of existing spots and revealed an additional 1,000 proteins that were not visible in gels of the untreated sample. A convenient, easy-to-use spin column format needing only a lab centrifuge allows parallel processing of different

samples. Detailed studies with the new system have shown it to be a highly useful tool for the detection of new biomarkers in blood, synovial fluid, CSF, or other protein samples that would otherwise be affected by excess albumin and IgG.

In February 2009, a consortium of European companies and universities were awarded a 3-year research grant by the European Commission to develop a plasma biomarker discovery platform called proactive. The goal of the initiative is to develop methods and reagents for multiplexed proximity ligation assays for detecting low-abundance proteins in plasma, along with statistical methods and data management tools.

### **3D Structure of CD38 as a Biomarker**

Human CD38 is a multifunctional protein involved in diverse functions. As an enzyme, it is responsible for the synthesis of two  $\text{Ca}^{2+}$  messengers, cADPR and NAADP; as an antigen, it is involved in regulating cell adhesion, differentiation, and proliferation. Besides, CD38 is a marker of progression of HIV-1 infection and a negative prognostic marker of B-CLL. 3D crystal structure CD38 has been determined, which may lead to important discoveries about how cells release calcium (Liu et al. 2005e). The findings also may offer insights into mechanisms involved in certain diseases, ranging from leukemia to diabetes and HIV-AIDS. For example, CD38 interrupts an interaction between the AIDS virus and its point of entry into cells – a protein receptor called CD4. CD38's 3D structure reveals that a peptide may play a role in interrupting the interface between CD4 and HIV-AIDS. Levels of the protein rise for unknown reasons during illness, making human CD38 a biomarker for these diseases.

### **BD<sup>TM</sup> Free Flow Electrophoresis System**

BD<sup>TM</sup> Free Flow Electrophoresis (FFE) System (BD Biosciences) is a novel separation concept that enables greater penetration of the proteome via separation of a wide variety of charged or chargeable analytes, ranging from small molecules to cells. Unlike traditional electrophoretic or chromatographic techniques, the method provides continuous electrophoretic separation, in the absence of solid-phase interactions, providing the simultaneous benefits of reproducibility, speed, a high-resolution fractionation gradient, and high recovery. Laboratory results demonstrate an improvement in the resolution on 2D PAGE by more than a factor of five, when compared to unfractionated samples. Additionally, digested fractions analyzed by LC-MS/MS yield hundreds of protein identifications per fraction. A further complementary benefit of this novel technology is the ability to rapidly and efficiently deplete albumin from human plasma during the fractionation process. Additionally, the FFE methods developed are appropriate for processing large numbers of samples, allowing high sample throughput without sacrificing resolution or protein concentration dynamic range, which should improve identification of novel protein biomarkers in plasma.

### **Isotope Tags for Relative and Absolute Quantification**

Life Technologies Corp has developed isotope tags for relative and absolute quantification (iTRAQ<sup>TM</sup>), which is potentially more cost-effective because up to four samples can be screened simultaneously with four isotopically distinguishable reagents. The iTRAQ method utilizes isobaric tags containing both reporter and balancer groups. The reporter is quantitatively cleaved during collision-induced dissociation (CID) to yield an isotope series representing the quantity of a single peptide of known mass from each of up to four different samples. This quantification group (the reporter) is “balanced” by a second group (the balancer) depleted of the same stable isotopes, which maintains each isotopic tag at exactly the same mass. Since the peptide remains attached to the isobaric tags until CID is conducted, the peptide is simultaneously fragmented for sequence identification.

The current generation of iTRAQ reagents labels lysine residues and the N termini of peptides, meaning that most peptides are multiply labeled (as with GIST). Therefore, iTRAQ suffers the same peptide overabundance problem and must be coupled with one or more dimensions of chromatographic or electrophoretic separation before MS analysis to limit the number of isobaric tagged peptides in the first MS dimension. The advantage of iTRAQ over these methods is that the label is cleaved in the tandem MS before quantification.

Because differences in peptide levels can only be determined after tandem MS, the first MS dimension cannot be used to pre-screen peptides for differential expression before tandem MS identity determination. Therefore, each and every peptide must be subjected to tandem MS analysis, making iTRAQ both time consuming and sample intensive for biomarker discovery applications. Furthermore, any untagged isobaric chemical noise may confound tandem-MS sequencing of the iTRAQ-labeled peptides.

Another issue with this method is the problem of protein variants. Any variant of the peptide of interest will not be isobaric with the same tagged peptides from control samples. Such non-isobaric peptides can be detected by their absence, but may be falsely interpreted as downregulation of the parent protein. Furthermore, such peptides may be isobaric with other peptides, confounding the interpretation of expression levels or sequences of other peptides. However, in target validation, patient profiling or toxicological screening applications, where the masses of the peptides are known, iTRAQ<sup>TM</sup> is potentially very cost-effective.

### **Plasma Protein Microparticles as Biomarkers**

Although current proteomic technologies enable detection and analysis of extremely small amounts of proteins (picomole to attomole level), it is difficult to detect and quantify proteins present at two to three orders of magnitude lower than the more abundant proteins. Microparticles are subcellular particles varying in size from 50 nm to 1  $\mu$ m that are released by essentially all cells upon activation. They have a variety of important physiological and pathophysiological functions including role as the main carrier of tissue factor in the blood and participating

in intercellular communications. In the plasma of healthy individuals, over 90% originate from platelets. Other sources of microparticles are cancer cells, endothelial cells including those from newly formed angiogenic vessels, leukocytes, and smooth muscle cells. Although this subproteome makes up less than 0.01% of the total plasma proteome, it is rich in proteins altered under a variety of pathological conditions. The proteome of microparticles isolated from the plasma of healthy individuals is slightly but significantly different from that of platelet microparticles.

The total number and the cellular origin of these microparticles are altered in a wide variety of pathological conditions including cardiovascular diseases. The protein composition of microparticles differs according to the disease. Normal biological fluids used for biomarker discovery, such as plasma or urine, contain a small number of proteins present at much higher amounts than the remaining proteins. For example, in the plasma, albumin and immunoglobulins are present at milligrams per milliliter, while proteins of interest for biomarker discovery may be present at micrograms to picograms per ml. Microparticle subproteome is being investigated as a potential source of biomarkers (Smalley and Ley 2008). Microparticle proteomics is developing microparticle-based biomarker screening and diagnostic tests for cardiovascular diseases.

### **Proteome Partitioning**

Protein depletion has been used for some years to remove most of the albumin and/or IgG from biofluids such as plasma and serum prior to analysis, but it is clear that this alone is insufficient to enable progress to be made in biomarker discovery. The presence of highly abundant proteins significantly complicates the discovery process by masking the presence and limiting the detection of low-abundance species. ProteomeLab IgY partitioning (Beckman Coulter) addresses this issue by reversibly capturing 12 of the more abundant proteins from human biofluids such as plasma and serum, yielding an enriched pool of low-abundance proteins for further study. The captured proteins can also be easily recovered for investigation if required – hence the term partitioning rather than depletion. IgY-12 selectively partitions the 12 highly abundant proteins and the partitioned fractions can be taken to the next stage of the discovery process, such as multidimensional fractionation using the ProteomeLab PF 2D system or profiling using 2D PAGE.

### **Stable Isotope Tagging Methods**

Stable isotope tagging methods provide a useful means of determining the relative expression level of individual proteins between samples in a mass spectrometer with high precision. Because two or more samples tagged with different numbers of stable isotopes can be mixed before any processing steps, sample-to-sample recovery differences are eliminated. Mass spectrometry also allows post-translational modifications, splice variations, and mutations (often unnoticed in immunoassays) to be detected and identified, increasing the clinical relevance of the assay and avoiding

the issues of non-specific binding and cross-reactivity observed in immunoassays. Several stable isotope tagging methods are available for use in proteomics research.

Baiting and affinity pre-enrichment strategies, which overcome the dynamic range and sample complexity issues of global proteomic strategies, are very difficult to couple to mass spectrometry. This is due to the fact that it is nearly impossible to sort target peptides from those of the bait since there will be many cases of isobaric peptides. Isotope-differentiated binding energy shift tags (IDBEST) have been developed by Target Discovery Inc. as a tagging strategy that enables such pre-enrichment of specific proteins or protein classes as the resulting tagged peptides are distinguishable from those of the bait by a mass defect shift of approximately 0.1 atomic mass units (Hall and Schneider 2004). The special characteristics of these tags allow resolution of tagged peptides from untagged peptides through incorporation of a mass defect element; high-precision quantitation of up- and downregulation by using stable isotope versions of the same tag; and potential analysis of protein isoforms through more complete peptide coverage from the proteins of interest.

### **Technology to Measure Both the Identity and the Size of The Biomarker**

While SELDI-TOF platforms provided MS approaches that could generate size and identity information, the inability of surface-based chromatography to provide a strategy for low-abundance analytes necessitates the development of new approaches of fragment-based analyte detection. One ideal format would be high-throughput MS technology coupled with true affinity chromatography whereby larger quantities of body fluids could be queried over a flow-through high-capacity surface. Therefore, the future of peptide-based diagnostics will require the invention and adoption of wholly new technologies that rapidly read both the identity and the exact size of the molecule. Immuno-MS provides a means to do this. Using this technology, a microaffinity antibody column, perhaps in a multiplexed microwell format, is first used to capture all species of molecules that contain the antibody recognition site, regardless of size. MS analysis of eluted peptides provides an extremely accurate mass determination of the entire population of captured peptides. Thus, in only two steps, immuno-MS can rapidly tabulate a panel of peptide fragments derived from a known parent molecule.

Conventional immunoassay and newer multiplexed technologies such as antibody arrays and suspension bead arrays cannot measure panels of peptide analytes that carry their diagnostic information based on two dimensions of both size and identity. Efforts are being made to develop immunoassay-based applications using a mass spectrometer as the detector rather than a fluorescence detector. The approach could provide a solution to the challenge of multiplexed fragment-based analyte measurements. Other types of configured formats, such as plasmon resonance-based affinity mass spectrometry, may also be successful in translating mass spectrometry into clinic applications.

## ***Biomarkers in the Urinary Proteome***

Plasma membrane proteins are likely present in urine by secretion in exosomes. Urine is a desirable material for the diagnosis and classification of diseases due to the convenience of collection in large amounts. However, the urinary proteome catalogs currently being generated have limitations in their depth and confidence of identification. Methods involving a linear ion trap – Fourier transform (LTQ-FT) and a linear ion trap – orbitrap (LTQ-Orbitrap) MS have been developed for the in-depth characterization of body fluids and applied to the analysis of the human urinary proteome (Adachi et al. 2006). More than 1,500 proteins were identified in the urine obtained from 10 healthy donors, and nearly half of these were membrane proteins. This analysis provides a high confidence set of proteins present in human urinary proteome and provides a useful reference for comparing data sets obtained with different methods. The urinary proteome is unexpectedly complex and may prove useful in biomarker discovery in the future.

## ***Peptides in Body Fluids and Tissues as Biomarkers of Disease***

Importance of peptides in physiology and pathophysiology is being increasingly recognized in recent years. Bodily fluids contain a vast array of low-molecular-weight (LMW) peptides generally produced from larger precursor proteins. The low molecular weight region of the serum peptidome contains protein fragments derived from two sources: (1) high-abundance endogenous circulating proteins and (2) cell and tissue proteins. MS-based profiling has indicated that the peptidome may reflect biological events and contain diagnostic biomarkers. Recent studies have established distinctive serum polypeptide patterns through MS that reportedly correlate with clinically relevant outcomes. Wider acceptance of these signatures as valid biomarkers for disease may follow sequence characterization of the components and elucidation of the mechanisms by which they are generated.

Peptidomics technologies provide new opportunities for the detection of low-molecular-weight proteome biomarkers (peptides) in body fluids. Improvements in peptidomics research are based on separation of peptides and/or proteins by their physicochemical properties in combination with mass spectrometric detection, identification, and sophisticated bioinformatics tools for data analysis (Schulte et al. 2005). These provide an opportunity to discover novel biomarkers for diagnosis and management of disease. An example of the application of Peptidomics<sup>®</sup> Technologies (BioVision AG) is the study of plasma samples of diabetic patients for biomarkers before and after oral glucose challenge and validation by immunoassays. Peptidomics analysis enabled display of >1,500 circulating peptides from 1 mL EDTA plasma and excellent sensitivity ranges reaching ~100 pmol/L (pg/mL). Quantitative changes were detected in the picomolar range. Known diagnostic markers were found in a differential peptide display and validated by ELISA measurement. Unknown, putative biomarkers could have been discovered and inter-individual differences were demonstrated.

## **Analysis of Peptides in Bodily Fluids**

Current methods for detecting enzyme abundance and activity are limited because (1) they lack the requisite sensitivity to detect the low-level activities that are typically present in bodily fluids and (2) they rely on prior knowledge of the identity of the enzymes present in the sample. Thus, new methods are being explored for better detection of enzymes that would lead to improved diagnostic and/or prognostic assays for various diseases.

Researchers at University of California (San Francisco, CA) have developed a novel method to characterize enzyme activity that can be used to diagnose diseases or monitor stability of biological samples. The method involves MS analysis of alterations in the endogenous peptide repertoire and/or custom exogenous probes. Advantages of this method are

- Inexpensive and rapid proteinase activity assay.
- Direct tool to differentiate between post-translational and transcriptional regulation of peptide production.
- Comprehensive, non-biased analyses that do not require prior knowledge of proteinase identity.
- Improved sensitivity with amplified detection of enzymatic activity pointing to the identity of the enzyme species in situation where the abundance of these molecules is below the detection limit of other methods.
- These biomarkers can form the basis of diagnostic test for human diseases that have proteolytic components (e.g., autoimmune diseases, cancer).

## **Serum Peptidome Patterns**

The peptidome information is archived in at least three dimensions: (1) the identity of the peptide (e.g., the peptide sequence or parent protein from which it was derived); (2) the quantity of the peptide itself; and (3) the state of the modified form (fragment size and cleavage ends, post-translational glycosylation sites, etc.). Peptide fragments are developed by the disease system and embody an integrated record of the system. Thus, taking a systems biology approach to measure panels of peptidome biomarkers can potentially overcome the failures of previous biomarkers to achieve adequate clinical sensitivity and specificity (Petricoin and Liotta 2006).

A study using a highly optimized peptide extraction and MALDI-TOF MS-based approach has shown that a limited subset of serum peptides (a signature) can provide accurate class discrimination between patients with three types of solid tumors and controls without cancer (Villanueva et al. 2006). This small but robust set of biomarker peptides has enabled a highly accurate class prediction for an external validation set of prostate cancer samples. This study provides a direct link between peptide marker profiles of disease and differential protease activity, and the patterns described may have clinical utility as surrogate markers for detection and classification of cancer. These findings also have important implications for future peptide biomarker discovery efforts.

## **SISCAPA Method for Quantitating Proteins and Peptides in Plasma**

Stable isotope standards and capture by anti-peptide antibodies (SISCAPA) is a proteomics platform aimed at biomarker quantification (Anderson et al. 2004). The SISCAPA technique was developed to measure protein levels in human plasma and is based on immunoaffinity enrichment of peptide surrogates of biomarker proteins and their identification by MS. Anti-peptide antibodies immobilized on 100 nL nanoaffinity columns are used to enrich specific peptides along with spiked stable-isotope-labeled internal standards of the same sequence. Upon elution from the anti-peptide antibody supports, electrospray MS is used to quantitate the peptides (natural and labeled). SISCAPA is thus limited to sequence-defined (predetermined) analytes, but offers the possibility of greatly increased sensitivity (by removing unwanted peptides from the set delivered to the MS). This technique can be used to develop proteomics-based assays for diagnosing Alzheimer disease.

## ***Verification for Interlaboratory Reproducibility of Protein Biomarkers***

Discovery proteomics often results in a list of tens or even hundreds of potential biomarkers, but because analyzing each biomarker can take up to several weeks and the sample numbers are low, the false discovery rates tend to run high. This is not necessarily because of technical variability but rather is the consequence of biological variability in the samples.

Verification of candidate biomarkers relies upon specific, quantitative assays optimized for selective detection of target proteins and is increasingly viewed as a critical step in the discovery pipeline that bridges unbiased biomarker discovery to preclinical validation. Such technology has been limited by the availability of well-characterized antibodies, a well-known problem in protein research. Developing high-quality immunoassays also is costly and time consuming. Although individual laboratories have demonstrated that multiple reaction monitoring (MRM) coupled with isotope dilution mass spectrometry can quantify candidate protein biomarkers in plasma, reproducibility, and transferability of these assays between laboratories have not been demonstrated.

A multilaboratory study has assess reproducibility, recovery, linear dynamic range, and limits of detection and quantification of multiplexed, MRM-based assays, conducted by NCI-CPTAC (Addona et al. 2009). Using common materials and standardized protocols, it was shown that these assays can be highly reproducible within and across laboratories and instrument platforms and are sensitive to low microgram per milliliter protein concentrations in unfractionated plasma. The study has provided data and benchmarks against which individual laboratories can compare their performance and evaluate new technologies for biomarker verification in plasma. Such methods, combined with protein- and peptide-enrichment strategies, are able to hit target values for limits of quantitation that are in the very bottom of

the nanogram per milliliter range for proteins in blood, where many biomarkers of clinical utility are located.

### ***Significance of Similar Protein Biomarkers in Different Tissues***

The specificity of proteins identified by proteomics as biomarkers for defined conditions or as components of biological processes and pathways is crucial. Usually 2DGE analysis of proteins expressed in two different but related samples, such as healthy and diseased organs leads to the assumption that those proteins expressed at different levels in the two organs are linked to the disease in some way and can be used as biomarkers for it. But this simple idea could well be wrong. Protein biomarkers discovered by one research group are often not confirmed by other groups. This may be due to laboratory errors but another explanation is that many of the differently expressed proteins are not actually specific for the disease or condition being investigated. Even if the researchers are correctly identifying differently expressed proteins in the two samples, the difference may have nothing to do with disease.

A study reported that reading of several 2DGE-based articles featuring lists of differentially expressed proteins reveals that the same proteins seem to predominate regardless of the experiment, tissue, or species (Petрак et al. 2008). The most frequently identified protein was a highly abundant glycolytic enzyme enolase 1, differentially expressed in nearly every third experiment on both human and rodent tissues. Heat-shock protein 27 (HSP27) and heat-shock protein 60 (HSP60) were differentially expressed in about 30% of human and rodent samples, respectively. Considering protein families as units, keratins and peroxiredoxins are the most frequently identified molecules, with at least one member of the group being differentially expressed in about 40% of all experiments. The authors wondered if these commonly observed changes represent common cellular stress responses or are a reflection of the technical limitations of 2DGE.

In another study, differentially expressed proteins from comparative proteomic studies identified by 2DGE followed by MS, especially with MALDI technique, were critically reviewed (Wang et al. 2009a). Based on 66 of those studies, a list of 44 proteins was presented as generally detected proteins regardless of species, in vivo or in vitro conditions, tissues and organs, and experimental objective. Similarly, a list of 28 generally detected protein families was presented. The enriched functions linked to these generally detected proteins reveal that there are some common biological features beyond the technical limitations. Cellular stress response can be the universal reason as to why these proteins are generally expressed differentially. Using those proteins as biomarkers for cellular processes other than stress response should be done with caution. Such disease biomarkers would be merely pinpointing stressed out cells with no useful application. In future proteomic studies more profound approaches should be applied to look beyond these proteins to find specific biomarkers.

## Glycomic Technologies

Mass spectrometry, in combination with modern separation methods, is one of the most powerful and versatile techniques for the structural analysis of glycans (oligosaccharides). NMR spectroscopy and computer graphic modeling can be used to display the dynamic structure of protein glycosylation. These methods can be used for intermolecular and intramolecular protein–oligosaccharide interactions and complement X-ray crystallography.

Recent advances in high-throughput glycomic techniques, glyco biomarker profiling, glyco-bioinformatics, and development of increasingly sophisticated glyco-arrays combined with an increased understanding of the molecular details of glycosylation have facilitated the linkage between aberrant glycosylation and human diseases, and highlighted the possibility of using glyco-biomarkers as potential determinants of disease and its progression (Alavi and Axford 2008).

## Metabolomic Technologies

Within the last few years, metabolomics has developed into a technology that complements proteomics and transcriptomics. In combination with techniques for functional analysis of genes, it is hoped that a holistic picture of metabolism can be formed. In addition to the genome analysis and proteome analyses, the exhaustive analysis of metabolites is important for a comprehensive understanding of cellular functions because the dynamic behavior of metabolites cannot be predicted without information regarding metabolome.

In view of the chemical and physical diversity of small biological molecules, the challenge remains of developing protocols to gather the whole “metabolome.” No single technique is suitable for the analysis of different types of molecules, which is why a mixture of techniques has to be used. In the field of metabolomics, the general estimations of the size and the dynamic range of a species-specific metabolome are at a preliminary stage. Metabolic fingerprinting and metabonomics with high sample throughput but decreased dynamic range and the deconvolution of individual components achieve a global view of the in vivo dynamics of metabolic networks. The technologies used include NMR, direct infusion mass spectrometry, and/or infrared spectroscopy. Gas chromatography (GC)-MS and LC-MS technology achieve a lower sample throughput but provide unassailable identification and quantitation of individual compounds in a complex samples.

However, it is important to note that each type of technology exhibits a bias toward certain compound classes, mostly due to ionization techniques, chromatography, and detector capabilities. GC-MS has evolved as an imperative technology for metabolomics due to its comprehensiveness and sensitivity. The coupling of GC to time-of-flight (TOF) mass analyzers is an emerging technology. High scan rates provide accurate peak deconvolution of complex samples. GC-TOF-MS capabilities provide an improvement over conventional GC-MS analysis in

the analysis of ultracomplex samples, which is particularly important for the metabolomics approach. Ultracomplex samples contain hundreds of co-eluting compounds that vary in abundance by several orders of magnitude. Thus, accurate mass spectral deconvolution and a broad linear dynamic range represent indispensable prerequisites for high-quality spectra and peak shapes. Modern GC-TOF-MS applications and incorporated mass spectral deconvolution algorithms fulfill these requirements.

The advantages of metabolomics technologies are

- Ability to analyze all bodily fluids such as blood, CSF, and urine as well as cultured or isolated cells and biopsy material.
- High-throughput capability enabling simultaneous monitoring of biological samples.
- Analysis of multiple pathways and arrays of metabolites simultaneously from microliter sample quantities.

### ***Mass Spectrometry-Based Kits for Discovery of Metabolic Biomarkers in Plasma***

MS-based kits are being developed for discovery of metabolic biomarkers. The best known of these is AbsoluteIDQ™ Kit (Biocrates Life Sciences), which works with Applied Biosystems API 4000 and API 4000 QTRAP® Mass Spectrometer. It enables accurate identification and quantification of more than 160 metabolites in over four compound classes from clinical and preclinical plasma samples within a few minutes. This is the first integrated technology platform (metabolomics, genomics, and proteomics), which automates proprietary standardized pre-analytical and analytical steps for metabolic analysis using bioinformatics. It provides a sample base with access to clinical data for discovery, validation, and evaluation of new biomarkers and drug targets.

### ***Urinary Profiling by Capillary Electrophoresis***

Metabolomic approaches have become particularly important for discovery of biomarkers in urine. The analytical technology for urinary profiling must be efficient, sensitive, and offer high resolution. Until recently these demands were commonly met by HPLC-MS, GC-MS, and NMR. The analytical armory for urinary profiling has now been extended to include cyclodextrin-modified micellar electrokinetic capillary chromatography (CD-MECC), which enables highly cost-effective, rapid, and efficient profiling with minimal sample volume and preparation requirements. The CD-MECC profiles typically show separation for over 80 urinary metabolites. These profiles have been visualized using novel advanced pattern recognition tools.

## ***Lipid Profiling***

Modern medicine has come to rely on a small suite of single biomarkers, such as plasma cholesterol or triglycerides, to assess the risk of certain diseases. However, such single-biomarker assessments overlook the inherent complexity of metabolic disorders involving hundreds of biochemical processes. Assessing the full breadth of lipid metabolism is what drives the field of lipomic profiling. However, unlike the other “-omic” technologies, in which only a small portion of the genes or proteins is known, lipid metabolic pathways are well characterized. Another limitation of “-omics” technologies is that they produce so many false-positive results that it is difficult to be sure that findings are valid. Metabolomics is not immune to this problem but, when practiced effectively, the technology can reliably produce knowledge to aid in decision making. Focused metabolomics platforms, which restrict their target analytes to those measured well by the technology, can produce data with properties that maximize sensitivity and minimize the false-discovery problem. The most developed focused metabolomics area is lipid profiling. TrueMass<sup>®</sup> (Lipomic Technologies) analysis produces lipomic profiles – comprehensive and quantitative lipid metabolite profiles of biological samples. With TrueMass, lipomics measures hundreds of lipid metabolites from each small quantity of tissue, plasma, or serum sample. Because the resulting data are quantitative, TrueMass data can be seamlessly integrated with pre-existing or future databases.

Data-dependent acquisition of MS/MS spectra from lipid precursors enables emulation of simultaneous acquisition of an unlimited number of precursor and neutral loss scans in a single analysis (Schwudke et al. 2006). This approach takes full advantage of rich fragment patterns in tandem mass spectra of lipids and enables their profiling by complex scans, in which masses of several fragment ions are considered within a single logical framework. No separation of lipids is required, and the accuracy of identification and quantification is not compromised, compared to conventional precursor and neutral loss scanning.

## ***Role of Metabolomics in Biomarker Identification and Pattern Recognition***

Metabolomics research has increased significantly over recent years due to advances in analytical measurement technology and the advances in pattern recognition software enabling one to visualize changes in levels of hundreds or even thousands of chemicals simultaneously. Multivariate metabolomic and proteomic data and time-series measurements can be combined to reveal protein–metabolite correlations. Different methods of multivariate statistical analysis can be explored for the interpretation of these data. The discrimination of the samples enables the identification of novel components. These components are interpretable as inherent biological characteristics.

Biomarkers that are responsible for these different biological characteristics can easily be classified because of the optimized separation using independent components analysis and an integrated metabolite–protein data set. Evidently, this kind of analysis depends strongly on the comprehensiveness and accuracy of the profiling method, in this case metabolite and protein detection. Assuming that the techniques will improve, more proteins and metabolites can be identified and accurately quantified, the integrated analysis will have great promise.

### ***Validation of Biomarkers in Large-Scale Human Metabolomics Studies***

A strategy for data processing and biomarker validation has been described in a large metabolomics study that was performed on 600 plasma samples taken at four-time points before and after a single intake of a high fat test meal by obese and lean subjects (Bijlsma et al. 2006). All samples were analyzed by a LC-MS lipidomic method for metabolic profiling. Such metabolomics studies require a careful analytical and statistical protocol. A method combining several well-established statistical methods was developed for processing this large data set in order to detect small differences in metabolic profiles in combination with a large biological variation. The strategy included data preprocessing, data analysis, and validation of statistical models. After several data preprocessing steps, partial least-squares discriminate analysis (PLS-DA) was used for finding biomarkers. To validate the found biomarkers statistically, the PLS-DA models were validated by means of a permutation test, biomarker models, and non-informative models. Univariate plots of potential biomarkers were used to obtain insight in up- or downregulation.

### **Lipidomics**

Lipidomics is the use of high-dimensional lipid analysis technologies that provide researchers with an opportunity to measure lipids on an unprecedented scale. The development of lipidomics has been accelerated by the concepts of systems biology and advances in the following areas (Han 2007):

- Proteomics, particularly ESI/MS. Shotgun lipidomics, based on multidimensional MS array analyses after multiplexed sample preparation and intrasource separation, has matured as a technique for the rapid and reproducible global analysis of cellular lipids.
- Recognition of the role of lipids in many common metabolic diseases.
- Recognition that metabolism of lipids is linked to and requires the simultaneous analysis of multiple lipid classes, molecular species, and metabolic networks.

At its current stage, this technology enables us to analyze more than 20 lipid classes and thousands of individual lipid molecular species directly from lipid extracts of biologic samples (Gross and Han 2007). Substantial progress has been made in the application of lipidomics in biomarker discovery. Notable are the applications in the following areas:

- Substantial alterations have been found in myocardial cardiolipin content in early stages of diabetes.
- Obesity has been related to increase in the content of lysophosphatidylcholine molecular species independently of genetic influences and is related to insulin resistance.
- Neurolipidomics is important for discovery of biomarkers of diseases of the nervous system as lipids make up half of the human brain in dry weight.

Lipidomics is moving toward more methodical and structured approaches to biomarker identification. Experimental designs focusing on well-defined outcomes have a better chance of producing biologically relevant results and strategies for improving the quality of data analysis have been reviewed (Wiest and Watkins 2007).

## Fluorescent Indicators for Biomarkers

In earlier years scientists tagged samples – whether nucleic acid, protein, cell, or tissue – with radioactive labels, and captured images on film. Safety concerns, convenience, and sensitivity, spurred the development of alternative techniques, and today, researchers can choose from a range of options, including fluorescence. Fluorescence occurs when light is absorbed from an external (excitation) source by a fluorescent molecule (fluorophore) and subsequently emitted. The cycle of excitation and emission will continue until the excitation source is turned off or the fluorophore is consumed in a chemical reaction. A fluorescent probe is any small molecule that undergoes changes in one or more of its fluorescent properties as a result of non-covalent interaction with a protein or other macromolecular structure. There are many different fluorescent molecules, both naturally occurring and synthetic, each molecule having distinctive spectroscopic properties. This variety of molecules can be exploited to enable highly localized, sensitive detection of chemical reactions, for tagging and identification of molecular structures, for monitoring of cellular states, and even for monitoring multiple properties simultaneously. Due to their higher sensitivity than chromogenic dyes, the fluorescent probes can be used to effectively signal the presence of minute amounts of a specific protein, DNA, or RNA without the hazards associated with radioactive labels.

Simple water-soluble lanthanum and europium complexes are effective at detecting neutral sugars as well as glycolipids and phospholipids. In solutions at physiologically relevant pH the fluorescent lanthanum complex binds neutral sugars with apparent binding constants comparable to those of arylboronic acids. Interference

from commonly occurring anions is minimal. The europium complex detects sialic acid-containing gangliosides at pH 7.0 over an asialoganglioside. This selectivity is attributed, in large part, to the cooperative complexation of the oligosaccharide and sialic acid residues to the metal center, based on analogous prior studies. Lysophosphatidic acid (LPA), a biomarker for several pathological conditions including ovarian cancer, is selectively detected by the europium complex (Alpturk et al. 2006). LPA is also detected via a fluorescence increase in human plasma samples. The 2-sn-OH moiety of LPA plays a key role in promoting binding to the metal center. Other molecules found in common brain ganglioside and phospholipid extracts do not interfere in the ganglioside or LPA fluorescence assays.

## **Molecular Imaging Technologies**

Imaging has a long history in medicine. Several technologies are used for imaging *in vivo*. X-rays are the oldest form of *in vivo* imaging and form the basis of computer tomography imaging. The current focus in relation to imaging biomarkers is on molecular imaging, which is defined as the visualization, characterization, and measurement of biological processes at the molecular and cellular levels in humans and other living systems. Molecular imaging typically includes 2D or 3D imaging as well as quantification over time. The techniques used include radiotracer imaging/nuclear medicine, MRI/MRS, optical imaging, ultrasound, and others.

### ***Computer Tomography***

In computer tomography (CT) X-rays are emitted from a source and received by a set of detectors on the opposing side of the object. The source and detectors revolve around the patient continuously while the patient lies on a robotically controlled gantry. The resulting spiral CT scan is processed through interpolation and back-projection algorithms to form a series of cross-sectional images which, taken together, form a 3D data set representing the object of interest.

CT offers high resolution and fast acquisition of data and is suited for studies that require precise characterizations of bony structures, such as the spine, or for cases in which image acquisition duration is of crucial importance, such as trauma. High-resolution CT is also preferred for characterizing lesions and other abnormalities within the chest and vascular structures, as well as many applications within the gastrointestinal and genitourinary regions. In the context of drug development, CT is a valuable tool for quantifying lesion size and location, particularly within the chest and abdominal regions. The image resolution of CT is  $>0.5$  mm. More recently, by tracking the appearance of contrast-enhancing dyes as they wash into an organ, CT perfusion has gained greater applicability and utility.

## ***Magnetic Resonance Imaging***

In contrast to the use of ionizing radiation in CT, magnetic resonance imaging (MRI) uses radio-frequency pulses and magnetic fields to obtain signals from changes in nuclear magnetic moments. Specifically, as the alignment and relaxation of protons occur in response to pulsed radio-frequencies, characteristic relaxation times can be measured, most notably T1 (the longitudinal relaxation time) and T2 (the transverse relaxation time). Whereas CT images are characterized by a single parameter, namely the X-ray attenuation of the tissue along the propagation path, MRI is characterized by far more parameters, including proton density, T1, T2, flow, diffusion and susceptibility, among others. It is this flexibility that makes the use of MRI a double-edged sword. MRI is useful in numerous applications, but repeatability and standardization in multicenter clinical trials can be challenging. Although lower in resolution and requiring more time for data acquisition than that by CT, MRI offers superior soft tissue contrast, making MRI the modality of choice in the brain, in addition to specific applications in musculoskeletal and gastrointestinal systems. MRI, with and without contrast agents, is also used for numerous functional assessments, including tissue perfusion, diffusion, tumor permeability, and blood oxygenation level-dependent (BOLD) fMRI studies. A technique based on the same principle as MRI, but providing a greater degree of molecular characterization is magnetic resonance spectroscopy (MRS), where spectroscopic profiles of the chemical constituents within a sample are obtained. Real-time molecular imaging with MRI is possible now as demonstrated by imaging of the distribution of pyruvate and mapping of its major metabolites lactate and alanine within a time frame of  $\sim 10$  s (Golman et al. 2006).

## ***Positron Emission Tomography***

Radionuclide imaging uses bi-functional agents containing a radiolabel that confers detectability, and a chemical and/or pharmaceutical moiety that determines uptake and distribution in the body. In the case of positron emission tomography (PET), the emitted positron passes through tissue and is ultimately annihilated when combined with an electron, resulting in two 511 keV photons emitted in opposite directions. Detectors are arranged in a ring around the tissue of interest, and only triggering events that arrive near-simultaneously at diametrically opposite detectors are recorded. Tomographic methods are then used to produce the resulting PET images.

Numerous radioisotopes are used for nuclear imaging. These tracer isotopes can be substituted into drug compounds directly to mimic naturally occurring compounds or can be complexed with other molecules to form new compounds referred to as radiopharmaceuticals. 2-18F-fluoro-2-deoxy-D-glucose (FDG), for example, is an analogue of glucose labeled with a positron-emitting form of fluorine and is used in PET imaging of metabolic activities that involve glucose uptake.

PET provides a non-invasive view into a person's living biology as it tracks a range of biological processes from metabolism to receptors, gene expression, and

drug activity. This imaging tool examines the chemistry and biology of a person's body by monitoring ingested tracer molecules, and it is used to study the metabolism of the brain, the heart, and cancer.

### ***Advantages of Imaging Biomarkers***

Several characteristics of imaging biomarkers set them apart from other biomarkers. The advantages are

- Imaging has been in routine use for diagnosis and disease management for several decades, and the ability to identify a wide spectrum of pathophysiology using imaging methods is well established.
- Imaging biomarkers tend to be much more closely associated with the expressed phenotype of diseases, thus enabling direct associations between therapy and effect.
- Functional imaging provides a dynamic picture of the disease.
- Imaging offers tremendous versatility for providing continuous, structural, and functional assessments of therapy, offering snapshots of the bioactivity of drug compounds overtime.
- Imaging provides therapy assessments in animals and humans alike and is therefore an important tool for promoting translational research.
- Imaging has now entered molecular era with molecular imaging and nanoparticles as contrast media.

### ***Monitoring In Vivo Gene Expression by Molecular Imaging***

Molecular imaging is an emerging field of study that deals with imaging of disease on a cellular and molecular level. It can be considered as an extension of molecular diagnostics. In contradistinction to "classical" diagnostic imaging, it sets forth to probe the molecular abnormalities that are the basis of disease rather than to image the end effects of these molecular alterations. Radionuclide imaging, MRI, and PET can be used visualize gene expression. Work done at the Beckman Institute/California Institute of Technology (Pasadena, CA) deals with 3D MRI image of gene expression based on intracellular messenger concentration.

Several current in vitro assays for protein and gene expression have been translated into the radiologic sciences. Endeavors are under way to image targets ranging from DNA to entire phenotypes in vivo. The merging fields of molecular biology, molecular medicine, and imaging modalities may provide the means to screen active drugs in vivo, image molecular processes, and diagnose disease at a presymptomatic stage. Role of imaging in drug discovery and development is described in Chapter 4.

## ***Molecular Imaging In Vivo as a Biomarker***

Molecular imaging is now recognized as an in vivo biomarker. Some examples of uses of commercially available technologies are as follows:

- Bio-imaging markers (Bio-Imaging Technologies Inc.) include use of all imaging modalities: X-ray, CT, MRI, scintigraphy, SPECT, PET plus assays for specialized molecular biomarkers. These are used as end points in clinical trials for evaluation of drug effects.
- Fluorescent nanoparticles from Kodak, which emit strong signal but do not contain toxic heavy metal atoms, are used for multispectral imaging. Molecular images obtained in this manner are used as biomarkers for drug development in both animals and humans.
- Detection by fluorescence molecular tomography (VisEn Medical Inc.) is used for imaging of cancer, cardiovascular disease, skeletal disorders, and inflammation to detect in vivo biomarkers.
- Radiopharmaceuticals such as gadopentetate dimeglumine (Magnevist<sup>TM</sup>, Bayer Schering Pharma) is used as an injection along with MRI. This is used for detection of biomarkers in neurological disorders including Alzheimer disease.

## ***Challenges and Future Prospects of Molecular Imaging***

### **Basic Research in Molecular Imaging**

Research leads clinical practice, and one of the challenges to the medical and regulatory communities is to facilitate the introduction of new imaging techniques into patient management. As technology advances, scientists will further their ability to use different probes with the whole spectrum of molecular imaging modalities to identify new targets within cells and associated with cell membranes and receptors and to quantify treatment effects on the expression of these biomarkers. Traditional tracer-based nuclear medicine research will be expanded within the molecular imaging arena to include optical imaging and magnetic resonance spectroscopy. These new imaging technologies, particularly associated with new probe development, can provide new contrast to medical imaging. Optical imaging is currently limited by its spatial resolution and imaging depth, and thermoacoustic tomography or radio-frequency-based photoacoustic tomography is being developed to meet this challenge.

### **Imaging Intracellular NADH as a Biomarker of Disease**

Reduced nicotinamide adenine dinucleotide (NADH) is a major electron donor in the oxidative phosphorylation and glycolytic pathways in cells. NADH fuels a series of biochemical reactions that involve various enzymes to produce ATP, the major energy source in cells. In the event of disease or a metabolic disorder, these enzymes and their related reactions can become disabled, causing a buildup of unused NADH.

Intrinsic NADH fluorescence has been employed as a natural probe for a range of cellular processes that include apoptosis, cancer pathology, and enzyme kinetics. Two-photon fluorescence lifetime and polarization imaging of intrinsic NADH in breast cancer and normal cells has been reported for quantitative analysis of the concentration and conformation of this coenzyme (Yu and Heikal 2009). Using a newly developed non-invasive assay, the authors estimated the average NADH concentration in cancer cells to be approximately 1.8-fold higher than in normal breast cells. Excess amounts of intracellular NADH, a naturally fluorescent molecule found in all living cells, could serve as a natural biomarker for cancer. These quantitative studies demonstrate the potential of dynamics (rather than intensity) imaging for probing mitochondrial anomalies associated with neurodegenerative diseases, cancer, diabetes, and aging. This approach is also applicable to other metabolic and signaling pathways in living cells, without the need for cell destruction as in conventional biochemical assays.

### **Devices for Molecular Imaging**

An evolution in imaging technology is occurring and will continue as imaging capabilities continue to expand from the anatomical to the functional and to the molecular. The expansion of imaging capabilities will enable the identification of imaging probes specific for molecular processes, and new multimodality imaging technologies will be developed to appropriately utilize these new probes, focusing on normal and abnormal biological processes. The future will bring nanoparticle delivery vehicles to deliver gene therapy to patients, smart contrast agents, target-specific optical agents, and stem cell-based imaging therapy.

### **Imaging Biomarkers in Clinical Trials**

SNM (<http://www.snm.org/>), an international scientific and medical association dedicated to advancing molecular imaging and therapy, has created the Molecular Imaging Clinical Trials Network as a response to the need for streamlined processes to utilize imaging biomarkers in clinical trials and clinical practice. There is widespread agreement that the use of imaging biomarkers in the drug-development process can significantly reduce this burden and speed the timelines to clinical use. To specifically address this opportunity, SNM has designed a first-of-its-kind model for the use of imaging biomarkers in clinical trials that spans drug development, molecular imaging, radiolabeled probe development, and manufacturing and regulatory issues to integrate the use of investigational imaging biomarkers into multicenter clinical trials.

### ***Molecular Imaging in Clinical Practice***

Work will continue to examine and validate future clinical applications for FDG PET/CT for oncology (diagnosis and staging, treatment planning and response,

detection of recurrent or residual disease, restaging), for myocardial perfusion (coronary artery disease, myocardial viability), and for neurology and neurosurgery (brain tumors, medically intractable epilepsy, stroke, movement disorders, Alzheimer disease, and other dementias). Bioluminescence imaging, which enables visualization of genetic expression and physiological processes at the molecular level in living tissues, can identify specific gene expression in cancer cells and may be used to identify metastatic potential.

## **Nuclear Magnetic Resonance**

High-resolution nuclear magnetic resonance (NMR) spectroscopy is a quantitative technique that can report on hundreds of compounds in a single measurement. The ubiquitous presence of hydrogen in biomolecules as well as a favorable isotope distribution and magnetic sensitivity make  $^1\text{H}$  NMR the obvious choice for generating profiles; however, there are NMR-visible isotopes for most chemical elements, including  $^{13}\text{C}$ ,  $^{31}\text{P}$ , and  $^{15}\text{N}$ . Modern cryogenically cooled probes and capillary probes can push limits of detection by NMR to nanomolar levels or samples volumes as low as 1.5 mL. This enables metabonomics to be applied to a number of volume or mass-limited circumstances, such as the use of microdialysates or regular blood sampling from small animals without sacrificing them.

Sample preparation for routine biofluid analysis is minimal. The use of magic angle spinning NMR (MAS-NMR) enables intact tissues and cells to be examined with little or no preparation and on as little as 20 mg of material. Profiles generated via MAS-NMR can also reveal the effects of toxin treatment on the metabolite composition within different cellular environments either via diffusion and relaxation measurements in intact tissue or by comparing NMR profiles generated by various tissue isolates. The ability to localize biochemical changes to a specific tissue, cell type, or organelle provides valuable information pertaining to the mechanism of toxicity and can aid the interpretation of biofluid analyses. MAS-NMR may provide a means to validate *in vitro* systems in terms of *in vivo* metabolic responses to toxicity.

### ***Chemical Derivatization to Enhance Biomarker Detection by NMR***

A simple chemical reaction has been developed in Purdue University (Lafayette, IN) to improve the ability to detect important molecules in complex fluids such as blood, rendering the biomarkers for some genetically caused metabolic disorders up to 100 times more visible. The chemical derivatization method selects a class of metabolites from a complex mixture and enhances their detection by  $^{13}\text{C}$  NMR (Shanaiah et al. 2007). Acetylation of amines directly in aqueous medium with 1,1'-(13)C(2) acetic anhydride is a simple method that creates a high sensitivity and

quantitative label in complex biofluids with minimal sample pretreatment. Detection using either 1D or 2D  $^{13}\text{C}$  NMR experiments produces highly resolved spectra with improved sensitivity. Experiments to identify and compare amino acids and related metabolites in normal human urine and serum samples as well as in urine from patients with the inborn errors of metabolism tyrosinemia type II, argininosuccinic aciduria, homocystinuria, and phenylketonuria have demonstrated the usefulness of this method. The use of metabolite derivatization and  $^{13}\text{C}$  NMR spectroscopy produces data suitable for metabolite profiling analysis of biofluids on a time scale that allows routine use. Extension of this approach to enhance the NMR detection of other classes of metabolites has also been accomplished. The improved detection of low-concentration metabolites creates opportunities to improve the understanding of the biological processes and develop improved disease detection methods.

### ***Fluxomics by Using NMR***

Fluxomics is measurement of flux rates of biomarkers using NMR with stable isotope precursors. NMR can do non-invasive monitoring of live tissue and repetitive sampling (longitudinal) without sacrificing animals. The technology can be translated directly to humans. NMR and  $^1\text{H}$  spectroscopic imaging (MRSI) can make in vivo fluxomic measurements, whereas LC-MS can be used to identify less abundant metabolites. This approach has been used for human prostate to differentiate between cancer and benign hypertrophy. Advantages of this method are

- Fluxomics is more sensitive than concentration alone.
- Fluxomics completes the system biology equation permitting mechanistic modeling and anchoring proteomic and transcriptomic data.

### **Nanobiotechnology**

Nanotechnology (Greek word nano means dwarf) is the creation and utilization of materials, devices, and systems through the control of matter on the nanometer-length scale, i.e., at the level of atoms, molecules, and supramolecular structures. It is the popular term for the construction and utilization of functional structures with at least one characteristic dimension measured in nanometers – a nanometer is one billionth of a meter ( $10^{-9}$  m). This is roughly four times the diameter of an individual atom and the bond between two individual atoms is 0.15 nm long. Proteins are 1–20 nm in size. The definition of “small,” another term used in relation to nanotechnology, depends on the application, but can range from 1 nm to 1 mm. Nano is not the smallest scale; further down the power of ten are angstrom ( $=0.1$  nm), pico, femto, atto, and zepto. By weight, the mass of a small virus is about 10 attograms. An attogram is one-thousandth of a femtogram, which is one-thousandth of a picogram, which is one-thousandth of a nanogram. Applications of nanobiotechnology are described in a special report on this topic (Jain 2010c).

## *Nanomaterials for Biolabeling*

Nanomaterials are suitable for biolabeling. Nanoparticles usually form the core in nanobiomaterials. However, in order to interact with biological target, a biological or molecular coating or layer acting as an interface needs to be attached to the nanoparticle. Coatings that make the nanoparticles biocompatible include antibodies, biopolymers, or monolayers of small molecules. A nanobiomaterial may be in the form of nanovesicle surrounded by a membrane or a layer. The shape is more often spherical but cylindrical, plate-like, and other shapes are possible. The size and size distribution might be important in some cases, for example, if penetration through a pore structure of a cellular membrane is required. The size is critical when quantum-sized effects are used to control material properties. A tight control of the average particle size and a narrow distribution of sizes allow creating very efficient fluorescent probes that emit narrow light in a very wide range of wavelengths. This helps with creating biomarkers with many and well-distinguished colors. The core itself might have several layers and be multifunctional. For example, combining magnetic and luminescent layers one can both detect and manipulate the particles.

The core particle is often protected by several monolayers of inert material, for example, silica. Organic molecules that are adsorbed or chemisorbed on the surface of the particle are also used for this purpose. The same layer might act as a biocompatible material. However, more often an additional layer of linker molecules is required that has reactive groups at both ends. One group is aimed at attaching the linker to the nanoparticle surface and the other is used to bind various biocompatible substances such as antibodies depending on the function required by the application.

Water-soluble, biocompatible, fluorescent, and stable silver/dendrimer nanocomposites have been synthesized that exhibit a potential for labeling cells *in vitro* as cell biomarkers.

Efforts to improve the performance of immunoassays and immunosensors by incorporating different kinds of nanostructures have gained considerable momentum over the last decade. Most of the studies focus on artificial, particulate marker systems, both organic and inorganic. Inorganic nanoparticle labels based on noble metals, semiconductor quantum dots and nanoshells appear to be the most versatile systems for these bioanalytical applications of nanophotonics. The underlying detection procedures are more commonly based on optical techniques. These include nanoparticle applications generating signals as diverse as static and time-resolved luminescence, one- and two-photon absorption, Raman and Rayleigh scattering as well as surface plasmon resonance, and others. In general, all efforts are aimed to achieve one or more of the following goals:

- Lowering of detection limits (if possible, down to single-molecule level).
- Parallel integration of multiple signals (multiplexing).
- Signal amplification by several orders of magnitude.
- Prevention of photobleaching effects with concomitant maintenance of antigen-binding specificity and sensitivity.

## Quantum Dot Molecular Labels

Quantum dots (QDs) are crystalline semiconductors composed of a few hundred or thousand atoms that emit different colors of light when illuminated by a laser. Stable QDs are made from cadmium selenide and zinc sulfide. Because these probes are stable, they have the ability to remain in a cell's cytoplasm and nucleus without fading out much longer than conventional fluorescent labels. This could give biologists a clear view of processes that span several hours or even days, such as DNA replication, genomic alterations, and cell cycle control. Their longevity has also made QDs a molecular label, allowing scientists to study the earliest signs of cancer, track the effectiveness of pharmaceuticals that target the cellular underpinnings of disease, and understand the events that occur during stem cell differentiation.

One drawback to this approach, however, is that these QDs may release potentially toxic cadmium and zinc ions into cells. To solve this problem, QDs are coated with a protective layer of polyethylene glycol (PEG), which is a very non-reactive and stable compound that is used extensively by the pharmaceutical industry in drug formulation. This layer is designed to prevent the dots from leaking heavy metal ions into cells once they are inside. The tool used to test the safety of QDs is a gene chip packed with 18,400 probes of known human genes and it is a comprehensive method to measure the toxicity of nanoscale particles. This chip is designed to enable the researchers to expose the human genome to QDs and determine the extent to which the compound forces the genes to express themselves abnormally.

Use of a high-throughput gene expression test has helped to determine that specially coated QDs fluorescent nanoprobe affect only 0.2% of the human genome, dispelling the concern that the mere presence of these potentially toxic sentinels disrupts a cell's function (Zhang et al. 2006). The number of genes affected is very small given the large dose of QDs used in the study, which is up to 1,000 times greater than the dose that would typically be used in human applications. Moreover, the affected genes are not related to heavy metal exposure, which would be the case if the cells had been exposed to cadmium or zinc ions. Because of their protective coating, QDs have minimal impact on cells; the only gene changes are in transporter proteins, which is expected because the dots have to be transported into and within the cell.

PEG-coated QDs will be used for *in vivo* imaging of breast and prostate cancer. The cancers could be detected at very early stages and the molecular makeup can be characterized for effective treatment.

## Bioconjugated QDs for Multiplexed Profiling of Biomarkers

Bioconjugated QDs, collections of different sized nanoparticles embedded in tiny polymer beads, provide a new class of biological labels for evaluating biomarkers on intact cells and tissue specimens. In particular, the use of multicolor QD probes in immunohistochemistry is considered one of the most important and clinically relevant applications. The medical use of QD-based immunohistochemistry has been limited by the lack of specific instructions, protocols, for clinicians. Preliminary

results and detailed protocols for QD-antibody conjugation, tissue specimen preparation, multicolor QD staining, image processing, and biomarker quantification have been published (Xing et al. 2007). The results demonstrate that bioconjugated QDs can be used for multiplexed profiling of biomarkers, and ultimately for correlation with disease progression and response to therapy. This will increase the clinician's ability to predict the likely outcomes of drug therapy in a personalized approach to disease management. Bioinformatics and systems biology is used to link each individual's molecule profile with disease diagnosis and treatment decisions. The usefulness of these protocols was demonstrated by simultaneously identifying multiple biomarkers in prostate cancer tissue. In general, QD bioconjugation is completed within 1 day, and multiplexed molecular profiling takes 1–3 days depending on the number of biomarkers and QD probes used.

### **Magnetic Nanotags for Multiplex Detection of Biomarkers**

Magnetic nanotags (MNTs) are a promising alternative to fluorescent labels in biomolecular detection assays, because minute quantities of MNTs can be detected with inexpensive sensors. Probe sensors are functionalized with capture antibodies specific to the chosen analyte. During analyte incubation, the probe sensors capture a fraction of the analyte molecules. A biotinylated linker antibody is subsequently incubated and binds to the captured analyte, providing binding sites for the streptavidin-coated MNTs, which are incubated further. The nanotag-binding signal, which saturates at an analyte concentration-dependent level, is used to quantify the analyte concentration. However, translation of this technique into easy-to-use and multiplexed protein assays, which are highly sought after in molecular diagnostics such as cancer diagnosis and personalized medicine, has been challenging. Multiplex protein detection of potential cancer biomarkers has been demonstrated at subpicomolar concentration levels (Osterfeld et al. 2008). With the addition of nanotag amplification, the analytic sensitivity extends into the low femtomolar concentration range. The multianalyte ability, sensitivity, scalability, and ease of use of the MNT-based protein assay technology make it a strong contender for versatile and portable molecular diagnostics in both research and clinical settings.

### ***Nanoproteomics and Biomarkers***

Nanoproteomics – application of nanobiotechnology to proteomics – improves on most current protocols including protein purification/display and automated identification schemes that yield unacceptably low recoveries with reduced sensitivity and speed while requiring more starting material. Low abundant proteins and proteins that can only be isolated from limited source material (e.g., biopsies) can be subjected to nanoscale protein analysis – nanocapture of specific proteins and complexes, and optimization of all subsequent sample handling steps leading to mass analysis of peptide fragments. This is a focused approach, also termed targeted proteomics, and involves examination of subsets of the proteome, e.g., those proteins

that are either specifically modified, or bind to a particular DNA sequence, or exist as members of higher order complexes, or any combination thereof. This approach is used to identify how genetic determinants of cancer alter cellular physiology and response to agonists.

### **High-Field Asymmetric Waveform Ion Mobility Mass Spectrometry**

An ion mobility technology – high-field asymmetric waveform ion mobility mass spectrometry (FAIMS) – has been introduced as online ion selection methods compatible with electrospray ionization (ESI). FAIMS uses ion separation to improve detection limits of peptide ions when used in conjunction with electrospray and nanoelectrospray MS. This facilitates the identification of low-abundance peptide ions often present at ppm levels in complex proteolytic digests and expand the sensitivity and selectivity of nanoLC–MS analyses in global and targeted proteomics approaches. This functionality likely will play an important role in drug discovery and biomarker programs for monitoring of disease progression and drug efficacy.

### ***Nanoparticles for Molecular Imaging***

Although developments in molecular imaging have been dominated by nuclear medicine agents in the past, the advent of nanotechnology led to MRI molecular agents that enable detection of sparse biomarkers with a high-resolution imaging. A wide variety of nanoparticulate MRI contrast agents are available, most of which are superparamagnetic iron oxide-based constructs. Perfluorocarbon (PFC) nanoparticulate platform is not only effective as a T1-weighted agent but also supports  $^{19}\text{F}$  magnetic resonance spectroscopy and imaging. The unique capability of  $^{19}\text{F}$  permits confirmation and segmentation of MRI images as well as direct quantification of nanoparticle concentrations within a voxel.

Incubation of alpha(nu)beta(3)-expressing cells with targeted nanoparticles has been shown to significantly inhibit binding to a vitronectin-coated surface, confirming the bioactivity of the targeted nanoparticles (Schmieder et al. 2005). This study lowers the limit previously reported for detecting sparse biomarkers with molecular MRI in vivo. This technique may be employed to non-invasively detect very small regions of angiogenesis associated with nascent melanoma tumors and to phenotype and early-stage melanoma in a clinical setting.

Ultrasmall superparamagnetic iron oxide (USPIO) is a cell-specific contrast agent for MRI. An open-label phase II study has tested the potential of USPIO-enhanced MRI for macrophage imaging in human cerebral ischemic lesions (Saleh et al. 2004). USPIO-induced signal alterations throughout differed from signatures of conventional gadolinium-enhanced MRI, thus being independent from breakdown of the blood–brain barrier (BBB). USPIO-enhanced MRI may provide an in vivo surrogate marker of cellular inflammation in stroke and other CNS pathologies. USPIO has favorable properties that result from its intravascular retention and

lack of extravasation, allowing optimal contrast between the vessel and the adjacent tissue for several minutes postinjection.

### ***Nanoparticles for Discovering Biomarkers***

Most of the applications of nanoparticles have focused on imaging systems and drug delivery vectors. The physicochemical characteristics and high surface areas of nanoparticles also make them ideal candidates for developing biomarker harvesting platforms. Given the variety of nanoparticle technologies that are available, it is feasible to tailor nanoparticle surfaces to selectively bind a subset of biomarkers and sequestering them for later study using high sensitivity proteomic tests (Geho et al. 2006). Biomarker harvesting is an underutilized application of nanoparticle technology and is likely undergo substantial growth.

Functional polymer-coated nanoparticles can be used for quick detection of biomarkers and DNA separation. Multiarray of hepatocyte spheroids on a micro-fabricated polymer-brush surface can maintain the hepatocyte viability and liver-specific functions, offering a new high-throughput screening method of pharmacologically and toxicologically active compounds for drug discovery (Nishiyama and Kataoka 2006).

### ***Nucleoprotein Nanodevices for Detection of Cancer Biomarkers***

DNA Y-junctions have been used as fluorescent scaffolds for EcoRII methyltransferase-thioredoxin (M•EcoRII-Trx) fusion proteins and covalent links were formed between the DNA scaffold and the methyltransferase at preselected sites on the scaffold containing 5FdC (Singer and Smith 2006). The resulting thioredoxin-targeted nanodevice was found to bind selectively to certain cell lines but not to others. The fusion protein was constructed so as to permit proteolytic cleavage of the thioredoxin peptide from the nanodevice. Proteolysis with thrombin or enterokinase effectively removed the thioredoxin peptide from the nanodevice and extinguished cell line-specific binding measured by fluorescence. Potential applications for devices of this type include the ability of the fused protein to selectively target the nanodevice to certain tumor cell lines and not others suggesting that this approach can be used to probe cell surface receptors as biomarkers of cancer and may serve as an adjunct to immunohistochemical methods in tumor classification.

### ***Future Prospects of Application of Nanobiotechnology for Biomarkers***

Nanobiotechnology is progressing rapidly and the impact will be felt on discovery of biomarkers. Nanotechnology offers the possibility to create devices which

can screen for disease biomarkers at very fast rates. Tools will be developed by identifying biomarkers for particular diseases that can then lead to diagnostic tests.

Scientists at the California Institute of Technology (Pasadena, CA) are pursuing an approach to early detection of cancer that is based on tiny circuits with nanosize transistors. Each transistor can be attached to an antibody, a biological molecule specially designed to attach to a biomarker. When the antibody binds to the biomarker, the transistor's ability to conduct electricity changes slightly, signaling the biomarker's presence. The long-term goal is to build a circuit analogous to a microscopic computer chip that can detect hundreds or thousands of biomarkers in a single test that could catch early cancers that would otherwise go undetected.

## **Bioinformatics**

The deluge of genomic and proteomic data in the past two decades has driven the development of bioinformatics as well as creation of tools that search and analyze biomolecular sequences and structures. Bioinformatics is highly interdisciplinary, using knowledge from mathematics, statistics, computer science, biology, medicine, physics, chemistry, and engineering. Machine learning techniques are being employed to select biomarkers that show discriminating power and applies statistical, mathematical, or computational tools for the derivation of the patients' information. The aim is identification of biomarkers that can predict progression of a disease and would aid in targeting aggressive therapy to those patients that could benefit the most from treatment that would slow or stop progression of a disease. Some of the other approaches are described here.

### ***Biomarker Workflow Guide***

Biomarker Workflow Guide (Ingenuity Systems Inc.) is used to identify potentially useful biomarkers during the preclinical exploratory phase via analysis of gene expression array data or proteomics data. With the workflow guide, researchers can more easily prioritize biomarker leads by associating them with biology relevant to the observed phenotype. IPA can also help researchers select biomarker panels from the most significant clusters by grouping clusters in the context of pathways, functions, or diseases. Computational pathways analytics was used to develop network signatures from gene expression data and to identify osteopontin as a biomarker candidate for mesothelioma (Pass et al. 2005). It is expected to become a standard protocol for biomarker discovery.

### ***Analysis of Microarray Data for Selecting Useful Biomarkers***

DNA microarrays are used for studying gene expression. One potential application of this technology is the discovery of biomarkers for more accurate diagnosis

and prognosis, and potentially for the earlier detection of disease or the monitoring of treatment effectiveness. Because microarray experiments generate a tremendous amount of data and because the number of laboratories generating microarray data is rapidly growing, new bioinformatics strategies that promote the maximum utilization of such data are necessary. Bioinformatic is used in the pattern analysis of serum biomarkers of breast and ovarian cancers.

### ***Role of Bioinformatics in Discovery of Proteomic Biomarkers***

Developments in proteomic technology offer tremendous potential to yield novel biomarkers that are translatable to routine clinical use but major hurdles remain for translation into clinical application. There is a need for rigorous experimental design and methods to validate some of the unproven methods used currently. There is an ongoing debate on where the burden of proof lies: statistically, biologically, or clinically. There is no consensus about what constitutes a meaningful benchmark. It has been pointed out that statistical and machine learning methods are not a crutch for poor experimental design nor can they elucidate fundamental insight from poorly designed experiments. It is now clear that SELDI-TOF MS instrumentation used in the earlier proteomic pattern studies had insufficient resolution to enable the unambiguous identification of the putative marker molecules, which is needed if they are to be validated for forming the basis of a simplified, more widely adopted diagnostic. There is a need for calibration style benchmarking where the linearity of instrument responsiveness is established, to the ultimate benchmark – real clinical usage – as well as for many challenges in between, such as data normalization, peak detection, identification and quantification, and, at some point, classification.

For non-hierarchically organized data in proteome databases, it is difficult to view relationships among biological facts. Scientists at Eli Lilly & Co have demonstrated a platform where such data can be visualized through the application of a customized hierarchy incorporating medical subject headings (MeSH) classifications (Gaylord et al. 2006). This platform gives users flexibility in updating and manipulation. It can also facilitate fresh scientific insight by highlighting biological impacts across different hierarchical branches. They have integrated biomarker information from the curated Proteome database using MeSH and the StarTree visualisation tool.

A novel framework has been presented for the identification of disease-specific protein biomarkers through the integration of biofluid proteomes and inter-disease genomic relationships using a network paradigm (Dudley and Butte 2009). This led to the creation of a blood plasma biomarker network by linking expression-based genomic profiles from 136 diseases to 1,028 detectable blood plasma proteins. The authors also created a urine biomarker network by linking genomic profiles from 127 diseases to 577 proteins detectable in urine. Through analysis of these molecular biomarker networks, they found that the majority (>80%) of putative protein biomarkers are linked to multiple disease conditions and prospective disease-specific protein biomarkers are found in only a small subset of the biofluid

proteomes. These findings illustrate the importance of considering shared molecular pathology across diseases when evaluating biomarker specificity. The proposed framework is amenable to integration with complimentary network models of biology, which could further constrain the biomarker candidate space and establish a role for the understanding of multiscale, interdisease genomic relationships in biomarker discovery.

### ***Role of Bioinformatics in Detection of Cancer Biomarkers***

Cancer biomarkers are described in Chapter 6. Bioinformatics is applied for the exploration of cancer-related biomarkers, such as predisposition markers, diagnostics markers, prognostics markers, and therapeutics markers (Kato and Kato 2006). Because quite large amounts of data are produced by the whole-genome SNP scanning, bioinformatics is necessary for the identification of SNPs associated with cancer predisposition. Individual cancer risks can be estimated accurately by detecting multiple SNPs affecting critical cancer-associated genes. Because expression profiles are determined by the signaling networks in cancer, network analyses promote the exploration of diagnostic biomarkers. Network analysis software using gene set enrichment program is developed by a variety of companies or groups; however, network analyses driven by human intelligence of experts is still powerful and more accurate. In the postgenomic era, bioinformatics is utilized for the identification of novel prognostic markers, recurrence prediction markers, and metastasis markers out of large amounts of genomics, transcriptomics, and proteomics data. Bioinformatics is utilized for the exploration of cancer-related biomarkers to select therapeutic optional choice among surgical operation, radiation therapy, and chemotherapy. Biomarkers for localized tumors with low metastatic potential support the selection of a surgical procedure, while biomarkers for infiltrating tumors with high metastatic potential support the selection of radiation therapy and/or chemotherapy. Therapeutic biomarkers for squamous carcinoma of esophagus, lung, and cervical uterus have been characterized by using bioinformatics on transcriptomics data.

### ***Biomarker Databases***

References are given to several genomic, proteomic, and metabolomic databases in this report and these are relevant to biomarkers. The value of such databases is recognized. Building of reliable biomarker databases and integration of information from the genome programs expand the scientific frontiers on etiology, health risk prediction, and prevention of environmental disease. Biomarker validation may be performed in a number of ways: bench-side in traditional labs; web-based electronic resources such as gene ontology; literature databases; and clinical trials (Phan et al. 2006). Biomarker databases have potential value for pharmaceutical research and development.

The biomarker database of GVK Bio (Hyderabad, India) holds information from published literature relating to clinical and preclinical biomarkers and may be used in biomarker design and validation research. Under an agreement, the FDA uses this database as part of its Voluntary Exploratory Data Submission Program and in internal research projects.

### ***Gene Networks as Biomarkers***

Biomarkers are always parts of critical regulatory circuits either as players or as end point readouts. These roles as well as biological flexibility are generally accounted for by looking at groups of biomarkers that are linked to the situation observed but different subsets of biomarkers are found, which make it difficult to decide whether these results indicate the same basic situation. An innovative approach taken by Genomatix Software (Munich, Germany) elucidates the regulatory networks associated with a particular situation and then projects all the observed changes onto this network. This enables determination whether the two overlapping but distinct sets of biomarkers found are actually part of the same network. The gene network is thus a robust biomarker. Individual pathways in cancer are like lines that run through this network, which represents more stable manifestations than individual pathways. This approach has been applied to comparison of microarray data from freshly frozen and FFPE samples of prostate cancer from the same patient. Although these data sets look different, both belong to the same network. Another advantage of this approach is that analysis of pertinent data may enable detection of changes in the precancerous phase.

## **Pitfalls in the Discovery and Development of Biomarkers**

Merely identifying a putative biomarker is very different from showing that it can be validated. The former might require only a matter of weeks and a few patient samples, yet the latter may require multiple multicenter trials. Past failures show how even seemingly straightforward biomarkers can turn out to be false.

A trial for chronic granulomatous disease was originally planned to last only long enough to evaluate whether patients' white blood cells overcame the disease's characteristic defect – an inability to generate a bacteria-killing oxygen burst. Finally a longer trial plan was adopted and showed that the drug had no detectable effect on oxygen production or bacterial death, but still reduced the rate of recurrent serious infections by 70%.

Techniques to discover and interpret biomarkers produce results that vary considerably. For example, mass spectra can be analyzed through machine learning and various computer algorithms to identify sets of peaks that distinguish between, late-stage and early-stage cancer. However, the same biomarker proteins tend to show up in multiple unrelated diseases, making them useless for accurate diagnosis. There is a need for strong data tying biomarker to clinical outcome.

The technology to discover the ideal biomarkers does not exist yet. Humans make hundreds of thousands of proteins and peptides, some of which are a trillion-fold more common than others, and the most informative biomarkers are probably among the least abundant. The techniques necessary to enrich and fractionate low-abundance proteins are immature. Artifacts of sample preparation and processing are frequently indistinguishable from true peptide profiles. Current attempts to validate biomarkers are inefficient and require improvement.



<http://www.springer.com/978-1-60761-684-9>

The Handbook of Biomarkers

Jain, K.K.

2010, XX, 492 p., Hardcover

ISBN: 978-1-60761-684-9

A product of Humana Press