CANCER GENOMICS & PROTEOMICS 3: 253-264 (2006)

Review

Proteomics in Cancer Research and Diagnosis; An Update

VASILIKI IFANDI[1,2], SUSAN E. SLADE[3] and MICHAEL KHAN[1,2]

[1]Biomedical Research Institute and [3]Biological Mass Spectrometry and Proteomics Facility, Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL;
[2]Clinical Sciences Research Institute, Warwick Medical School, Walsgrave Hospital, Coventry, CV2 2DX, U.K.

Abstract. *The availability of reference genomes for, amongst others, man and mouse, coupled with the ability to manipulate the expression of single genes/proteins in model organisms, has fuelled a revolution in biomedical research. Non-invasive diagnosis and treatment, disease prediction and personalised medicine are all expected developments to result from the genomics revolution. Building on this resource, we are now able to employ high-throughput techniques for profiling the expression of thousands of genes/proteins in tissues, body fluids and even individual cells and have evolved a whole new science-systems biology that seeks to reveal the biological information in such data-rich studies. Given that proteins are the ultimate biological effectors within the cell and, due to various post-translational regulatory steps, are more varied and numerous than the protein-encoding genes, researchers are increasingly progressing from genome and gene transcript analyses to powerful proteomics techniques that offer a more complete picture of cell behaviour. Cancer research and clinical oncology, in particular, have been at the forefront in adapting and exploiting the new biology. The road map to cancer is being described for many individual cancers, tumours are profiled for prognostic and therapeutic purposes and biomarkers are sought to help with early diagnosis and disease monitoring. This review will focus on proteomics and its application to cancer research and diagnosis.*

Information flow in cell biology has long been appreciated, though with the recent identification of small regulatory RNAs, we may have to adjust our position somewhat. Broadly, the genome is transcribed into mRNA and the mRNA is subsequently translated into proteins. The human genome codes for only around 30,000 genes (less than twice the number in nematode worms), but, due to alternative splicing and other intermediate steps, these are transcribed into at least twice that number of mRNAs, which in turn, following translation and then post-translational modifications (PTMs), direct the formation of an estimated >100,000 proteins; at each level, many variations occur making the system very complex.

The term "proteome" was first coined by Mark Wilkins and colleagues (1, 2) and subsequently proteomics was introduced as " the study of all the protein forms expressed within an organism, as a function of time, age, state, external factors, *etc.*" (3). Given, that the cellular phenotype is dictated by proteins, biologists across the disciplines have long awaited techniques to allow the detailed description of changes in cellular behaviour at the level of protein expression. In particular, cancer researchers, already at the forefront in applying post-genome era high-throughput techniques such as gene arrays, have been quick to see the potential of proteomics for providing a more complete understanding of cellular behaviour (4). The Human Genome Project has created great opportunities, but even in genetic diseases such as cancer, it was clear that even a complete description of a given cancer cell genome would not allow a complete understanding of the disease. At a simple level, not all genes are expressed in all cells and, particularly when even in clonal diseases such as cancer, aberrant patterns of gene expression may be inherited by subsequent generations of cancer cells by epigenetic factors, such as loss of imprinting and promoter methylation, which are not detectable by genome sequencing. Moreover, even when we directly profile gene expression, by measuring all the mRNAs, in a cell (or more usually a group of cells or tissue) this does not always accurately correlate with the protein expression, because numerous factors operate after preparation of the mRNA transcripts which all influence the levels and even the final make-up of the proteins that are produced. This is very relevant as proteins are the ultimate

Correspondence to: Vasiliki Ifandi, Biomedical Research Institute, Department Biological Sciences, University of Warwick, Coventry, CV4 7AL, U.K. e-mail: V.Ifandi@warwick.ac.uk

Key Words: Proteomics, cancer diagnosis, biomarkers, review.

arbiters of biological function. It must always be borne in mind that regulation of key cellular behaviours including growth, survival, replication and motility, by various intracellular signalling pathways is mediated by the addition or removal of phosphate groups from the tyrosine or serine/threonine residues of various proteins – self-evidently these can not be directly determined by looking at gene expression (though indirectly the activation of such a signalling pathway may be inferred by changes in gene expression known to be regulated in this way). In fact, other than phosphorylation, proteins can be present in many modified forms, including ubiquitination, sumoylation, prenylation, glycation and so on, all of which have major effects on stability and level, activity and location. Each post-translationally-modified protein is a different protein because it has a different function/location. Hence, the roughly 30,000 genes probably encode for over a million functionally-different proteins. Therefore, to a large extent, the tissue complexity within a given system and its responses to changes can ultimately only be fully represented by studying the proteins themselves. It is worth repeating again that genetic mutations, giving rise to initiation and clonal evolution of cancer, are actioned at the protein level, involving aberrant protein expression, function and information flow within the cancer cell and the tumour microenvironment.

Therefore, proteomics can be considered as the missing link between gene expression and diseases (5). In relation to cancer, proteomics encompasses the identification and quantification of proteins and the effect of their modifications, movements, interactions, activities and function, during the initiation and progression of cancer and during treatment. Hopefully, in the future, proteomics may be employed for screening (biomarkers) and disease monitoring.

In this review, the current tools and techniques used in proteomics are briefly outlined and some examples of their application to cancer research are discussed.

Proteomic Tools and Techniques

Traditionally, proteomics has been divided into three separate areas, namely expression proteomics, functional proteomics and structural proteomics (Figure 1). Expression proteomics focuses on the identification of total and differentially-expressed proteins and aims to quantitatively study protein expression and, ultimately, identify novel proteins. Functional proteomics is mainly aimed at protein interactions, by direct analysis of protein complexes, leading to an understanding of the functional organizational levels within a cell (6). Finally, structural proteomics focuses on mapping out structural protein complexes in relation to their localization within the cell (5). Cancer proteomics mainly utilize expression and functional proteomics; for example, using expression proteomics to study the differential expression of various proteins in tissues or serum may lead to the identification of disease markers and, therefore, be important in early detection and diagnosis. Functional proteomics, on the other hand, can provide important information regarding protein networks and complexes in relation to internal and external signals, thus increasing our understanding of the biological processes involved leading to cancer (7).

Several powerful analytical tools for proteomics have been developed and optimised, offering an unprecedented capability to analyse the expression of hundreds of proteins in a single experiment. Moreover, as many of these are high-throughput and readily automated, they offer unique opportunities in the area of cancer screening and disease monitoring. First, developments in the ionisation of larger biomolecules have made possible the characterisation of biomolecules including peptides and proteins (5) and, second, the field of bioinformatics has provided proteomics with invaluable tools. These include the analysis of protein expression changes, processing of mass spectrometry (MS) data and the creation of protein databases and programmes that allow a variety of MS-generated datasets to be used in the interrogation of the databases to identify the proteins under investigation.

Two-dimensional gel electrophoresis (2DGE). It has been 30 years since O'Farrell (8) first described 2DGE. Subsequently, this powerful technique for separating individual proteins in two dimensions has evolved and is proving invaluable across a wide range of research areas. In 2DGE, the first dimension consists of protein separation according to the isoelectric point (pI) of each protein, using isoelectric focusing (IEF) and usually performed on immobilised pH gradient gel strips (9); whilst in the second dimension, proteins are separated according to their size by means of sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). One of the most important factors for successful protein separation in 2DGE is the sample preparation. When dealing with such a heterogeneous solution, which includes proteins of widely differing hydrophobicities and aggregates which result from cell lysis, it is imperative to choose an appropriate combination of buffers, which should include protease inhibitors, detergents, reductants and denaturing agents (10). Various factors such as pI, protein solubility and size should also be considered in the selection of a suitable sample bufffer. Good sample preparation is imperative to ensure maximum protein solubility and homogeneity of the starting material, resulting in significantly improved gel-to-gel reproducibility. It may be necessary to reduce the sample complexity through a combination of pre-

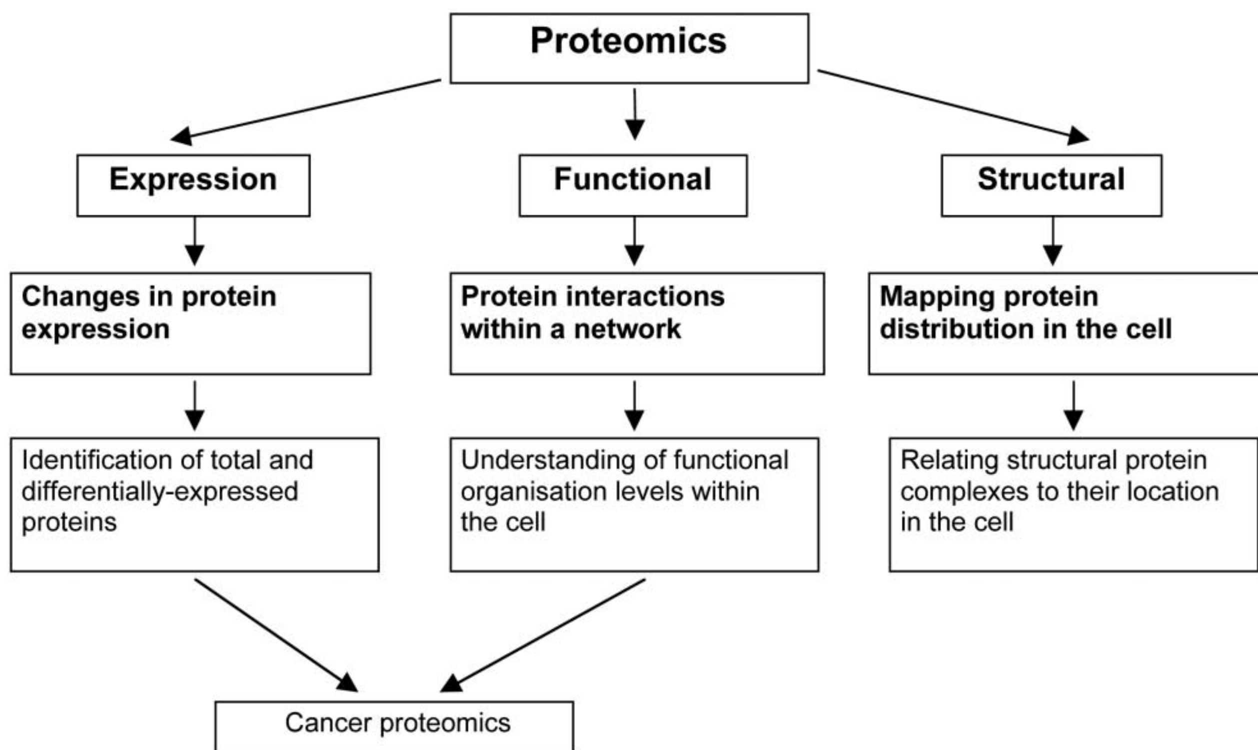

Figure 1. *Various areas of proteomics.*

fractionation techniques in order to resolve single proteins. 2DGE has been successfully used in a significant number of studies and, though the resolution and reproducibility continue to improve with the introduction of a wider range of immobilised pH gradients, wider applicabilities of solubilising buffers and automated gel analysis software, 2DGE is still a laborious and time-consuming process, which may prove to be an obstacle for its routine usage in clinical research. Other factors which limit the applicability of 2DGE for such studies include the limited availability of clinical samples; sample to sample variability between patients that may not be linked to the medical condition under investigation; the fact that many sample sources may not be amenable to 2DGE through their lack of solubility; that a significant number of proteins have a pI that lies outside the current range of commercial IEF strips; that additional gels (and sample) are required to analyse proteins of high or low molecular weight and the recurring difficulty of identifying proteins present in very low abundance. Nevertheless, for general research purposes it remains an invaluable tool.

Protein detection and quantification following 2DGE is a very important step in proteomic analysis. Protein visualisation is usually achieved with the use of various dyes such as Coomassie™ (ICI Ltd.) blue, silver and fluorescent stains. The latter have proven to be most important in protein quantification, because they are both sensitive and provide a greater linear dynamic range (limit of detection for SYPRO™ (Molecular Probes, Inc.) Ruby is 1 ng/band of bovine serum albumin with a dynamic range of 10^3) compared to Coomassie™ blue or silver. This has been particularly useful in differential gel electrophoresis (DiGE) (11). Briefly, the protein samples to be compared are covalently-labelled with various fluorescent dyes. There are two approaches to DiGE, namely the 'minimal' and 'saturation' labelling approaches (12). The most commonly used is the minimal approach, in which the ratio of protein to dye is kept at high levels so that the proteins that are visualised on the gel contain one single molecule of the dye (13). Hence, there is a linear relationship between the fluorescent dye and protein concentration, making comparative quantification relatively accurate. The advantage of this technique is reduced ambiguity with regards to spot matching, since the samples for comparison are present in the same gel. Nevertheless, gel-to-gel variation can still be problematic with this approach, as replicates of samples need to be performed. DiGE has been successfully employed in a variety of cancer studies including breast (14), colon (15) and lung cancer (16), to profile protein expression.

In general, the number of proteins that can be resolved using 2DGE is between 1000-3000 spots, depending on

factors such as the sample preparation, strip length and pH gradient used for the first dimension and the acrylamide/ crosslinker concentrations for the second dimension.

Following protein visualisation and sample comparison, spots of interest are excised from the gels and subjected to proteolytic digestion for identification by means of mass spectrometry (Figure 2).

Biological mass spectrometry (MS). The mass spectrometer is comprised of: a sample inlet; a source region which produces ions from the sample; at least one mass analyser to separate the various ions based on their mass-to-charge (*m/z*) ratio and a detector for the ions emerging from the analyser (17). Briefly, the protein and peptide molecules can be singly or multiply protonated and the positively-charged ions are extracted into the instrument prior to mass selection in the analyser and ultimate detection. Although there are a variety of ionisation techniques suitable for MS, the two most widely used techniques for biological samples are electrospray ionisation (ESI), where the sample is introduced as ionised droplets into the mass spectrometer (18) and matrix-assisted laser desorption ionisation (MALDI), where the analyte is present in an involatile solid deposit and is prepared by the co-crystallisation of the sample with a UV-absorbing chemical matrix (19). The deposit is irradiated and both sample and matrix ions are desorbed near the surface of the deposit, forming a gaseous plume. The ions are then pulse-extracted and focused through the instrument and into the analyser. Once the ions have been produced, they are separated according to their *m/z* ratio, detected and acquired by the software. There are various types of mass analysers, but some are not suitable for biological applications. The following discussion is restricted to those most widely used in proteomics applications.

Quadrupole mass analysers (20) consist of four parallel rods. One set of opposite rods is at a positive electrical potential and acts as a high mass filter, whilst the other set is at a negative potential and acts as a low mass filter. When a radio frequency (RF) and direct current (DC) is applied to the pairs of rods, ions undergo an oscillating movement as they traverse the quadrupole. Only ions of a certain *m/z* will have a stable trajectory and will emerge from the mass analyser. The remaining ions are destabilised and lost through collisions with the quadrupole and the walls of the chamber. A mass spectrum is collected by scanning the RF using a fixed DC, thus enabling ions of a wide *m/z* range to be detected.

A quadrupole ion trap (20) consists of hyperbolic electrodes to which various potentials are applied, trapping the pulsed incoming ions in a stable oscillating trajectory in the centre. Gradual variation in the electrode potentials allows ions of increasing *m/z* to be sequentially ejected from the trap through to the detector.

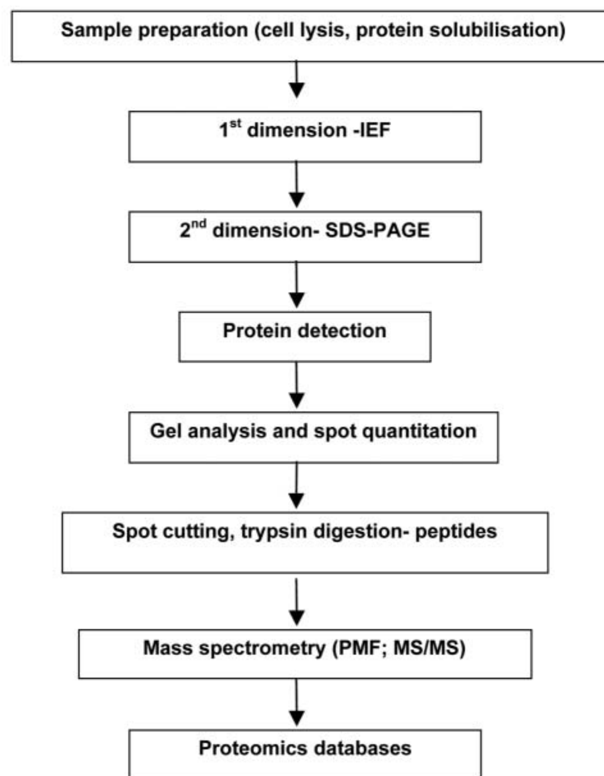

Figure 2. *Schematic representation of a typical 2DGE experiment.*

TOF analysers (21) can be integrated with both MALDI and ESI ion sources. An acceleration pulse is applied to the ions resulting in them having similar initial kinetic energies, but differing *m/z*, as they enter a field free drift region. The ions traverse a given distance with the lower *m/z* ions arriving first at the detector. Peak broadening can occur due to variations in the kinetic energy of ions of a discrete *m/z*, resulting in their arrival at the detector in a distribution rather than a narrow packet of ions. Incorporation of a post-flight tube reflectron, consisting of a series of ring electrodes which form a decelerating field for the ions, allows refocusing of the ions. Those exhibiting higher kinetic energies penetrate deeper into the reflectron before being turned around and accelerated towards a slightly off-axis detector, resulting in a narrower distribution of discrete *m/z* ions.

Fourier transform ion cyclotron resonance mass spectrometers, despite having excellent mass resolution and sensitivity, require extensive infrastructure and expertise and, as a result, are in use in only a small number of MS facilities worldwide (22).

Some instruments combine several mass analysers in sequence allowing tandem mass spectrometry (MS/MS) experiments to be undertaken. In collisionally-induced decomposition experiments (CID), a precursor ion is

selected by one mass analyser and allowed into a collision cell containing an inert gas (23). The resulting fragment (product) ions formed from the collisions are then analysed by another mass analyser and detected. Mass spectrometers capable of trapping ions can perform these experiments by ejecting every ion except the precursor, subsequent fragmentation being achieved by a number of ion excitation methods including CID with the product ions being ejected and detected. Sequential fragmentation and ejection cycles can be performed on a single component resulting in comprehensive structural characterisation termed MS^n.

Protein identification by means of mass spectrometry. The commercial development of MALDI and ESI, combined with substantial improvements in the detection limits of modern MS instrumentation (typically femto to attmole levels), has revolutionised proteomics experiments. The ability to accurately measure the molecular weight of proteins and peptides (often within an error tolerances of tens of parts per million) has facilitated the identification of millions of proteins. Improvements in automation have increased the throughput of proteomics experiments, meaning that the MS analysis is often not the rate-limiting step.

While MS can analyse intact proteins and peptides, the mass of intact proteins is of restricted use in protein identification. Frequently, there are conflicts of amino acid sequence in the protein databases resulting in a variety of molecular weights. PTMs result in proteins with different molecular weights to those predicted from the sequence. Therefore, intact proteins are digested using proteolytic enzymes and the identification is achieved on the analysis of the resultant peptides.

Protein identification can be achieved by means of peptide mass fingerprinting (PMF) and/or tandem mass spectrometry. In PMF the mass spectrometer (usually MALDI-MS) is used to measure the *m/z* ratios of the generated peptides following proteolytic digestion of the sample, usually with trypsin. The protein is subsequently identified by matching the experimental peptide masses with the theoretical peptide digests of established databases, reviewed in (24). Although rapid and sensitive, PMF has a number of limitations including: the requirement for the sequence to be present in the database (or a close homologue); the ambiguities of identifying a mixture of proteins in one sample; proteins of low molecular mass typically yielding few tryptic peptides; PTMs reducing the number of identified peptides as the theoretical peptide masses are generated solely from the amino acid sequence; the fact that as protein databases increase in size, especially those containing sequences from proteins of very high molecular weight, the number of false-positives increases.

Tandem MS is utilised when PMF has failed to unambiguously identify the protein under investigation. In order to substantially increase the sensitivity of the analysis, tandem MS experiments are usually performed using an in-line capillary liquid chromatography (LC) separation step prior to ESI-MS/MS. The tryptic peptides are partially or fully resolved using a narrow bore (75 μm) C18 reverse phase column with an increasing organic solvent gradient prior to infusion into the instrument. The eluting peptides are detected by the software and suitable precursor ions are selected for CID experiments, in real time. Typically tens of peptides will undergo CID during the course of one chromatographic separation. The result of these experiments is a list of precursor and product ion masses from each peptide selected during the experiment.

Protein identification from tandem MS experiments can be achieved by either interrogating a database with the precursor-product ion list (an uninterpreted search) (25) or interpreting the MS/MS spectrum and inferring an amino acid sequence from it. The sequences can then be used in homology searches using BLAST (26), if required to identify homologous proteins, or the information can be used to design primers for gene sequencing.

One of the most recent advances in MS, that has been commercially developed and is rapidly gaining ground in proteomics, is surface enhanced laser desorption ionisation (SELDI) MS. This technique can utilise a variety of chemical and biochemical surfaces or chips with MALDI-TOF MS analysis. These surfaces have been engineered in such as way that they allow the differential capture of proteins from a mixture, based on the properties of the proteins (27). There are surfaces that bind proteins according to their chemical properties, such as hydrophobicity, and metal affinity, or biological properties, such as antibody affinity, antigen binding fragments *etc.* (28). Solubilised cells/tissue or body fluids are directly applied to the surface and, following several washes, the remaining proteins that are bound are subsequently laser desorped and ionised for MS analysis. The use of SELDI-TOF MS is rapidly increasing because this technique is very well suited for the detection of small molecules; it is indicative that a simple search on Medline has produced over 100 reports where SELDI-TOF MS has been utilized in cancer research and, in particular, in biomarker discovery in various types of cancer, including lung (29), ovarian (30), prostate (31), gastric (32), breast (33), melanoma (34) and lymphoma (35), to name but a few.

Liquid chromatography. A limitation of traditional 2DGE is that there is poor representation of basic and hydrophobic proteins and many proteins may precipitate at their pI. Additionally, low abundance proteins are not very well represented on 2D gels. While membrane proteins are of immense interest in cancer research due to changes observed on the cell surface proteins and receptors, low

abundance proteins are also very important especially with regard to searching for cancer biomarkers. Recent developments in LC coupled with MS analysis have been very promising.

An alternative to gel-based approaches is the use of in-solution protein separation methods. These may include initial chromatofocusing steps, followed by secondary separations using reverse phase chromatography (36). Commercial automated 2D-LC systems are available with higher protein loading capacities than immobilised IEF gradients, which should increase the protein available for MS-based identification after chromatographic resolution. This approach shows promise for the study of low abundance proteins in complex systems and, in addition, proteins at the extremes of pI can be collected for further analysis. Post-separation the proteins are available for MS-based characterisation by intact molecular mass analysis and tryptic digestion with protein identification. Preliminary indications are that the methodology may have an application in the analysis of membrane-associated proteomes.

Multidimensional protein identification technology (MudPIT) was first introduced in 1999 (37) and, since then, it has proved to be most effective in investigating global protein changes. MudPIT is used for the separation of complex protein and peptide mixtures, and incorporates high-pressure liquid chromatography (HPLC, LC/LC), MS/MS and database searching. Briefly, MudPIT uses a column that contains strong cation exchange (SCX) gradients coupled to a column with reverse phase (RP) chromatography, resulting in peptides being sequentially eluted from the RP and then being analysed using MS/MS (38). Usually the sample to be analysed is prepared in a similar way as for 2DGE, *i.e.*, tissue/cells are lysed but then the resulting cell lysate is digested to obtain peptides, which are then applied to the columns for concurrent separation and analysis by MS/MS.

Recently, MudPIT has been successfully used to identify proteins released from pancreatic cancer cells (39); this study found that several proteoglycans, including versican and syndecan 1 and 4, are released from the pancreatic cancer cells – proteins that have never before been associated with this type of cancer. These findings enforce the view that one of the major advantages of using MudPIT is that it combines high resolution techniques with a decrease in sampling losses, while covering a wider range of proteins that are 'missed' by 2DGE.

Isotope-coded affinity tags (ICAT™, Applied Biosystems) technology is another technique that combines chromatography with MS. ICAT™ is a protein profiling technology that allows the identification and quantification of proteins (40) that might not be detected using 2DGE, especially if they are present in low abundance. In this method, the ICAT™ reagent is used to label the samples to be compared. The reagent is composed of three elements, an affinity biotin tag that is responsible for isolating the labelled peptides; a linker that incorporates the stable isotopes; and a reactive group with cysteinyl specificity. The reagent exists in two forms, the heavy form that contains eight deuteriums and the light form that does not have any deuteriums. Briefly, the samples that are to be compared are treated with isotopically heavy and light ICAT™ reagents and the ICAT™ reagent is covalently attached to each cysteinyl residue in every protein. The samples are then combined and fractionated, with the biotin-labelled peptides being isolated using high performance liquid chromatography (HPLC). Finally, the peptides are analysed and identified using LC-MS/MS; the original amount of proteins from each sample is maintained in the peptide fragments and therefore quantification is achieved by comparing the ratios of the peptide pairs (5, 14, 40).

The ICAT™ technology has been effectively used to study protein expression in a variety of systems, including breast carcinoma (reviewed in 41). ICAT™, coupled with laser capture microdissection (LCM) and 2D-LC-MS/MS, was also used for qualitative and quantitative analysis of hepatocellular carcinoma (42); the study resulted in the identification of up to 644 proteins, many of which were involved in the cellular structure, apoptosis and transcription.

Despite the limitation regarding the number of proteins identified using ICAT™ and the requirement for cysteine residues in the proteins under investigation, when compared with other methods such as 2DGE, this technique has proven to be a very promising proteomic approach that has the ability to provide high throughput with reproducibility.

The iTRAQ™ (Applied Biosystems) stable isotope labelling approach for protein quantitation does not require specific residues to be present in a protein sample. All N-termini peptides from each sample are labelled using an isobaric peptide tag with the ability to multiplex up to four samples per experiment, with each identical peptide from the four sources co-eluting as a single unresolved precursor ion. Greater confidence in quantitation is achieved due to the increased number of labelled peptides per sample, at the expense of sample complexity. Unlike ICAT™, quantitation is accomplished using diagnostic reporter ions in the LC-MS/MS spectra from each of the samples (43).

Protein microarrays. One area that is of intense interest is the development of protein microarrays, based on the same principles used for gene arrays. This technology would have the ability to provide us with information regarding protein changes at pathway levels, and global protein expression changes without the use of specialized equipment such as MS-based proteomics. There are two basic designs, namely forward phase arrays (FPA) and reverse phase arrays (RPA) (44, 45). In the former (FPA), antibodies are arrayed

and then probed with the cell lysate, while in the latter (RPA), the reverse takes place, the cell lysate being arrayed and then probed with antibodies. There are a variety of surfaces where the proteins are arrayed such as solid surfaces, capillary systems or immobilised beads. However, despite the promising potential of this technique, there is one main drawback, which is that proteins are less structurally stable than nucleic acids and they display extensive differences in biochemical properties. It is difficult to preserve the proteins in their biologically active conformational state prior to the analysis; also various proteins display different affinities for the targets on the arrays. Nevertheless, they have started appearing in clinical studies (46-48). In particular, in a recent study on profiling differentially-expressed proteins in normal and breast cancer tissue (48), tissue microarrays in conjunction with immunohistochemistry enabled the researchers to monitor the expression of 26 selected proteins in more than 1600 cancer samples from 552 cancer patients. However, despite a number of studies that have established the value of this technique, its low sensitivity and high inconsistency, coupled with the high costs involved in their production, indicate that, at this stage, their use in biomedical research is still far from being routine.

Post-translational modifications (PTM). Following translation, the function, destination and stability of a cellular protein is governed by the influence of various post-translational modifications, such as phosphorylation, acetylation, alkylation, isoprenylation, glycosylation, ubiquitination and sumoylation. It has long been appreciated that the activity of most, if not all, proteins involved in signal transduction is regulated by changes in phosphorylation. It has become abundantly clear that other modifications play a crucial role in all key cellular functions. Particularly in cancer biology, key regulatory steps in the cell division cycle are regulated not only by regulation of the gene expression and then by phosphorylation and de-phosphorylation of key proteins, but also by the stability of the proteins – in this respect the ubiquitination of regulatory proteins targets them for degradation in the proteasome and is an essential part of normal cell cycle regulation. In fact, all these modifications play an important role and are essential for our understanding of protein kinetics, regulation and function. It is important to note that major changes in cellular processes can take place without changes in gene expression; thus, many cell cycle regulators, DNA damage response proteins and others are 'switched' from active to inactive states by changes in phosporylation and by changes in stability of the protein; therefore, understanding the role of such modifications is an essential part of understanding the biology of protein pathways and networks.

The phosphorylation and dephosphorylation of proteins, particularly on serine/threonine or tyrosine residues, is regulated by serine/threonine and tyrosine kinases and phosphatases, respectively. These events are critically involved in cell signalling and, not surprisingly, in those directly implicated in aberrant behaviour of the cancer cell. Thus, the deregulated activities of receptor or non-receptor tyrosine kinases are very common contributors to excessive replication, angiogenesis and invasion and are useful disease and prognostic markers, as well as targets for many of the new specific cancer therapies, such as Gleevec™ (Novartis Pharmaceuticals) and Herceptin® (Genentech, Inc.). Various approaches have been employed to identify aberrantly-phosphorylated proteins. These include genome level screening for constitutively active mutant receptor and non-receptor tyrosine kinases, such as *Neu* or *Src,* respectively; *Ras* and regulators of the Ras pathway involved in triggering the MAPkinase cascade and others or, alternatively, transcriptional profiling for the presence of gene expression patterns suggesting aberrant activity of various signalling pathways. However, increasingly we will be able to employ various direct analyses of the proteins themselves using traditional techniques such as immunochemistry and Western blotting (7, 49, 50) and, in the future, also 2DGE and MS.

Laser capture microdissection (LCM). LCM is a relatively new technique, yet it has been well established and widely used in studies where tissue heterogeneity creates a problem in terms of sample collection. It is a method of collecting cells from specific regions of tissue placed on microscope slides. LCM has proved to be invaluable, especially in cancer studies where the diseased cells of interest, such as precancerous cells, would have been impossible to isolate within the heterogeneity of the tissue structure. Briefly, a stained slide is placed under the microscope and, when the desired area for isolation has been found, a laser beam activates a specific transfer film that is applied to the surface of the tissue section; this film then binds specifically to the selected section and is lifted off, leaving the rest of the unwanted tissue behind (7). LCM can successfully be combined with 2DGE, MS and SELDI, despite the fact that the number of cells needed is quite high (51-53). Nevertheless, LCM may prove to be an invaluable tool in the investigation of biomarkers in cancer research. This has been demonstrated in the case of pancreatic ductal adenocarcinoma (54) where LCM was successfully used to enrich both the normal and malignant pancreatic ductal epithelial samples. More recently, LCM was used to help compare cells microdissected from normal ductal epithelium and breast metastatic ductal carcinoma (55). Following LC-ESI-MS/MS analysis, the study identified 76 proteins, such as mitochondrial isocitrate dehydrogenase and 14-3-3 ζ/δ,

found to be significantly up-regulated in breast cancer tumour cells.

Applying Proteomics to Cancer

Diagnosis and treatment monitoring. Despite truly remarkable progress in many areas of molecular medicine, genomics and proteomics, cancer deaths have not declined significantly in the last few decades, in striking contrast to the enormous progress made in reducing cardiovascular diseases over the same period. An important explanation may be that, in general, we can not readily identify patients at risk of cancer in order to prioritise preventative measures, and that established cancers are detected rather late when surgery, drug or other treatments are less likely to be effective. Early detection represents one of the most promising approaches to reducing the growing cancer burden (56), and has been revolutionised with the advent of post-genome era technologies that can identify cellular changes at the level of the genome or proteome, including new developments in data analyses and modelling. The ability to simultaneously detect the expression of multiple genes or proteins offers unparalleled opportunities for discovering novel biomarkers that can be used not only to diagnose diseases, but also to monitor therapeutic response and progression. Much of this effort has been driven by the availability of microarray-based and proteomics technologies able to identify changes in the expressions of large numbers of genes/proteins within cancer tissues. In fact, these technologies have had a significant impact on the field of oncology; gene expression profiling of various human tumour tissues has led to the identification of expression patterns related to disease outcome and drug resistance, as well as to the discovery of new therapeutic targets and insights into disease pathogenesis (57-59). However, as these approaches require removal of cancer tissues, they are not ideal for achieving earlier diagnosis or for general screening, except in cancers which have identifiable potential precursor lesions that are readily accessible, such as skin or oral cancer. For most cancers, a local or systemic non-invasive test would have numerous advantages. Therefore, considerable efforts are now directed at finding 'biomarkers' in complex biological fluids, such as the serum, plasma, urine, saliva and in nipple aspirate. These types of sample can be obtained relatively non-invasively and rapidly and could much more readily be employed in screening large numbers of individuals (60, 61). Their role could also be extended into surgical surveillance for potentially operable disease and post-operative follow-up for disease recurrence. In breast cancer, many autoantigens have been cloned by immuno-screening cDNA expression libraries with breast cancer sera or identified using proteomics (reviewed in 62, 63). However, as with most cancers, with the notable exception of the prostate specific antigen for prostate cancer, no biomarkers have been validated and incorporated into clinical practice for the early diagnosis of breast cancer. However, the recent application of genomics, proteomics and array technologies has resulted in new potential biomarkers being identified, and some of these are in the process of being validated prospectively in large cohorts of patients with breast cancer. Unfortunately, to date, these novel candidate biomarkers have not been translated from bench to bedside, but this will hopefully change in the future. Various proteomics techniques are being adopted in order to identify highly sensitive and specific biomarkers for early cancer detection, disease stratification and monitoring. Several studies have shown that the low molecular weight (LMW) serum proteome may contain the information necessary to achieve these goals (reviewed in 61, 64). Various techniques have shown promise, but generating data of sufficient breadth and depth remains elusive. Although to a degree the LMW proteome can be described by modern MS techniques, it is likely that further progress will be needed; it is hoped that developments in nanotechnology for harvesting "biomarker-containing materials" as well as higher sensitivity MS techniques will further increase the power of this approach. However, a number of techniques have been employed with varying degrees of success. For instance, when tumour samples of human gliomas of various grades were directly analysed by MALDI-MS, protein patterns that correlated with the tumour histology and patient survival were identified (65). The use of SELDI-TOF has improved the ability to identify serum biomarkers and to analyse small tumour samples (63).

The use of activity-based protein profiling and multidimensional protein identification technologies (ABPP-MudPIT) for the streamlined analysis of human cancer samples is also showing promise. In fact, in a recent paper in Nature Methods (66), a novel approach using a rapid initial phase, in which enzyme activity signatures are generated for functional classification of samples is described. This was followed by in-depth analysis of representative members from each class and over 50 enzyme activities in human breast tumours were identified, many of which had not previously been characterized and, importantly, would not have been recognised by mutational analyses.

Research. In response to the development of techniques, as exemplified by gene microarrays and proteomics, which can generate vast amounts of information from single experiments, new disciplines have emerged to deal with these data-rich sciences. Systems biology, and sub-disciplines such as functional genomics, is a highly complex new area of research interest which has been referred to as the "science of collaboration", because progress depends on the coordinated activity of multiple scientific disciplines, including mathematics, statistics, informatics, genetics, physiology,

structural biology, and molecular and cell biology. There has been an explosion of published work in this area in cancer biology, which has been ably reviewed elsewhere (64).

Proteomics studies have begun to illustrate the complexity of many different human and animal models of cancer, but have also helped identify some common features, including the ability to identify specific individual gene mutations in cancer on the basis of expression signatures derived from global gene expression arrays and/or proteomics. Moreover, several novel cancer-relevant genes/proteins have been identified by employing these high-throughput techniques and some have subsequently been validated as cancer-relevant by functional studies using inhibitory RNAs in cell lines *in vitro,* or by use of genetically-altered organisms *in vivo*. For example, proteins involved in the oncogenic potential or apoptotic activity of the oncoprotein c-MYC have been identified by this means (reviewed in 67-69).

Conclusion and Future directions

There is a clear need to identify highly sensitive and specific biomarkers for early cancer detection, disease stratification and monitoring. However, with the notable exception of the prostate specific antigen for prostate cancer, no novel biomarkers have been validated and incorporated into clinical practice. Recent developments in genomics, proteomics and array technologies have identified numerous novel biomarkers, many of which are in the process of being validated prospectively in large cohorts of patients. It is likely that further progress will result from developments in nanotechnology for harvesting ‘biomarker-containing materials’, as well as higher sensitivity MS techniques for analysing small tissue samples. The discovery of disease biomarkers has implications beyond diagnosis and disease monitoring, because some of the molecules identified to date have known roles in cell proliferation, DNA damage responses and apoptosis and, thus, also provide clues as to the pathoetiology of the relevant cancer. Moreover, if the promise of "patient-tailored therapy" becomes a reality, then one might anticipate that future therapies will be individualised on the basis of the specific ‘cancer molecular roadmap’ of each particular patient.

References

1 Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF and Williams KL: Progress with proteome projects; why all the proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev *13*: 19-50, 1996.

2 Williams KL and Hochstrasser DF: Introduction to the proteome. *In*: Proteome Research: New Frontiers in Functional Genomics. Wilkins MR, Williams KL, Appel RD, Hoschstrasser DF (eds.). Springer, Berlin, Germany, pp. 13-33, 1997.

3 Reynolds T: For proteomics research, a new race has begun. J Natl Cancer Inst *94*: 552-554, 2002.

4 Kolch W, Mischak H and Pitt AR: The molecular make-up of a tumour: proteomics in cancer research. Clin Sci *108*: 369-383, 2005.

5 Plebani M: Proteomics: the next revolution in laboratory medicine? Clin Chim Acta *357*: 113-122, 2005.

6 Yanagida M: Functional proteomics; current achievements. J Chromatogr B Analyt Technol Biomed Life Sci *771(1-2)*: 89-106, 2002.

7 Wu W, Hu W and Kavanagh JJ: Proteomics in cancer research. Int J Gynecol Cancer *12*: 409-423, 2002.

8 O’Farrell PH: High resolution two-dimensional electrophoresis of proteins. J Biol Chem *250(10)*: 4007-4021, 1975.

9 Görg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R and Weiss W: The current state of two-dimensional electrophoresis with immobolised pH gradients. Electrophoresis *21*: 1037-1053, 2000.

10 Rabilloud T: Solubilisation of proteins for electrophoretic analysis. Electrophoresis *17*: 813-829, 1996.

11 Van den BG and Arckens L: Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel based proteomics. Curr Opin Biotechnol *15*: 38-43, 2004.

12 Somiari RI, Russell S, Somiari SB *et al*: Difference gel electrophoresis in a high throughput environment. *In*: The Proteomics Protocol Handbook. Walker JM (ed.). Humana Press Inc., NJ (USA), pp. 850, 2005.

13 Tonge R, Shaw J, Middleton B, Rowlinson R, Rayner S, Young J, Pognam F, Hawkins E, Currie I and Davison M: Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics *1*: 377-396, 2001.

14 Somiari RI, Sullivan A, Russell S, Somiari S, Hu H, Jordan R, George A, Katenhusen R, Buchowiecka A, Aeciero C, Brzeski H, Hooke J and Shriver C: High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast. Proteomics *3(10)*: 1863-1873, 2003.

15 Friedman DB, Hill S, Keller JW, Merchant NB, Levy SE, Coffey RJ and Caprioli RM: Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. Proteomics *4(3)*: 793-811, 2004.

16 Seike M, Kondo T, Fujii K, Okano T, Yamada T, Matsumo Y, Gemma A, Kudosh S and Hirohashi S: Proteomic signatures for histological types of lung cancer. Proteomics *5(11)*: 2939-2948, 2005.

17 de Hoffmann E and Stroobant V: Mass Spectrometry: Principles and Applications, Second Edition. Wiley, Chichester, England, pp. 1-10, 2002.

18 Fenn JB, Mann M, Meng CK, Wong SF and Whitehouse CM: Electrospray ionization for mass spectrometry of large biomolecules. Science *246(4926)*: 64-71, 1989.

19 Karas M and Hillenkamp F: Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analyt Chem *60(20)*: 2299-2301, 1988.

20 Paul W and Steinwedel H: Ein neues massenspektrometer ohne magnetfeld (a new mass spectrometer without magnetic field). Zeitschrift fur Naturforschung *(8A)*: 448-450, 1953.

21 Wiley WC and McLaren IH: Time-of-flight mass spectrometer with improved resolution. Rev Scientific Instruments *26*: 1150-1157, 1955.

22 Marshall AG: Milestones in fourier transform ion cyclotron resonance mass spectrometry technique development. Int J Mass Spectrom *200(1-3)*: 331-356, 2000.

23 Jennings KR: Collision-induced decomposition of aromatic molecular ions. Int J Mass Spectrom Ion Phys *1*: 227-235, 1968.

24 Henzel WJ, Watanabe C and Stults JT: Protein identification: the origins of peptide mass fingerprinting. J Am Soc Mass Spectrom *14(9)*: 931-942, 2003.

25 Eng JK, McCormack AL and Yates JR: An approach to correlate tandem mass spectra data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom *5(11)*: 976-989, 1994.

26 Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ: Basic local alignment search tool. J Molec Biol *215(3)*: 403-410, 1990.

27 Rodland KD: Proteomics and cancer diagnosis: the potential of mass spectroscopy. Clin Biochem *37*: 579-583, 2004.

28 Issaq H, Veenstra T, Conrads T and Felschow D: The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. Biochem Biophys Res Commun *292*: 587-592, 2002.

29 Yang SY, Xioa XY, Zhang WG, Zhang LJ, Zhang W, Zhou B, Chen G and He DC: Application of serum SELDI proteomic patterns in diagnosis of lung cancer. BMC Cancer *5(1)*: 83, 2005.

30 Woong-Shick A, Sung-Pil P, Su-Mi B, Joon-Mo L, Sung-Eun N *et al*: Identification of haemoglobin-alpha and beta subunits as potential serum biomarkers for the diagnosis and prognosis of ovarian cancer. Cancer Sci *96(3)*: 197-201, 2005.

31 Malik G, Ward, MD, Gupta SK, Trosset MW, Grizzle WE, Adam BL, Diaz JI and Semmes OJ: Serum levels of an isoform of apolipoprotein a-II as a potential marker for prostate cancer. Clin Cancer Res *11(3)*: 1073-1085, 2005.

32 Ebert MP, Meuer J, Wiemer JC, Schulz HU, Reymond MA, Traugott U, Malfertheiner P and Rocken C: Identification of gastric cancer patients by serum protein profiling. J Proteome Res *3(6)*: 1261-1266, 2004.

33 Laronga C, Becker S, Watson P, Gregory B, Cazares L, Lynch H, Perry RP, Wright GL Jr, Drake RR and Semmes OJ: SELDI-TOF serum profiling for prognostic and diagnostic classification of breast cancers. Dis Markers *19(4-5)*: 229-238, 2003-2004.

34 Wilson LL, Tran L, Morton DL and Hoon DS: Detection of differentially expressed proteins in early-stage melanoma patients using SELDI-TOF mass spectrometry. Ann NY Acad Sci *1022*: 317-322, 2004.

35 Lin Z, Jenson SD, Lim MS and Elenitoba-Johnson KS: Application of SELDI-TOF mass spectrometry for the identification of differentially expressed proteins in transformed follicular lymphoma. Mod Pathol *17(6)*: 670-678, 2004.

36 Hamler RL, Zhu K, Buchanan NS, Kreunin P, Kachman MT, Miller FR and Lubman DM: A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification. Proteomics *4(3)*: 562-577, 2004.

37 Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM and Yates JR III: Direct analysis of protein complexes using mass spectrometry. Nat Biotech *17*: 676-682, 1999.

38 Paoletti AC, Zybailov B and Washburn MP: Principles and applications of multidimensional protein identification technology. Expert Rev Proteomics *1(3)*: 275-282, 2004.

39 Mauri P, Scarpa A, Nascimbeni AC, Benazzi L, Parmagnani E *et al*: Identification of proteins released by pancreatic cancer cells by multidimensional protein identification technology: a strategy for identification of novel cancer markers. FASEB J *19(9)*: 1125-1127, 2005.

40 Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH and Aebersold R: Quantitative analysis of complex protein mixtures using isotope coded affinity tags. Nature Biotechn *17*: 994-999, 1999.

41 Somiari RI, Somiari S, Russell S and Shriver CD: Proteomics of breast carcinoma. J Chromatogr B Analyt Technol Biomed Life Sci *815(1-2)*: 215-225, 2005.

42 Li C, Hong, Y, Tan Y-X, Zhou H *et al*: Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma, using laser capture microdissection coupled with isotope coded affinity tag and two-dimensional liquid chromatography mass spectrometry. Mol Cell Proteomics *3*: 399-409, 2004

43 Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A and Pappin DJ: Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3(12)*: 1154-1169, 2004.

44 Espina V, Woodhouse EC, Wulfkuhle J, Asmussen HD, Percoin EF III and Liotta LA: Protein microarray detection strategies: focus on direct detection technologies. J Immunol Meth *290*: 121-133, 2004.

45 Posadas EM, Simpkins F, Liotta LA, MacDonald C and Kohn EC: Proteomic analysis of the early detection and rational treatment of cancer – realistic hope? Annals Oncol *16*: 16-22, 2005.

46 Hudelist G, Pacher-Zavisin M, Singer CF *et al*: Use of high-throughput protein array for profiling of differentially expressed proteins in normal and malignant breast tissue. Breast Cancer Res Treat *86*: 281-291, 2004.

47 Wulfkuhle JD, Aquino JA, Calvert VS *et al*: Signal pathway profiling of ovarian cancer from human tissue specimens using reverse phase protein microarrays. Proteomics *3*: 2085-2090, 2003.

48 Jacquemier J, Ginestier C, Rougemont J, Bardou VJ *et al*: Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. Cancer Res *65(3)*: 767-779, 2005.

49 Yan JX, Packer NH, Gooley AA and Williams KL: Protein phosphorylation: technologies for the identification of phosphoamino acids. J Chromatogr A *808*: 23-41, 1998.

50 Kaufmann H, Bailey JE and Fussenegger M: Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. Proteomics *1*: 149-199, 2001.

51 Banks RE, Dunn MJ, Forbes MA, Stanley A, Pappin D, Naven T *et al*: The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomics analysis – preliminary findings. Electrophoresis *20*: 689-700, 1999.

52 Craven RA and Banks RE: Laser capture microdissection and proteomics: possibilities and limitation. Proteomics *1*: 1200-1204, 2001.

53 Craven RA, Totty N, Harnden P, Selby PJ and Banks RE: Laser capture microdissection and two-dimensional polyacrylamide gel electrophoresis: evaluation of tissue preparation and sample limitations. Am J Pathol *160*: 815-822, 2002.

54 Shekouh AR, Thompson CC, Prime W, Campbell F, Hamlett J, Herrington CS, Lemoine NR, Crnogorac-Jurcevic T, Buechler MW, Friess H, Neoptolemos JP, Pennington SR and Costello E: Application of laser capture microdissection combined with two-dimensional electrophoresis for the discovery of differentially regulated proteins in pancreatic duct adenocarcinoma. Proteomics *3*: 1988-2001, 2003.

55 Zang L, Palmer-Toy D, Hancock WS, Sgroi DC and Karger NL: Proteomics analysis of ductal carcinoma of the breast, using laser capture microdissection, LC-MS and $^{16}O/^{18}O$ isotopic labelling. J Proteome Res *3*: 604-612, 2004.

56 Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, Radish J, Anderson G and Hartwell L: The case for early detection. Nat Rev Cancer *3(4)*: 243-252, 2003.

57 van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH and Bernards R: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med *347(25)*: 1999-2009, 2002.

58 van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerhoven RM, Roberts C, Linsley PS, Bernands R and Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature *415(6871)*: 530-536, 2002.

59 Wulfkuhle J, Espina V, Liotta LA and Petricoin E: Genomic and proteomic technologies for individualisation and improvement of cancer treatment. Eur J Cancer *40(17)*: 2623-2632, 2004.

60 Liotta LA, Ferrari M and Petricoin EF: Clinical proteomics: written in blood. Nature *425(6961)*: 905, 2003.

61 Calvo KR, Liotta LA and Petricoin EF: Clinical proteomics: from biomarker discovery and cell signalling profiles to individualized personal therapy. Biosci Rep *25(1-2)*: 107-125, 2005.

62 Fernandez Madrid F: Autoantibodies in breast cancer sera: candidate biomarkers and reporters of tumourigenesis. Cancer Lett *230(2)*: 187-198, 2005.

63 Ludwig JA and Weinstein JN: Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer *11*: 845-856, 2005.

64 Conrads TP, Hood BL, Petricoin EF III, Liotta LA and Veenstra TD: Cancer proteomics: many technologies, one goal. Expert Rev Proteomics *2(5)*: 693-703, 2005.

65 Schwartz SA, Weil RJ, Thompson RC, Shyr Y, Moore JH, Toms SA, Johnson MD and Caprioli RM: Proteomic-based prognosis of brain tumour patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. Cancer Res *65(17)*: 7674-7681, 2005.

66 Jessani N, Niessen S, Wei BQ, Nicolau M, Humphrey M, Ji Y, Han W, Noh DY, Yates JR III, Jeffrey SS and Cravatt BF: A streamlined platform for high-content functional proteomics of primary human specimens. Nat Methods *2(9)*: 691-697, 2005.

67 Pelengaris S, Khan M and Evan G: c-MYC: more than just a matter of life and death. Nat Rev Cancer *2(10)*: 764-776, 2002.

68 Shiio Y, Donohoe S, Yi EC, Goodlett DR, Aebersold R and Eisenman RN: Quantitative proteomic analysis of Myc oncoprotein function. EMBO J *21(19)*: 5088-5096, 2002.

69 Shiio Y, Suh KS, Lee H, Yuspa SH, Eisenman RN and Aebersold R: Quantitative proteomic analysis of Myc-induced apoptosis: a direct role for Myc induction of the mitochondrial chloride ion channel, mtCLIC/CLIC4. J Biol Chem *281*: 2750-2756, 2005.

Received March 3, 2006
Accepted March 20, 2006