2

Evolution of Proteomic Methods for Analysis of Complex Biological Samples – Implications for Personalized Medicine

Amanda Nouwens and Stephen Mahler
The University of Queensland
Australia

1. Introduction

We are on the threshold of a paradigm shift for proteomics, moving from largely a qualitative discipline, to now being capable of quantification of a protein within a complex sample at great sensitivity. The potential application of such advanced proteomic technology is enormous as we will be able to detect and quantify low levels of expressed proteins in complex samples, and so move comparative proteomics to a new level.

The evolving practice of personalized medicine will be dependent on devising new techniques and methodologies that will allow the detection and quantification of proteins that are implicated in contributing to the diseased state. There is perhaps somewhere over 5000 genes that are linked to disease states and complex networks of interactions of these expressed genes ultimately lead to these disease states. The myriad of single nucleotide polymorphisms (SNPs) contribute to such phenomena, as an individual's SNP profile play a major role in susceptibility to disease and in adverse reactions to drugs, for example. Coupled with mutations that occur throughout life, the complex "disease state" proteome will contain mutant proteins at low levels that need to be identified and quantified, so that therapeutic intervention based on rational scientific hypotheses can be investigated.

Plasma and serum contain an unknown number of proteins with amounts ranging from pg-g/L levels (i.e. very high dynamic range). As we know one of the major problems faced by proteomic studies of plasma or serum, or indeed any complex protein sample, is that a relatively small number of abundant proteins accounts for the great majority of protein content of the sample. The upshot is that the proteins of interest, which may have regulatory function, are masked by these abundant proteins, and non-targeted methods of proteomic analysis bias at the top end of the abundance scale. The development of new methods for quantifying low abundance proteins has evolved rapidly, concomitant with the evolution of powerful mass spectrometers of increasing sensitivity. The use of antibodies for targeting peptides prior to mass spectrometry analysis is becoming prominent, as a means of partitioning low abundance peptides away from peptides in the bulk sample.

This review will provide a broad overview of the evolution of proteomic methods to analyse biological samples, including Differential In-Gel Electrophoresis (DIGE), Isotope-Coded Affinity Tag (ICAT), Isobaric tags for relative and absolute quantification (iTRAQ), Stable isotope labeling with amino acids in cell culture (SILAC), Unique ion signature Mass

Spectrometry, Selected reaction monitoring (SRM)- based targeted mass spectrometry and Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). Examples of the application of these methods to the identification of proteins involved in a variety of disease states and their implication for personalized medicine will be provided.

2. Overview of personalized medicine

Personalized medicine is designed medicine based on the genotype, or more specifically the SNP profile of individuals. Personalized medicine facilitates the selection of treatments best matched to the individual and disease phenotype (Marko-Varga et al., 2007). What are the main factors which contribute to genotype diversity? SNP is the overriding factor, reflecting past mutation, and occurs wherever there is more than one nucleotide when comparing two sequences. It is the spread of SNPs within genomes which contribute to our individuality, with an estimated 93% of genes containing an SNP (Chakravati, 2001). The individual "fingerprint" of SNPs reflects differences in susceptibility to disease and our varying response to drugs. Pharmacogenetics is the study of how these differences in genotype are manifested in inter-individual variation in response to drugs. The convergence of traditional pharmacogenetics with the relatively new discipline of human genomics has resulted in the evolution of pharmacogenomics (Weinshilbaum et al., 2004). The Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB) is an entity associated with the cataloging genes involved in modulating the response to drugs. The Pharmacogenomics Research Network (PGRN) is a collaboration of scientists studying the effect of genes on responsiveness to a wide variety of medicines (Altman, 2007). The PGRN is linked to the PharmGKB integrative database containing genetic and clinical information on participants in studies (http:www.pharmgkb.org). Thus the integration and availability of data associated with information at the genomic and transcriptomic levels are well developed and a valuable resource for researchers involved in the development of personalized medicine.

The incorporation of proteomics in the further development of the concept of personalized medicine is a more recent phenomenon, and has given rise to the area of pharmacoproteomics, which in essence studies how the proteome changes in response to a drug, and is a logical extension of the pharmacogenetics and pharmacogenomics (Jain, 2004). While pharmacogenetics and pharmacogenomics provide information at the level of the genome and transcriptome, pharmacoproteomics yields information on function (i.e. translational level), although it should be pointed out that small nuclear (sn) RNAs have relatively recently been shown to contribute to regulation of cellular processes and thus have a functional role (Mercer et al., 2009; Mattick, 2009). To this end proteomics and proteomic profiling of individuals' serum and tissues are becoming increasingly important in patient diagnosis and assessment, and together with pharmacogenetics and pharmacogenomics, will provide a more complete picture of the status of an individual, particularly at the functional level.

Identification of disease states can be based on genomic analyses. For example, identification of mutations in breast cancer genes BRAC1 and BRAC2 can be used in the diagnosis of breast cancer (Miki et al., 1994; Wooster et al., 1995). However, at present DNA alone does not necessarily reflect the physiological state of functioning cells and thus analysis of gene products, both RNA and protein are required. RNA expression in comparison to protein, is easier to perform, but transcript levels in the cell do not always

reflect protein levels, nor account for post-translational modifications on proteins or alternative splicing events.

2.1 Significance

There are a number ways in which proteomics may be utilized in personalized medicine and more broadly in drug discovery, research and development. In its simplest application, proteomics has contributed to the discovery of disease biomarkers, clinical entities that define and /or predict normal and pathogenic states (Krejsa et al., 2006). Furthermore, the clinical response to treatment can be monitored through proteome profiling of relevant biomarkers. The clinical use of a biomarker is contingent on whether it is a validated biomarker, which ultimately depends on its clinical reliability and utility. Combinations of validated markers biomarkers can indicate surrogate endpoints that can predict clinical outcomes. On a more global protein expression level, comparative proteomics can generate patterns of protein expression or expression profiles, which may be utilized to define a specific physiological state, or diseased state.

One area where proteomics can yield information not possible by other means is the identification and localization of proteins in various cellular compartments and extracellular space. The paradigm that proteins have fixed locations within cells has recently proven to be simplistic, and that proteins have diverse functions depending on their cellular location. The identification of a protein outside of its known functional zone in cellular preparations was once thought to be due to rupture of cells/cell organelles and leakage of the protein into other fractions. However it is now known that proteins translocate between intracellular and extracellular compartments (Butler et al., 2009). This has enormous implications in drug targeting as the presence of a target in multiple locations may complicate therapy. For example, chaperone proteins, including HSP 10, 70 and 90 have now been shown to exist in extracellular locations, where it was once thought that the chaperone proteins were exclusively located intracellularly, to aid protein folding and carry out chaperone function. Inhibitors of HSP90α are in clinical trials for treatment of cancer (Banerji, 2009), however inhibition of the extracellular wound-healing function of HSP90α21 could be an undesired adverse effect.

2.2 Personalized medicine for cancer

Due to the great diversity of cancer types, and individual variation within specific tumours, cancer perhaps shows the greatest potential for development of personalized therapy. Cancer accounts for about 13% of all deaths annually world wide, and is a major cause of morbidity and mortality (Krause-Heuer et al., 2009). Notwithstanding the emergence of new chemotherapeutic drugs and novel therapies for cancer, significant challenges remain for understanding tumourigenesis and tumour cell biology, and in developing new, effective strategies for cancer treatment. (Mozafari et al., 2009). Some of these challenges include;

- Identification of new tumour targets
- Drug potency, due to inadequate concentration at the cell surface.
- Non-selective nature of cytotoxic agents and a low therapeutic index.
- Development of multi-drug resistance (MDR)

The development of therapeutic monoclonal antibodies (mAbs) has shown promise in treatment of cancer amongst other indications. At present there are around 30 approved therapeutic mAbs predominantly for treating cancer and diseases associated with

inflammation (Walsh, 2010). Anti-cancer antibodies are designed to target tumour cell surface antigens, with subsequent eliciting of an immune response on tumour binding. Most commonly this is termed antibody dependent cellular cytotoxicity (ADCC), and activated natural killer cells are recruited to attack the tumour. Currently approved anti-cancer therapeutic mAbs targeting the tumour cell surface are specific for antigens including EGFR, HER2/*neu*, CD20, CD52 and CD33. Other targets which are receiving much attention are the mucins, principally MUC1 (there is no existing approved antibody for MUC1), IGFR and CEA and cancer stem cell antigen CD44. The future development of monoclonal antibody (and other) cancer therapies will be contingent on the identification, development and validation of new tumour targets. However identification of new tumour biomarkers that reliably and accurately diagnose early stage cancer has not been met with great success. As an example, a group of researchers, as part of the Early Detection Research Network (EDRN), tested a group of recently discovered putative biomarkers for ovarian cancer, however none were superior to CA-125, which has been used extensively for 30 years. Notwithstanding the significant challenges in the discovery and development of clinically useful biomarkers, proteomics will be central for the discovery of new, novel biomarkers for early detection and diagnosis; some of these biomarkers may be suitable for development as novel drug targets (Pastwal et al., 2007).

Recently the International Cancer Genomes Consortium (ICGC) was formed, with a charter to co-ordinate and integrate large-scale cancer genome sequencing projects, focusing on 50 different types of cancers (The International Cancer Genome Consortium, 2010). The expanded studies will consist of investigating around 25,000 specific cancers (biopsy material from individuals). The primary objectives of the consortium were made public in April 2008 and were released in April 2008 (http://www.icgc.org/files/ICGC_ April_29_2008. pdf). However these studies will be at the genomic, epigenomic and transcriptomic levels. At the proteomic level, the Human Proteome Organization (HUPO, http://www.hupo.org/) instigated the Human Proteome Project (HPP), a co-ordinated global initiative to map the protein-based molecular architecture of the human body. This initiative, will aid in the discovery and cataloging of new tumour associated antigens and potential targets.

2.3 Evolution of methods of analysis using proteomics

The term proteomics, encompassing the analysis of and tools used to examine proteins expressed by a genome was coined only about 15 years ago (Wilkins et al., 1996). The progression and development of proteomics since this time however, has naturally afforded a refinement in techniques and methods for simultaneously detecting, identifying and quantifying proteins in biological samples. Fundamental to any identification or further characterization is the ability to first separate proteins from complex samples. This is particularly important for samples such as blood where it is estimated that the dynamic range of proteins is greater than 10exp9. With such a vast dynamic range, just 22 proteins account for 99% of protein content in blood (reviewed in Simpson et al., 2008). With respect to the search for protein biomarkers, the situation is also complicated by the notion that most biomarkers will be low abundance proteins. As protein function and levels of abundance are often altered in disease states, identification of such changes by comparison of healthy and disease samples will allow a greater understanding of the disease, provide new therapeutic targets, as well as identify markers of disease status. Establishment of what could be defined as the "healthy phenotype", will depend on detailed characterization of the proteomes of healthy and diseased states of cells/tissue. One school of thought suggests that

by creating complex "proteomic fingerprints" of healthy and diseased states (and transitions thereof), one may recognize perturbations from the healthy state phenotype before manifestation of the disease state. Therapeutic intervention during the transition-to-disease state may instigate a reversion to the healthy phenotype. Thus a systems biology approach to studying the proteomes of cells in normal and diseased states, and also the network of protein-protein interactions, should enhance the opportunity for attaining this goal.

Proteomics and the tools to identify and quantify proteins have evolved substantially since its conception. The availability of genomic information, particularly the complete human genome sequence has in many ways pushed the bottleneck from the genomic to proteomic arena. Developments include tools for protein separation, protein identification, quantification and automated processes. The following sections provide a summary of some of the major approaches to protein separation, identification and quantification using both gel and gel-free proteomic methods. Separation of proteins can be based on one or more physical or biochemical parameter including size, pI, sub-cellular location, or other depletion / enrichment strategies, with separation involving two or more 'orthogonal' approaches providing greater separation than a single dimension alone.

2.3.1 Two-dimensional gel electrophoresis

One of the earliest approaches to protein separation was based on the use of two-dimensional gel electrophoresis (2DGE), in which proteins are first separated by charge in a pH gradient (isoelectric focusing), followed by separation based on size in SDS-PAGE gels (O'Farrell, 1975). The gel is subsequently stained to visualize proteins, with each protein spot-volume representative of abundance of the protein(s) within it. This approach, particularly when combined with other upstream fractionation steps (e.g. sub-cellular) can provide separation and resolution of a large number of proteins (Cordwell et al., 2000) and has been used in various projects aiming to identify differentially expressed proteins – potentially biomarkers, by comparing control vs. diseased samples via gel-analysis software and statistical tests. Using 2DGE, identification of candidate biomarkers has been achieved for a range of diseases including atherothrombotic ischemic stroke (Brea et al., 2008), pancreatic cancer (Park et al., 2011) and breast cancer (Lee, et al., 2011) with many studies identifying new potential markers.

2.3.2 Differential In-Gel Electrophoresis (DIGE)

Improvements to 2DGE include the addition of small fluorescent tags (CyDyes) on protein samples prior to separation, thus allowing multiple samples to be combined in the same physical gel (Tonge et al., 2001). This approach, known as Differential In-Gel Electrophoresis or DIGE, circumvents some of the problems of quantifying proteins across different samples in different gels as the distinct fluorescent tags on different samples allow the researcher to detect proteins from control vs. disease samples simultaneously. The inclusion of a pooled internal standard representing a mix of all samples can help circumvent some of the technical difficulties (e.g. gel warping and spot matching) that arise from single-sample gels. An advantage of DIGE is that due to the sensitive fluorescent nature of the CyDye labels low amounts of sample (as little as 10 μg) are required for analysis. However, a caveat is that although protein separation and quantification can be achieved, the identity of the proteins still remains unknown unless the protein spots are excised and further analysed, which can be difficult due to the low amounts of protein used in the analysis. Limitations also exist in

the dynamic range of protein detection, estimated at 10exp4 (reviewed in Rabilloud, 2002) Typically a separate, unlabelled gel with larger amounts of protein loaded are required to be run and subsequently cross matched with the original DIGE experiment, as described by Matigian et al., 2010. Despite these difficulties, 2DGE has a proven track record in the separation and identification of proteins, with numerous differentially expressed proteins and potential biomarkers uncovered. 2DGE has been applied to a wide range of sample types including tissues (e.g. breast, skin, brain) and fluids such as cerebrospinal fluid, serum, urine, and tears targeting diseases such as various cancers, Alzheimers disease and dementia, cardiovascular diseases, infections such as HIV, conjunctivitis and toxoplasmosis. Overall, 2DGE with or without fluorescent labeling of samples does provide good separation of proteins, but is time consuming and a laborious process both in the gel procedures, and analysis of 2D protein spot profiles.

2.4 Mass spectrometry-based approaches

Mass spectrometry (MS) has formed the basis for standard protein identification for many years, typically in a 'bottom-up' approach (in which proteins are digested, usually with trypsin) and the resulting peptides analyzed to determine protein identity, but also in some cases by 'top-down' approaches where intact proteins form the basis of analysis. MS-based approaches hold some advantages over traditional 2DGE methods in that samples can potentially be analyzed and identified simultaneously through methods such as two dimensional LC-MS/MS using a combination of strong cation exchange followed by reverse phase separation of peptides. In comparison to gel-based approaches, MS analyses appear to be more effective at identifying low abundance proteins, as well as those with extreme physical properties such as molecular weights (low or high) or pI values, which are often difficult to resolve on gels. MS-based analyses also offer better prospects for automation of separation, analysis and identification of proteins.

With the ability to rapidly identify large numbers of proteins via MS, the emphasis has since shifted to also quantifying those proteins detected by MS. Broadly speaking, the approaches to quantifying proteins via mass spectrometry can be based on labeled or label-free methods. Each approach has its own advantages and disadvantages. Label-based methods include Isotope-Coded Affinity Tag or ICAT, and Isotope Tag for Relative and Absolute Quantitation or ITRAQ, in which multiple samples can be labeled, mixed and then analyzed simultaneously via MS to avoid technical issues relating to reproducibility that may be encountered with label-free approaches.

2.4.1 ITRAQ, ICAT and SILAC

ICAT and ITRAQ differ in their labeling chemistries and site of attachment. For ICAT, cysteine (cys) residues are targeted and selected for via avidin affinity chromatography. The enrichment of only those cys-containing peptides provides one avenue to quantify samples without the complexity of analyzing all peptides in a sample. However, ICAT becomes problematic for analysis of proteins which lack any cys residues. Furthermore, as reviewed in Patton et al., 2002, approximately 70% of proteins contain four or less cys residues thus limiting the usefulness of this approach.

ITRAQ utilizes lysine residues for labeling (Ross et al., 2004), thus avoiding the problem of limited cys residues encountered with ICAT. ITRAQ is a multiplexed approach, where tags are based on isobaric reagents. This means that up to eight different samples can be labeled

with unique tags. The physical properties of the tags differ only in the isotopes used in their synthesis, meaning during LC separations, and in MS mode are identical. Only upon fragmentation in MS/MS mode are the isobaric tags distinguishable. The result is that proteins can be identified via MS/MS and due to unique reporter ions from the ITRAQ tag, the protein can also be quantified. The initial ITRAQ labels were designed to label up to four different samples, although tags to label and detect up to eight different samples are now available. Limitations for ITRAQ lie with the difficulty of identify proteins and quantify them when uniquely expressed in only one sample type eg a protein expressed only in the diseased state. Some technical difficulties, in particular with 8plex tags, resulting in a reduction in identification efficiency have also been reported (Thingholm et al., 2010).

Other label-based strategies such as Stable Isotope Labeling with Amino acids in Cell culture (SILAC) also exist. This method is based on labeling proteins in culture with heavy and light forms of amino acids. The approach is a useful way of comparing two samples, but is limited only to cells grown in culture or in some cases, animal models (Zanivan et al., 2012) and would not be applicable to human or clinical studies due to the use of radioactivity.

2.4.2 Selected reaction monitoring / multiple reaction monitoring

Selected reaction monitoring (SRM), also known as Multiple Reaction Monitoring (MRM) provides a targeted approach to quantifying proteins in a sample. It advances the 'global' approach to quantification by simply targeting those proteins specifically of interest to the researcher. Typically MS instruments such as triple quadrupoles are used, where a precursor mass representing a peptide (typically tryptic) from the protein of interest is selected, fragmented and specific product ions unique to that peptide, monitored. Generally for each protein of interest a number of precursor ions, and subsequent product ions are monitored to ensure specificity. This information is then used to identify and quantify the proteins present. The main limitation of SRM/MRM analyses is that it is essential to know the proteins of interest beforehand, so that appropriate precursor / product ion can be monitored.

SRM / MRM is a targeted approach, where identify of the protein(s) of interest must be known beforehand. The usefulness of SRM/MRM analyses is thus as a downstream technology after discovery-phase experiments have concluded, and candidate proteins of interest requiring quantification already identified. The advantage of SRM / MRM assays is the ability to simultaneously monitor numerous potential biomarkers in a single analysis and quantify protein levels and is thus currently a popular area of investigation. To help with SRM / MRM analyses, a consortium called SRMAtlas has been established (www.mrmatlas.org) to quantify proteins in complex samples by MS. As well as human entries, mouse and yeast information is also contained, and provides both web-interface and command line tools to search for assays. This readily available information means researchers can potentially circumvent some method development steps as optimal coordinates for SRM / MRM transitions of numerous proteins are available.

2.4.3 Stable isotope standards and capture by anti-peptide antibodies

Stable isotope standards and capture by antipeptide antibodies (SISCAPA) is a method which allows the quantification of peptides from complex digests. Originally described by Anderson et al., 2004 the method utilizes stable-isotope-labeled internal standards for

comparison with (unlabelled) peptides that are enriched via anti-peptide antibodies, with subsequent quantification performed by electrospray mass spectrometry. The approach offers increased sensitivity over non-enriched methods, particularly when coupled with SRM / MRM assays. SISCAPA also offers potential in the verification of diagnostic protein panels from large samples as well as increased efficiency in assay time for the bind/elute process over conventional reverse phase separations. There are also distinct advantages over traditional techniques such as ELISA in development time for biomarker assays (Whiteaker et al, 2009). The main disadvantage of SISCAPA is the need for preselected targets, as well as the cost in producing the internal peptide standards and generation of the peptide binding antibodies. However, given the sensitivity of the assay (only low fmol – pmol amounts are required), and the fact the antibody itself can be recycled and used again means on-going costs can be reduced. Since the original design of SISCAPA, refinements in the assay have been developed to reduce loss of low abundance peptides, automated processing steps, and improvement in antibody sources i.e. from polyclonal to monoclonal (Anderson et al., 2009; Schoenherr 2009). As this method is only a relatively recent development, no biomarkers have as yet been published as validated with this approach, although proof-of-principle experiments have been performed with established biomarkers such as tropinin I (cTnI) (Kuhn, et al., 2009) and thus SISCAPA remains a promising tool.

2.4.4 Alternative strategies

In addition to the above technologies, other strategies have been developed to complement gel and MS approaches to detect biomarkers through improved sample preparation methods. For example, hexapeptide libraries, based on combinatorial peptide libraries offer a way to deplete samples of highly abundant proteins (Guerrier et al., 2008). In this innovative technique, a large collection of specific hexapeptides (hexapeptide library) is attached to beads. The complex protein sample of interest is mixed with the hexapeptide-bead library. The peptide library is of high diversity and so it would be expected that a specific peptide(s) in the library would have affinity for each individual protein in the complex sample. After separation of the beads from the mixture, the adsorbed proteins are eluted from the beads. As each hexapeptide is equally represented within the library, the end result is that the abundant proteins are depleted, while proteins of low abundance are concentrated. This approach is particularly useful for biological fluids (serum, saliva, urine etc) which have particularly large dynamic range of protein abundance. For example, hexapeptide enrichment of urine has uncovered an additional 251 proteins that were not previously known to be present in this fluid (Castagna et al., 2005). Although depletion/enrichment strategies may not, in their own right, uncover biomarkers, their usefulness lies in the ability to mine deeper into the proteome of these highly complex samples so that low abundance proteins can be identified.

Depletion / enrichment strategies can be problematic if the abundant protein is a carrier for low abundance molecules and the use of depletion strategies must be done with caution. For example it has been shown that the depletion of albumin from human plasma can also remove low abundance proteins such as cytokines (Granger et al., 2005). More recent studies (Bellei et al., 2010) have also concluded that removal of high-abundance proteins can result in a loss of non-targeted, less abundant proteins. Obviously unintentional and unknowing loss of low abundance proteins is a cause for concern in the search for biomarkers of disease.

Other approaches for analysis of samples include Surface Enhanced Laser Desorption Ionisation (SELDI) and Matrix Assisted Laser Desorption Ionisation (MALDI), which have both been utilised particularly in examination of body fluids such as serum for biomarker discovery. This is effectively a "protein pattern recognition" approach (reviewed in Zhan & Desiderio, 2010) which compares profiles from control versus disease samples to identify those proteins differentially expressed. This approach has been used in particular for analysis of cancer patients.

2.4.5 Post-translational modifications

The majority of the above technologies focus on protein expression and differential expression in control vs. disease states. However, greater emphasis in the future on protein post-translational modifications (PTMs) such as glycosylation and phosphorylation will be needed. Already, perturbations in modifications of proteins by the glycan O-linked B-N-acetylglucosamine (O-GlcNAc) has been implicated in a range of diseases, including Alzheimers and diabetes (reviewed in Dias & Hart 2007). Similarly, differential phosphorylation has been identified in diseased states such as cancer compared to control patients (Semaan et al., 2011).

2.4.6 The Human Proteome Project

Fundamental to rational design for disease treatment and prevention is the understanding of genes present, and the expression and function of gene products, including proteins involved in the disease process. The Human Genome Project (HGP) was established to map all genes encoded in the human genome. Surprisingly, the total number of protein coding genes, only approximately 20,300, was substantially lower than expected, with increased complexity presumably due to splice variants, and post-translational modifications. Following on from this ground-breaking work, is the recent establishment of the Human Proteome Project which aims to map the human proteome (Legrain et al., 2011). At present, of the protein-coding genes in humans identified in the human genome, approximately one third have not been detected at the protein level, while for many others, basic information such as abundance, sub-cellular localization, or function are unknown. Mapping of the proteome will be valuable in understanding human biology, and downstream applications in developing diagnostics, prognostics and new therapies to treat diseases. The HPP will have a 'gene-centric' approach to map information about proteins back to gene loci. HPP will aim to address three parts (HUPO Views, 2010):

- Identification and characterization of proteins from every gene.
- Distribution of proteins in all normal tissues and organs.
- Mapping of pathways and protein networks and interactions.

With respect to sample type, bodily fluids relatively easily attainable, such as urine, saliva, tears as well as those requiring more slightly more invasive methods for collection such as serum, plasma and CFS have all been analyzed for a variety of diseases. Fluids as opposed to solid tissues would generally form a better basis for determining personalized signatures and biomarker detection due to their ease of attainment. There has been some question over whether blood is the best choice for searching for biomarkers. The rationale has been that specific proteins are secreted by the body from different organs, and that these can represent a biological "fingerprint" of physiological state (reviewed in Simpson et al., 2008).

Using model systems such as mice, researchers have shown changes in the plasma proteome prior to any clinical evidence of breast cancer (Pitteri et al., 2011). A separate study in humans (Li, 2011) also suggests that it may be possible to observe proteome plasma changes prior to diagnosis. Plasma proteomics has also been used in the search for pre-diagnostic markers in other diseases such as coronary heart disease (Prentice et al., 2010). The ability to identify proteome changes prior to manifestation of disease phenotypically, will potentially improve patient outcome, particularly for those diseases such as cancer where early diagnosis strongly correlates with survival rates.

3. Conclusion

The ability to define proteomic "signatures"' for individuals will vastly enhance the ability of the medical community to diagnose and treat diseases, as well as potentially identify disease before symptoms appear. Early treatment will in turn prolong life, as well as potentially address healthcare costs through the application of more refined and defined therapies suitable for individual patients. The heterogeneity of some diseases such as breast cancer, in which specific proteins, e.g. progesterone receptor, estrogen receptor and HER-2 may or may not be expressed, make it difficult to broadly treat patients, as a 'one size-fits all' approach does not always apply, and emphasizes the need for individualized and personalized medicine. By examining the proteome, it is possible to gain a better understanding of the heterogeneity present in an individual and potentially can help determine best choice of therapies, as well as indicate disease status and progression. Given the complication of genetic factors and environmental influences on an individual, personalized medicine strategies will require complementation of proteomic data with other areas and strategies for analysis and compile this information to determine diagnostic approaches and tailor therapeutic strategies for the individual.

As yet, despite the excitement of biomarker discovery, and the vast number of publications claiming detection of biomarkers for a specific disease, the majority of candidate biomarkers are yet to be validated or used in clinical settings. However, once candidate biomarkers are confirmed, the emphasis will be on high-throughput approaches to expand analyses to greater numbers of samples. Clearly the proteomic tools available to detect and characterize samples, particularly in a high-throughput quantitative fashion are now a reality. Thus, personalized medicine is not far off the horizon. We anticipate a new era of therapeutic approaches and more refined medicinal treatments for diseases which will be more targeted and precise, not just for the disease, but for the individual, based on establishment of "proteomic fingerprints". In addition to greater confidence in diagnoses, proteome signatures would allow a more individualized and targeted approach to therapy. Potentially, such signatures may also provide better insight into future recurrences of the disease.

Besides the quest for discovery, research and development of new and unique biomarkers, other facets of biomarker research incorporate aims such as improving reliability, increasing the speed of detection and reducing the amount of sample needed for analysis. However, the search for biomarkers is particularly important for those diseases such as breast cancer, for which there are no current clinical biomarkers available, and for which mortality is tightly related to disease stage in the initial surgery (Bohm et al., 2011). Using proteomics however, a biomarker signature for non-metastatic breast cancer has been uncovered. This study (Bohm et al., 2011) found using serum samples and SELDI-TOF and MALDI-

TOF/TOF analyses, a combination of 14 biomarkers that can identify breast cancer patients from controls, with a specificity of 67%. It is unlikely at this stage that this panel or signature will entirely replace imaging diagnostics, but does have the potential to aid current diagnostic approaches, particularly when cancer survival rates are greatly improved with early detection, while tumors <5 mm are normally not detected.

One of the fundamental problems of assigning 'biomarker' status for a protein found to be differentially expressed in a disease is the overlap of these differentially expressed proteins across different diseases. A number of proteins have been implicated across a number of different diseases, making the notion of a single biomarker to indicate a specific disease more difficult. For example, serum amyloid A has been proposed as a prognostic marker for melanoma (Findeisen et al., 2009), breast cancer (Schaub et al., 2009), atherothrombotic ischemic stroke (Brea et a., 2009). Potentially, for greater confidence in disease diagnosis or prognosis, it may be required that a suite of biomarkers, be needed to provide greater specificity and confidence.

The significance of the future development of personalized medicine is far reaching, and will allow/facilitate the following:

- Predicting a patient's response to drugs.
- Development of customized' prescriptions.
- Minimizing, or in some cases eliminating adverse events.
- Improving rational drug development.
- Improving drug R&D and the approval of new drugs - better designed clinical trials based on genomic/proteomic information.
- Screening and monitoring certain diseases e.g. advanced diagnosis before disease symptoms.
- Reducing the overall cost of healthcare.

If the concept of routine personalized medicine is to become a reality in the future, the development of new proteomic techniques and methodologies will be vital, and will build on current methodologies now available.

4. References

Altman, R. (2007). PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. *Nature Genetics*, Vol.39, pp. 426, ISSN 1061-4036

Anderson, N.L.; Anderson, N.G.; Haines, L.R.; Hardie, D.B.; Olafson, R.W.; & Pearson T.W. (2004). Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-peptide Antibodies (SISCAPA) *Journal of Proteome Research,* Vol.3, pp. 235-244, ISSN 1535-3893

Anderson, N.L.; Jackson, A.; Smith, D.; Hardie, D.; Borchers, C. & Pearson, T.W. (2009) SISCAPA Peptide Enrichment on Magnetic Beads Using an In-line Bead Trap Device. *Molecular & Cellular Proteomics,* Vol.8, pp.995-1005, ISSN 1535-9476

Banerji, U. (2009). Heat shock protein 90 as a drug target: some like it hot. *Clinical Cancer Research,* Vol.15, pp. 9–14, ISSN 1078-0432

Butler, G.S. & Overall, C.M. (2009). Proteomic Identification of Multitasking Proteins in Unexpected Locations Complicates Drug Targeting *Nature Reviews Drug Discovery,* Vol.8, pp. 935-948, ISSN 1474-1776

Bellei, E.; Bergamini, S.; Monari, E.;Fantoni, L.I.; Cuoghi, A.; Ozben, T.; & Tomasi, A. (2010). High-abundance proteins depletion from serum proteomic analysis: concomitant removal of non-targeted proteins. *Amino Acids* Vol.40, pp.145-156 ISSN 0939-4451

Bohm, D.; Keller, K.; Wehrwein, N.; Lebrecht, A.; Schmidt, M.; Kolbl, H. & Grus, F.-H. (2011). Serum proteome profiling of primary breast cancer indicates a specific biomarker profile. *Oncology Reports,* Vol.26, pp. 1051-1056, ISSN 1021-335X

Brea, D.; Sobrino, T.; Blanco, M.; Fraga, M.; Agulla, J.; Rodriguez-Yanez, M.; Rodriguez-Gonzalez, R.; Perez de la Ossa, N.; Leira, R.; Forteza, J.; Davalos, A. & Castillo, J. (2009) Usefulness of haptoglobin and serum amyloid A proteins as biomarkers for atherothrombotic ischemic stroke diagnosis confirmation. *Atherosclerosis,* Vol.205, pp. 561-567, ISSN 0021-9150

Buchen, L. (2011). Missing the Mark: Why Is It So Hard to Find a Test to Predict Cancer *Nature,* Vol.471, pp. 428-432, ISSN 0028-0836

Butler, G.S. & Overall, C.M. (2009). Proteomic Identification of Multitasking Proteins in Unexpected Locations Complicates Drug Targeting *Nature Reviews Drug Discovery,* Vol.8, pp. 935-948, ISSN 1474-1776

Castagna, A.; Cecconi, D.; Sennels, L.; Rappsilber, J.; Guerrier, L.; Fortis, F.; Boschetti, E.; Lomas, L.; & Righetti, P.G. (2005). Exploring the Hidden Human Urinary Proteome via Ligand Library Beads. *Journal of Proteome Research,* Vol.4, pp. 1917-1930, ISSN 1535-3893

Chakravati, A. (2001). To a Future of Genetic Medicine (2001) *Nature,* Vol.409, pp. 822-823, ISSN 0028-0836

Cordwell, S.J.; Nouwens, A.S.; Verrills, N.M.; Basseal, D.J.; & Walsh, B.J (2000). Sub-proteomics based upon protein cellular location and relative solubilities in conjuction with composite two-dimensional electrophoresis gels. *Electrophoresis,* Vol. 21, pp. 1094-1103, ISSN 0173-0835

Dias, W.B. & Hart, G.W. (2007) *O*-GlcNAc modification in diabetes and Alzheimer's disease. *Molecular Biosystems,* Vol.3, pp. 766-772, ISSN 1742-206X

Filiou, M.; Turck, C. & Martins-de-Souza, M. (2011). Quantitative proteomics for investigating psychiatric disorders. *Proteomics Clinical Applications,* Vol.5, pp. 38-49, ISSN 1862-8354

Granger, J.; Siddiqui, J.; Copeland, S.; & Remick, D. (2005). Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics* Vol.5, pp. 4713-4718, ISSN 1615-9853

Guerrier, L.; Righetti, P.G.; & Boschetti, E. (2008). Reduction of Dynamic Protein Concentration Range of Biological Extracts for the Discovery of Low-Abundance Proteins by Means of Hexapeptide Ligand Library. *Nature Protocols,* Vol.3, pp.883-890, ISSN 1750-2799

HUPO Views (2010). A Gene Centric Human Proteome Project. *Molecular and Cellular Proteomics,* Vol.9, pp. 427-429, ISSN 1535-9476

Jain, R. (2004). Role of Pharmacogenomics in the Development of Personalized Medicine *Pharmacogenomics*, Vol.5, pp. 331-336, ISSN 1462-2416

Krause-Heuer, A.M.; Grant, M.P.; Orkey, N. & Aldrich-Wright JR. (2008). Drug Delivery Devices and Targeting Agents for Platinum(ii) Anticancer Complexes. *Australian Journal of Chemistry,* Vol.61, pp. 675-81, ISSN 0004-9425

Krejsa, C.; Rogge, M. & Sadee, W. (2006). Protein therapeutics: new applications for pharmacogenetics. *Nature Reviews Drug Discovery,* Vol.5, pp. 507-521, ISSN 1474-1776

Kuhn, E.; Addona, T.; Keshishian, H.; Burgess, M.; Mani, D.R.; Lee, R.T.; Sabatine, M.S.; Gerszten, R.E.; Carr, S.A. (2009). Developing Multiplexed Assays for Troponin I and Interleukin-33 in Plasma by Peptide Immunoaffinity Enrichment and Targeted Mass Spectrometry. *Clinical Chemistry,* Vol.55, pp. 1108-1117, ISSN 0009-9147

Lee, S.; Terry, D.; Hurst, D.R.; Welch, D.R.; & Sang, Q-X.A. (2011). Protein signatures in human MDA-MB-231 breast cancer cells indicating a more invasive phenotype following knockdown of human endometase / matrilysin-2 by siRNA. *Journal of Cancer,* Vol.2, pp. 165-176, ISSN 1837-9664

Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C.H.; Corthals, G.L.; Costello, C.E.; Deutsch, E.W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G.H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlén, M.; Wu, C.H.; Yamamoto, T.; Paik, Y.K. & Omenn, G.S. (2011) The human proteome project: current state and future direction. *Molecular and Cellular Proteomics,* Vol.10, M111.009993 ISSN 1535-9476

Li, C.I (2011). Discovery and Validation of Breast Cancer Early Detection Biomarkers in Preclinical Samples. *Hormones and Cancer,* Vol.2, pp. 125-131, ISSN 1868-8497

Marko-Varga, G.; Ogiwara, A.; Nishimura, T.; Kawamura, T.; Fujii, K.; Kawakami, K.; Kyono, Y.; Tu, H.; Anyoji, H.; Kanazawa, M.; Akimoto, S.; Hirano, T.; Tsuboi, M.; Nishio, K.; Hada, S.; Jiang, H.; Fukuoka, M.; Nakata, K.; Nishiwaki, Y.;Kunito, H.; Peers, I.; Harbron, C.; South, M.; Tim Higenbottam, T.; Nyberg, F.; Kudoh, S and I and Kato, H. (2007). Personalized Medicine and Proteomics: Lessons from Non-Small Cell Lung Cancer. *Journal of Proteome Research,* Vol.6, pp. 2925-2935, ISSN 1535-3893

Matigian, N.; Abrahamsen, G.; Sutharsan, R.; Cook, A.; Vitale, A.; Nouwens, A.; Bellette, B.; An, J.; Anderson, M.; Beckhouse, A.; Bennebroek, M.; Cecil, R.; Chalk, A.; Cochrane, J.; Fan, Y.; Féron, F.; McCurdy, R.; McGrath, J.; Murrell, M.; Perry, C.; Raju, J.; Ravishankar, S.; Silburn, P.; Sutherland, G.; Mahler, S.; Mellick G.; Wood, S.; Sue, C.; Wells, C. & Mackay-Sim, A. (2010). Disease-specific, neurosphere-derived cells as models for brain disorders. *Disease Models and Mechanisms,* Vol.3, pp. 785-798, ISSN 1754-8403

Mattick, J.S. (2009). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Annals of the New York Academy of Sciences,* Vol.1178, pp. 29-46, ISSN 0077-8923

Mercer, T.R.; Dinger, M.E.; Sunkin, S.M.; Mehler, M.F. & Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics,* Vol.10, pp. 155-159, ISSN 1471-0056

Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.; Harshman, K,; Tavtigian S.; Liu, Q.; Cochran, C.; Bennett, L.M.; Ding, W.; et al., (1994). A strong candidate for the breast

and ovarian cancer susceptibility gene BRCA1. *Science,* Vol.266, pp. 66-71, ISSN 0036-8075

Mozafari, M.R.; Pardakhty, A.; Azarmi, S.; Jazayeri, J.A.; Nokhodchi, A. & Omri, A. (2009). Role of nanocarrier systems in cancer nanotherapy. *Journal of Liposome Research,* Vol.19, pp. 310-21, ISSN 0898-2104

O'Farrell, P.H. (1975). High resolution two-dimensional electrphoresis of proteins. *Journal of Biological Chemistry,* Vol.250, pp. 4007-4021

Park, J.Y.; Kim, S.-A.; Chung, J.W.; Bang, S.; Park, S.W.; Paik, Y.K. & Song, S.Y. (2011). Proteomic analysis of pancreatic juice for the identification of biomarkers of pancreatic cancer. *Journal Cancer Research and Clinical Oncology,* Vol.137, pp. 1229-1238, ISSN 0171-5216

Pastwal, E.; Somiari, S.; Czyz, M. & Somiari, R. (2007). Proteomics in human cancer research. *Proteomics Clinical Applications,* Vol.1, pp. 4-17, ISSN 1862-8354

Patterson, S.; Van Eyk, J. & Banks, R. (2010). Report from the Wellcome Trust/EBI ''Perspectives in Clinical Proteomics'' retreat – A Strategy to Implement Next-Generation Proteomic Analyses to the Clinic for Patient Benefit: Pathway to Translation. *Proteomics Clinical Applications,* Vol.4, pp. 883-887, ISSN 1862-8354

Patton, W.F.; Schulenberg, B. & Steinberg, T.H. (2002) Two-dimensional gel electrophoresis: better than a poke in the ICAT? *Current Opinion in Biotechnology,* Vol.13, pp.321-328, ISSN 0958-1669

Pitteri, S.J,; Kelly-Spratt, K.S.; Gurley, K.E,; Kennedy, J.; Busald Buson, T.; Chin, A.; Wang, H.; Zhang, Q.; Wong, C.-H.; Chodosh, L.A ; Nelson, P.S.; Hanash, S.M.; & Kemp, C.J. (2011). Tumor Microenvironment-Derived Proteins Dominate the Plasma Proteome Response during Breast Cancer Induction and Progression. *Cancer Research,* Vol.71, pp. 5090-5100, ISSN 0008-5472

Prentice, R.L.; Paczesny, S.J.; Aragaki, A.; Amon, L.M.; Chen, L.; Pitteri, S.J.; McIntosh, M.; Wang, P. ; Buson Busald, T,; Hsia, J. ; Jackson, R.D. ; Rossouw, J.E. ; Manson, J.E. ; Johnson, K.; Eaton, C. & Hanash, S.M. (2010). Novel proteins associated with risk for coronary heart disease or stroke among postmenopausal women identified by in-depth plasma proteome profiling. *Genome Medicine,* Vol.2, pp. 48-60, ISSN 1756-994X

Rabilloud, T. (2002). Two-dimensional gel electrophoresis in proteomics : Old, old fashioned, but it still climbs up the mountains. *Proteomics* Vol.2 pp. 3-10, ISSN 1615-9853

Remily-Wood, E.; Liu, R.; Xiang, Y.; Chen, Y.; Thomas, C.; Rajyaguru1, N.; Kaufman, L.; Ochoa, J.; Hazlehurst, L.; Pinilla-Ibarz, J.; Lancet, J.; Zhang, G.; Haura, E.; Shibata, D.; Yeatman, T.; Smalley, K.; Dalton, W.; Huang, E.; Scott, E.; Bloom, G.; Eschrich, S. & Koomen, J. (2011). A database of reaction monitoring mass spectrometry assays for elucidating therapeutic response in cancer. *Proteomics Clinical Applications,* Vol.5, pp. 383-396, ISSN 1862-8354

Ross. P.L.; Huang, Y.N.; Marchese, J.N.; 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. & Pappin, D.J. (2004). Multiplexed Protein Quantitation in Saccharomyces cerevisiae using amine-reactive isobaric

tagging reagents. *Molecular and Cellular Proteomics,* Vol.3, pp. 1154-1169, ISSN 1535-9476

Schoenherr, R.M.; Zhao, L.; Whiteaker, J.R.; Feng, L.-C.; Li, L.; Liu, L.; Liu, X. & Paulovich, A.G. (2010) Automated Screening of Monoclonal Antibodies for SISCAPA Assays using a Magnetic Bead Processor and Liquid Chromatography-selected Reaction Monitoring-mass Spectrometry. *Journal of Immunological Methods,* Vol.353, pp. 40-61, ISSN 0022-1759

Simpson, R.J.; Berhhard, O.K.; Greening, D.W. & Moritz, R.L. (2008). Proteomics-driven cancer biomarker discovery: looking to the future. *Current Opinion in Chemical Biology*, Vol.12, pp. 72-77, ISSN 1367-5931

Semaan, S.M. ; Wang, X. ; Stewart, P.A.; Marshall, A.G. & Sang, Q.X.A (2011). Differential phosphopeptide expression in a benign breast tissue, and triple-negative primary and metastatic breast cancer tissues from the same African-American woman by LC-LTQ/FT-ICR mass spectrometry. *Biochemical and Biophysical Research Communications,* Vol.412, pp. 127-131, ISSN 0006-291X

Sherman, J.; McKay, M.; Ashman, K. & Molloy, M.P. (2009). Unique Ion Signature Mass Spectrometry, a Deterministic Method to Assign Peptide Identity. *Molecular and Cellular Proteomics,* Vol.8.9, pp. 2051-2062, ISSN 1535-9476

The International Cancer Genome Consortium (2010). International Network of Cancer genome Projects. *Nature,* Vol.464, pp. 993-998, ISSN 0028-0836

Thingholm, T.E.; Palmisano, G.; Kjeldsen, F. and Larsen, M.R. (2010). Undesirable Charge-Enhancement of Isobaric Tagged Phosphopeptides Leads to Reduced Identification Efficiency. *Journal of Proteome Research*, Vol.9, pp. 4045-4052

Tonge, R.; Shaw, J.; Middleton, B.; Rowlinson, R.; Rayner, S.; Young, J,; Pognan, F.; Hawkins, E.; Currie, I. & Davison, M. (2001). Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics,* Vol.1, pp. 377-396, ISSN 1615-9853

Walsh G. (2010). Biopharmaceutical Benchmarks. *Nature Biotechnology,* Vol.28, pp. 917-924.

Weinshilbaum, R. & Wang, L. (2004). Pharmacogenomics: Bench to Bedside. *Nature Reviews Drug Discovery,* Vol.5, pp. 38-49, ISSN 1474-1776

Whiteaker, J.R.; Zhao, L.; Anderson, L. & Paulovich, A.G. (2010). An Automated and Multiplexed Method for High Throughput Peptide Immunoaffinity Enrichment and Multiple Reaction Monitoring Mass Spectrometry-based Quantification of Protein Biomarkers. *Molecular and Cellular Proteomics,* Vol.9, pp. 184-196, ISSN 1535-9476

Wilkins, M.R.; Pasquali, C.; Appel, R.D,; Ou, K,; Golaz, O.; Sanchez, J.-C.; Yan, J.X.; Gooley, A.A.; Hughes, G.; Humphery-Smith, I,; Williams, K.L. & Hochstrasser, D.F. (1996). From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. *Biotechnology,* Vol.14, pp. 61-65, ISSN 0733-222X

Wooster, R.; Bignell, G.;, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. et al., (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature,* Vol.375, pp. 789-792, ISSN 0028-0836

Zanivan, S.; Krueger, M. & Mann, M. (2012). In vivo Quantitative Proteomics: The SILAC mouse. *Methods in Molecular Biology,* Vol.757, pp. 435-450, ISSN 1064-3745

Zhan, X. & Desiderio, D.M. (2010). The use of variations in proteomes to predict, prevent, and personalize treatment for clinically nonfunctional pituitary adenomas. *EPMA Journal,* Vol.1, pp. 439-459, ISSN 1878-5077
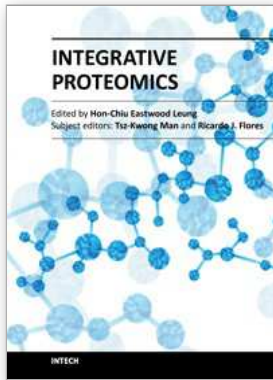
Integrative Proteomics

Edited by Dr. Hon-Chiu Leung

Proteomics was thought to be a natural extension after the field of genomics has deposited significant amount of data. However, simply taking a straight verbatim approach to catalog all proteins in all tissues of different organisms is not viable. Researchers may need to focus on the perspectives of proteomics that are essential to the functional outcome of the cells. In Integrative Proteomics, expert researchers contribute both historical perspectives, new developments in sample preparation, gel-based and non-gel-based protein separation and identification using mass spectrometry. Substantial chapters are describing studies of the sub-proteomes such as phosphoproteome or glycoproteomes which are directly related to functional outcomes of the cells. Structural proteomics related to pharmaceutics development is also a perspective of the essence. Bioinformatics tools that can mine proteomics data and lead to pathway analyses become an integral part of proteomics. Integrative proteomics covers both look-backs and look-outs of proteomics. It is an ideal reference for students, new researchers, and experienced scientists who want to get an overview or insights into new development of the proteomics field.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Amanda Nouwens and Stephen Mahler (2012). Evolution of Proteomic Methods for Analysis of Complex Biological Samples – Implications for Personalized Medicine, Integrative Proteomics, Dr. Hon-Chiu Leung (Ed.), ISBN: 978-953-51-0070-6, InTech, Available from: http://www.intechopen.com/books/integrative-proteomics/evolution-of-proteomic-methods-for-analysis-of-complex-biological-samples-implications-for-personali

INTECH

open science | open minds