We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Validation of Protein Biomarkers to Advance the Management of Autoimmune Disorders

David S. Gibson[1,2], Sorcha Finnegan[1], Stephen Pennington[3], Ji Qiu[4],
Joshua LaBaer[4], Madeleine E. Rooney[1] and Mark W. Duncan[2,5]
[1]Centre of Infection and Immunity, Queen's University of Belfast,
[2]Division of Endocrinology, Metabolism and Diabetes, University of Colorado, Denver,
[3]Proteome Research Center, Conway Institute, University College Dublin, Dublin,
[4]Center for Personalized Diagnostics, Arizona State University, Tempe,
[5]Obesity Research Center, King Saud University,
[1]United Kingdom
[2,4]United States of America
[3]Ireland
[5]Saudi Arabia

1. Introduction

Despite the anticipated boom stemming from proteomic investigations, the rate at which novel protein biomarkers are introduced into clinical practice has remained static over the past 20 years. The reality is that approaches to both discover and validate protein biomarkers remain inadequate, and consequently, many areas of medicine, including the broad field of autoimmune disorders, remain deprived of the tools essential for the optimal management of patients. Most importantly, there is a huge backlog of candidate biomarkers that are yet to undergo thorough investigation and validation to assess their clinical utility. A recent assessment of the situation has estimated that although many tens of thousands of publications claim biomarker discoveries, there are roughly only 100 routinely used in clinical practice (Poste, 2011).

This chapter reviews the potential applications of protein biomarkers to manage autoimmune diseases with a special focus on the transition from the biomarker discovery through to validation phases using proteomic strategies. We emphasize the importance of careful review of the discovery data, the critical roles of protein isoform verification, and the essential features of targeted and thorough validation. Ultimately, when these factors are appropriately considered and implemented, we are optimistic that autoimmune disorders can be transformed by omics technologies and personalized practice can become a reality.

1.1 Biochemical markers and their potential role in autoimmune disease

Biological markers are widely used in medicine and can provide an objective measure of normal and pathogenic processes or pharmacologic responses to a therapeutic intervention. By the term '*biological markers*' (or biomarkers) we mean an objective molecular indicator or surrogate of pathological processes which possess diagnostic, prognostic or predictive

utility. These are distinct from '*clinical markers*' which rely on physical variables or symptoms such as joint count, pain assessment or radiological findings. Biomarker examples include: cardiovascular risk assessment through cholesterol checks; pituitary and target gland hormone determinations to assess endocrine function and dysfunction; hemoglobin A1c (HbA1c) evaluations to monitor blood glucose levels in diabetic patients; liver function tests (LFT) in liver disease; and, prostate-specific antigen (PSA) determinations to assess prostate cancer risk. Not surprisingly, there is considerable interest in developing additional clinical biomarkers in medicine; however, the path from their discovery to routine adoption is painstakingly complex and slow.

1.2 Why target protein biomarkers

Although genomic and transcriptomic methods are powerful, they cannot predict downstream events. Specifically, they can't predict what protein forms will be expressed in a particular tissue or biofluid, nor can they reliably estimate expressed protein levels. Because it is the gene products that contribute directly to physiological or pathological change, they alone provide the best clues to function in health and dysfunction in disease. Just as importantly, protein modifications including a plethora of post-translational changes are not evident upstream. As discussed later, proteins may require cleavage of a specific sequence to become biologically active, additional sidegroups may be added at specific amino acids including phosphorylation to propagate signal transduction or glycosylation to transport the molecule to a specific site. The analysis of hundreds if not thousands of proteins – proteomics – is therefore potentially the most illuminating of all multiplexed strategies.

Proteomics, however, is an imperfect science, and although its methods are rapidly evolving, it is important to acknowledge that all existing approaches have serious limitations. Notably, the available discovery tools offer poor precision, sensitivity, specificity and low throughput. An example of a low throughput proteomic platform is classic liquid chromatography separation of a single complex biological sample, coupled with electrospray ionization to generate of mass spectra. These limitations place severe constrains on study design (*e.g.*, small sample numbers) and they can lead investigators to place a disproportionate confidence in the data generated in a discovery setting. Regardless of these shortcomings, proteomic methods have been widely adopted and have generated many potential candidate markers. As we will illustrate, candidates identified using these techniques should be considered a set of ideas or leads from which the investigator can generate testable hypotheses. The study design and process of testing or validating these candidates is especially important if the test is to be applied in a clinical setting.

This chapter focuses on several key areas of the biomarker development process and uses rheumatoid arthritis as a case model of autoimmune disease for discussion. We first discuss the issue of clinical need, *i.e.*, how biomarkers *are currently* used in the practice of rheumatology what biomarkers might offer in *the future* in a clinical setting, then highlight some clinical scenarios describing considerations for study design. Because discovery methods have been reviewed elsewhere (Gibson & Rooney, 2007; Hu et al., 2006; Tilleman et al., 2005) our emphasis here is on describing and evaluating the targeted proteomic methods that are essential to candidate validation. It is the authors' belief that the task of validation has not received adequate attention. We are overwhelmed with discovery studies and data arising from them; validation approaches are infrequently reviewed and relatively little

validation data have been reported. The significance of major barriers to their widespread clinical adoption is also briefly discussed. These challenges to new marker development and commercialization include factors such as study design, pre-analytical variables, data interpretation, bioinformatics, validation, ethics and commerce. While our focus is on protein biomarkers and proteomic methods in rheumatology, the principles we discuss are generally applicable to other analytes and areas.

1.3 Biomarkers currently used in the practice of rheumatology

The biomarkers currently used in the practice of rheumatology can be subdivided into three distinct molecular classes: nucleic acid, proteins, and metabolites. Some of these tests and their applications are included in Table 1. This is not an exhaustive list and while an invaluable armamentarium, it is limited in number, clinical utility and performance. With the exception of anti-CCP autoantibody and S100 proteins, the list has remained essentially static for the last 5-10 years. Additional potential biomarkers have been reported, but few, if any, have been adopted in routine practice. There is, however, no dispute that better tools are urgently required for both objective diagnosis and optimal management of rheumatoid arthritis.

1.4 The future: additional markers & rheumatology practice

Several areas of medicine lack objective, quantitative measures of a disease and its response to therapy and the inflammatory forms of arthritis are no exception. Below, we define three distinct stages in the clinical progression to chronic inflammatory disease exemplified by some autoimmune disorders and outline where and how biomarkers could aid in managing the RA patient.

i. *The pre-symptomatic stage*: Here asymptomatic individuals, especially those who are genetically susceptible to rheumatoid arthritis (RA), need to be screened for early indications of disease onset. Biochemical markers could remove uncertainty in detecting the disease in its early stages and would allow early and appropriate intervention. Such early detection could help clinicians minimize or even halt disease progression, reduce morbidity and mortality, and markedly lower the costs of health care delivery in rheumatology.

ii. *The early clinical stage*: In the early stages of rheumatic diseases patients may be symptomatic but there are insufficient clinical or laboratory findings to confirm the diagnosis. A few biomarkers exist which are helpful in predicting outcome at a population level, such as rheumatoid factor and, more recently, anti-cyclic citrullinated peptide antibodies (anti-CCP) for the diagnosis of rheumatoid arthritis (See Table 1). Clinicians believe that early diagnosis and treatment are essential for the best outcome. In the past, the presence of the biomarker, IgM rheumatoid factor (RF), has helped to identify patients likely to have more aggressive disease (Vocovsky et al., 2003). In the last decade, anti-CCP antibodies have been shown to predict a more aggressive disease course in very early disease (van Venrooij, et al., 2008). However, the relatively low sensitivity and specificity of their assay offers relatively low diagnostic sensitivity and this means that a significant minority of patients with aggressive disease cannot be identified (Kastbom et al., 2004; Lindqvist et al., 2005). Furthermore, neither RF nor anti-CCP antibodies are of any use in the other inflammatory arthritides of adulthood, such as psoriatic arthritis and ankylosing spondylitis, or for the vast majority of children with juvenile idiopathic arthritis.

Marker	Molecular Class	Example Application
Creatinine	Metabolite	*Creatinine* is used together with liver enzyme determinations (AST and ATL) as an index of drug toxicity and kidney function (Anders et al., 2002).
C-reactive protein (CRP)	Protein	A plasma protein routinely measured as a non-specific index of acute inflammation. CRP levels are typically integrated into clinical response scores, such as DAS28, which help to titer drug dosage (Pepys & Hiirschfield, 2003).
S100 proteins	Protein	Phagocyte derived proteins found in a variety of inflammatory diseases , with the ability to discriminate between diseases (Wittkowski et al., 2008).
Anti-nuclear antibody (ANA)	Protein- Auto-antibody	Used along with other tests, ANA helps with the diagnosis of arthritides. ANA titers and nuclear stain patterns can vary between patients depending on the condition ($\approx$ 95% of systemic lupus erythrematosis (SLE)) (Wilk, 2005). Specific subsets of ANA's can be used to distinguish the type of autoimmune disease, *e.g.,* Sjögren's syndrome. ANA positivity increases the risk of eye disease in juvenile idiopathic arthritis patients.
Rheumatoid Factor (RF)	Protein- Auto-antibody	RF antibodies target the Fc region of IgG and are detected in 60-80% of RA patients, though are present in other inflammatory and connective tissue diseases (Chen et al. , 1987).The presence of RF early in the course of RA is associated with more active disease (Nakamura, 2000).
Anti-cyclic citrullinated peptide (CCP)	Protein- Auto-antibody	Anti-CCP antibody positivity predicts the development of RA and may occur long before the onset of symptoms (Nielen, 2004). Anti-CCP is associated with severe erosive disease and can predict disease progression in RA patients (Meyer et al., 2003).
Anti-Ds DNA antibodies	Protein-Auto-antibody	Anti dsDNA antibodies are highly specific for the diagnosis of SLE with a specificity of 95% but with a low sensitivity of <60% (Kavanaugh et al., 2002).
HLA-DR4 or HLA-DRB	Nucleic acid-Gene	Human leukocyte antigen genes have been found to be associated with RA. This association is particularly strong for HLA-DRB-1 alleles which share a similar amino acid sequence known as the shared epitope (van der Horst-Bruinsma et al., 1999). The presence of these alleles both increases the risk of RA and associate with more severe disease (Wagner et al., 1997).
PTPN22	Nucleic acid-Gene	The protein tyrosine phosphatase, non-receptor 22 (PTPN22) allele is a major risk factor for several autoimmune diseases. The protein product increases the tyrosine dephosphorylation of T-cell receptor resulting in decreased signaling via this pathway (Vang et al., 2005).
TRAF1/C5	Nucleic acid-Gene	TNF-receptor-associated factor-1 is one of many single nucleotide polymorphisms involved in the pathway of tumour necrosis factor alpha. The protein encoded by TRAF1 mediates signal transduction from the family of TNF receptors (Kurreeman et al., 2007). Increased susceptibility to and severity of RA is associated with this SNP, by influencing TRAF1/C5 function.
TLR/ TNFR/ NF-κB	Nucleic Acid-Transcript	Pharmacogenomic tests which use cellular transcript measurements to predict drug response are yet to be implemented in the clinic. Interesting data has emerged on the association between anti-TNF antagonist response and genetic variation (Bowes et al., 2009; Potter et al., 2010).

Table 1. Biomarkers routinely used in the diagnosis & treatment of arthritides and some potential future markers-arthritides.

iii. *The disease management stage*: Early commencement of effective therapy is essential if joint damage and other complications are to be avoided. Historically, monitoring response to treatment is a composite of clinical findings and laboratory markers such as erythrocyte sedimentation rate (ESR), C-reactive protein (CRP) and disease activity score (e.g. DAS28) . Treatment is modified according to these parameters. However disease response can take many months, indeed years. Thus by the time the patients disease is deemed unresponsive substantial joint damage can have occurred. The identification of biomarkers that would predict disease response would have an enormous impact on outcome. Treatment may also be discontinued due poor tolerability. Identifying such patients in advance would improve patient care and reduce stress. Biologic drugs, as third line therapy such as anti-TNF have revolutionised the treatment of rheumatic diseases and systematic reviews have confirmed their efficacy and relative safety (Alonso-Ruiz et al. 2008). However these drugs are extremely expensive. Months of treatment can be required before the clinician knows whether they are effective. This is both costly and inefficient. The identification of biomarkers that would predict the response of individual patients to these expensive agents would help patients, clinicians and funding agencies alike. Finally there are concerns associated with the use of such targeted therapy i.e. the risk of life threatening infections using and more worryingly the long term risk of malignancy (Bongartz et al. 2006). The ability to identify such patients in advance would protect them from such serious adverse reactions.

2. The proteomics biomarkers development pipeline

2.1 Overview

There is no dispute that new biomarkers would advance the diagnosis and management of autoimmune disorders. The ongoing challenge, however, is how to discover candidate markers and how to validate them, *i.e.,* define their performance characteristics when adopted in a routine clinical setting.

A major impetus for increased interest in biomarkers has been the introduction of the omics technologies. In a single study these allow interrogation of hundreds (or thousands) of independent variables, such as genes, mRNA, metabolites or proteins and given the volume of information generated from such studies, many have anticipated candidate biomarkers would flow quickly from each new investigation. The reality, however, is otherwise. Comparing the levels of hundreds (or thousands) of data points in several distinct groups, especially when the sample numbers are small, gives rise to many apparent differences, only some of which are related to biology. Chance alone gives rise to many apparent "distinguishing features" - the trick is identifying the biologically relevant differences and ignoring the others.

For example, consider a proteomics experiment in which 200 proteins are measured simultaneously in a control and a test sample. At $P < 0.05$, 10 proteins will appear to be different by chance alone and unrelated to the treatment or condition. This consideration is not presented to undermine the value or utility of omics research, but rather, to underscore the importance of verifying any observed differences in follow-up studies. In more conventional scientific studies it is typical to examine a single dependent variable, run replicates, and to use standard statistical approaches to analyze the outcomes.

Omics studies are very different. Data sets are typically of high dimensionality, but the sample size is small. There are typically very few if any replicates and any interesting trends are hidden within the combination the variables. Under these circumstances the probability of finding associations by random chance is high. Although multivariant statistics can help with the analysis of these complex data sets, there is no easy way out. Any single omics study is best considered as an observational investigation that aids in generating novel hypotheses that can direct their future studies. Contrary to what was hoped, the omics methods do not provide a fast-track to biomarkers or shortcut the scientific process. They do, however, allow an investigator to operate independent of existing knowledge and to be less dependent on insight, instinct or experience. A single omics study can provide data from which dozens of testable hypotheses can be formulated or, put another way, it can identify dozens of biomarker *candidates*. Accordingly, the validation of each candidate biomarker is analogous to hypothesis testing where the investigator sets out to falsify (or disprove) the claim that candidate "x" is a valid biomarker in defined clinical scenario "y".

In the sections that follow some of the other important components of the biomarker development pipeline are discussed and we highlight the primary concerns that are necessary to optimize the success of biomarker development.

2.2 The clinical objective and study design considerations

Several types of biomarkers can be developed, but in each instance the process requires a different study design and unique sample sets. Biomarker development in autoimmune disease must incorporate a cross-section of patients representing the full spectrum of the specific disorder, and given its complex and heterogeneous nature, a panel of markers, not a single marker, will likely be necessary to reflect all relevant clinical parameters. Depending on the groups incorporated into the study and the comparisons made, integrated panels of individual biomarkers may be identified that provide valuable screening/diagnostic, predictive or prognostic information.

Screening biomarkers: Biomarkers that can be used to screen and identify disease before the onset of any symptoms are the Holy Grail of autoimmune disease; however, their discovery and validation presents substantial challenges. While some of the biomarkers evident in symptomatic disease *may* be present early on in asymptomatic individuals, it is more likely that these will be low abundance and masked by more abundant, non-specific biomarkers of inflammatory and secondary processes related to the disease. Therefore, reliable identification of the biochemical events that earmark early stage, asymptomatic disease requires access to biobanks with adequate numbers of samples to give statistical power collected from affected individuals well before disease onset. (In the case of juvenile RA these samples should ideally be collected from birth onwards.) When samples are available from the same individual both before and after disease onset, patients can serve as their own controls and therefore changes characteristic of the disease can be measured against a relatively constant biochemical background. Sample sets for these studies are, however, difficult to come by and require substantial long-term logistical and financial investment. Consequently, a compromised study design incorporating a relevant control group and early stage (symptomatic) subjects is more commonly adopted to meet this objective.

Diagnostic biomarkers: Typically, a case-control approach is used in this setting, such that a cohort of disease-free controls is compared to a similarly-sized cohort of diseased subjects. Comparisons of this type require careful design and implementation because observed

differences may be non-specific and associated with the consequences of end-stage disease, rather than related to RA itself. Although widely used, for this reason alone the case-control study design is frequently problematic. Case and control samples need to be age- and lifestyle-matched, detailed clinical histories must be available on both cohorts, and strict inclusion/exclusion criteria are required. The value of any markers identified by this approach, and especially those that are not specific to autoimmune disorders alone, can only be assessed through large-scale hypothesis-driven studies performed in a defined clinical setting.

Prognostic or predictive biomarkers: For a prognostic or predictive biomarker study, suitable samples must be available both before and after the measured outcome from each subject. Often these studies use samples collected from completed studies that addressed a different clinical question, but are then used to identify and track a molecular signature (biomarker profile) that may have predictive/prognostic value. (*i.e.,* They employ retrospective samples.) However, in all instances, prospective (purpose-driven) sample collection is preferred. Although logistically difficult, this approach affords greater control over pre-analytical variables including storage time, storage conditions and use of additives. With foresight and planning, a randomized controlled trial with longitudinal sample collection can incorporate multiple nested outcome studies relating, for example, to therapeutic response, disease progression or disease recurrence.

2.3 Samples and sampling considerations

It is important to state the obvious: a biomarker study can only be as good as the clinical samples and their associated records. If there are errors with annotation, if patient details are inaccurate, or if the samples themselves have not been collected and stored properly, the exercise of biomarker development may be futile (Poste, 2011). Although there are limited numbers of biobanks available at this time, thankfully, more and more investigators, research centers and commercial entities are committing to establishing and maintaining high-quality sample repositories linked to accurate clinical records.

With respect to samples, factors such as the timing from sample collection to freezer, the complexity and reproducibility of any sample handling steps, length of storage time, storage temperature and freeze-thaw cycles, may affect the stability of some analytes. For example, samples sourced from different cohorts at different locations may 'carry forward' pre-analytical background signals with discriminating features unrelated to the biology of the disease (Addona et al., 2009; Davis et al., 2010; Ransohoff, 2010). Ideally, samples should be processed immediately, then aliquoted into airtight tubes and frozen in liquid nitrogen or -70°C freezers. Similarly, multiple freeze-thaw cycles can affect the stability of potential biomarkers (Flower et al., 2000; Rai et al., 2005). Protease inhibition may help preserve sample integrity, but this approach is not without its complications. For example, irreversibly-binding to sample components can have undesirable downstream consequences.

Fluids or tissues proximal to the sites of pathology can act as biomarker sinks. As a result, these should also be considered, when available, alongside plasma or serum as a means of focusing the search on pathologically-relevant candidates. For example, in the case of arthritis, synovial fluid, cartilage and synovium are potential sources for biomarker discovery. Here protein variants unique to the site and pathology can be measured before they escape into plasma (Gibson et al., 2009).

Studies indicate that plasma is likely a better substrate for proteome analysis than serum due to the obfuscation of results associated with the high proportion (>40%) of clot-related proteins and peptides in serum (Haab et al., 2005; Rai et al., 2005; Tammen, 2005). Less invasive samples also amenable to protein biomarker discovery include urine, saliva and tear fluid. Although putative biomarkers can come from discovery work, candidates can also come from literature searches and genomic or transcriptomic mining. All candidates, however, must undergo subsequent verification and validation (Pepe et al., 2008).

2.4 Discovery strategies

Discovery strategies allow many analytes to be measured simultaneously (*i.e.*, multiplexed analysis). The objective is to identify qualitative and/or quantitative differences across distinct clinical phenotypes that are reproducible and can then be adopted in a clinical setting. As discussed previously, however, what is observed in a discovery setting could be an artifact of statistical chance or experimental bias, and any findings must be rigorously validated. Discovery is typically costly, slow (low-throughput) and labor-intensive. Further, because the methods are not optimized for any single analyte, their performance characteristics are compromised (*i.e.*, limited sensitivity, selectivity and precision). The methods are therefore only suitable to survey – they are not suited to efficient, precise and accurate quantification. When proteins are the targets of the discovery process then there are two orthogonal strategies that are adopted: peptide-centric and protein-centric.

Peptide-centric (bottom-up or shotgun) strategies: This approach begins with proteolytic digestion of proteins to peptides and the 'digest' is then subjected to fractionation (HPLC) and tandem mass spectrometry (Duncan et al., 2010; Aebersold & Mann, 2003; Chait, 2006). The tandem mass spectra are then converted to peptide sequences and their precursor proteins are "assumed" by computational approaches. Refinements of this approach sometimes incorporate fractionation prior to digestion and/or multiple stages of fractionation post digestion (Washburn et al., 2001; Wolters et al. 2001). The principal assumption of a bottom-up strategy is that the identity of intact proteins can be ascertained from their constituent peptide fragments. As discussed elsewhere, this assumption is frequently invalid (Duncan et al., 2010).

Protein-centric (top-down) strategies: With protein-centric approaches intact proteins are first separated, typically by 2D gel electrophoresis, then the proteins are isolated and identified by mass spectrometry. Typically identification involves enzymatic cleavage of each individual protein to peptides and then either: (a) the masses of the peptide products of each pure protein are determined (*via* single stage mass spectrometry); or (b) the tandem mass spectrum (fragmentation pattern) of one (or more) of the peptides is determined (*via* tandem mass spectrometry). One or both these data sets is/are then used to interrogate a database and identify the protein. Relative protein amounts can be determined from the gel by staining. Because a top-down approach retains the protein integrity, modifications and sequence variations can be investigated.

As we will illustrate, the discovery findings should be considered a set of leads that require meticulous validation, especially with respect to the utility of the biomarker(s) in a routine clinical setting.

2.5 Biostatistical considerations

Because proteomic studies of clinical samples can generate cumbersome data sets, bioinformaticians are frequently involved in study design (*e.g.*, patient selection and study

size calculation) and the hunt for significant and reproducible patterns in the data. Their objective is to find reproducible differences which correlate with a defined clinical outcome and that are independent of the influence of experimental bias, over-fitting and statistical chance.

The incorporation of a randomization strategy in sample analysis reduces bias by accounting for the day-to-day variations in the analytical technique. Similarly, it is prudent to calibrate and record the performance characteristics of the instruments used in the analyses. Calibration in proteomic analyses entails, for example in mass spectrometry, initialising the mass accuracy to a standard mixture of purified proteins or peptides of known mass. Routine calibration of sensitive instruments subject to 'drift' in measurement over a period of time should become part of good laboratory practice. Further, in the discovery phase, the objective is to have sufficient sample numbers to provide confidence that the list of protein candidates is worthy of follow-up during the validation phase. Typically in this phase of biomarker development, the sample size is small due to the cost and time of analysis, and sometimes because of the difficulty associated with obtaining samples. However, the number of proteins (independent variables) measured in each sample is typically very large. This ratio of samples to variable size is contrary to the traditional application of multivariate statistics and leads to some unique considerations that have been discussed by others (Dowsey et al., 2009; Karp & Lilley, 2007). Conversely in the validation phase this relationship is inverted so that patient cohorts are much larger (typically 100's -1000's) and the number of biomarker candidates carried over from discovery are reduced depending on the strength of their relationship to the clinical outcome or measure being assessed. The costs incurred by the validation phase therefore sit in a multi-million dollar range far exceeding the costs of discovery. The financial implications alone may account for the relative dearth of publications on this phase as the main players, large pharmaceutical or diagnostic corporations, having invested large amounts of time and money likely strive to protect the resultant intellectual property prior to further clinical testing and pre-market approval.

Bias, or any discrimination occurring due to a non-biological signal, can potentially confound discovery. For example, spurious results may arise because of differences in how patient samples are collected, *e.g.,* type of blood collection tube, time taken to freeze sample, or the order in which the samples are analyzed. Over-fitting can occur when regression analysis tools are used to 'fit' (too) many variables to a limited set of outcomes. The discriminating 'pattern' or 'signature' then becomes an artifact of the patient cohorts. To resolve issues of bias, statistical analyses must consider the biology of the system being analyzed and take into account the assumptions and limitations of the methods (Ransohoff, 2009). Statistical tests capable of gauging the level of false positives across multiple comparisons include the student's unpaired t-test (for two group comparisons), ANOVA (for three or more group comparisons) and linear regression (for quantitative or correlative studies) (Dowsey et al., 2009). Alternatively, if the data are not normally distributed, non-parametric Mann Whitney and Kruskal-Wallis tests should be substituted (Karp & Lilley, 2007). These methods can be used to analyze the features one-at-a-time and then to compile a ranked list of them based on a combination of p values and effect size. As noted earlier, a longitudinal design can minimize the potential for bias relative to a typical case-control study. Nevertheless, false biomarker leads are common and therefore rigorous validation essential.

The false discovery rate (FDR) can also be calculated (Benjamini et al., 2003; Storey, 2003; Strimmer, 2008). By setting the FDR level, it is possible to diminish the risk of a false positive identification for a differentially-expressed protein, i.e., at $P \leq 0.05$, we expect only 5% false positives. However, by doing so, the process of discovery may be compromised by overly stringent criteria. Although proteins displaying the most dramatic changes may appear to be useful biomarkers, it is important to attempt to rationalize their changes to the pathology. For example, acute phase proteins are frequently identified in plasma or serum-based studies as 'specific' biomarkers of a wide range of chronic disorders, including arthritis and cancer but clearly they are not specific to any one disease (Addona et al., 2009).

Cross-validation procedures can be used to reduce false positives. In this instance one data set is used to build the model (called training) and a second data set generated from an independent patient cohort is used to assess the predictive accuracy of the model (called testing). Another commonly-used validation strategy is known as K-fold cross-validation where the analysis is repeated over many random splits of the data. For each analysis, a subset of the data is used to build K number of predictive models, with the remaining subset available for a test of predictive accuracy. Although useful initially after discovery, validation based on splitting a single data set is of limited use because confounding factors can introduce systematic biases into both training and test splits.

Given the issues noted above, it advisable to validate intial 'discoveries' on independent sample sets, perhaps incorporating analysis by orthogonal methods which are more amenable to the requirements of clinical throughput and precision (Dupuy & Simon, 2007). Re-analysis or meta-analysis using raw data coming from other research groups is another possibility, although data standards, such as the 'minimum information about a proteomics experiment' (MIAPE) (Taylor et al., 2007), often do not extend into the initial design of clinical studies. Consequently, detailed clinical data may not be captured and reported consistently for clinical proteomics experiments, limiting the ability of investigators to independently verify, combine or correlate data from multiple experiments.

For thorough validation, the number of patient samples required should be determined through the use of statistical tools that take into account the imprecision of the analytical method, inter-patient variability and the acceptable threshold of difference that is deemed significant for a given biomarker application (Ye et al., 2009). Patient numbers (biological replicates) and other statistical considerations of power have also been discussed in detail elsewhere (Cairns et al., 2009).

2.6 Feature selection and classifier assessment

Several multivariate analysis tools are available for the analysis of large multidimensional data sets and some of these have been arranged into commercial software packages. Visual tools, including principle component analysis, hierarchical cluster analysis and heat maps which display variance, relatedness and patterns in data (respectively), are also available and are useful preliminary aids in data analysis. These analyses stive to represent variance in a graphical fashion and give for example an overall view protein expression prevalence within outcome groups in the case of heat maps or 'relatedness' of expression levels between different proteins with hierarchical trees (Marengo et al. 2006; Marengo et al. 2008). Emphasis however, should be placed on using supervised or semi-supervised methods such as distribution free learning (kernel- based or Bayesian analysis) or support vector machine (SVM) which allow for advanced categorization and classification of multidimensional

proteomic data with respect to clinical data. This process known as 'feature selection' and leads ultimately to the creation of a 'classifier' or biomarker driven algorithm specific to the disease and outcome being measured (Liu et al., 2009; Zhu et al., 2009). Since most protein expression profiles will likely not be correlated to a specific outcome, supervised methods screen out uninformative proteins and select protein combinations to develop a 'classifier'. A recent application of the SVM principle has been used to guide feature selection of exhaled peptides as potential biomarkers of asthma (Bloemen et al., 2011).

Depending on whether proteome studies are focused on biomarkers for (i) diagnosis (class discovery), (ii) prognosis (outcome-related) or (iii) prediction (supervised prediction), various rationales should be employed to generate and assess the reliability a classifier. Class discovery methods are best suited for grouping proteins into subsets that elucidate pathways with similar expression profiles across patient subgroups. In outcome-related studies, the goal is to identify which proteins have expression levels that correlate with outcomes grouped into discrete classes: for example, in arthritis patients with a good versus a bad prognosis. When prediction of patient outcome is the aim, supervised prediction methods that use a selected proteome profile are used to generate an algorithm based on individual profiles. In supervised class prediction studies, a totally independent cohort should be used for cross-validation purposes when rigorous testing of a predictive model is desired (Dupuy & Simon, 2007).

The statistical significance of a selected proteome 'classifier' gives an incomplete estimate of its predictive ability and potential clinical utility. The number of true and false positives or negatives should be presented allowing the calculation of sensitivity and specificity. This reveals clinically-relevant information on how the classifier performs in each outcome category. List of statistical tools and recommendations for their application have been reported (Dupuy & Simon, 2007; Karp & Lilley, 2007; Marengo et al. 2006). Depending on the clinical question there may however be multiple outcome measures that are not amenable to a simple binary classification system. Statistical evidence of prevalence and analytical limits of detection of a specific group of isoforms can then direct the study towards validation of candidate biomarkers in a much larger group of multi-center patient populations.

3. The biomarker development process

Three distinct phases can be delineated within a typical development pipeline: discovery, verification and validation. These can be further subdivided so there is a reduced number of candidates at each stage, each with an increased probability of utility [51]. In the subsequent sections we aim to clearly segregate these phases in the biomarker 'pipeline' and further expand on the vastly different requirements of each (Figure 1). This process is prefaced by a brief overview of pre-analytical factors which can introduce unwanted bias or variation.

3.1 The discovery phase

In the discovery phase, proteomic platforms are unsupervised and are used to highlight qualitative and/or quantitative differences in multiple proteins across distinct clinical phenotypes. The process of discovery is focused on assessing many candidates, while minimizing the probability of false positives and negatives.

Discovery by definition requires an analytical approach which does not preempt the identity of the biomarker candidates. Generally speaking as most discovery methods prioritise the

measurement of as many proteins as possible they have inherently low throughput, are labor intensive and offer a low dynamic range. These characteristics preclude their use in later phases of the biomarker pipeline. It is also important to realize that as yet there is no single method available for looking at the complete complexity of the proteome within a given clinical sample. Because we are working with a relatively blunt set of tools in discovery we need to transition to more precise methods for validation.

A two-step approach to the discovery phase, though widely used, is not well defined in the literature. In the initial pilot exploration of a low number of individuals the aim is to gain a grasp of the variability of whole proteome being measured across the cohort, selecting a suitable sample type, optimizing the separation and quantification platform and ultimately calculating appropriate patient numbers to power a second (discovery) round with greater statistical confidence.

Phase	Aim	Platform	Precision/ Accuracy	Sample Turnover
Candidate Discovery	Relative, Unsupervised	2DE DIGE LC-MS	Low	Low
Specificity Verification	Qualititative, PTM Isoform Targeted	MS/MS	High	Low
Sensitivity Qualification	Quantitative, Prevalence, Classifier Test	ELISA MRM	High	Medium
Clinical Validation	Classifier Validated, Standardized, FDA Approved IVDMIA	MRM High Content Array MALDI profiling	High	High

Patient Samples: 10-20, 80-100, 200-500, 100-100,000
Candidate Proteins: 1000's, 100's, 10-50, 10

Fig. 1. Biomarker pipeline

Table describes the aim, the likely analytical platform and associated characteristics of each phase in an ideal biomarker discovery pipeline through verification to validation and final pre-market approval. The schema represents the increase in patient sample and decrease in candidate protein numbers as a biomarker study moves from discovery (two-step) through to validation phases; 2DE- 2-dimensional gel electrophoresis, DIGE- difference in-gel electrophoresis, LC-MS- liquid chromatography associated with mass spectrometry, ELISA- enzyme linked immuno-adsorbant assay, MRM- multiple reaction monitoring mass spectrometry, IVDMIA-in vitro diagnostic multivariate index assay.

3.2 Verification of protein modifications

Protein modifications are common but are frequently overlooked, especially during the discovery phase. Amongst the most significant modifications are covalent alternations to amino acids (*e.g.,* phosphorylation, nitration or redox changes) and covalent addition of large groups (*e.g.,* glycosylation). These modifications can have dramatic effects on protein function and may play a significant role in a range of arthritides and autoimmune disorders. Because most biomarker candidate identification strategies rely on peptide surrogate based mass spectrometry, there is added potential to characterize low abundance PTM variants. MALDI-TOF is an example mode of mass spectrometry can scrutinize multiple variants of a given protein in a concurrent, swift and relatively sensitive fashion. Several criteria determine accurate structural assignment and the quantification of specific modifications via

a peptide-centric approach (Duncan et al., 2010), including spectra search criteria, sequence coverage and database completeness.

Accordingly, changing levels of a modified protein may represent a better biomarker than changes in the total expression levels of a given protein. For example, alterations in the levels of naturally-occurring glycosylation motifs can serve as a marker of inflammation, lymphocyte tolerance and senescence in arthritis (Garcia et al., 2005), *viz.* increased branching of sugar moieties on alpha-1 acid glycoprotein can act as biomarkers of inflammation, whereas decreased branching of T-cell receptor affects the development of Th1/Th2 cells increasing susceptibility to autoimmunity (Havenaar et al., 1998; Morgan et al., 2004).
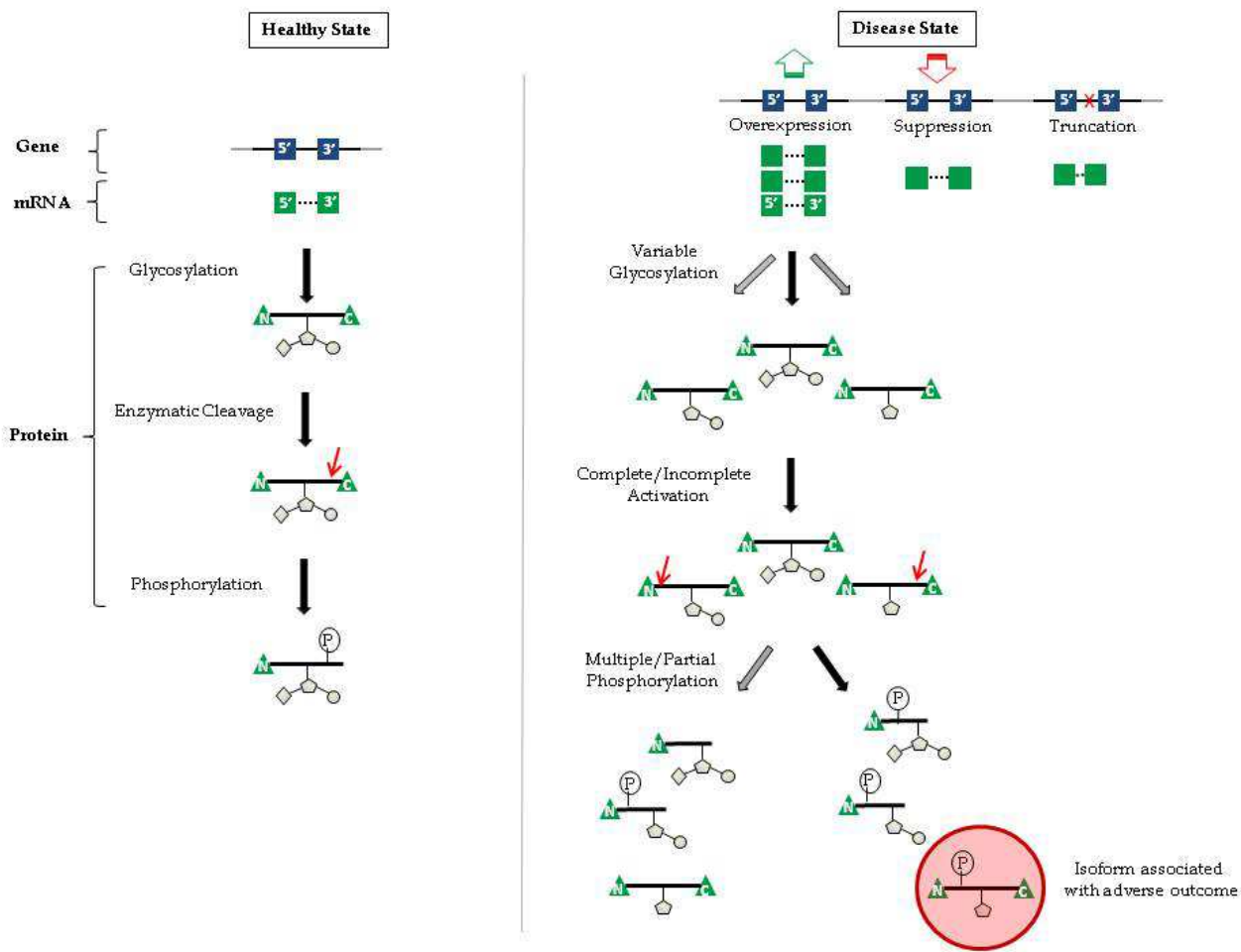
Fig. 2. Protein isoform verification

A depiction of possible qualitative and quantitative changes in protein isoforms between health and a disease state. The illustration of an isoform of a given protein associated with a specific adverse outcome demonstrates that it can only be detected by high 'resolution' proteomic strategies which can detect variance in post translational modifications. Conventional genomic and antibody based methods will only pick up on a change in expression of recognized transcripts or epitopes, giving a high likely hood of missing the significance of the isoform prevalent in a particular disease outcome.

Recent evidence suggests that oxidative modifications to the proteins S100A8 and S100A9 shifts function from macrophage and neutrophil activation in inflammatory arthritis towards a protective role (Lim et al., 2009). In this case, the modification appears to serve as a regulatory switch. Citrulination of arginine side chains has the potential to alter structure, antigenicity and protein function (Wegner et al., 2010). In fact, synthetic peptides modified to mimic possible neo-antigens which trigger an autoimmune response have been used to identify novel diagnostic/prognostic autoantibodies (McLaren et al., 2005; Papini et al., 2009).

Before disease becomes apparent, it is likely that a particular disease pathology 'specific' protein isoform combination has been expressed for some time, impacting normal physiological pathways. These disease 'specific' proteins may also be expressed in a benign or developing state of the disease devoid of clinical symptoms and may contain a sub pool of surrogate markers of chronic inflammation. An example from the world of autoimmune disease is presented by a study of systemic lupus erythematosus patients in whom autoantibodies were detected prior to clinical symptoms (Eriksson et al., 2011). Susceptibility to develop several other auoimmune diseases including diabetes and rheumatoid arthritis can be predicted by long periods of pre-clinical autoantibody expression (Bastra et al., 2001; Rantapaa-Dahlquist et al, 2003). Another recent study indicates that galactosylation of IgG precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis patients (Ercan et al., 2010).

Evidently these preclinical biomarker 'screening' studies are unique in that they rely heavily on concerted biobanking of samples in a prospective fashion, generally have focused on more easily retrieved antibodies and may incur long 'wait times' until a specific disorder may occur. They do however offer a fascinating glimpse of what could be occurring at the protein level prior to disease onset, which arguably could offer a window of opportunity to diagnose earlier, manage the pathology before it becomes clinically symptomatic and possibly prevent aberrant processes all together. Alterations in protein isoforms therefore may also comprise part of the milieu of pathological changes and thereby serve as biomarkers. Studies aimed at full length characterization of proteins indicate that preliminary discovery stages may therefore not reflect the full extent of protein variants due to the low cohort sizes (and low throughput techniques) typical of this stage. For example, a study of diabetes patients revealed that, within a cohort of 96 individuals, an average of 3 variants of each protein were observed; a further 8 variants were observed across 1000 individuals (Borges et al., 2010). This highlights the importance of accounting for protein micro-heterogeneity across patient populations and correlation of prevalence with specific disease outcome sub-groups (Figure 2). Statistical evidence of prevalence and analytical limits of detection of a specific group of isoforms should then direct the study towards validation of candidates in a much larger group of multi-center patient populations.

4. Emerging tools for targeted biomarker validation

The biggest challenge in proteomics remains independent validation of changes 'discovered' in observational investigations. Traditionally, validation has been undertaken by antibody-based approaches, including Western blotting, ELISA and immunohistochemistry (IHC). However, despite major efforts to generate proteome-scale panels of suitable antibodies (most notably the impressive Human Protein Atlas initiative [http:// www. proteinatlas.

org/ index.php]), this remains a slow process. It requires antibody generation and characterization to establish specificity and utility in different assay formats.

4.1 Multiple reaction monitoring

Antibody-independent strategies are highly desirable. The most popular of these is based on peptide-centric, multiple reaction monitoring (MRM). MRM is a technology that has unique potential for reliable quantification of analytes of low abundance in complex mixtures. In an MRM assay, a predefined precursor ion and one of its fragments are selected by the two mass filters of a triple quadrupole instrument and monitored over time for precise quantification. A series of transitions (precursor/fragment ion pairs) in combination with the retention time of the targeted peptide can constitute a definitive assay (Lange et al., 2008). The combination of MRM, chemistry and software to aid with the selection of suitable proteotypic peptides, has provided the opportunity to rapidly develop quantitative multiplexed assays of protein expression and post-translational modification that are both highly specific and sensitive (Scheiss et al., 2009). In recent years, significant advances have been made in the measurement of protein expression using MRM on triple quadrupole (QQQ) mass spectrometers (Pan et al., 2009). In this system, one or more peptide ions of unique and known mass are preselected in the first quadrupole (Q1), induced to fragment in the second quadrupole (Q2), and some of the resulting 'product ions' (or fragments) are selected for transmission to the detector in the third quadrupole (Q3) (Figure 3A). MRM supports the simultaneous measurement of multiple proteotypic peptides and synthetic mass variants of them (usually spiked into samples in known amounts). The strategy enables the absolute quantification of multiple proteins (Keshishan et al., 2007; Kuzyk et al., 2009). When MRM is combined with immunoaffinity purification and internal peptide standards, for example SISCAPA, detection is in the sub-femtomolar range (Whiteaker et al., 2010).

In a relatively early demonstration of peptide MRM, assays were developed to simultaneously quantify the expression of sixteen cytochrome P450 enzymes - proteins important in determining susceptibility to adverse drug reactions (Jenkins et al., 2006). Previously, a method was described for the MRM assay of C-reactive protein (CRP) as a means of differentiating erosive from non-erosive RA patients (Kuhn et al., 2004). The same research team then applied the same MRM technique to measure elevated levels in synovial fluid of six additional members of the S100 calcium-binding proteins associated with an erosive subtype of RA (Liao et al., 2004).

4.2 Nucleic acid programmable protein arrays

The production of antibodies against self-antigens (autoantibodies) is a characteristic feature of many autoimmune diseases. At a clinical level, tests for specific autoantibodies, such as ANA positivity, are routinely employed to aid the diagnosis and track the progress of these diseases. Traditionally, autoantibodies have been identified with a one-antigen-at-a-time, hypothesis-driven approach using methods such as immunofluorescence and ELISA.

Microarrays provide a particularly effective platform for the systematic study of thousands of proteins in parallel because they are sensitive and require low sample volumes (MacBeath & Schreiber, 2000; Zhu et al., 2001). Protein microarrays involve the display of thousands of different proteins with high spatial density on a microscopic surface. Protein microarrays have been applied to autoimmune biomarker studies focused on pre-symptomatic screening and diagnosis, clinical outcome prognosis and therapeutic response prediction (Hueber et

al., 2005; Quitana et al., 2004) With particular relevance to the remit of this chapter, conventional printed arrays have been used to study rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, hepatitis and encephalomyelitis (Fattal et al., 2010; Hueber et al. 2009; Li et al., 2005; Somers et al., 2009; Song et al., 2010).

A In protein multiple reaction monitoring (MRM), one or more peptides of unique and known mass (proteotypic peptides) are preselected in the first quadrupole (Q1), induced to fragment in Q2 by collisional excitation with a neutral gas in a pressurized cell and some of the resulting 'product ions' (fragments) are selected for transition to the detector in the third quadrupole (Q3). **B1** Nucleic acid programmable protein array (NAPPA) spotted with genes of interest; All proteins are tagged at the c-terminus to ensure only full length translated proteins can be captured in situ by co-spotted anti-tag antibodies. NAPPA has consistent protein amounts displayed at each spot; most are within two fold of the average (Ramachandaran et al., 2008). Proteins are expressed "just-in-time" for assay, which eliminates concern of protein stability. **B2** Image of NAPPA with randomly selected 768 genes probed with a synovial fluid sample from a patient with juvenile arthritis. Antibodies in patient samples bind to their antigen targets on the array and are detected by Alexa647-conjugated goat anti-human IgG. **B3** Scatterplot of reactivity on NAPPA between paired plasma and synovial fluid samples from arthritis patients. Median correlation is 0.982. **C1** Matrix assisted laser desorption ionization- time of flight (MALDI-TOF) mass spectrometry whereby proteins or peptides imbedded in a crystallized matrices are ionized by a high frequency laser beam and accelerated through a flight tube by electrical field; ions 'fly' and reach the detector plate with respect to their mass:charge ratio. **C2** A spectra is generated which reflects the energy of a given ion vs the mass:charge ratio (m/z). **C3** A birds eye view representation of the spectra reveals distinguishing peaks (*) from the six samples analysed.
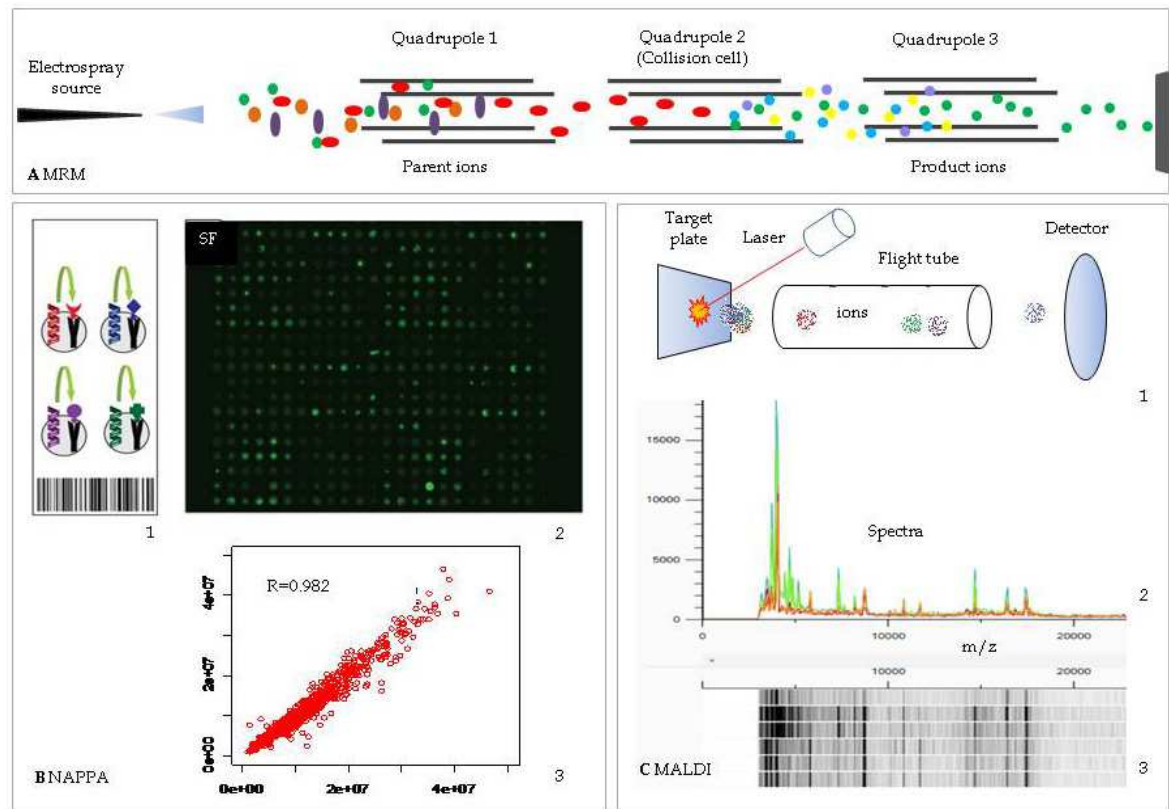

Fig. 3. Targeted identification methods

Nucleic Acid Programmable Protein Array (NAPPA) is an innovative method to produce protein microarrays, where cDNAs encoding proteins of interest are spotted onto activated surfaces and proteins are produced *in situ* using mammalian *in vitro* expression systems (Ramachandran et al., 2004; Ramachandran et al., 2008). The freshly made protein is captured by co-spotted antibodies specific for a 'tag' encoded at the end of the amino acid sequence. This approach circumvents the labor and cost considerations associated with conventional spotting of labile recombinant proteins into arrays. NAPPA technology recently revealed that ankylosing spondylitis patients' autoantibody responses were targeted towards connective, skeletal and muscular tissue, unlike those of RA patients (Wright et al., 2010). In a recent pilot study, a strong correlation was observed between 768 autoantibodies in paired plasma and synovial fluid samples from patients with juvenile arthritis (Figure 3B).

4.3 Proteomic profiling methods

Intact protein profiling across clinical cohorts gives a glimpse into the degree of variation evident in a single gene product (Borges et al., 2008a). The same approach may be useful in the study of arthritis. Mass spectrometry-based techniques can potentially distinguish these physical and structural variations and allow the relative abundance of one isoform to be determined (Duncan et al., 2010). By contrast, these variants would be overlooked by conventional ELISA methods (Figure 2). A brief description and recent application of such techniques follows.

MALDI / SELDI Profiling (Immuno-MALDI): Matrix assisted laser desorption ionisation (MALDI) mode of mass spectrometry allows the 'soft' ionization of complete proteins which are liable to fragment under conventional ionization methods. The type of a mass spectrometer most widely used with MALDI is the time-of-flight (TOF), mainly due to its large mass range (Figure 3C). Purifying a protein from a clinical sample by immunoprecipitation can greatly reduce the complexity of the proteome being analysed. In one approach, purified polyclonal antibodies that capture the target protein isoforms can be immobilized onto sepharose beads packed within a pipette tip or 'fret' (Borges et al., 2008b). Eluted proteins can then be spotted on a MALDI target plate and spectra obtained. For example, some recent MALDI profiling applications have demonstrated the ability to diagnose early RA and hypertension and distinguish active SLE (Dai et al., 2010; Long et al., 2010; Reid et al., 2010). Glycosylation heterogeneity of selected inflammation associated molecules such as serum amyloid and vitamin D binding protein have been investigated in cancer and diabetic patients (Rehder et al., 2009; Weiss et al., 2011).

As a modification of MALDI, surface-enhanced laser desorption ionization (SELDI) methods can be used to target lower molecular weight proteins (<20 KDa) to differentiate arthritides and therapeutic response (de Seny et al., 2008; Miyame et al. 2005). The technology is currently being developed to affinity capture the protein of interest directly to the mass spectrometry target plate (Brauer et al., 2010).

4.4 Biomarker research and grant funding

Although proteomics has been full of promise, few validated biomarkers have made their way into the public domain and even fewer influence clinical practice. There is little doubt that validation is a serious bottleneck in the biomarker development process. While there is abundant discussion of approaches to discovery, the tools for validation and their

applications have received little attention. It is very often difficult to receive funding from traditional grant programs to validate markers: funding agencies balk at the prospect of funding a 're-measurement' of the same entity in larger independent cohorts. Additionally, the continuum from discovery through to validation is tedious and extends well beyond the time-frame of a typical research grant. In fact, the time from initial discovery to routine use can take up to a decade (Anderson, 2010; Wilson et al., 2007). A recent example illustrates the seven year journey from discovery to FDA approval for the multivariate diagnostic test OVA1, used to screen ovarian cancer patients (Fung, 2010).

Similarly, when validation fails it is difficult for academic investigators to publish these 'negative' results; when validation succeeds, the emphasis frequently shifts to commercialization rather than publication.

5. Conclusions

While there is widespread recognition of the value of biomarkers, scientific progress is slow. Over the years, biomarkers have sometimes been the center of excessive "hype", prompting excessive or unreasonable expectations. In addition, the use of biomarkers as surrogate endpoints have led to some public failures when they were felt to be falsely reassuring, creating general skepticism amongst scientists and clinicians alike (e.g. Petricoin et al., 2002). In addition to limited validation, resistance still hindering biomarker acceptance includes:

- Resistance to sharing data across independent efforts – Organizations may work on similar research or discover keystone advances yet resist sharing knowledge because they feel that doing so will jeopardize their competitive advantage. However, sharing information could help companies achieve greater overall progress and reduce costs.

- Need for new R&D models with greater precision and flexibility – The industry needs an R&D model with greater precision to improve pipelines, leveraging active clinical knowledge to offset the declining success in new drug development. Some research and development leaders are concerned that using an approach that targets treatment for only a subset of patients decreases profits and increases research costs. Others recognize that this direction has already created value beyond costs and are building these capabilities into their business strategy. For example, Herceptin® is considered an effective targeted treatment for breast cancer. Targeted treatments could actually increase both the medical and economic success of a therapeutic.

- Insufficient interoperability – Traditional data resides in disparate places that often do not easily connect. Factor in imaging biomarkers constituted by terabytes of data and you have a complex mix of data from which it is difficult to extract new insights (Poste, 2011). The path forward – interoperability – is a design and intent to have systems share information that relies on data standards, and more importantly, semantics. Semantics use common vocabularies and business rules to relate clinical terms reported across different sources to find common meaning.

Tools to support development, medical care, health policy such as the FDA's critical path, and BioPharma investment decisions. The biomarker development and validation process is necessary but costly for one company. Innovation takes place in many organizations and, as such, stakeholders work redundantly on the same effort. Many collaborative forums exist but these usually involve sharing "safe" information that really does not hasten overall progress. Consequently, most existing biomarkers have taken decades to become part of medical practice.

Currently there are few FDA-approved proteomic tests for autoimmune disease. Although there is little doubt that such tests could help the diagnosis and treatment of arthritis, it is a major clinical and financial challenge to develop, validate and market them. Robust validation data including evidence of sensitivity, specificity and correlation to the existing limited set of clinical or laboratory criteria are necessary to support clinical utility. Disease activity scores (DAS-CRP and DAS28), for example, combine inflamed joint count and ESR/CRP to document levels of disease activity at a static time point. The measurement of specific proteins that flag a particular patient's status add objectivity in circumstances where the clinician currently relies on clinical judgment alone.

From a clinician's perspective, it is important to address several questions in a timely fashion for a given patient presenting with autoimmune disease. In each instance, the clinician is attempting to minimize underlying disease and adverse outcomes, such as joint damage in arthritis. Key questions that can currently only be partially answered by clinical observation and patient history include: (a) is this true autoimmune-driven arthritis (i.e., diagnosis), (b) how severe or at what stage is the disease process, (c) what is this patient's likely outcome (i.e., prognosis) and (d) which drugs could abrogate that outcome (i.e., prediction)? Decision-making also extends to selection of therapy: (e) what is the patient-specific titer, (f) which disease subgroups will benefit from a specific therapeutic strategy and (g) when should treatment be terminated?

This chapter has addressed and discussed three key areas for consideration, which if addressed after initial discovery work could provide solid evidence of their clinical utility and commercial viability: (i) limiting bias in study design, (ii) thorough protein isoform verification and (iii) modes of orthogonal and targeted validation.

6. Glossary- the language of biomarker and proteomic research

Bias- In statistics, bias is systematic favoritism present in data collection, analysis or reporting of quantitative research

Biomarker- or *biological marker*, is a molecular characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Classifier- in statistics is the formula or criteria for identifying a sub-population based on quantitative information on one or more measurements, traits or characteristics.

Development pipeline- represents the process from candidate discovery, through verification, validation and final pre-market approval.

Diagnostic- in the context of medicine is any test performed or criteria applied to aid to determine and/or identity a possible disease or disorder.

Discovery- in the context of biomarkers, describes the initial process of observation, identification and quantification of one or more biological molecules which may act as a classifier.

Isoform- describes the biological phenomenon of several different structural forms of the same protein which may arise by alternate gene splicing and single-nucleotide polymorphisms before messenger RNA translation and chemical modifications e.g. phosphorylation or glycosylation which occur post-translation of proteins.

Multiplex- in the context of protein assay is a method or platform which permits the simultaneously measururement of multiple analytes (dozens or more) in a single test.

Omics- this suffix, often used in modern biological research, refers to the lofty aim of observing, identifying and quantifying the totality of a particular class of molecules i.e. genomics, proteomics.

Orthogonal method- describes the ideal of using alternate types of analyses to corroborate the original findings by independent means.

Peptide-centric- or *bottom-up*, proteomics is a common method used to identify proteins by proteolytic digestion of proteins prior to analysis by mass spectrometry.

Protein-centric- or *top-down*, proteomics is a method of intact protein identification e.g. an ion trapping mass spectrometer used to store an isolated protein ion for mass measurement and tandem mass spectrometry analysis.

Power analysis- to calculate the number of samples required for a study to reach statistically sound conclusions.

Predictive model- in the context of medicine, is created or chosen to try to predict the probability of a clinical outcome by use of one or more *classifiers*.

Prognostic- is a clinical test which can forecast the likely course or outcome of an illness.

Sensitivity- measures the proportion of true positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition).

Specificity- measures the proportion of true negatives which are correctly identified (e.g. the percentage of healthy people who are correctly identified as not having the condition).

Throughput- refers to the rate of analysis of samples by a particular method e.g. analysis of a single protein by Western blotting is relatively low throughput compared to ELISA.

Validation- later stage in the biomarker pipeline is defined as the documented act of demonstrating that putative biomarker classifiers will consistently lead to the expected results i.e establish *sensitivity* and *specificity* performance in large populations and begin to optimize the assay for commercial use.

Verification- intermediate phase in biomarker pipeline bridging *discovery* and *validation*, which typically reduces the number of candidates, confirms specific protein isoforms within a *classifier* and begins to assess the sensitivity in expanded populations.

7. Acknowledgement

8. References

Alonso-Ruiz A., Pijoan J.I., Ansuategui E., Urkaregi A., Calabozo M. & Quintana A. Tumor necrosis factor alpha drugs in rheumatoid arthritis: systematic review and metaanalysis of efficacy and safety. *BMC Musculoskelet. Disord.*, 9, (2008), pp 52.

Anders H.J., Rihl M., Vielhauer V. & Schattenkirchner M. Assessment of renal function in rheumatoid arthritis: validity of a new prediction method. *J. Clin. Rheumatol.*, 8(3), (2002), pp 130-3.

Addona T.A. Abbatiello S.E., Schilling B., Skates S.J., Mani D.R., Bunk D.M., Spiegelman C.H., Zimmerman L.J., Ham A.J., Keshishian H., Hall S.C., Allen S., Blackman R.K., Borchers C.H., Buck C., Cardasis H.L., Cusack M.P., Dodder N.G., Gibson B.W., Held J.M., Hiltke T., Jackson A., Johansen E.B., Kinsinger C.R., Li J., Mesri M., Neubert T.A., Niles R.K., Pulsipher T.C., Ransohoff D., Rodriguez H., Rudnick P.A., Smith D., Tabb D.L., Tegeler T.J., Variyath A.M., Vega-Montoto L.J., Wahlander A., Waldemarson S., Wang M., Whiteaker J.R., Zhao L., Anderson N.L., Fisher S.J., Liebler D.C., Paulovich A.G., Regnier F.E., Tempst P. & Carr S.A.. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.*, 27, (2009), pp 633-641.

Aebersold R. & Mann M. Mass spectrometry-based proteomics. *Nature*, 422, (2003), pp 198-207.

Anderson N.L. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin. Chem.*, 56(2), (2010), pp 177-85.

Batstra M.R., Aanstoot H.J. & Herbrink P. Prediction and diagnosis of type 1 diabetes using beta-cell autoantibodies. *Clin. Lab.*, 47(9-10), (2001), pp 497-507.

Bloemen K., Van Den Heuvel R., Govarts E., Hooyberghs J., Nelen V., Witters E., Desager K. & Schoeters G. A new approach to study exhaled proteins as potential biomarkers for asthma. *Clin. Exp. Allergy*, 41(3), (2011), pp 346-56.

Bongartz T., Sutton A.J., Sweeting M.J., Buchan I., Matteson E.L. & Montori V. Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *JAMA*, 295(19), (2006), pp 2275-85.

Borges C.R., Jarvis J.W., Oran P.E. & Nelson R.W. Population studies of Vitamin D Binding Protein microheterogeneity by mass spectrometry lead to characterization of its genotype-dependent O-glycosylation patterns. *J. Proteome Res.*, 7(9), (2008), pp 4143-53.

Borges C.R., Jarvis J.W., Oran P.E., Rogers S.P. & Nelson R.W. Population studies of intact vitamin D binding protein by affinity capture ESI-TOF-MS. *J. Biomol. Tech.*, 19(3), (2008), pp 167-76.

Borges C.R., Rehder D.S., Jarvis J.W., Schaab M.R., Oran P.E. & Nelson R.W. Full-length characterization of proteins in human populations. *Clin. Chem.*, 56(2), (2010), pp 202-11.

Bowes J.D., Potter C., Gibbons L.J., Hyrich K., Plant D., Morgan A.W., Wilson A.G., Isaacs J.D.,Worthington J. & Barton A. Investigation of genetic variants within candidate genes of the TNFRSF1B signalling pathway on the response to anti-TNF agents in a UK cohort of rheumatoid arthritis patients. *Pharmacogenet. Genomics*, 19(4), (2009), pp 319-23.

Brauer H.A., Lampe P.D., Yasui Y.Y., Hamajima N. & Stolowitz M.L. Biochips that sequentially capture and focus antigens for immunoaffinity MALDI-TOF MS: a new tool for biomarker verification. *Proteomics*, 10(21), (2010), pp 3922-7. Cairns D.A., Barrett J.H., Billingham L.J., Stanley A.J., Xinarianos G., Field J.K., Johnson P.J., Selby P.J. & Banks R.E. Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. *Proteomics*, 9(1), (2009), pp 74-86.

Chait, B. T. Chemistry. Mass spectrometry: bottom-up or top-down? *Science*, 314, (2006), pp 65-66.

Chen P.P., Robbins D.L., Jirik F.R., Kipps T.J. & Carson D.A. Isolation and characterization of a light chain variable region gene for human rheumatoid factors. *J. Exp. Med.*, 166(6), (1987), pp 1900-1905.

Dai Y., Hu C., Wang L., Huang Y., Zhang L., Xiao X. & Tan Y. Serum peptidome patterns of human systemic lupus erythematosus based on magnetic bead separation and MALDI-TOF mass spectrometry analysis. *Scand. J. Rheumatol.*, 39(3), (2010), pp 240-246.

Davis M.T., Auger P.L. & Patterson S.D. Cancer biomarker discovery via low molecular weight serum profiling--are we following circular paths? *Clin. Chem.*, 2010, 56(2), pp 244-247.

de Seny D., Fillet M., Ribbens C., Marée R., Meuwis M.A., Lutteri L., Chapelle J.P., Wehenkel L., Louis E., Merville M.P. & Malaise M. Monomeric calgranulins measured by SELDI-TOF mass spectrometry and calprotectin measured by ELISA as biomarkers in arthritis. *Clin. Chem.*, 54(6), (2008), pp 1066-1075.

Dowsey, A.W., Morris, J.S., Gutstein, H.B. & Yang, G.Z. in Proteome Bioinformatics (eds Hubbard, S. J. & Jones, A. R.) 239 (Humana Press, New York, 2009).

Duncan M.W., Aebersold R. & Caprioli R.M. The pros and cons of peptide-centric proteomics. *Nat. Biotechnol.*, 28(7), (2010), pp 659-64.

Dupuy A. & Simon R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J. Natl. Cancer Inst.*, 99(2), (2007), pp 147-157.

Ercan, A. Cui J., Chatterton D.E., Deane K.D., Hazen M.M., Brintnell W., O'Donnell C.I., Derber L.A., Weinblatt M.E., Shadick N.A., Bell D.A., Cairns E., Solomon D.H., Holers V.M., Rudd P.M. & Lee D.M. Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis. *Arthritis Rheum.*, 62, (2010), pp 2239-2248.

Eriksson C., Kokkonen H., Johansson M., Hallmans G., Wadell G. & Rantapaa-Dahlqvist S. Autoantibodies predate the onset of Systemic Lupus Erythematosus in northern Sweden. *Arthritis Res. Ther.*,. 13(1), (2011), pp R30.

Fattal I., Shental N., Mevorach D., Anaya J.M., Livneh A., Langevitz P., Zandman-Goddard G., Pauzner R., Lerner M., Blank M., Hincapie M.E., Gafter U., Naparstek Y., Shoenfeld Y., Domany E. & Cohen I.R. An antibody profile of systemic lupus erythematosus detected by antigen microarray. *Immunology*, 130(3), (2010), pp 337-343.

Flower L., Ahuja R.H., Humphries S.E. & Mohamed-Ali V. Effects of sample handling on the stability of interleukin 6, tumour necrosis factor-alpha and leptin. *Cytokine*. 12(11), (2000), pp 1712-1716.

Fung E.T. A recipe for proteomics diagnostic test development: the OVA1 test, from biomarker discovery to FDA clearance. *Clin. Chem.*, 56(2), (2010), pp 327-329.

Garcia G.G., Berger S.B., Sadighi Akha A.A. & Miller R.A.. Age-associated changes in glycosylation of CD43 and CD45 on mouse CD4 T cells. *Eur. J. Immunol.*, 35(2), (2005), pp 622-31.

Gibson D.S. & Rooney M.E. The human synovial fluid proteome: A key factor in the pathology of joint disease. *PROTEOMICS - CLINICAL APPLICATIONS*, 1, (2007), pp 889-899.

Gibson, D.S. Blelock S., Curry J., Finnegan S., Healy A., Scaife C., McAllister C., Pennington S., Dunn M. & Rooney M.. Comparative analysis of synovial fluid and plasma proteomes in juvenile arthritis--proteomic patterns of joint inflammation in early stage disease. *J. Proteomics*, 72, (2009), pp 656-676.

Haab B.B, Geierstanger B.H., Michailidis G., Vitzthum F., Forrester S., Okon R., Saviranta P., Brinker A., Sorette M., Perlee L., Suresh S., Drwal G., Adkins J.N., Omenn G.S. Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. *Proteomics*, 5(13), (2005), pp 3278-3291.

Havenaar E.C., Axford J.S., Brinkman-van der Linden E.C., Alavi A., Van Ommen E.C., van het Hof B., Spector T., Mackiewicz A. & Van Dijk W.. Severe rheumatoid arthritis prohibits the pregnancy-induced decrease in alpha3-fucosylation of alpha1-acid glycoprotein. *Glycoconj. J.*, 15(7), (1998), pp 723-729.

Hu, S., Loo, J.A. & Wong, D.T. Human body fluid proteome analysis. *Proteomics*, 6, (2006), pp 6326-6353.

Hueber W., Kidd B.A., Tomooka B.H., Lee B.J., Bruce B., Fries J.F., Sønderstrup G., Monach P., Drijfhout J.W., van Venrooij W.J., Utz P.J., Genovese M.C. & Robinson W.H. Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis Rheum.*, 52(9), (2005), pp 2645-2655.

Hueber W., Tomooka B.H., Batliwalla F., Li W., Monach P.A., Tibshirani R.J., Van Vollenhoven R.F., Lampa J., Saito K., Tanaka Y., Genovese M.C., Klareskog L., Gregersen P.K. & Robinson W.H. Blood autoantibody and cytokine profiles predict response to anti-tumor necrosis factor therapy in rheumatoid arthritis. *Arthritis Res. Ther.*, 11(3), (2009), pp R76.

Jenkins R.E., Kitteringham N.R., Hunter C.L., Webb S., Hunt T.J., Elsby R., Watson R.B.,Williams D., Pennington S.R., Park B.K. Relative and absolute quantitative expression profiling of cytochromes P450 using isotope-coded affinity tags. Proteomics, 6(6), (2006), pp 1934-1947.

Karp N.A. & Lilley K.S. Design and analysis issues in quantitative proteomics studies. *Proteomics*, 7 Suppl 1, (2007), pp 42-50.

Kastbom A., Strandberg G., Lindroos A. & Skogh T. Anti-CCP antibody test predicts the disease course during 3 years in early rheumatoid arthritis (the Swedish TIRA project). *Ann. Rheum. Dis.*, 63(9), (2004), pp 1085-1089.

Kavanaugh A.F., Solomon D.H. & American College of Rheumatology Ad Hoc Committee on Immunologic Testing Guidelines. Guidelines for immunologic laboratory testing in the rheumatic diseases: anti-DNA antibody tests. *Arthritis Rheum.*, 47(5), (2002), pp 546-555.

Keshishian H., Addona T., Burgess M., Kuhn E. & Carr S.A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics*, 6(12), (2007), pp 2212-2229.

Kuhn E., Wu J., Karl J., Liao H., Zolg W. & Guild B. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring

mass spectrometry and 13C-labeled peptide standards. *Proteomics*, 4(4), (2004), pp 1175-86.

Kurreeman F.A., Padyukov L., Marques R.B., Schrodi S.J., Seddighzadeh M., Stoeken-Rijsbergen G., van der Helm-van Mil A.H., Allaart C.F., Verduyn W., Houwing-Duistermaat J., Alfredsson L., Begovich A.B., Klareskog L., Huizinga T.W. & Toes R.E. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med.*, 4(9), (2007), pp e278.

Kuzyk M.A., Smith D., Yang J., Cross T.J., Jackson A.M., Hardie D.B., Anderson N.L. & Borchers C.H. Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics*, 8(8), (2009), pp 1860-1877.

Lange V., Picotti P., Domon B.& Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol .Syst. Biol.*, 4, (2008), pp 222.

Li Q.Z., Xie C., Wu T., Mackay M., Aranow C., Putterman C. & Mohan C. Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. *J. Clin. Invest.*, 115(12), (2005), pp 3428-39.

Liao H., Wu J., Kuhn E., Chin W., Chang B., Jones M.D., O'Neil S., Clauser K.R., Karl J., Hasler F., Roubenoff R., Zolg W. & Guild B.C. Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. *Arthritis Rheum.*, 50(12), (2004), pp 3792-803.

Lim S.Y., Raftery M.J., Goyette J., Hsu K. & Geczy C.L. Oxidative modifications of S100 proteins: functional regulation by redox. *J. Leukoc. Biol.*, 86(3), 2009, pp 577-587.

Lindqvist E., Eberhardt K., Bendtzen K., Heinegard D. & Saxne T. Prognostic laboratory markers of joint damage in rheumatoid arthritis. *Ann. Rheum. Dis.*, 64(2), (2005), pp 196-201.

Liu Q., Sung A.H., Chen Z., Liu J., Huang X. & Deng Y. Feature selection and classification of MAQC-II breast cancer and multiple myeloma microarray gene expression data. *PLoS One*, 4(12), (2009), pp e8250.

Long L., Li R., Li Y., Hu C. & Li Z. Pattern-based diagnosis and screening of differentially expressed serum proteins for rheumatoid arthritis by proteomic fingerprinting. *Rheumatol. Int.*, (2010).

MacBeath G. & Schreiber S.L. Printing proteins as microarrays for high-throughput function determination. *Science*, 289(5485), (2000), pp 1760-1763.

Marengo E., Robotti E., Bobba M., Liparota M.C., Rustichelli C., Zamò A., Chilosi M. & Righetti P.G. Multivariate statistical tools applied to the characterization of the proteomic profiles of two human lymphoma cell lines by two-dimensional gel electrophoresis. *Electrophoresis*, 27(2), (2006), pp 484-494.

McLaren M,. Waring A., Galarraga B., Rudd A., Morley K. & Belch J.J. Investigation of platelet glycoprotein IIIa polymorphism using flow cytometry in patients with rheumatoid arthritis. *Scand. J. Rheumatol.*, 34, (2005), pp 437-440.

Meyer O., Labarre C., Dougados M., Goupille P., Cantagrel A., Dubois A., Nicaise-Roland P., Sibilia J. & Combe B.Anticitrullinated protein/peptide antibody assays in early rheumatoid arthritis for predicting five year radiographic damage. *Ann. Rheum. Dis.*, 62(2), (2003), pp 120-126.

Miyamae T., Malehorn D.E., Lemster B., Mori M., Imagawa T., Yokota S., Bigbee W.L.,Welsh M., Klarskov K., Nishomoto N., Vallejo A.N. & Hirsch R. Serum protein profile in

systemic-onset juvenile idiopathic arthritis differentiates response versus nonresponse to therapy. Arthritis Res. Ther., 7(4), (2005), pp R746-755.

Morgan R., Gao G., Pawling J., Dennis J.W., Demetriou M. & Li B. N-acetylglucosaminyltransferase V (Mgat5)-mediated N-glycosylation negatively regulates Th1 cytokine production by T cells. *J. Immunol.*, 173(12), (2004), pp 7200-7208.

Nakamura RM. Progress in the use of biochemical and biological markers for evaluation of rheumatoid arthritis. *J. Clin. Lab. Anal.*, 14(6), (2000), pp 305-313.

Nielen M.M., van Schaardenburg D., Reesink H.W., van de Stadt R.J., van der Horst-Bruinsma I.E., de Koning M.H., Habibuw M.R., Vandenbroucke J.P. & Dijkmans B.A. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum.*, 50(2), (2004), pp 380-6.

Pan S., Aebersold R., Chen R., Rush J., Goodlett D.R., McIntosh M.W., Zhang J., Brentnall T.A. Mass spectrometry based targeted protein quantification: methods and applications. *J. Proteome Res.*, 8(2), (2009), pp 787-797.

Papini, A.M. The use of post-translationally modified peptides for detection of biomarkers of immune-mediated diseases. *J. Pept. Sci.*, 15, (2009), pp 621-628.

Pepe M.S., Feng Z., Janes H., Bossuyt P.M. & Potter J.D. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J. Natl. Cancer Inst.*, 100(20), (2008), pp 1432-1438.

Pepys M.B. & Hirschfield G.M. C-reactive protein: a critical update. *J. Clin. Invest.*, 111(12), (2003), pp 1805-1812.

Petricoin, E.F. Ardekani A.M., Hitt B.A., Levine P.J., Fusaro V.A., Steinberg S.M., Mills G.B., Simone C., Fishman D.A., Kohn E.C. & Liotta L.A. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 359, (2002), pp 572-577. Potter C., Cordell H.J., Barton A., Daly A.K., Hyrich K.L., Mann D.A., Morgan A.W., Wilson A.G., the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGSS) & Isaacs J.D.. Association between anti-tumour necrosis factor treatment response and genetic variants within the TLR and NF{kappa}B signalling pathways. *Ann. Rheum. Dis.*, 69(7), (2010), pp 1315-1320.

Poste G. Bring on the Biomarkers. *Nature*, 469, (2011), pp 156-157.

Quintana F.J., Farez M.F., Viglietta V., Iglesias A.H., Merbl Y., Izquierdo G., Lucas M., Basso A.S., Khoury S.J., Lucchinetti C.F., Cohen I.R. & Weiner H.L. Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis. *Proc. Natl. Acad. Sci. USA*, 105(48), (2008), pp 18889-18894.

Quintana F.J., Hagedorn P.H., Elizur G., Merbl Y., Domany E. & Cohen I.R. Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proc. Natl. Acad. Sci. USA*, 101 Suppl. 2, (2004), pp 14615-14621.

Rai A.J., Gelfand C.A., Haywood B.C., Warunek D.J., Yi J., Schuchard M.D., Mehigh R.J., Cockrill S.L., Scott G.B., Tammen H., Schulz-Knappe P., Speicher D.W., Vitzthum F., Haab B.B., Siest G. & Chan D.W. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics*, 5(13), (2005), pp 3262-3277.

Ramachandran N., Hainsworth E., Bhullar B., Eisenstein S., Rosen B., Lau A.Y., Walter J.C. & LaBaer J. Self-assembling protein microarrays. *Science*, 305(5680), (2004), pp 86-90.

Ramachandran N., Raphael J.V., Hainsworth E., Demirkan G., Fuentes M.G., Rolfs A., Hu Y. & LaBaer J. Next-generation high-density self-assembling functional protein arrays. Nat. Methods, 5(6), (2008), pp 535-538.

Ransohoff D.F. & Gourlay M.L. Sources of bias in specimens for research about molecular markers for cancer. *J. Clin. Oncol.*, 28(4), (2010), pp 698-704.

Ransohoff D.F. Promises and limitations of biomarkers. *Recent Results Cancer Res.* 181, (2009), pp 55-59.

Ransohoff D.F. Proteomics research to discover markers: what can we learn from Netflix? *Clin. Chem.* 56, (2010), pp 172-176.

Rantapää-Dahlqvist S., de Jong B.A., Berglin E., Hallmans G., Wadell G., Stenlund H., Sundin U. & van Venrooij WJ. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum.*, 48(10), (2003), pp 2741-2749.

Rehder D.S., Nelson R.W. & Borges C.R. Glycosylation status of vitamin D binding protein in cancer patients. *Protein. Sci.*, 18(10), 2009, pp 2036-2042.

Reid J.D., Holmes D.T., Mason D.R., Shah B. & Borchers C.H. Towards the development of an immuno MALDI (iMALDI) mass spectrometry assay for the diagnosis of hypertension. *J. Am. Soc. Mass Spectrom.*, 21(10), (2010), pp 1680-6.

Reiner A., Yekutieli D. & Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, 19(3), (2003), pp 368-375.

Robinson W.H., DiGennaro C., Hueber W., Haab B.B., Kamachi M., Dean E.J., Fournel S., Fong D., Genovese M.C., de Vegvar H.E., Skriner K., Hirschberg D.L., Morris R.I., Muller S., Pruijn G.J., van Venrooij W.J., Smolen J.S., Brown P.O., Steinman L. & Utz P.J. Autoantigen microarrays for multiplex characterization of autoantibody responses. Nat. Med., 8(3), (2002), pp 295-301.

Robinson W.H., Fontoura P., Lee B.J,. de Vegvar H.E., Tom J., Pedotti R., DiGennaro C.D., Mitchell D.J., Fong D., Ho P.P., Ruiz P.J., Maverakis E., Stevens D.B., Bernard C.C., Martin R., Kuchroo V.K., van Noort J.M., Genain C.P., Amor S., Olsson T., Utz P.J., Garren H. & Steinman L. Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis. *Nat. Biotechnol.*, 21(9), (2003), pp 1033-9.

Schiess R., Wollscheid B. & Aebersold R. Targeted proteomic strategy for clinical biomarker discovery. *Mol. Oncol.*, 3(1), (2009), pp 33-44.

Smith M.P., Banks R.E., Wood S.L., Lewington A.J. & Selby P.J. Application of proteomic analysis to the study of renal diseases. *Nat. Rev. Nephrol.*, 5(12), (2009), pp 701-712.

Somers K., Govarts C., Stinissen P. & Somers V. Multiplexing approaches for autoantibody profiling in multiple sclerosis. *Autoimmun. Rev.*, 8(7), (2009), pp 573-579.

Song Q., Liu G., Hu S., Zhang Y., Tao Y., Han Y., Zeng H., Huang W., Li F., Chen P., Zhu J., Hu C., Zhang S., Li Y., Zhu H. & Wu L. Novel autoimmune hepatitis-specific autoantigens identified using protein microarray technology. *J. Proteome Res.*, 9(1), (2010), pp 30-39.

Storey J.D. & Tibshirani R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA*, 100(16), (2003), pp 9440-9445.

Strimmer K. A unified approach to false discovery rate estimation. *BMC Bioinformatics*, 9, (2008), pp 303.

Tammen H., Schulte I., Hess R., Menzel C., Kellmann M. & Schulz-Knappe P. Prerequisites for peptidomic analysis of blood samples: I. Evaluation of blood specimen qualities and determination of technical performance characteristics. *Comb. Chem. High Throughput Screen.*, 8(8), (2005), pp 725-733.

Taylor C.F., Paton N.W., Lilley K.S., Binz P.A., Julian R.K. Jr, Jones A.R., Zhu W., Apweiler R., Aebersold R., Deutsch E.W., Dunn M.J., Heck A.J., Leitner A., Macht M., Mann M., Martens L., Neubert T.A., Patterson S.D., Ping P., Seymour S.L., Souda P., Tsugita A.,Vandekerckhove J., Vondriska T.M., Whitelegge J.P., Wilkins M.R., Xenarios I., Yates J.R. 3rd & Hermjakob H. The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.*, 25(8), (2007), pp 887-93.

Tilleman K., Deforce D. & Elewaut D. Rheumatology: a close encounter with proteomics. *Rheumatology (Oxford)*, 44, (2005), pp 1217-1226.

van der Horst-Bruinsma I.E., Visser H., Hazes J.M., Breedveld F.C., Verduyn W., Schreuder G.M., de Vries R.R., Zanelli E. HLA-DQ-associated predisposition to and dominant HLA-DR-associated protection against rheumatoid arthritis. Hum. Immunol., 60(2), (1999), pp 152-158.

van Venrooij W.J., van Beers J.J. & Pruijn G.J. Anti-CCP Antibody, a Marker for the Early Detection of Rheumatoid Arthritis. *Ann. N. Y. Acad. Sci.*, 1143, (2008), pp 268-285.

Vang T., Congia M., Macis M.D., Musumeci L., Orrú V., Zavattari P., Nika K., Tautz L., Taskén K., Cucca F., Mustelin T. & Bottini N. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat. Genet., 37(12), (2005), pp 1317-1319.

Vencovsky J., Machacek S., Sedova L., Kafková J., Gatterová J., Pesáková V. & Růzicková S. Autoantibodies can be prognostic markers of an erosive disease in early rheumatoid arthritis. *Ann. Rheum. Dis.*, 62(5), (2003), pp 427-30.

Wagner U, Kaltenhauser S, Sauer H, Arnold S, Seidel W, Häntzschel H, Kalden JR, Wassmuth R. HLA markers and prediction of clinical course and outcome in rheumatoid arthritis. Arthritis Rheum., 40(2), (1997), pp 341-351.

Washburn M.P., Wolters D. & Yates J.R.,3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19, (2001), pp 242-247.

Wegner N., Lundberg K., Kinloch A., Fisher B., Malmström V., Feldmann M. & Venables P.J. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol. Rev.*, 233, (2010), pp 34-54.

Weiss N.G., Jarvis J.W., Nelson R.W. & Hayes M.A. Examining serum amyloid P component microheterogeneity using capillary isoelectric focusing and MALDI-MS. *Proteomics*, 11(1), (2011), pp 106-113.

Whiteaker J.R., Zhao L., Anderson L. & Paulovich A.G. An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. *Mol. Cell. Proteomics*, 9(1), (2010), pp 184-196.

Wiik A.S. Anti-nuclear autoantibodies: clinical utility for diagnosis, prognosis, monitoring, and planning of treatment strategy in systemic immunoinflammatory diseases. Scand. *J. Rheumatol.*, 34(4), (2005), pp 260-268.

Wilson C., Schulz S. & Waldman S.A. Biomarker development, commercialization, and regulation: individualization of medicine lost in translation. *Clin. Pharmacol. Ther.*, 81(2), (2007), pp 153-155.

Wittkowski H., Frosch M., Wulffraat N., Goldbach-Mansky R., Kallinich T.,Kuemmerle-Deschner J., Frühwald M.C., Dassmann S., Pham T.H., Roth J. & Foell D. S100A12 is a novel molecular marker differentiating systemic-onset juvenile idiopathic arthritis from other causes of fever of unknown origin. *Arthritis Rheum.*, 58(12), (2008), pp 3924-3931.

Wolters D.A., Washburn M.P. & Yates J.R.,3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* 73, (2001), pp 5683-5690.

Wright C., Sibani S., Trudgian D., Fischer R., Kessler B., Labaer J. & Bowness P. Detection of multiple autoantibodies in patients with ankylosing spondylitis using nucleic acid programmable protein arrays. *Mol. Cell. Proteomics*, (2010).

Ye X., Blonder J. & Veenstra T.D. Targeted proteomics for validation of biomarkers in clinical samples. *Brief. Funct .Genomic Proteomic*, 8(2), (2009), pp 126-135.

Zhu H., Bilgin M., Bangham R., Hall D., Casamayor A., Bertone P., Lan N., Jansen R., Bidlingmaier S., Houfek T., Mitchell T., Miller P., Dean R.A., Gerstein M. & Snyder M. Global analysis of protein activities using proteome chips. Science, 293(5537), (2001), pp 2101-2105.

Zhu Y., Shen X. & Pan W. Network-based support vector machine for classification of microarray samples. *BMC Bioinformatics*, 10 Suppl 1, (2009), pp S21.
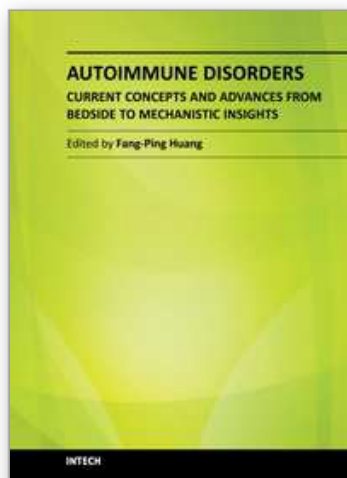
Autoimmune Disorders - Current Concepts and Advances from Bedside to Mechanistic Insights
Edited by Dr. Fang-Ping Huang

Autoimmune disorders are caused due to break down of the immune system, which consequently fails in its ability to differentiate "self" from "non-self" in the context of immunology. The diseases are intriguing, both clinically and immunologically, for their diversified clinical phenotypes and complex underlying immunological mechanisms. This book offers cutting-edge information on some of the specific autoimmune disease phenotypes, respective diagnostic and prognostic measures, classical and new therapeutic options currently available, pathogenesis and underlying mechanisms potentially involved, and beyond. In the form of Open Access, such information is made freely available to clinicians, basic scientists and many others who will be interested regarding current advances in the areas. Its potential readers will find many of the chapters containing in-depth analysis, interesting discussions and various thought-provoking novel ideas.

INTECH
open science | open minds