UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

Breast cancer biomarker discovery

by

Bashar Zeidan

Thesis for the degree of Doctor of Philosophy

January 2013

ABSTRACT

Several environmental and genetic factors are involved in breast cancer development and prognosis. It is clear that mortality rate of breast and other cancers increase with advanced clinical and pathological stage. Early detection thus holds the best cure and identification of prospective markers for breast cancer early detection, as well as understanding the mechanisms of its tumourigenesis and metastatic spread are prerequisites for more effective disease management. This report describes our investigations in breast cancer biomarkers discovery. Here, serum samples from healthy volunteers and patients with breast tumours were analysed and compared to reveal diagnostic and prognostic breast cancer biomarkers.

Biobanks hold a large number of samples that could provide statistically powerful cohorts with a wealth of lengthy follow up data; useful for pre disease/treatment biomarker discovery. However, ambiguous handling standards and source variability have limited their analysis. Conducting a two centre study, we illustrated that archival serum samples can be reliably analysed with high reproducibility. This highlights the utility of such samples for validation of markers discovered in recent studies. In addition, these samples could select for “real world” biologically stable marker entities. This work suggested that this is a potentially useful proteomic arena; however the corner stone for any future archival discovery projects remains dependent on multi centre immunovalidation.

Breast cancer biomarkers including ER/PR and HER2 status; have led to targeted patient stratification and therapy. A more complex molecular sub classification could explain the different outcome within the disease sub groups. However, clinically reliable early detection and markers remain missing. Here, we investigated differentially expressed serum markers between non metastatic breast cancer, benign breast disease and healthy volunteers. Three validated candidate biomarkers (ANX A3, Apo C1 and a 6.4kDa biomarker) differentiating the three groups. Such breast cancer markers can be used as adjuncts to mammography. Further validation of these markers is ongoing and will be followed by elucidation of potential related molecular pathways.

Finally, using a novel proteomic profiling platform, we identified and validated three prediction markers of post treatment outcome in early onset breast cancer. Here, ANX A2, Apo C1 and NOS2 were confirmed as serum prognosticators, and further validation and elucidation of their biological role in the disease holds promise for improved and personalised treatment regimes.

Table of Contents

List of figures

List of tables

DECLARATION OF AUTHORSHIP

I, Bashar A. Zeidan

Declare that the thesis entitled Breast cancer biomarker discovery and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission, or [delete as appropriate] parts of this work have been published as: [please list references]

Signed: B A Zeidan

Date: 31/08/2012

Acknowledgements

It would be impossible to thank all of the people who have so far contributed to this work in some way or another. I would like to sincerely thank Professor Paul Townsend and Professor Graham Packham for everything especially for being the most supportive advisors during and after my project. I feel blessed to have been able to join a group where creativity was encouraged. I would also like to thank the National Institute for Health Research (NIHR), the Wessex Medical Research and the Wessex Cancer Trust for funding this project. An extra special thank you goes to Mr Roger Allsopp for his help and tremendous enthusiasm in instigating this project. Special thanks go also to Professor Ian Fentiman, Diane Allen and Dr Nick Murray for their help in the archival sample analysis and the Wessex cohort projects.

Dr Gary Coulton in the Medical Biomics Centre at the St. George's University of London has always been a great collaborator, many thanks for your support. Special thanks to Dr Spiros Garbis who has been an exceptionally helpful collaborator in the Biomedical Research Foundation of the Academy of Athens. To my colleagues who have made working on a challenging project an enjoyable experience; Dr. Sarah Bailey, Dr. Sam Larkin and Aneta Rybka; thanks for your help.

Very special thanks to Professor John Primrose who has been a great support through my project and academic surgical career, and inspired me as an academically driven surgeon. I would like to thank the University of Southampton, Faculty of Medicine, the Wessex School of Surgery and their members for their endless help and support.

Finally, and most importantly, I would like to thank my family especially my parents Atef and Sahar, my wife Dana and my little angel Celine for their support, encouragement, patience and love.

Definitions and abbreviations

ACN : acetonitrile

ADH : atypical ductal hyperplasia

ANX : Annexin

Apo C1 : Apolipoprotein C1

AUC : area under the curve

BBD : benign breast disease

BSA: bovine serum albumin

CHCA : α-Cyano-4-hydroxycinnamic acid

CPLL : combinatorial peptide ligand libraries

CV : coefficients of variations

DCIS : ductal carcinoma in situ

DTT : dithiothreitol

EAM : energy absorbing molecule

EBC : early onset breast cancer

ELISA : enzyme-linked immunosorbent assay

ER : Oestrogen receptor

FFE : free flow- electrophoresis

FM : focus mass

HER2: Human Epidermal Growth Factor Receptor 2

HILIC: hydrophilic interaction liquid chromatography

HMW : high molecular weight

HPLC: high-performance liquid chromatography

IAA : iodoacetamide

IDC : invasive ductal carcinoma

IDC : invasive ductal carcinoma

IEF : isoelectric focusing

kDa : kilo Dalton

LE : laser energy

LMW : low molecular weight

m/z : mass to charge ratio

MA : matrix attenuation

MALDI : matrix assisted laser desorption ionisation

MR : mass range

MS : mass spectrometry

MudPIT : multidimensional protein identification technology

NAF : nipple aspiration fluid

OPN : Osteopontin

PAGE : polyacrylamide gel electrophoresis

PBS: phosphate buffered saline
PCA: principal component analysis
PR : Progesterone receptor
PTM : post translational modification
RPM: round per minute
RT: room temperature
SAX : strong anion exchange
SEC: size exclusion chromatography
SELDI : surface enhanced laser desorption ionisation
SPA : sinapinic acid
TFA : trifluoroacetic acidoacetonitrile
TOF : time of flight
WB : western blotting
WCX : weak cation exchange

Chapter 1

Introduction

1. Introduction

1.1 General Introduction

One of the major challenges in the battle against cancer is the lack of reliable predictive, diagnostic and prognostic biomarkers that can aid in cancer diagnosis, treatment and response/outcome monitoring[1, 2]. Therefore, cancer biomarker discovery plays a critical role in the current anticancer research era and the future of cancer diagnosis and therapy relies on targeting such markers and their related pathways. Once the molecular markers and pathways involved in cancer development and spread are defined, various aspects related to tumour progression and therapy would be attainable. Ideally, identification of such markers would lead to robust early detection strategies, and personalised effective cancer therapies improving patient outcome(s).

Recent evidence supports the hypothesis that genomic expression profiling is a reliable method for cancer classification and prognostication[3, 4]. However, genes lack any catalytic or signalling capabilities; they instead exert their effects through translation into active proteins. Hence genetic biomarker information alone cannot forecast tumourigenesis and its response to treatment. Consequently, the post genomic project correlating protein expression profiles to cancer is essential for comprehensive representation of cancer biology. Moreover, targeting specific protein pathways involved in tumourigenesis currently present a real hope in the war against cancers, as proteins exert their effects through specific pathways rather than functioning individually[1, 5].

Clinical proteomics is currently experiencing rapid advances in technology that promises the discovery of new means to improve early diagnosis, patient stratification, and treatment response. The main focus of novel proteomic tools aims to create a means of proteomic biomarker discovery. The instigation of any biomarker discovery effort should however involve precise study design, as this will determine the classes and reliability of markers mined in different cohorts.

Moreover, the energetic field of clinical proteomics provides a promising inquest to identify new disease biomarkers in biological fluids, cells and tissue. Well designed, optimised and validated efforts have recently shown promising evidence leading to useful biomarker clinical application[6-8]. To date, there are two main streams in clinical proteomic research; the top-down and bottom up proteomic approaches. The top down approach applies high-throughput technologies in order to identify novel markers using a non-targeted proteomic profiling approach. Markers discovered here require further identification and characterisation with respect to sensitivity, specificity, and function. The bottom-up approach focuses on formerly identified proteins and pathways and mainly focuses on protein-protein or metabolite interactions, or interactions related to specific pathways[9]. Such targeted proteomic analysis can be complicated by overwhelming data sets resulting from shot gun MS analysis rendering global profiling of complex samples a challenge. Yet, the introduction of multi-dimensional protein identification technologies (MudPIT) has paved the way to a more global bottom up profiling proteomic studies[10, 11].

Current oncoproteomic studies focus on the quantitative and qualitative profiling of proteins and peptides in clinical specimens in an effort to identify and/or target key pathological players. A major challenge in this field rises from the proteins' ability to exist in multiple forms due to post-translational modifications (PTMs) and/or degradation processes that affect protein structure, function, and localisation. It is possible that in addition to the proteome, proteolytic degradation products termed "degradomes", may also contain disease-specific information and lead to the identification of disease specific biomarkers. Although high throughput top down MS approaches have adopted adjunct methods to overcome the complexity and dynamic range issues, they generally remain less efficient in identifying such multi form complex proteomic events. Thus, there has been a growing interest in bottom up "shot gun" proteomic biomarker discovery efforts. Several models aiming to achieve the best in-depth analysis of biological samples were recently introduced. The main focus of these models aim to overcome three inherent biological hurdles; (a) complexity of the proteome, (b) the dynamic range of fluid proteomics and (c) mining PTMs[9, 12].

Many efforts have focused on "Diagnostic markers" for early and accurate diagnosis of cancer. Such biomarkers represent the main stream in cancer biomarker discovery as cancer mortality does not arise from the lack of available therapies, per se, but rather from the later diagnosis when treatment is usually less effective. "Stratification markers" are used on the other hand to classify a disease into different sub groups. This subset of markers is key to patient tailored treatment and could aid outcome forecast. Conversely, markers of treatment response and outcome prediction "prognostic biomarkers" aim to estimate the effectiveness of therapy and long term outcome of the disease.

Breast cancer is the most commonly diagnosed malignancy and the second most common cause of cancer related mortality in women[13, 14]. In the United Kingdom the post treatment 5 year survival for good prognostic groups is the same as in the general population[15], but the outlook is worse for the poorer prognostic groups regardless of advances in treatment[16-18]. Current standard diagnostic techniques for symptomatic breast disease are well documented[19-21], and whilst mammographic screening reduces mortality from breast cancer[22-24], the sensitivity of mammography for breast cancer detection is variably reported at 63%-87%[25]. Clearly, breast cancer biomarkers are essential. In this project, the differential breast cancer serum protein expression profiles are explored in an effort to expose novel diagnostic and prognostic biomarkers.

Despite advances in the field of biomarker discovery, the majority of studies failed to result in clinically applicable tests[26-30]. This can be attributed to poor study design, the complexity and dynamic range of samples studied, platform limitations, and more importantly challenging pre analytical variability interfering with true biomarker identification[9, 30]. Thus, enhanced MS technologies were applied aiming to translate biomarker discovery in this project to a clinical setting.

In this project, serum top down proteomic profiling was conducted to identify diagnostic markers of breast cancer. A multi-centre multiplex proteomic profiling approach was used to identify and validate markers discriminating invasive breast cancer (IDC), benign breast disease (BBD) and healthy controls.

A novel bottom up approach was also used to identify prognostic serum markers of early onset breast cancer (patients diagnosed below the age of 40). Here, a novel 3D MudPIT platform was used to analyse closely matched serum samples representing poor and good outcome patients from the POSH cohort.

1.2 Breast cancer biology

The mammary gland is composed of epithelial (luminal, basal myoepithelial and alveolar) and mesenchymal (adipocytes, fibroblasts, blood) cells[31]. The luminal cells line the ducts as a single epithelial layer, whereas the basal component consists of myoepithelial cells which are in direct contact with the adjacent stroma, and the alveolar cells line the alveolar lumen and are involved in milk production[32]. At puberty, mammary morphogenesis is controlled by ovarian steroids (oestrogen and progesterone), pituitary hormones (growth hormone and prolactin), local growth factors (ErbB1-4, IGF, FGF and TGFβ) and cytokines[33]. At each stage of mammary development, both epithelial and stromal cells respond to different signals that control proliferation, differentiation, and apoptosis[32]. This controlled complex interaction between the different mammary cell lines and different biological “controllers” ensure mammary functional requirement at puberty. This complex interaction also exposes the mammary gland to the risk of tumourigenesis, due to the imbalance in interpreting developmental as opposed to cancer progression signals[33, 34].

Breast cancer is a complex and heterogeneous disease that may result from the malignant transformation of mammary stem, progenitor and/or fully differentiated cells[35]. This heterogeneity is manifested by its distinctive molecular subtypes, treatment response, disease relapse and overall outcome. Both, clinical parameters such as tumour size, lymph node involvement, histological grade, and classical clinicopathological biomarkers (ER, PR and HER2) provided incomplete explanation for this variable disease behavior[36]. Research over the last decade has dissected much of the biological complexity of the disease revealing four intrinsic subtypes of breast cancer (Luminal A, Luminal B, ErbB2 over-expressing, Basal like)[37]. These are being advocated to complement the classical clinicopathological markers in providing more personalised disease management[37-40]. Despite improved molecular understanding and early disease detection provided by screening programs; 20-30% of patients develop incurable metastatic disease, and over 100,000 breast cancer patients in the western world are expected to die from the disease annually[36, 41]. This can be partly explained by other unresolved breast cancer molecular codes. In fact, ER, PR and HER2 are not accurate surrogates for intrinsic breast cancer subtypes, which indicate the presence of unexplored disease molecular pathways[40].

To date, breast cancer management faces many contests including the search for sensitive and specific biomarkers that can be exploited to detect and stratify early disease, progression and treatment response pathways. At present, conventional hormonal markers are far from perfect[42]. Other markers including carcinoembryonic antigen (CEA) and carbohydrate antigen (CA) 15-3 show poor sensitivity for early disease detection, and are mainly used as progression indicators. Recent molecular signatures suggested discriminative breast cancer molecular sub types, differentiation and prognosis by mRNA-based gene expression profiles. However, these markers have limited clinical application due to laborious analysis and variable results. Furthermore, the expression levels of mRNAs are not linearly related to the biological effectors “proteins”. Thus, proteomic biomarker discovery research is crucial to aid completing the biological picture, early diagnosis and stratification of this challenging disease

1.3 Proteomic aspects in biomarker discovery

Proteomics is the study of the proteome, including the protein components of cells, microvesicles, tissues, secretome and biological fluids. Proteomics encompasses the study of modifications, interactions and behaviour in correlation to a pathologic or a physiologic state. The proteome is dynamic and can fluctuate depending on various internal and external factors.

Clinical proteomics aim to reveal the molecular face of diseases by identifying explicit alterations in the proteome in correlation with specific pathologies. It is currently accepted that no single molecular feature alone, neither gene nor protein, could be a reliable biomarker in cancer diagnosis. The molecular complexity of cancer necessitates multi component classifying markers to build potent diagnostic, prognostic and monitoring clinical tests[43]. Although still not fully understood, multi-gene signatures identified for breast cancer have proved their diagnostic power[44, 45]. Similar efforts to identify cancer specific proteomic barcodes are needed to improve the detection power of genomic classifiers and establish a more comprehensive understanding of disease pathways. Proteomics is a powerful approach to identify signalling proteins and to decode the complex signalling circuitry involved in cancer development. Together with genomics and other sister "omics", proteomics is well on the way to resolve the molecular complexity of the different types of breast tumours[46-48], and thus provide the opportunity for new therapeutic targets to be identified.

1.4 Mass spectrometry (MS) proteomics

The aim of proteomics are centralised around the identification of the proteins in a given environment; the determination of protein interactions; and, finally, the understanding of the mechanism of their function. An ultimate result of proteomics is therefore the understanding of complex biological systems, which can lead to new diagnostics and therapy. The first step toward this is the identification of all proteins in a given system by protein mapping. Typically, this is considered the discovery phase of proteomics and involves the comparison of different biological states such as diseased and normal or drug-treated and untreated. Traditionally, this profiling has been accomplished by two-dimensional gel electrophoresis. Although 2D PAGE has successfully resolved many proteomes, it is labour-intensive, time-consuming, irreproducible and, at best, semi quantitative. In addition, 2D PAGE has been limited in exploring low molecular weight (LMW) analysis despite decades of research. The LMW (<10 kDa) component of proteome is a promising source of previously undiscovered novel biomarkers. Although critical signalling and proliferative proteomes may cluster in this compartment, classical proteomic approaches have lacked the ability of consistent profiling in this region. Since this proteomic population is below the limit of effective resolution of conventional gel electrophoresis, mass spectrometric analysis appears to provide a promising solution in these settings[49-52]. MS is quoted as "the most comprehensive and versatile tool in large scale proteomics"[12]. So far, this technology has revealed multi-peptide sets of markers that have potential prognostic and predictive values in different cancers[7, 52-54]. Primarily, top down mass spectrometry

MS based approaches have proven effective at identifying differentially expressed proteins upon resolution by 2D PAGE. More recently, progress has been made to bypass 2D PAGE dependent identification. Quantitative or semi quantitative proteomic tools using isotopic labelling and label free methods combined with high-throughput MS showed promising results[43, 52, 55, 56]. These ground breaking MS derived protein patterns can be used to form diagnostic, prognostic and follow up disease “fingerprints”. Contrary to most classical proteomic tools, this approach does not explicitly assay “identified” biomarkers, but rather focuses on detecting fine changes in the relative intensities of one or a subset of peptide/protein ion peaks. Although this “unbiased proteomics” technology heralded substantial promise, studies using this approach should be carefully designed and results should be stringently critiqued to ensure the true biological meaning of the results is obtained. On the biological side, the main concern following pattern characterisation is the identity and the origin of biomarkers observed in the proteomic patterns. Whether a biomarker is released from tumour cells, an acute phase response, or a differentially expressed/modified host protein, it should be validated as a candidate marker. Although the action of metalloproteinases, exoproteases and other enzymes is responsible for many differential markers; tumour cells can drive these and other activities that results in proteolysis and differential expression of serum proteins solely observed in affected patients[57]. In studies that have identified the diagnostic features within proteomic patterns, there is strong evidence to support this hypothesis[57-59].

1.5 Breast cancer proteomics

Despite advances, current treatment remains less effective for late-stage breast cancer and early detection holds the best promise for patients. Screening mammography is the most effective tool for breast cancer detection and yet; a significant number of the patients are misdiagnosed[25, 60-62]. Moreover, fast growth rates of tumours in younger patients results in a higher incidence of late diagnosis in this crucial group. Novel markers aiding screening and early detection of the disease are thus a must. Several studies have already addressed the possibility of applying MS analyses of blood proteome in diagnostics of breast cancer. Many of these have revealed patterns specific for patients with breast cancer at either an early or a late clinical stage[63-65]. Among the peptides identified in such differentiating patterns were fragments of C3adesArg and a C-terminal truncated form of C3adesArg, FPA, fibrinogen, ITIH4, apoA-IV, bradykinin, factor XIIIa and transthyretin[57, 64]. In addition, mass spectrometry analyses of the blood proteome allowed the identification of patterns specific to breast cancer patients with different outcomes and differential responses to therapy[66-68]. Interestingly, several markers were repeatedly detected in different studies illustrating the potential of this approach to identify robust markers[53]. Protocol(s) standardisation however remains a major challenge limiting cross study analysis and validation[9].

Breast cancer proteomics is now entering into the field of biomedicine with the aim of identifying pathological markers and therapeutic targets. Current proteomic tools allow large-scale, high-throughput analyses for the detection, identification, and functional investigation of LMW low-abundance proteins. However, major limitations of breast cancer proteomic investigations remain the complexity of biological structures and physiological processes. This has filled the path of exploration of related pathologies with various difficulties and pitfalls. The case of breast cancer illustrates the major challenge facing modern post genomic analyses; tackling the complexity and heterogeneiity of the disease. This has been addressed by establishing clear molecular sub classifications matching the genomic findings[46, 47]. Four molecular breast cancer sub groups have been confirmed both by genomic and proteomic profiling. These can be summarised as follows:

(a) Luminal A type tumours with the highest expression of the ERα gene, GATA binding protein 3, X-box binding protein 1, trefoil factor 3, hepatocyte nuclear factor 3α, and oestrogen-regulated LIV-1. These tumours seem to be involved in biological processes including fatty acid metabolism and steroid hormone mediated signaling[38].

(b) Luminal subtype B which shows similar ERα gene expression with PR positive or HER2 positive tendencies[37].

(c) ERBB2 positive group is characterised by high expression of several genes in the ERBB2 amplicon at 17q22.24 including ERBB2 and GRB[37].

(d) Basal-like subtype which is characterised by high expression of keratins 5 and 17, laminin, and fatty acid binding protein 7. The most over-enriched processes in this sub type include many cancer "hallmark" genes, such as the cell cycle, cell proliferation (MCM3, MCM4, MCM7 and MAD2L1) and differentiation, protein phosphorylation, B-cell-and antibody-mediated immunity[38].

There are two main expected outcomes from proteomic analyses of breast cancer. The first is to discover new molecular markers of breast tumours. The second is to decipher the intracellular signalling pathways leading to the initiation and progression of breast tumours. Such data should provide the knowledge base for the identification of new therapeutic targets and the development of innovative strategies against breast cancer.

To date, progression from bench side biomarker discovery to clinical application remains slow. This lag can be partly attributed to a lengthy biomarker identification and validation process. The lack of standardised profiling protocols and the population difference in various studies are also important limiting factors[9]. It is anticipated that standardised multicentre efforts would enhance the likelihood of validating clinically reliable breast cancer biomarker(s). Such biomarkers or biomarker proteomic profiles may improve cancer screening, diagnosis and outcome prediction as well as patient tailored treatment strategies.

1.6 Serum proteomics challenges

To date, a major element in the search for proteomic biomarkers has focused on bio-fluids; where an ideal biomarker could be detected from samples obtained using simple and minimally invasive methods. Serum and plasma represent very rich sources of biomarkers as they perfuse all tissues, resulting in a rich source of novel markers. These can be shed from tumour cells undergoing necrosis, or generated as a host response to the tumour burden. Unfortunately, most of these valuable proteins in this environment will be masked by a complex mixture of highly abundant proteins such as albumin and immunoglobulins[69]. This dynamic obstacle limits the identification, validation, and subsequently, the clinical translation of such potential markers in the presence of higher abundance proteins. Therefore, comprehensive analysis of these inherently complex and versatile serum samples is exceptionally challenging. Simple comparison can result in the detection of differentially expressed protein(s) simply due to differences in patients' age, gender, diet, health status, ethnicity etc., leading to false biomarker discovery. Adding to the challenge is the fact that discovered potential markers might not present a "real world" stable protein/peptide leading to its invalidation. Clearly, conquering the real world proteome is a major challenge.

Reducing the complexity of the circulating proteome and enhancing the detection power of low abundance proteins is thus vital for biomarker discovery[12, 70]. This can be achieved through pre fractionation using chromatographic and electro-focusing techniques, immuno-depletion of the most abundant circulating proteins, and enrichment of "trace" proteomic signals using combinatorial peptide ligand libraries (CPLL). Targeting the sub proteome of organelles of interest and/or isolating specific proteomic sub-groups, including the phospho and glycoproteome, is also an alternative approach, which complements the results gained by these methods[71-76]. Above all, appropriate study design, protocol standardisation and group matching are obvious, yet vital steps towards meaningful biomarker discovery. These are critical steps towards a unified proteomic approach vital to clinical advances in breast cancer diagnostics and therapy.

1.7 Proteomic Patterns

Pathway based biomarker discovery focuses on proteins secreted in different biological fluids to study molecular networks. Blood has been frequently studied in the biomarker mining process over the last decades. Nonetheless, only a few markers (e.g. CEA, CA125, PSA, CA19.9) are currently used with limited clinical usefulness, due to their low diagnostic and prognostic power[2, 77]. For example, PSA is probably the most renowned screening marker to come into clinical use in the last three decades, yet, the use of PSA for prostate cancer screening is controversial[78]. Although serum PSA is usually raised in prostate cancer, the specificity to cancer is limited and other non-malignant prostate pathologies (BPH and prostatitis) are also associated with raised PSA.

A commonly used approach that takes into consideration the “fingerprint” of the whole proteome regardless of the identity of differentially expressed proteins is denoted “proteome pattern analysis”[79]. Here, protein signatures generated by MS or 2D gel electrophoresis are used to differentiate two physiological and/or pathological groups regardless of the proteomic identity of these classifiers. This approach adopts a strategy similar to the hunt for multi gene signatures in functional genomics. Multi component sets of peptide/protein ions at defined mass to charge (m/z) values are used to characterise and classify samples even though their individual components may lack the differentiating effect when analysed individually. The use of proteomic patterns to classify cohorts without detailed tandem mass analysis of blood is a growing method in clinical proteomics. So far, exploiting differentially expressed profiles that represent cancer specific fingerprint(s) have revealed few early results in cancer research[67, 68, 80]. In breast and other cancers, a shortfall of this strategy is illustrated by the variable and discrepant candidate markers identified in different MS studies[63, 64, 81]. However; this can be related to inconsistent pre and/or peri-analytical variability including differences in sample populations, collection and storage methods, preparation and bioinformatic analysis. Yet, commonly validated biomarkers are still remarkable and pattern analysis remains a valid approach[63, 81, 82]. Another major shortfall in this approach also lies in the variability related to protein isoforms and PTM, which can falsely alter unique proteomic barcodes and classification results. Moreover, such patterns are largely dependent on the original profiling method of analysis, which may be clinically impractical. Overall, this remains a valid approach in proteome profiling despite the shortfall of missing biological information due to the absence of protein identification. Yet, these non-sequenced signatures should be validated with high sensitivities and specificities prior to translation into clinically useful tests.

1.8 Proteomic pathway analysis

The future of cancer therapy relies on using targeted molecular approaches. Thus, characterisation of cancer related cellular pathways, and understanding the impact of disease and therapy on cellular networks is crucial. Although chemo-resistance continues to be a hurdle, oncogene targeted therapies have already been shown to be effective[83, 84]. These remedies usually target proteins which are altered in cancer. Since protein groups are the basis of biological functions, including cancer development, there is a trend to explore and target novel protein pathways instead of individual proteins[1, 5]. Once these pathways are known and validated, cancer progression and therapy should be studied, focusing on interactive molecular pathways rather than individual genes or proteins. Pathway biomarker(s) discovery should be regarded as a key component in any proteomic study. This approach is however limited by certain factors. As circulatory proteins are more likely to reveal such pathways, blood remains an attractive mining field for these pathways. This is a challenging era given the complexity, dynamic range and variability of both serum and plasma[53]. An alternative approach in the hunt for these markers is to study the local environment fluids or the tumour secretome. This has the advantage of a higher concentration and a narrower dynamic range. Nonetheless, biomarkers revealed by this strategy would lack the essential and early elements of the host response. In particular, many acute phase proteins, constituting a critical part of cancer pathways, originate in the blood. Moreover, an intense local environment biomarker signal is threatened by the dilutional and proteolytic effects commonly occurring in the blood leading to failure of further validation efforts. Regardless of the method, a highly desirable feature in pathway marker discovery is that they can be monitored in the blood. Mass spectrometry combined with pathway analysis methods have been the principle tool in recent cancer biomarker research and are expected to lead the way in this area[47, 85, 86]. The mining of breast cancer proteomic pathways and network markers remains to date an attractive field for research.

1.9 Mass Spectrometry in cancer proteomics

Proteomics is a rapidly evolving strategy for biomarker discovery. Several advanced proteomic profiling tools have shown utility for identifying biomarkers for the diagnosis of several tumour types including breast cancer[7, 52, 63, 65, 80, 87-94]. MS specifically has been successful in effectively resolving low mass/abundance proteins using minimal amounts of samples, two of the main challenges facing discovery in classical proteomics. This technology was shown to be reproducible and can be used to build promising cancer classification algorithms[63, 95-97]. A new inherent challenge has evolved with the advent of the new era of MS based discovery; consistent data analysis. The amount and complexity of data generated in these experiments is overwhelming, and can limit the generation of reproducible and relevant results and/or new working hypotheses. Clearly, experience rather than insight in bioinformatics remains essential to allow the integration of MS data into the molecular biological context.

1.10 Top-down or bottom-up MS proteomic profiling?

Standard proteomic analysis begins with the sample preparation in which proteins are either digested into peptides (bottom-up analysis) or analysed intact (top-down analysis)[98-100].

Top-down methods use masses of intact proteins and/or their fragments for successful biomarker identification. Some of the benefits of the top-down approach include higher sequence coverage of target proteins and better characterisation of posttranslational modifications[101]. Compared with bottom-up approaches, the lower sequence coverage abilities of top-down experiments can provide high throughput abilities, reduce ambiguities of the peptide-to-protein mapping, allowing the identification of specific protein isoforms[102]. Moreover, the top-down approach improves the reliability of protein quantification as protein abundances are measured directly instead of extrapolation from peptide signals. However, there are several technological limitations, including the need for separation of intact proteins prior to analysis, the lack of efficient methods to fragment large proteins, and the demanding transition from protein mass to identification in certain platforms. Thus, the scope of the top-down approach has been limited[12, 103].

The bottom-up (shotgun) approach is the most popular method for targeted proteomic analyses[12]. Shotgun proteomics is an approach in which proteins are digested into peptides prior to MS analysis, and the ensuing peptide masses and sequences are used to identify corresponding proteins[104]. The main advantage of this method lies in the instant identification of potential biomarkers by the end of the analysis, as well as the high resolution provided by many tandem mass platforms. On the other hand, bottom-up proteomics results in overwhelming peptide data from complex biological samples (such as serum), loss of some PTMs secondary to limited protein sequence coverage, and ambiguity of the origin for some redundant sequences. Thus, modifications of this approach involving multi-dimensional chromatographic fractionation are essential to reduce the potential shortfalls associated with this approach[103].

Clearly, there is not a perfect approach for proteomic profiling, and the decision of using either is dependent of the study design, samples analysed and more importantly, the question to be answered.

1.11 MS serum cancer proteomic biomarkers: biology or artefact?

To date, mining the sera for novel cancer biomarkers using promising mass spectrometry (MS) based profiling technologies has not only been a source of optimism, but also an unremitting debate[44, 105]. As tumourigenesis involves the transformation and proliferation of altered cell types that produce high levels of specific proteins and enzymes such as proteases, it not only modifies the barcode of existing serum proteins, but also their metabolic products. It is well established that human serum contains thousands of active proteolytic pathways, yet it remains uncertain whether this complex peptidome may provide a real correlation to biological events such as cancer. Recent reports have advocated the use of MS-based serum peptide profiling to determine qualitative and quantitative disease patterns[57, 106, 107]. Yet, *ex vivo* proteolytic processes constantly question the biological relevance of many MS reported markers in relation to tumourigenesis.

Variations in sample collection, storage, and preparation conditions could lead to biased activation of different proteolytic pathways and could significantly affect the results obtained in clinical proteomic studies[108]. This is a major concern over the use of MS and other profiling techniques for the identification of cancer biomarkers in serum collections. This is principally relevant studying archival samples stored in different tissue banks worldwide. Although such collections have great potential value for the identification of pre-symptomatic and other clinically valuable disease biomarkers, inconsistency of pre analytical conditions limits their utility. Sample handling, storage temperature, collection tubes, coagulation temperature, and freeze thaw cycles have been studies, and found to have a dramatic effect on the proteome[108, 109]. Yet, investigating the usefulness of analysing samples stored for decades remains ambiguous. Thus far, it is unknown if biomarkers reported in recent standardised studies can be reliably detected and retain predictive power in older tissue and serum banks. Recently it was found that samples from tissue banks can be a useful resource for biomarker studies[110]. The hypothesis that protein biomarkers existing in serum can be reliably and repeatedly attainable, and may distinguish cancer from non-cancer state, even decades after collection was also tested. Historic serum samples have been shown to give highly reproducible spectra[111]. Validation of currently reported biomarkers using archival samples will be the testament for their prospect in future studies. This would also give an answer to the utility of some controversial markers, and whether they represent a real biology or just a pre/peri-analytical artefact.

1.12 Breast cancer: the "real world" markers

As the leading cause of cancer related deaths in women, breast cancer research and biomarker discovery in particular, is attracting more interest as researchers explore the molecular face of this disease. MS serum proteomics has generated a rapidly growing interest in novel biomarkers that can differentiate breast cancer patients from healthy individuals, and into distinct sub groups[44, 103, 106]. However, despite a wealth of research, authentic real world biomarkers remain elusive. This could be due to several factors: Are we using controls that are too "ideal"? Is it the fact that global standardised protocols remain missing? Or is it that such protocols are unattainable in the virtual clinical setting? These could all be contributing factors to the prolonged search for reliable markers, but it could also be attributed to the heterogenic behaviour of breast cancer. This heterogeneous disease pattern results from various genetic alterations, and explains at least in part, the wide variety of clinical response and outcome in breast cancer. Deeper molecular understanding is necessary to identify relevant targets for early detection, stratification and novel treatments. A molecular taxonomy of breast cancer has been defined based on genomic microarray analysis[46]. This divides the disease into four molecular subtypes: luminal A and B, ERBB2-overexpressing, basal-like disease. As expected, these different sub groups have a distinct clinical course and response to therapeutic agents[112, 113]. Overall, the luminal group (oestrogen receptor ER positive representing 60% of breast cancers) generally have a good prognosis, although subtype B shows a lower ER and higher proliferative profile, thus it has a poorer prognosis in comparison with subtype A). ERBB2 over-expressing is ER-negative with ERBB2 over-expression, presenting 20-30% of the cases. The basal-like sub group also manifests the ER negative and HER-2-negative states; this group affects 10-20% of breast cancer patients. Both the ERBB2 and basal are considered as poorer prognosis subtypes[4, 46, 114].

Mammography is the gold standard for the detection of breast cancer yet it has concurrent critical limitations. Firstly, albeit minimal, there is risk associated with the mammography radiation. Repeated exposure can theoretically increase carcinogenesis risk[115, 116]. Secondly, a breast tumour must be at least a few millimetres in size to be detected. Given that a single cell can progress to the development of a whole tumour (clonal origin of cancer); it is already somewhat late when the tumour is detected by mammography. Finally, evidence has illustrated that a higher rate of over diagnosis and benign biopsies with mammography potentially compromising subsequent disease diagnosis[60, 117]. In clinical practice, after the surgical removal of a tumour, its characterisation as malignant or benign is made by histology. Such parameters as tumour size, grade and node involvement are then used to decide treatment and prognosis. However, breast cancer is not a homogeneous disease and there are different molecular subtypes dictating its behaviour and outcome. Depending on the cellular and histological origin of the cancer cells and on the evolution of the disease, a more personalised diagnostic and therapeutic approach aided by molecular markers is required.

The classical biological markers used for therapeutic and prognostic purposes are oestrogen and progesterone receptors. These are used for the selection of patients potentially responding to treatment with anti-oestrogen treatment[118]. In addition, ErbB2 (A tyrosine kinase receptor over-expressed in about 20% of breast tumours) is also utilised as a prognostic marker in breast cancer. The recent understanding of the molecular role of this marker has opened new aspects for breast cancer treatment. Herceptin, a truncated blocking antibody directed against ErbB2, has been successfully developed and has now entered into clinical practice[119]. However, breast cancer can be regulated by other molecular factors that either stimulate or inhibit proliferation, migration, and differentiation, acting to promote tumour growth and metastasis[120]. This may explain the different responsiveness and outcome to tamoxifen and Herceptin in patients within the same breast cancer sub group. The identification of other decisive “real world” markers, and corresponding targets, remains a significant objective for future studies.

1.13 Surface enhanced laser desorption ionisation - time of flight mass spectrometry (SELDI) in breast cancer biomarkers discovery

1.13.1 Background

The critical need to find new molecular markers for detection, typing and treatment of breast cancer has initiated a wide range of genomic, transcriptomic, proteomic, metabolomic and interactomic studies. The discovery of new markers and therapeutic targets as well as the corresponding drugs using proteomics and other sister "oncomic" studies is highly anticipated. The mammary gland and breast tumours are complex and dynamic structures and full understanding of breast carcinogenesis with genomic and proteomic tools is under extensive research. Among the various proteomic tools used in the breast cancer markers hunt, Surface enhanced laser desorption ionisation - time of flight mass spectrometry (SELDI-TOF MS) -thereafter called SELDI- emerged as a promising discovery tool[47, 65, 66, 121-124]. By blending together the principles of retention chromatography and mass spectrometry, it provides a rapid, high-throughput, and a relatively sensitive screening method capable of detecting and analysing complex protein samples. This technology is also capable of the rapid separation, detection and analysis of proteins at the femtomole level directly from biological samples. Per se, it is a tool that makes multi-analyte discovery possible, and enables analysis of large numbers of different samples with the simultaneous study of multiple biological variables. This approach is a modified version of the matrix assisted laser desorption ionisation (MALDI) platform, where the matrix absorbs laser energy and transfers it to the acidified analyte, and rapid laser heating causes desorption of matrix and M+H+ ions of analyte into the gas phase[12]. MALDI ionisation also has the advantage of high sample throughput, which allows the generation of sufficient data to adequately power statistical tests.

The principles and clinical applications of SELDI have been reviewed, and have been successfully used in biomarker discovery on various tissues and samples[94, 121, 125-127]. Briefly, SELDI can be summarised as extra surface affinity array fractionation of various protein and peptide molecules prior to MALDI style analysis. Only a few microlitres of the sample of interest are deposited on a selective chromatographic surface (array), which retains proteins from biological mixtures according to their physicochemical properties. Consequently, the proteins of interest are captured on the chromatographic surface by adsorption, partition, electrostatic interaction or affinity depending on the arrays' properties. Mass testing and comparison between protein ion peaks can locate significant protein abundance differences between samples. This "unbiased" top down proteomic analysis allows the discovery of early stage breast cancer biomarkers in cellular or biological fluids unlike earlier classical techniques which identify single proteins. Such early stage diagnostics are invaluable in breast cancer where ductal carcinoma *in situ* (DCIS) is potentially curable[128, 129] and when currently, only 63% of breast cancer cases are localised at the time of diagnosis[128].

1.13.2 Why serum proteomic profiling in breast cancer?

Serum has been widely utilised for study by proteomics as it is minimally invasive, reproducible and cheap to obtain. During carcinogenesis, differentially expressed proteins in the blood can originate from diverse sources. Communication of cancer cells with the blood may lead to the release of tissue specific proteins upon cell damage or death. Serum profiles could also present cell death tissue specific or cancer excreted proteins, host response, acute phase reactants or even altered housekeeping protein levels in response to tumour burden. Since these profiles could represent altered phenotypic events during neoplastic transformation and progression, it is hypothesised that a characteristic proteomic profile would be present in the presence of cancer. This could then be considered as a new "screening" methodology or be utilized in the ongoing clinical mamangement of patients.

1.13.2.1 Serum MS proteomic profiling in breast cancer prediction

One of the earliest pilot studies of SELDI proteomic profiling in breast cancer was a prospective analysis of pre-treatment serum samples by Vlahou *et al.*[89]. Three discriminatory proteomic peaks differentiating cancer from normal groups were discovered. In addition, three different peaks segregated cancer serum sample proteomic profiles from the benign group. In support, Laronga *et al*[65] reproduced the previous work showing consistency between the studies[64].

Further studies using SELDI screening revealed a consistent significance of a panel of three markers (BC1: 4.3 kDa, BC2: 8.1 kDa, BC3: 8.9 kDa) in the serum of breast cancer or healthy controls[63, 64, 81, 124, 130]. The peaks were identified to be: BC1; a truncated form of interalpha-trypsin inhibitor heavy chain H4 (ITIH4), BC3 and BC2 were part of the serologically abundant complement components C3adesArg and a C-terminal truncated form of this component (C3adesArg) respectively.

These findings raised one of the concerns consistent with reported proteomic biomarkers non-specific acute phase molecules qualifying as discriminatory classifiers. In addition, the scarcity of samples representing DCIS was a noticeable limitation. Moreover, the reproducibility and robustness of these findings were under scrutiny as the markers were partially or discrepantly validated. Yet, such conflicting evidence could be related to inconsistent collection, storage, handling, sample preparation techniques, analytical and bioinformatic settings as well as population differences.

Despite these unpredictable findings, these markers remain the focus of ongoing translational efforts (United States Patent Application 20080311673).

A further large and validated study involving 310 subjects[131] explored the ability of mass spectrometry to characterise stage-I breast cancer patients from healthy subjects. Mammography was used to exclude cancer in the control group, and patients with family history of breast cancer were excluded from the control group. Seven ion peaks were identified that discriminated between the groups with a sensitivity and specificity of at least 95% and 85% respectively, both in the testing and validation sets. Although the identities of these markers remained anonymous, this demonstrated the potential of SELDI profiling to generate a robust detection signature of early stage breast cancer, and a major landmark in pattern analysis studies.

1.13.2.2 Serum proteomic biomarkers in BRCA1 mutation patients

BRCA1 mutations only account for 7-10% of breast cancers[132] with a propensity towards early development and poorer outcome. It is estimated that 80% of patients with BRCA1 mutations ultimately develop breast cancer. Targeted diagnostic and therapeutic strategies in this patient group are crucial. Hence predictive markers to identify those who are less likely to develop breast cancer would enable more personalised management strategies. BRCA1 mutation carrier monitoring is another promising utilisation of proteomic profiling. Studies have shown that differentiation between BRCA1 cancer and BRCA1 carrier groups could be achieved by two biomarker peaks[65, 66]. Although the cohort studies were small (n=30), an over-expressed 8.1kDa and an under-expressed 5.9kDa in the cancer group were found to reliably discriminate the two states (sensitivity 100%, specificity 87%). Whether the distinction in protein profiles found between the two BRCA1 mutation groups represents the early detection of a precancerous state or an occult malignancy marker is yet unknown. Once validated, such markers could thus be used for surveillance and/or early disease detection. Differentiation between BRCA1 cancer and sporadic breast cancer groups was also possible with a differentiation sensitivity of 94% and specificity of 100%. In addition, the separation between BRCA1 cancer and the control group was achieved with 87% sensitivity and 94% specificity. Despite the small number samples, such reproducibility of SELDI proteomic profiling published by multiple researchers is worthwhile validating at different sites to confirm these significant findings. Yet, to date there have been no published follow up studies targeting this question. This could be explained by the lengthy recruitment and follow up processes required to adequately power and fulfil the requirement for this specific cohort.

1.13.2.3 Proteomic profiling and biomarker discovery pre and post- surgery

Analysing paired samples from breast cancer patients' pre and post-surgery can yield significant markers and pathways elucidating disease development, progression and outcome. Reports covering this area have previously found three significantly differentially expressed ion peaks between the pre and post-surgery groups in serum with a sensitivity and specificity of 75% and 87% respectively[65]. Investigating whether samples from breast cancer patients following surgery could be differentiated from healthy controls is a key target in this area. This is of interest as, if removing the tumour lead to restoration of the healthy proteomic profiles, it would be reasonable to link the characteristic peaks to the proteins shed or inhibited by the tumour. In contrast however, the SELDI analysis clearly differentiated protein profiles between the controls and breast cancer patients post-surgery, with a final separation sensitivity of 93% and a specificity of 73%. Interestingly, the post treatment group comparison with control samples showed a partial retention of cancer profiles even after surgery. This can be explained by an ongoing systemic process from which these markers originated. In addition, the presence of residual tumourigenic processes may also explain the retention of malignancy markers. As the sample size limitation (n=84) remains a major shortfall in this study, these findings will need further evaluation and identification of candidate biomarkers is essential to explain this phenomenon.

1.13.2.4 Serum proteomic biomarker profiles and clinicopathological variables

Study of the association between biomarkers and other clinicopathological factors may lend insight into the biological significance of disease profiles. Profiles strongly associated with estrogen receptor (ER) status for instance might be expected to be due to proteins involved in hormonal signalling. It is also possible, however, that the absence of such a correlation with known clinicopathological variables might be due to the identification of new independent neoplastic molecules. Reports of clinicopathologically correlating markers have suggested the possible differentiation between lymph node positive and negative groups using six MS ion peaks, which could prevent unnecessary extensive surgery in some patients[65]. However, the most frequently reported serum breast cancer markers in SELDI studies (BC1, BC2 and BC3) do not correlate with clinicopathological features such as tumour size, nodal involvement, grade or ER/PR status[64, 81, 124]. No correlation was found between clinical or histological parameters including age, nodal status, metastasis, vascular invasion, ER/PR, and Ca15.3 in different studies[65, 124]. These findings illustrating the lack of correlation between many candidate proteomic markers and current prognostic factors could be related to the heterogeneity of this disease and the different subgroup entities under which the patients fall. In depth research of such correlations remains missing and may lead to better understanding of the correlation between the molecular and histopathological aspects of breast cancer.

1.13.3 Nipple aspirate fluid (NAF) MS proteomics

Intraductal sampling and diagnostic techniques are minimally invasive and allow direct access to the ductal system; the origin of most breast malignancies. This could be particularly useful in young women where mammographic sensitivities are reduced (i.e. in younger patients and in obese patients with high breast density). Breast duct fluid is a rich source of protein from the immediate tumour environment. Furthermore, compared to serum, breast ductal fluids could be a superior source of breast cancer markers due to close vicinity of breast ductal cells, particularly in early stage *in situ* disease, where the ductal fluid may contain tumour markers that are excluded from the circulation[53]. NAF has been the focus of breast cancer marker(s) profiling using SELDI and other sister proteomic technologies[133-136].

Work analysing NAF from both breasts from patients newly diagnosed with unilateral breast cancer and high-risk controls showed clear separation between the cancer bearing breast and control NAF profiles. The main discriminator between the two groups was shown by a peak at 15.9kDa identified as β chain of haemoglobin[137]. In a follow on study, these findings were not validated and other peaks at 5.2 and 33.4kDa differentiated DCIS from benign breast disease and invasive carcinoma (IDC)[136, 137]. Importantly, these ion peaks were consistent in differentiating the DCIS from the benign group, which may potentially provide an early predictor insight of the development of invasive tumours. Overexpression of the protein peaks at 5.2, 13.88 and 33.4kDa was also shown to be associated with both DCIS and IDC but not with ADH or benign disease. These biomarkers were integrated in a prediction model including age and parity, and achieved a sensitivity and specificity of 40% and 94% respectively[136]. Although the specificity achieved by this classification is outstanding, such low sensitivity classification is of limited use in future clinical applications.

In a landmark study, Li *et al*[123] identified three potential biomarkers which differentiated cancer NAF from controls. The 3.375, 3.447 and 3.490kDa peaks were identified as human neutrophil peptides (HNP) 1-3. This was also verified using SELDI immune-capture assays on breast cancer samples which originally revealed higher HNP1-3 levels. Moreover, Li *et al.*[123] used ELISA to quantify the three biomarkers providing further validation of their findings using SELDI. This high HNP level in cancer can be explained by tumour cell invasion and release of neutrophils and eosinophils, or by direct secretion of these peptides. Interestingly, HNP had been previously detected in tissue samples from several other cancers. This work added new evidence to the significance of immunological pathways in breast cancer[59, 88, 138, 139]. Targeting the pathways involving these markers in cancer progression is thus essential and would confirm the role of HNP in breast and possibly other cancers. Following this Li's group illustrated that an alpha1- anti trypsin fragment (ΔAAT) was associated with IDC, ADH and DCIS but not healthy NAF controls from paired samples and healthy controls [140]. The limited number of samples (IDC: 31, ADH: 5, DCIS: 6 and controls: 31) remains a drawback of this study, however, the consistent validation of this marker indicates its potential reliability as an early diagnostic/predictive marker.

Whilst comparison between two samples from the same individual is an attractive approach providing an internal control for hormonal and environmental effects, findings showing no significant differences in protein profiles between ipsilateral affected breast and the unaffected contralateral breast NAF in patients with breast cancer are confusing[141]. Adding to the confusion, comparisons of NAF from the unaffected breast in patients with breast cancer and NAF from healthy controls showed clear differential expression between the two groups[141]. The absence of any discriminatory peaks between the NAF from the ipsilateral affected breast and the contralateral unaffected breast is remarkable. These results support the assumption that breast cancer field changes occur across both breasts as part of cancer development. Such phenomenon indicates a potential systemic progression of cancer rather than local tumourigenesis and could signify inevitable contralateral breast cancer development at some stage. Although this is a striking finding, further evidence is yet needed as several previous reports contradicted these findings[140, 142, 143].

1.13.4 Cytosolic protein analysis in breast cancer

Tumour biomarker proteins related to cancer cells or their microenvironment might be attenuated, altered or fragmented once shed in the blood. This may have an effect on the profiles obtained by MS. Therefore, cytosolic extract represents an alternative source, with higher, more relevant protein concentrations and less non-biological variability effects. In addition, cellular profiles have a higher chance of detecting intracellular pathways and interactions. Intracellular proteomics from this environment includes cytosolic protein mRNA complexes which could affect gene transcription and be linked to cancer development. Mitochondrial protein behaviour could be linked to cancer development and treatment resistance and is closely monitored in cytosolic profiling.

Cytosolic protein profiling was performed by Ricolleau *et al.*[144] to identify patients with a high risk of relapse, who would benefit from systemic therapies, and those who were unlikely to benefit could avoid these therapies and their concomitant side effects and costs. Two candidate biomarkers at 8.5 kDa and 19.8 kDa (ubiquitin and ferritin light chain (FLC) respectively) showed a prognostic predictive power where the 8.5kDa peak was found at lower levels in patients with metastatic disease and the 19.8kDa peak was found at higher levels in these patients. Interestingly, combining the St. Gallen risk profile (age <35, size >20mm, negative hormone receptors, Grade 3) to the 2 peaks led to improved discrimination between the groups of patients who suffered a systemic relapse and those who did not. These promising findings could aid in risk subgroup stratification of patients that may benefit from, or safely avoid, chemotherapy.

A key study in this field by Brozokova and co-workers[47] focused on linking proteomic patterns to tumour clinicopathological characteristics. They found a clear clustering of biomarker patterns into five further sub groups associated with tumour type, hormonal receptor status and nuclear grade. Hormone receptor expressing luminal groups, HER2/neu positive and the basal high proliferation gene expression subtypes were generated by hierarchical clustering. Strikingly, this molecular tumour classification was identical to those generated by cDNA genomic expression profiling[145]. Two potential markers were identified (heat shock protein -HSP 27- and Annexin V) within these tumour sub-group classifiers. Both proteins were mainly over expressed in the luminal sub-cohort.

Other reports focusing on cellular proteomic indicated that ubiquitin may be differentially expressed in axillary lymph node metastasis tissue indicating a potential prognostic role[146]. This was supported by evidence correlating ubiquitin and S100A to apoptotic mammary activities and disease prognostication[147, 148].

These studies illustrate the importance of this post genomic approach in rectifying and expanding the molecular understanding of breast cancer tumour behaviour. This could lead to further understanding of the molecular role of new proteomic markers in patients' sub groups and outcome.

1.13.5 Saliva and tear proteomic biomarkers in breast cancer

Disease states can alter the proteomic environment of any biological fluid including the saliva. For example, reports showed that the levels of kallikreins, CA125, EGF and cErb2 differed in the saliva of healthy compared to diseased subjects[149-153]. Profiling the salivary and tear proteome has attracted more focus as these samples are safe, simple, inexpensive, easily repeatable and most importantly non-invasive to collect. Moreover, the lower concentrations of abundant proteins such as albumin in these samples enables easier analysis compared to serum. Researchers have previously studied the possibility that saliva may be used to diagnose systemic diseases[154-156]. MS was previously reported to detect plausible breast cancer markers in saliva and tears[153, 157, 158]. The sensitivity and threshold of MS in detecting significant differentially expressed peaks in these fluids between breast cancer groups was promising in these efforts. Larger discovery, and blinded validation studies are crucial to confirm any possible discriminatory peaks differentiating breast cancer groups.

1.13.6 Urinary breast cancer proteomic biomarkers

Human urine can be collected in a non-invasive fashion and has been extensively used mainly for the study and monitoring of renal conditions[159-161]. Normal urine contains proteins and peptides which are derived from either glomerular filtration of plasma, GPD-linked proteins or exosomal secretion by epithelial cells[162]. Tumour-related proteolytic activity might result in disease specific signatures of proteolytic fragments which are likely to be present at very low concentrations[163], and in actual fact masked by abundant serum proteins. As discussed earlier, several LMW enrichment tools can be used to maximise the detection power of these markers in the serum. However, each of these methods have its drawbacks including; co-depletion of LMW markers, poor fractionation abilities, irreproducibility and skewed protein/peptide quantification. A natural fractionation tool, eliminating the large protein masking effect, is provided by the nephrones through urinary secretion. Thus, exploring such fragments in urine, the natural ultra-filtrate of serum, is an attractive approach. This methodology can be enhanced by the use of the LMW detection abilities of MS.

To date, MS urinary proteomic biomarker discovery has mainly focused on urological/nephrological diseases including prostate, renal and bladder cancers as well as diabetes[161, 164-166]. In both scenarios, urinary protein secretion could be related to either local urothelial events, reflecting the disease microenvironment, or a nephron deficient process as shown in diabetic patients[167, 168]. Several studies indicated the early diagnostic potential of urinary breast cancer markers[169-171]. Metabolomic evidence in the urine suggested diagnostic and prognostic value of markers in this easily accessible biological fluid[172]. It should be noted however, that several factors could influence the usability/reliability of urinary MS profiling. Firstly, progressive degradation of proteins at room temperature and to a lesser extent the number of freeze-thaw cycles. This effect can be alleviated by the addition of protease inhibitors only up to 2 hours from the collection and standardised sample handling[162]. The timing of collection over the day as well as the type of matrix used could influence the intensity of peaks. Finally, confounding urological and/or nephrological pathologies might skew the profiling process.

Despite these obstacles, encouraging evidence from colorectal and ovarian cancer urine biomarker discovery advocates mining the urine proteome for markers of other non-urological solid tumours[173, 174]. The breast cancer urinary proteome is an under-explored venue which warrants further research.

1.13.7 Limitations in SELDI proteomic profiling[53]

With advances in proteomic research and mining techniques, the limited number of proteins detected previously in serum by traditional proteomic tools can be increased. SELDI as a high throughput solid phase extraction technology has provided a useful means of protein extraction, fractionation and detection. This technology presents a good potential provided parameters, such as sample handling, analytical settings and data analysis, are strictly controlled. SELDI was perceived as a suitable tool to enhance the power of profiling studies through mounting the analysis numbers of both samples and proteomic signals. Yet, a low molecular mass detection preference, selective surfaces, and semi quantitative abilities are factors that could limit biomarker detection. Adjoined to technical aspects including low resolution, non-sequencing function, and slow progression to validation, identification and clinical utility have contributed to sceptical SELDI views.

Variability and the lack of sequencing abilities currently limit wider applications of this tool. In addition, discrepant reports failing to validate earlier SELDI studies have encouraged conservative views of this method[175]. This may be related to small sample size and poor study design. In addition, biased non biological variability can result from trivial non standardised step(s) in any proteomic profiling work; starting from sample collection, and ending by data processing and analysis[108, 109, 176]. Studies testing the effects of different variables including storage tubes, clotting time, incubation temperature, storage temperature and handling proved the importance of uniform handling to exclude systemic pre-analytical inconsistency and false discovery[177, 178]. However, sub-standardised protocols in different validation studies generated deficient loops of conflicting results including clear variation in the discriminatory power and direction of several putative cancer biomarkers[53]. This, in addition to, pre-existing reproducibility and identification concerns has pessimistically affected the views on SELDI[179]. In a real world setting however, unifying patient cohorts and sampling protocols, and to a lesser extent, analysis settings are challenging tasks. A valid system to reduce the effects of such inconsistency is the use of adequately powered upstream studies analysing large cohorts. This would attenuate non-biological variability effect, and select for potential clinically valid biomarker patterns. Moving from detection to identification of biomarkers is a well-known limitation of SELDI and other top down biomarker discovery tools. Nonetheless, potent cancer biomarker signatures can be transiently used in multi centre validation providing a proof in support of or against their relevance. Identification of existing markers can then be established using other conventional downstream methods.

The identification of acute phase proteins or proteins resulting from exoproteases, as candidate biomarkers in SELDI reports also raised some concerns about this profiling method[57]. Nonetheless, it cannot be assumed that such findings imply underperformance. After all; any signal, including proteolytic products and acute response players, can reflect a milestone in tumourigenesis[57]. As long as candidate biomarkers are validated and are not merely false positive biomarkers, their role in cancer evolution remains a potential[53, 57].

The small size and poor design of some studies makes validation of many biomarkers quite challenging[180, 181]. Evidently, standardised biomarker discovery protocols are crucial for future unbiased discovery. Candidate biomarker validation, verification and authentic identification should be the basis for any discovery effort, should discriminatory biomarkers be unveiled. Following similar biomarker discovery methods with strict standardised steps would reduce the rate of false discovery and recover some of the reduced confidence in this field. This is supported by recent findings indicating common biomarker findings despite the variability hurdle facing SELDI profiling; a promising sign motivating further focused and standardised profiling efforts[182].

1.14 Early onset breast cancer

Early onset breast cancer occurring in women below the age of 40 years accounts for 5-10% of all breast cancer cases[16, 183]. This group of patients generally show worse prognosis, especially in the case of BRCA1/2 mutation carriers[184]. Although adjuvant therapies are widely used in the younger breast cancer age groups, and despite an overall decrease in mortality from breast cancer, early onset breast cancer mortality rate has not decreased over the years[185, 186]. In contrary, early onset breast cancer is commonly associated with poor prognosis[187]. This paradoxical behaviour between early and later onset disease can be explained by early diagnosis in the older age group by targeted screening programs, and more importantly, a distinct biological behaviour between the two groups. This was demonstrated in a large study involving over 5,000 patients with early onset breast cancer where larger tumours, lymph node involvement and higher grade tumours were noted over a median of 11 year follow up period[185]. Most of the 45% deaths reported in the former study occurred within the first five years. Such poor prognosis has been previously attributed to higher grade tumours in this group of patients[188]. In agreement with these findings, reports showed that over 40% of patients with grade 3 tumours and 29% of patients with grade 2 tumours died as opposed to no mortalities recorded with grade 1 disease[185]. Other reports indicated that HER-2 positive and p53 over-expression were associated with younger breast cancer patients and that HER-2 over-expression correlated with lymph node metastasis[189]. Interestingly, tissue microarray studies in late onset disease have revealed a link between basal type tumours and Fhit, Wwox, AP2α[48]. But, no such effort was directed towards early onset tumours, which are a biologically distinct entity. The biological basis for the variable response to neo-adjuvant chemotherapy (anthracycline based) and surgical treatment in early onset and locally advanced breast tumours remain an unresolved issue. Although pre-treatment, tumour grade and nuclear proliferative activity in the general population were shown to be prognostic in terms of response to anthracycline based chemotherapy[190], inconsistent findings in different studies undermine results[191-193].

The distinct more aggressive behaviour of early onset breast cancer necessitates extensive investigation of the molecular pathways involved in the disease in order to predict and improve the outcome of patients with this disease. Studies should be focused not only on chemo-responsiveness, but also overall responsiveness and outcome to enable effective personalised treatment plans.

1.15 Project aims and investigation outline

This project focuses on the diagnostic and prognostic breast cancer biomarker discovery. This is a three-stage project aiming to contribute to the growing evidence in this field. The areas covered here in are detailed below:

1.15.1 Archival sample proteomic analysis

Firstly, we aimed to test the efficacy of archival sample analysis in proteomic studies, which could provide a potential rich resource for future proteomic discovery and validation studies. The following areas were explored:

1. Assessment of 30 year old urine and serum samples using different chromatographic arrays and SELDI combinations.
2. Strong anion exchange (SAX) fractionation and the optimisation of serum sample preparation using different chromatographic arrays, buffers, matrices and settings.
3. Multi-centre pilot analysis of archival serum using standardised SELDI analysis conditions and settings.

1.15.2 Breast cancer diagnostic serum biomarker discovery

A large blinded prospective study to identify differential serum proteomic profiles between invasive breast cancer, benign breast disease and healthy volunteers was achieved through the following stages:

1. A large cohort of fractionated serum samples (1041) representing the three different groups was used in this investigation in order to achieve acceptable power and statistical confidence. The highly dynamic and complex serum obstacle was tackled by strong anion exchange (SAX) chromato-focusing and fractionation. A further sample of urea/CHAPS denatured serum was analysed in parallel.
2. Assessment of fractions and optimisation of chromatographic, preparation and analysis conditions using the SELDI platform was conducted. This lengthy stage illustrates the most favourable conditions and templates for meaningful MS analysis and potentially successful biomarker discovery.
3. Robotic preparation and standardised SELDI analysis was performed using three different fractions as well as denatured serum samples. Quality control measures were maintained to reduce false positive observations.
4. Two independent bioinformatic analyses were conducted and validated candidate proteomic biomarkers were reported. Further biomarker pattern (multivariate) analysis establishing classification rules between different groups was tested.

5. An independent validation study in a partner centre using a different MS platform (MALDI-TOF MS) was conducted to select for the most stringent markers.
6. Biomarker purification and identification using several methods including resin/liquid chromatography, SDS PAGE, free flow electrophoresis, and tandem mass was completed.

1.15.3 Early onset breast cancer prognostic biomarker discovery

In an attempt to elucidate the biological pathways determining early onset breast cancer treatment response, a multi-centre analysis of sera from a large cohort is performed.

It was anticipated that markers from this effort will enable more personalised treatment and possibly improve patient outcome.

1. Identification of potential early onset breast cancer proteomic prognosticators was carried out using a novel MudPIT analysis platform[11], in collaboration with the Biomedical Research Foundation of the Academy of Athens (BRFAA).
2. This study involved 399 serum samples from the Prospective study of Outcomes in Sporadic versus hereditary breast cancer (POSH) cohort[194], representing closely matched good and poor outcome patients.
3. Qualifying differentially expressed biomarkers were validated by sandwich ELISA assays based on their functional and pathological pathway relevance.

Chapter 2

Materials and Methods

2. Materials and methods

2.1 Sample collections

2.1.1 The Guernsey cohort

The Guernsey samples were originally collected between 1977 and 1991 and involved the collection of serum and urine samples from women aged 35 years or older who lived on the island of Guernsey[195]. All participants gave voluntary written informed consent for the use of blood and urine samples for research purposes. Blood samples were collected using a butterfly needle and, two vacutainers of venous blood (from the non-dominant anti-cubital fossa) were taken; 50 ml for sera and blood clot, and 20 ml in an EDTA tube for plasma and cells. The samples were allowed to stand for one hour at room temperature (RT) to allow the blood clot to form. Samples were then centrifuged at 2000rpm for 20 minutes at RT. The serum fraction was removed by aspiration and aliquoted into 8x2ml labelled Sarstedt tubes and stored at -20°C. Fourteen serum samples from the Guernsey breast cancer cohort were randomly selected for our study[196]. Seven samples were from women who were healthy at the time of sample collection, and seven were from women who were diagnosed with breast cancer. The mean age of the participants with cancer at the time of sample collection was 59.4 years versus 59.6 years for the non-cancer control group. Although small, the two groups were otherwise well matched for menopausal status and had no family history of breast cancer. Each phase of the study was approved by the Guernsey Research Ethics Committee.

2.1.2 The Wessex cohort

Serum samples from a recently collected cohort of breast cancer and healthy volunteers obtained from the Wessex region were analysed. This collection was approved by the Southampton General Hospital NHS Trust Ethics Committee (R&D No. RHM CAN 0392, REC Reference: 05/Q1702/13) and informed consent was obtained from all participants in the study. Serum samples from this cohort were collected between 2005 and 2010 from volunteer females covered by the National Health Service Breast Cancer Screening Programme (aged between 47-73 years), samples were collected from 347 volunteers who were recalled for further investigations following routine three yearly mammographic screening). Serum samples were prepared after 45 minutes coagulation time, centrifugation at 4000rpm for 15 minutes and stored at -80°C within 90 minutes of collection. A maximum of three freeze thaw cycles were allowed for each sample.

2.1.3 The POSH cohort

The POSH cohort is a prospective study recruiting 3,000 women aged 40 years or younger at breast cancer diagnosis between 2001 and 2007. This study obtained ethical approval from the South and West Multi-centre Research Ethics Committee (MREC 00/6/69). Written informed consent was obtained from patients at study entry. Family history and known epidemiological risk data were collected alongside clinical information about diagnosis, treatment and clinical. Follow up data were collected annually after the first year. A blood sample (30 ml) was collected from each recruit. Whole blood (for DNA) and serum aliquots were then stored at -80°C. Sera from patients with none metastatic early (stage 1-2) invasive ductal carcinoma (IDC) patients were analysed. Serum samples (203 good outcome *vs.* 196 poor outcome) were collected under strict criteria; to explore prognosticators key to patient outcomes. Patients who underwent treatment for breast cancer and had a disease free survival (DFS) of at least 4 years following treatment were denoted as the good outcome group, and were compared to other patients who showed less favourable outcomes (uni/bilateral recurrence, metastasis or death) within 2 years of treatment.

2.2 Reagents and solutions

2.2.1 SELDI/MALDI analysis

2.2.1.1 Serum denaturing buffers

U9 buffer (9 M Urea, 2% CHAPS, TrisHCl pH 9)

U1 buffer (1 M urea, 0.2% CHAPS, Tris HCL pH 9)

2.2.1.2 Serum (SAX) fractionation (Bio-Rad Catalogue number: K10-00007)

Buffers and solutions:

ProteinChip U9 buffer (9 M urea, 2% CHAPS, 50 mM Tris-HCl, pH 9).

Rehydration buffer (50 mM Tris-HCl, pH 9).

Wash buffer 1 (50 mM Tris-HCl, 0.1% n-Octyl β-D-glucopyranoside , pH 9).

Wash buffer 2 (50 mM HEPES, 0.1% n-Octyl β-D-glucopyranoside, pH 7).

Wash buffer 3 (100 mM Na acetate, 0.1% n-Octyl β-D-glucopyranoside, pH 5).

Wash buffer 4 (100 mM Na acetate, 0.1% n-Octyl β-D-glucopyranoside, pH 4).

Wash buffer 5 (50 mM Na citrate, 0.1% n-Octyl β-D-glucopyranoside, pH 3).

Wash buffer 6 (33.3% isopropanol, 16.7% acetonitrile, 0.1% trifluoroacetic acid)

2.2.1.3 Protein array wash and binding buffers

The following buffers were used in array preparation for SELDI analysis:

Weak cation exchange (WCX) CM10 arrays

CM10 Wash/binding buffer: (50 mM Ammonium acetate, pH 4- 4.5)

Strong anion exchange (SAX) Q10 arrays

Q10 Wash/Binding buffer (50mM Tris HCl, pH 8-10)

Immobilised metal affinity (IMAC) arrays

IMAC Wash/binding buffer: (0.1 M sodium phosphate, 0.5 M sodium chloride, pH 7).

IMAC charging solution: (0.1 M cupric sulphate).

IMAC neutralising solution: (0.1 M sodium acetate, pH 4).

Reverse phase (RP) H50 arrays

H50 Wash/binding buffer: 10% acetonitrile, 0.1% TFA

2.2.1.4 Energy Absorbing Molecules (EAM)

50% Sinapinic acid (SPA) was prepared by adding 200 μl of 100% acetonitrile (ACN) and 200 μl of 1% triflourouacetic acid (TFA) to 5 mg SPA.

α-Cyano-4-hydroxycinnamic acid (CHCA) was prepared by adding 500 μl 100% ACN and 500 μl 1% TFA to 25 mg CHCA.

2.2.1.5 In-Gel digestion reagents

Wash solution: 50% (v/v) methanol and 5% (v/v) acetic acid.

Ammonium bicarbonate: 100 and 50 mM ammonium bicarbonate solutions were used.

Reducing solution: 10 mM DTT

Alkylating agents: 100 mM iodoacetamide (IAA)

Trypsin solution: Unless stated otherwise, Promega (V511A) Sequencing Grade Modified Trypsin was used for protein digestion. A protease: protein ratio of 1:30 (w/w) was used for protein sequencing. 20 ng/μl trypsin was prepared by adding 1.0 ml of ice cold 50 mM ammonium bicarbonate to 20 μg of sequencing-grade modified trypsin concentration. Trypsin solution was kept on ice until use. The trypsin used was a sequencing-grade enzyme that has been modified to inhibit autolysis and to minimise non-tryptic protease activities.

Extraction buffer: 50% (v/v) acetonitrile and 5% (v/v) formic acid.

2.2.1.6 In-Solution protein digestion solutions

All of the reagents were prepared immediately prior to use. The water used in all components of the procedure was HPLC type water. The urea, Tris, ammonium bicarbonate, and acetic acid were the highest grade available.

Tris HCl stock: This was prepared by dissolving 12.1g of Tris base in 200 ml of water. Adjustment of the pH of the solution to pH 7.8 was achieved with 6 M HCl. Water was added to give a final volume to 250 ml. The final Tris concentration used was 0.4 M. This solution was stored at 4 °C for a maximum of 30 days.

Reducing agent: 30 mg of DTT was dissolved in 750 μl of water and 250 μl of Tris stock was added and vortexed to achieve a final concentrations of 200 mM DTT and 100 mM Tris.

Alkylating reagent: 36 mg of iodoacetamide (IAA) was dissolved in 750 ml of water and 250 μl of the Tris stock was added and mixed to make a final concentration of 200 mM iodoacetamide and 100 mM Tris.

Trypsin solution: 25 *μ*l of ice-cold Tris stock and 75 μl of ice-cold water were added to 20 μg of sequencing-grade modified trypsin (Promega V5111). The final concentration of trypsin was 200 ng/μl and solution was kept on ice until use.

2.2.1.7 ZipTip C18 desalting tips (Millipore, TN072)

The following solutions were used in accordance with the manufacturer's recommendations:

Wetting solution: 50% methanol (MeOH) in (TFA) in HPLC water.

Equilibration and washing solution: 0.1 %TFA in HPLC water.

Sample preparation solution: 2.5% TFA in HPLC water (5X stock solution).

Peptides elution solution: 50% MeOH/ 0.1% formic acid in HPLC water.

Proteins elution solution: 75% MeOH/ 0.1% formic acid in HPLC water.

2.3 Methods

2.3.1 SAX Fractionation and pH gradient sequential elution for SELDI analysis

Sample Preparation

Wessex serum samples were thawed on ice and then spun at 20,000 g for 10 minutes at 4°C. Serum samples (20 µl) were added to each well of a standard V-bottom 96-well microplate. Protein Chip U9 buffer (30 µl) was then added to each well. Microplates were covered with adhesive sealing film and mixed on the MicroMix 5 (set at form 20, amplitude 7) for 20 minutes at 2-8°C.

The filtration plates were tapped several times to make sure that the dry Q ceramic HyperD F sorbent had settled to the bottom of the plate. Rehydration buffer (200 µl) was added to each well and plates were sealed with microplate sealing strips before being manually shaken for five minutes. Filtration plates were then mixed on the MicroMix 5 (form 48, amplitude 7) for 60 minutes at room temperature. Plates were then unsealed, put on a vacuum collar and placed on a vacuum manifold. Vacuum was applied to remove the buffer from the filtration plate and 200 µl of rehydration buffer was added to each well. The rehydration buffer was then removed from the filtration plate using vacuum. This was repeated for a total of three cycles.

Binding Sample With Sorbent

Samples (50 µl) from each well of the sample microplate were added to the corresponding well in the 96-well filtration plate. U1 (50 µl) buffer was then added to each well of the sample microplate and mixed for 5 minutes. Adhesive sealing film was applied on the microplate during the mixing step, removed before applying vacuum, and replaced with a new piece for each mixing to avoid cross-contamination.

Fraction collection

Vacuum was applied to the 96-well filter plate and the flow through collected into the F1-F6 plate. Wash buffer 1-6 corresponding to fractions F1-F6 respectively was added to each well (100 µl) of the filtration plate and mixed for 10 minutes on the MicroMix 5 (form 20, amplitude 7) at room temperature. Vacuum collection of the eluate into the F1-F6 plates containing pH 9- organic acid eluates was completed resulting in 6 fractions based on the *pI* of proteins within the sample.

2.3.2 SELDI array preparation

For consistency, a BioMek3000 (Beckman Coulter) liquid handling robot was used and samples were analysed in duplicate. Unfractionated serum samples were mixed with U9 (1:10 dilution) and incubated on ice for 30 min for denaturation and then immediately stored at -80°C.

The following steps were used in array preparation using the appropriate buffers mentioned previously for each corresponding array:
Arrays were pre-wetted with 3μl HPLC water for 5 minutes at room temperature after which 150μl of wash/binding buffer was added to each spot, and then mixed on the Micromix5 DPC shaker for 10 minutes. The wash/binding buffer was then removed and discarded. This step was repeated for a total of three times.
Note: IMAC arrays were activated with 50 μl charging solution and then washed with 50 μl neutralising solution followed by 200 μl HPLC water prior to this step. Duplicates of 10μl (mixed with 90μl wash/binding buffer to generate 1:10 dilutions) of each serum/fraction sample were then incubated on Micromex 5 DPC (form 20, amplitude 7) for 60 minutes. Samples were then dispensed and arrays washed with 200 μl wash/binding buffer for a total of three times. Arrays were then washed with HPLC water for 3 minutes on the Micromex 5 DPC (form 20, amplitude 7) and water was then dispensed. Arrays were left to air dry for 20 minutes. Two applications of 50% Sinapinic acid matrix (SPA) were added to each spot in 10 minutes intervals and arrays were left to air dry. Arrays were immediately analysed using the SELDI enterprise platform.

2.3.3 SELDI Analysis

Serum samples were analysed using the Enterprise 4011 SELDI analysis platform. Arrays were analysed using the Protein Chip Data Manager (PCDM v3.0.7) Software. The following settings were used: focus mass of 5kDa, matrix attenuation at 2.5kDa with a mass range between 0 and 100kDa for the Guernsey pilot study and 0 and 200kDa for the Wessex serum profiling study. External calibration of the SELDI was performed using a protein standard calibration kit, comprising recombinant hirudin (6.96kDa), equine cytochrome *C* (12.23kDa), equine myoglobin (16.95kDa), and carbonic anhydrase (29.0kDa). Mass accuracy was calculated to be approximately 0.02% of actual mass value. Noise definitions were adjusted to eliminate the chemical noise from the energy absorbing matrix (SPA) in the low mass range, and the area below the detector blinding setting (m/*z* 2,500) was excluded.

Peaks with a signal-to-noise (S/N) ratio of 5 or higher and a valley depth greater than or equal to 3 were considered for clustering. Qualified peaks which were present in at least 10% of the spectra, were used to generate peak clusters. Unlabelled spectra were then labelled at the average mass of the cluster so that a peak intensity value was obtained for each spectrum. The mass window for each cluster was set at 0.3% of the peak mass for spectra optimised for low mass (0-30kDa), and at 2% of the peak mass for spectra optimised for high mass (30-100kDa) and (30-200kDa) for the Guernsey (G4) pilot and the main Wessex studies respectively. To ensure that no spurious peaks were ultimately used as candidate biomarkers, peaks were evaluated and relabelled as required prior to the final statistical analysis. Qualified mass peaks (S/N >5) within m/*z* range of 2.5-100kDa or 2.5-200kDa were then auto detected. Peak clusters were completed using a second pass peak selection (S/N >2, within 0.3% mass window) and estimated peaks were then added. The coefficient of variation for the QC spectra was calculated to be between 20-30%.

2.3.4 SELDI data processing, peak detection and statistical analysis

All spectra had baseline subtracted and were then normalised to the total ion current within the m/*z* range of 3-100 and 3-200kDa for the Guernsey pilot and Wessex cohort analysis respectively. Normalisation factors were then examined to identify outlier spectra indicating possible variation due to sample processing. Spectra that had normalisation factors greater than two-fold higher than the average were excluded from the analysis. A cluster graph was then generated using the Protein Chip Data Manager (PCDM) v3.0.7 Software, where each peak intensity was plotted against m/*z* for the different sample groups (cancer, benign disease and healthy control). Comparisons between groups were made using the Mann-Whitney U test. The Benjamini and Hochberg formula was adopted to adapt for false discovery rate (FDR) associated with multiple testing following which a *P*-value of <0.05 for a given peak indicated the significant differential expression of peaks. The coefficient of variation (CVs) of the pooled reference (QC) samples were calculated using the formula:
$CV = \sqrt{}\ ((CV_1^2+CV_2^2+CV_n^2)/n)$ where n represents the number of peaks within the spectra.

2.3.5 SELDI Instrument performance evaluation

QC and performance checks for the SELDI analysis included calibration and alignment of the Biomek 2000 robotic platform, which were performed monthly. Mass accuracy, resolution and sensitivity of the spectrometer were evaluated bimonthly using insulin and bovine IgG standard chips (Bio-Rad). A normal phase array, NP20 was run weekly, loaded with the All-in-1 Protein standard II (Bio-Rad) for external calibration of the spectra. To minimise slight systematic shifts in the time-of-flight data from one spot to another the PCDM software was used to calculate a spot-to-spot correction factor. The correction factor was calculated from 4-6 spectra of All-in-I Peptide Standard (Bio-Rad) on NP20 Protein Chip.

2.3.6 Principal Component Analysis (PCA) of Spectra

For an informative statistical analysis of the acquired proteopeptidomic profiles, PCA was used, as is commonly employed in microarray profiling cluster analysis. PCA is designed to capture variances in the given data sets in terms of their principal components, meaning a set of variables which defines a projection encapsulating the maximum amount of variation in a data set. This is a quick unsupervised statistical tool to identify separation between different groups, which is a good indicator for differential expression.

2.3.7 Receiver Operating Characteristic (ROC) Analysis

A statistical evaluation of intensity changes associated with all protein peaks generated by MS was performed to validate candidate protein peaks differential expression, thus testing their potential as cancer biomarkers. ROC analysis was used in test situations where diagnostic tests reveals numerical results that can be compared to an independent diagnosis, confirming either the presence or absence of disease [197]. The Protein Chip Data Manager (PCDM) software applies ROC analysis to establish two parameters: sensitivity and specificity. The PCDM software calculates the area under the ROC curve (AUC) which is a combined measure of sensitivity and specificity and numerically describes the performance of a particular analysis/biomarker(s). The area under the curve (AUC) indicates the overall performance of a given test and can be interpreted as the average value of sensitivity for all possible values of specificity and can take any value between 0 and 1 since both x and y axes can have values between these ranges. The closer AUC is to 1, the better the overall performance of the test or biomarker. A test with an AUC of 0.9 and higher is considered highly accurate, while a moderately accurate test will have an AUC value between 0.7-0.89. An AUC value that is lower than 0.7 is considered inaccurate [197]. This is a useful tool for the prioritisation of candidate markers for further evaluation and validation.

2.3.8 SDS PAGE gel preparation

Large format 25x20.5cm (W x L), 1.0mm thick SDS PAGE gels were prepared using the proportions in Table 1, addition of freshly prepared ammonium persulphate (APS) was leaved until last to prevent premature gel setting. Assembly of the equipment, casting and running of the gels was carried out according to the Bio-Rad Protean^{R} II xi Cell instructions. Briefly, the resolving gel mixture was prepared and then pipetted between clamped plates, pre-cleaned with dH_2O and 70% methanol, set in a casting system. The gels were then overlaid with water to ensure a straight edge. The gels were left to set for ninety minutes before being transferred to the fridge overnight to ensure complete polymerisation. Stacking gels were made according to the proportions detailed in Table 1, pipetted in between the plates on top of the resolving gel, and a comb inserted and then left to polymerise for 30-60 minutes prior to loading. 30 μl of each sample was combined with 50 μl of sample buffer (see Table 2), before being heated to 95°C for 1 minute to further denature the proteins. The gels were then attached to the gel tank cooling core and running buffer (see Table 3) was added to the top and bottom chambers before loading the samples into the wells. Samples were run in between two wells containing 20 μl of protein marker. A current of 100 mA was then applied for approximately 6 hours to move the proteins down the gels. This was done with cooling, provided by running water through the cooling core, to minimise uneven heat across the gels. Gels were then rinsed three times, for five minutes each, in distilled water. Following this, Biosafe Coomassie (Bio-Rad) or silver staining (Invitrogen) was used to stain the gels prior to destaining following the manufacturer protocols. Protein band molecular weights were determined by scanning the gels using a GS-800 densitometer in conjunction with Quantity OneR software (Bio-Rad).

2.3.9 Immunoblotting

Protein concentrations were measured using the Bradford protein assay (Bio-Rad). Briefly, samples were loaded on 12% SDS-PAGE gels, transferred to PVDF membranes and blocked with 5% milk in TBST buffer (20 mM Tris-HCl, 120 mMNaCl, 0.1% Tween) for 1 hour. Membranes were then incubated at 4 °C with primary (OPN, ANX A3) antibodies overnight. After washing, the blots were incubated with horseradish peroxidase conjugated anti-rabbit IgG and anti-mouse IgG secondary antibodies (GE Healthcare) for two hours. Immunoblot signals were quantified using Quantity One image analysis software (Bio-Rad).

2.3.10 Serum protein quantification

Total protein quantitation was carried out using a Qubit™ fluoremeter and the Quant-ITTM Protein assay kit (Invitrogen) based on the manufacturer instructions. Serum was diluted 1/100 and 5μl quantitated. Quant-iT reagent and buffer were combined at a ratio of 1:200 and 190μL for each of the three standards and 195μl for each sample was then aliquoted into QubitTM assay tubes. The standards and samples were then added to the corresponding tubes, vortexed for 2-3 seconds, and left to incubate at room temperature for 15 minutes. The tubes were then inserted into the QubitTM uorometer, calibrating with the three standards first and the QF number for each serum sample noted. This number was then converted into actual serum protein concentration using the equations below.

[Sample] = (QF value x 200/x) _ dilution factor
Where:
QF value = value given by QubitTM fluorometer in μg/ml
x = μL sample added to assay tube
So:
[Serum μg/ml] = (QF value x 200/5) x 100

Chemical	**10% resolving gel**	**4% stacking gel**
	Values in ml	**Values in ml**
dH2O	49	6.657
1.5 M Tris-HCl, pH 8.8	25	-
0.5 M Tris-HCl, pH 6.8	-	2.25
20% SDS	0.5	0.045
40%Acrylamide/Bis-acrylamide (37:1)	25	1
TEMED	0.05	0.01
10% APS	0.5	0.038

Table 1. PAGE composition - Adequate for two gels

Chemical	**Final Concentration**	**To make 10 ml Volume in ml**
Glycerol	25%	2.5
20% SDS	20%	1
1 M Tris-HCl, pH 6.8	62.5mM	0.625
0.02%Bromophenol Blue	0.01%	5
dH2O	-	0.875
Made up as above, DTT added as required		
Dithiothrietol (DTT)	350 mM	54 mg/ml

Table 2. PAGE sample buffer

Chemical	**Final concentration**	**To make 1L**
Tris	25 mM	3 g
Glycine	192 mM	14.4 g
20% SDS	0.1%	5ml
dH2O	-	to 1L

Table 3. SDS PAGE running buffer pH 8.3

2.3.11 CM10 carboxymethyl (WCX) spin columns

CM spin columns were lightly tapped to settle the sorbent, and then centrifuged at ~80 x g (1,000 rpm) for 30 seconds to remove the buffer. CM10 wash/binding buffer and HPLC water were used for the preparation and washing of the columns in an identical protocol to the optimised CM10 array preparation protocols. Samples were then added and vortexed for 60 minutes. Samples were eluted twice using serial ionic strength CM10 buffer pH 4.5 (strength between 200mM - 1M). To ensure that the CM Ceramic HyperD F sorbents possessed similar protein binding properties as the WCX protein chip surface, WCX SELDI mass were compared with raw spectra from raw serum, the CM10 serum/fractions and the proteins isolated using the CM Ceramic HyperD F sorbents from the low molecular weight serum fraction.

2.3.12 Array preparation and protein retentate harvesting

To identify proteins directly from the WCX arrays, 10 μl aliquots of samples were applied to two Protein Chip arrays (i.e., 16 spots) and prepared as described in the SELDI experimental section. To harvest the retained peptides/proteins, each array spot surface was extracted five times using 10 μl of 0.5% formic acid, 50% acetonitrile (ACN) at room temperature. The five extractions were combined and divided into two aliquots. The first aliquot was digested in solution as described below. The second aliquot was vacuum-dried and resuspended in 10 μl of Laemmli or Tris Tricine sample buffer, and resolved by SDS-PAGE (4-15% -Bio-Rad-) or Tris Tricine (10-20% -Bio-Rad) gel electrophoresis respectively.

2.3.13 Serum pre-fractionation by free-flow-electrophoresis (FFE)

Pre-fractionation of serum samples by isoelectric focusing was performed on a BD™ FFE Kit IEF 3-10 free flow- electrophoresis (FFE) instrument (Becton, Dickinson and Company) according to separation principles described[198, 199]. Before application, serum samples were equilibrated with an equal volume of FFE separation medium containing traces of 2-(4-sulfophenylazo)-1,8- dihydroxy-3,6-naphthalenedisulfonic acid (SPADNS) (Sigma- Aldrich, Taufkirchen, Germany) to ease optical monitoring of sample migration within the separation chamber. The FFE-fractionation chamber was coated with FFE separation medium (19.6% w/w Prolytes™ pH 3-10 (BD Bioscience, Munich, Germany), 40.3% w/w urea, 3.8% w/w mannit) containing 0.09% w/w hydroxyl propyl methyl cellulose (HPMC). Subsequently, free flow isoelectric focusing (FF-IEF) was performed at 10°C in a separation chamber positioned horizontally (0.4 mm spacer) with a working voltage of 500 V (18 mA) and a total flow rate of 60 ml/h. Following media were used: Counter flow medium (Inlet 8; 40.3% w/w urea, 3.8% w/w mannit); anodic stabilization medium (Inlet 1-2; 94.2% w/w counter flow medium, 100 mM

H2SO4, 150 mM aminobutter acid, 100 mM nicotinamin, 15 mM diglycine); separation medium (Inlet 3-5), cathodic stabilization medium (Inlet 6-7; 90.3% w/w counter flow medium, 150 mM NaOH, 75 mM EtOH-Amine, 75 mM AMPSO, 150 mM TAPS, 30 mM HEPES); anodic circuit electrolyte (100 mM H2SO4); cathodic circuit electrolyte (100 mM NaOH). Serum samples were applied into the separation chamber via the middle sample inlet at a flow rate of 0.8 ml/h. Fractions were collected in 96 well plates (fraction numbers 1 (anode) to 96 (cathode)) and immediately stored at - 20°C until use. For monitoring separation, aliquots of selected fractions were analysed by MS using NP20 arrays and SELDI.

2.3.14 In-Gel tryptic digestion

Candidate silver stained bands representing the marker of interest, blank gel pieces from a spot-free region, and reference spots (known marker proteins from the gel ladder) were excised and placed in a 96-well plate. In-gel digestion was performed automatically using a ProGest robotic system (DIGILAB) following standard digestion protocols[200]. In brief, gel plugs were destained by alternating between a 25 mM ammonium bicarbonate (ABC) (Fisher) and 100% acetonitrile (ACN) solutions. Plugs were then incubated with 10 mM DTT (Bio-Rad, Hercules, CA) in 25 mM ABC for 30 minutes; 56°C for 10 minutes and then allowed to cool to room temperature for the last 20 minutes. The reducing solution was then replaced with 100 mM iodoacetamide (IAA) (Fluka, Buchs, Switzerland) in 25 mM ABC and incubated for 45 minutes. The gel plugs were then washed twice with alternating solutions of 25 mM ABC and 100% ACN. Gel plugs were dried by two additions of 100% ACN. 10 μl of modified sequencing-grade trypsin at a concentration of 10 mg/ml was then added. The gel plugs were incubated in the trypsin solution for 10 minutes and then 15 μl of 25 mM ABC were added to each well. The gel plugs were digested for 4 hours at 37°C and the reaction was stopped by adding 7 μl of 3% formic acid to each well. The supernatant was removed and the extracted peptides were reconstituted in 25 μl 0.1% TFA for tandem mass analysis.

2.3.15 In-Solution Protein Digestion

Protein samples were evaporated and resuspended in 6 M urea, 100 mM Tris buffer at 10 mg/ml. 5 μl of the reducing reagent (DTT) was added and mixed with the sample by gentle vortexing for 1 hour at room temperature. Alkylating reagent (IAA) (20 μl) was then added and mixed with the sample by gentle vortex for 1 hour. DTT (20 μl) was then added any excess iodoacetamide. Samples were mixed gently and allowed to react at room temperature for 1 hour. The urea concentration was reduced by diluting the reaction mixture with 775 μl of water. This dilution reduces the urea concentration to ~ 0.6 M, a concentration at which the trypsin retains its activity.

Trypsin solution (20 μg/μl) was added and mixed, then digestion was allowed overnight at 37 °C. This amount of trypsin gives a protease-to-substrate rati of 1:50. The reaction was then stopped by adjusting the pH of the solution to <4 by adding concentrated acetic acid as needed. The digest was then either analysed directly or concentrated by evaporation.

2.3.16 Desalting of tryptic peptide mixtures

ZipTip C18 (Millipore) was used for purifying and concentrating femtomole to picomole amounts of protein and peptide samples prior to MS analysis. The protocol followed was based on the manufacturer's instructions to facilitate protein or peptide binding, salt and detergent removal, and sample elution for direct MS/MS analysis. Tips were pre-wetted repeatedly with 10 μl wetting solution and ZipTips were washed three times with the equilibration solution. Samples were then aspirated and dispensed seven times. Tips were washed and dispensed to waste using seven cycles of wash solution. Samples were then eluted three times using 3 μl of elution solution into a clean vial using a standard pipette tip. Eluates were dried using a Speed-Vac centrifuge at 35 °C. If necessary, samples were stored at -20 °C (< 24 hr) or -80 °C (for longer periods). Prior to MS/MS analysis desalted peptide lyophlates were solubilised with 1 μl matrix solution (10 μg/μl α-cyano-4-hydroxy-cinnamic acid in 0.1% trifluoroacetic acid: acetonitrile, 1:1, v/v). Half a microliter of sample mixture was loaded on a MALDI target plate (384 spot Bruker UltraflexIII) and allowed to dry in the dark at room temperature. The samples were then analysed using a MALDI TOF/TOF mass spectrometry. Alternatively, samples were analysed using ESI LC MS/MS.

2.3.17 MS/MS analysis by Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI-TOF/TOF)

For peptide profiling by MALDI TOF/ TOF MS we utilized the Ultraflex III platform (Bruker Daltonics, Bremen, Germany) and a 384 format plates. Samples freshly desalted and eluted from ZipTip beads were diluted 1:1 with TFA (0.1% (v/v)). 0.5μl sample and 0.5μl α-cyano-4-hydroxycinnamic acid (CHCA) were mixed on each spot. Sample/matrix solution was air dried in the dark. Mass spectra were acquired by an Ultraflex MALDI-TOF/ TOF mass spectrometer (Bruker Daltonics) equipped with a smartbeam™ solid-state laser in a reflectron positive-ion mode. In MS-mode 400 single-shot spectra were accumulated. An external calibration of the MS-spectra was performed by acquiring calibrant spectra from the calibrant positions on the target. The MS-spectra were automatically processed by a baseline subtraction and a chemical noise filtering in the flexAnalysis (Bruker Daltonics) software. Only compounds with S/N>7 were selected for automatic MS/MS measurements. MALDI MS/MS spectra were recorded on the selected compounds. For each MS/MS spectrum, 300 laser shots were recorded for the parent signal and 800 laser shots were recorded for the fragments. The MS/MS spectra were processed by smoothing and baseline correction using the flexAnalysis software. After acquisition of the MS/MS spectra a Mascot database search was triggered via the BioTools software (Bruker Daltonics). The database used for the search was the MS/MS DB database. The search parameters used were: taxonomy human, no enzyme specificity, no fixed or variable modifications, MS-tolerance 50 ppm, MS/MS tolerance 0.7 Da, and instrument type MALDI-TOF-TOF. Samples were measured from a minimum of four independent MALDI spots in the mass range of 200-1,800 Da.

2.3.18 Liquid Chromatography Electrospray Ionisation tandem mass (LC-ESI-MS/MS) analysis

To identify candidate markers from the Wessex profiling analysis, peptides were analysed by LC-MS/MS using a Surveyor LC system and LCQ Deca XP Plus (ThermoScientific). Briefly, peptides were resolved by reverse phase chromatography (Biobasic column, ThermoScientific; 180uM x 15mm) over a 30 minute ACN gradient at a flow rate of 2 µl/minute. Peptides were ionised by electrospray ionisation and MS/MS was acquired on ions dependant on their charge state and intensity. Mass accuracy and sensitivity of the MS was confirmed with the direct infusion of glufibrinopeptide (2.5 pmoles/µl) and LC/MS/MS performance was assessed with a digest of BSA. The tryptic peptide analysis for accuracy, sensitivity, retention time reproducibility, and protein sequence coverage were all within the specified ranges. BSA quality control checks were performed prior to the analysis of the sample and post-acquisition.

2.3.19 LC MS/MS Data processing

The data files (.raw) were converted into mascot generic files using the MassMatrix File Conversion Tool (Version 2.0; http://www.massmatrix.net) for input into the Mascot searching algorithm (Matrix Science). The data files were searched against SwissProt (v. 2010_06) NCBInr (v. 20080527) using the following search criteria: tryptic peptides with up to one missed cleavage and carbamidomethylation of cysteines and oxidation of methionines.

2.3.20 Biomarker immunoprecipitation (IP)-MS

Serum fractions containing annexin A3 (ANX A3) and apolipoprotein C1 (ApoC1) were immuno-depleted using the µMACS protein A/G microbeads kit (130-042-601, Miltenye Biotec) as per manufacturer instructions. In brief, either 2 µl monoclonal ANX A3 antibody (sc-134260, Santa Cruz) or 3 µl monoclonal ApoC1 antibody (ab5400, Abcam) was added to 100 µl of protein A/G and then mixed with 200 µl of sample before incubating on ice for 30 minutes. The mix was loaded onto a magnet mounted μMACS column followed by 4 washes with PBS-T. Purified protein was eluted by ACN/TFA/Isopropranol (16.7%/0.1%/33.3%-50%).

2.3.21 3D MudPIT analysis of the POSH cohort

2.3.21.1 Analysis outline

A novel quantitative proteomic and phosphoproteomic analysis was conducted based on previous breast cancer tissue and benign prostate hyperplasia serum studies[11, 201]. Briefly, triplicate analysis of all samples was performed using high-performance size-exclusion chromatography (SEC) for the pre-fractionation of serum proteins followed by dialysis exchange and solution phase trypsin proteolysis. The tryptic peptides were then fractionated offline with hydrophilic interaction chromatography (HILIC). Each fraction was then analysed online with reverse phase nano ultra-performance chromatography to nanoelectrospray ionization - tandem mass spectrometry (RP nUPLC - nESI MS/MS) using an ion trap mass analyser. For the spectral processing, SpectrumMill and InsPecT software programs and a novel label free software tool, for the spectral counting with spectral intensity determination, were used for the qualitative and quantitative assessment of the proteins and phosphoproteins. The DAVID gene ontology tool and the Ingenuity Pathway Analysis program were used for the functional annotation of the quantitative proteome and phosphoproteome and associated biochemical and molecular biology pathways and networks.

2.3.21.2 Size Exclusion Chromatography

Six aliquots of 70 μl from each pooled serum specimen were subjected to SEC protein pre-fractionation using two serially connected SEC HPLC columns (Shodex Protein KW 804, 8 mm ID × 300 mm L) on the Dionex P680 HPLC Pump with PDA-100 photodiode Array Detector. A Shodex SEC guard column was also retrofitted to this configuration for column care and contamination protection. The eluent used for the separation of serum proteins was composed of 10 % methanol (Sigma, 99.9 %), 50 mM KH_2PO_4 (Fluka, 99.5 %), 10 mM Tris-HCl (Biorad), 50 mM Ammonium Acetate (Fluka, 99 %), 0.3 M NaCl (Carlo erba, 99.9 %) and 6 M Guanidine HCl (Gibco BRL) at pH 5.3. The injection volume was 500 μl and each separation was carried out applying an isocratic method at flow rate 0.2 μl/minute at 15 °C. The fractions for all separations were collected based on the retention time of four distinct time segments into 50 μl tubes. These segments were pooled together from the six experiments to achieve a serial enrichment for each segment. The collected fractions were subjected to dialysis exchange using the Slide-A-Lyzer G2 Dialysis Cassettes 2 KDa MWCO (Thermo Scientific) with three volumes of 3 L ammonium bicarbonate buffer solution freshly changed overnight. The purified fractions were concentrated to 400 µl using a speedvac concentrator (Eppendorf concentrator 5301).

2.3.21.3 In solution trypsin digestion of serum proteins

For each SEC segment, six aliquots of 20 µl containing 100 µg of protein, measured by Bradford assay, were prepared for parallel trypsin digestion. The proteins were reduced by the addition of 2 µl tris-2-carboxymethyl phosphine (TCEP) 50 mM, followed by incubation for 1 hour at 60 °C in heating block (WEALTEC, HB-2). The cysteins were blocked by the addition of 1 µl of 200 mM methyl methanethiosulfonate (MMTS) in isopropanol followed by incubation at room temperature for 10 minutes. Trypsin digestion was performed by the addition of 6 µl of freshly prepared trypsin solution (Roche); (500 ng/µL) and 14 µl ultra-pure water into the protein sample which was then incubated for 12 hours at 37 °C in a heating bath. The six digested aliquots were then pooled and lyophilised under vacuum.

2.3.21.4 HILIC peptide fractionation

The tryptic peptides were fractionated with HILIC using a poroshell column configuration (Phenomenex Kinetex 150 × 2.1 mm, 2.6 µm particle). Mobile phase (A) was 100% ACN, 0.1 % formic acid, 15 mM ammonium formate and mobile phase (B) was 100% H_2O, 0.1 % formic acid, 15 mM ammonium formate, pH 3.3. The peptide samples were dissolved in 20 µl of 50% mobile phase (B) and were centrifuged at 15,000 rpm for 10 minutes. The supernatant was injected into a 20 µl sample loop and the separation method was as follows: 10 minutes isocratic 5% (B), for 90 minutes of gradient up to 30% (B), 20 minutes gradient up to 50% (B), 15 minutes gradient up to 95% (B), 60 min isocratic 95% (B) and finally 10 minutes gradient up to 5 % (B). The flow rate was 0.065 ml/min, column temperature was set at 40 °C and signal UV response was monitored at 280 and 254 nm. The fractions were collected in a time dependent manner every 3 minutes, they were dried for 2-3 hours with a speedvac concentrator and stored at -20 °C until the LC-MS/MS analysis.

2.3.21.5 LC-MS-MS analysis

All LC-MS/MS experiments were performed on an iontrap system (Agilent Technologies, 6330 XCT) retrofitted to a 1200 nano-HPLC system equipped with a micro well plate autosampler (Agilent Technologies, Karlsruhe, Germany). Individual lyophilised HILIC fractions were freshly reconstituted in 30 µl mobile phase A (3% ACN, 0.1% Formic acid). A 3 µl volume of the resulting sample solution was injected and then eluted at 150 nl/min onto a 0.075 × 150 mm reverse phase capillary column (Zorbax C18, 300 Å pore, 1.8 µm particle, Agilent Technologies, Karlsruhe, Germany) retrofitted onto the nanoelectrospray source and connected to a 1P-4P coated, 8 µm tip × 360 μm OD × 75 μm ID PicoTip nanoelectrospray emitter (New Objective, Dingoes, NJ). An 0.30 × 5 mm reverse phase guard column (Zorbax C18, 300 Å pore, 5.0 μm particle, Agilent Technologies, Karlsruhe, Germany) was connected between the pump outlet tubing and the capillary column. The 280 fractions collected from all HILIC separations were analysed in triplicate. The nano-HPLC separation started with an isocratic composition of 10% mobile phase B composed of 97% ACN, 0.1% Formic acid for 30 minutes, followed by a gradient up to 70% (B) for 50 minutes, 10 minutes isocratic elution at 70% (B) then 20 minutes re-equilibration. The iontrap analysis parameters were as follows: selection of the 4 most abundant multiply charged precursors above an absolute intensity threshold of 10^4 during the MS survey scan within a mass range of 300-2200 amu with 150 ms accumulation time and Enhanced Resolution. MS^2 experiments were conducted with Ultra-Scan Resolution with on flight adjustment of collision energy using the smart fragmentation operation. Precursors were excluded after the acquisition of 2 spectra and released again after 0.2 minute. The MS-MS mass range window was 100-2200 amu.

2.3.21.6 Data processing and Spectral counting

All RP-nUPLC-nESI-MS^2 raw data files were initially subjected to data extraction/peak picking/deisotoping and searching using the SpectrumMill software program (Agilent Technologies, Karlsruhe, Germany). Specifically, the data extraction and signal processing parameters were tailored for the quadrupole ion trap mass analyser that included a signal-to-noise threshold for the precursor MS peak of >25. The SpectrumMill results were further processed with the Scaffold software tool (version 3.0, Proteome Software, Portland, OR) in order to extract all spectra in .mgf format. For the final tryptic peptide/protein result, the Scaffold processed .mgf data files were processed by the InsPecT software tool (of the Center for Computational Mass Spectrometry at http://proteomics.ucsd.edu/LiveSearch/) that included the contaminant list option and M oxidation, STY phosphorylation, C methylthiolation and N deamidation as the permitted modifications. The specific data processing parameters chosen were as follows: instrument, ESI-ION-TRAP; protease, trypsin; parentmass tolerance, 2.5 Da; ion tolerance, 0.7 Da; maximum number of PTM allowed per peptide, 2; PTM, delta mass: +15.999400 Da, affected: MW, type: opt; PTM, delta mass: +79.979 900 Da, affected: STY, type: opt; PTM, delta mass: +46 Da, affected: C, type: opt; PTM, delta mass: +1 Da, affected: N, type: opt; sequence, Human SwissProt, common contaminants; results were filtered at a spectrum-level p-value of 0.05, measured by hits to the decoy database.

The above processing scheme yielded two peptide/protein lists for the good and the poor outcome serum samples. For the semi-quantitative comparison of these two lists, the spectral counting approach was applied. For each protein identified in both samples, a ratio (poor outcome/good outcome) of the two total intensity sums of all matching peptide spectra for a given protein, as reported by Inspect output, was calculated using a devised algorithm of R programming language (BRFAA). For the discrimination of the differentially expressed proteins various distributions where fitted using the 'lmom' R-package for the Log^2 of the calculated ratios (R package, version 1.5. URL: http://CRAN.R-project.org/package=lmom), which utilises the 'L Moments method' for parameter estimation. All the distributions were tested for 'goodness of fit' using a the two-sided Kolmogorov-Smirnov test (available in R) with a 95% confidence interval. The distribution presenting the best 'goodness of fit' was the 'Generalised Logistic' distribution. Having the distribution parameters the *P*-value of each log^2 ratio was calculated.

2.3.22 Biomarker ELISA validation

2.3.22.1 ANX A2 ELISA validation

Validation of the serum ANX A2 as a prognosticator in early onset breast cancer was validated using an independent validation set. A commercially available ELISA kit (E91944Hu) was used as per manufacturer’s instructions. All samples and standards were run in duplicate with absorbance measured on the ThermoMax plate reader and data analysed with GraphPad Prism4.

2.3.22.2 ANX A3 ELISA validation

ANX A3 measurements in serum fractions were performed using a commercially available ANX A3 sandwich ELISA kit according to the manufacturer’s protocol (E94786Hu, Uscn Life Science). All samples and standards were run in duplicate with absorbance measured on the ThermoMax plate reader and data analysed with GraphPad Prism4.

2.3.22.3 Apo C1 ELISA validation

Measurements of Apo C1 levels in an independent validation set from the POSH cohort were performed using a commercially available ApoC1 sandwich ELISA kit according to the manufacturer’s protocols (E90252Hu, Uscn Life Science). All samples and standards were run in duplicate with absorbance measured on the ThermoMax plate reader and data analysed with GraphPad Prism4.

2.3.22.4 NOS2 ELISA validation

Validation of serum NOS2 level as a prognosticator in early onset breast cancer was validated using an independent validation set. A commercially available ELISA kit (E90237Hu) was used as per manufacturer’s recommendations. All samples and standards were run in duplicates with absorbance measured on the ThermoMax plate reader and data analysed with GraphPad Prism4.

2.3.22.5 DNMT1 ELISA validation

Validation of serum DNMT1 as a prognostic biomarker in early onset breast cancer was validated using an independent validation set from the POSH cohort. A commercially available ELISA kit (E98244Hu) was used as per manufacturer's recommendations. All samples and standards were run in duplicates with absorbance measured on the ThermoMax plate reader and data analysed with GraphPad Prism4.

Chapter 3

Results

3. Results

3.1 Proteomic profiling of archival samples: The Guernsey cohort

3.1.1 Background

The application of MS based proteomic biomarker discovery is a challenging area. Although current available technologies aid high through put experiments, the biological validity and reproducibility of such analysis is highly dependent on a meticulous lengthy optimisation process. This is paramount to minimise the risk of false discovery and its associated loss of time and resources.

SELDI proteomic analysis is a complex method in which several variables and experimental conditions interact to create a meaningful proteomic profile[86]. Adopting affinity chromatography based arrays improved MS resolution, other technical and biological factors can also affect the interpretation of its profiles[86,169]. Thus, it is crucial to refine the experimental settings for each project in order to ensure biologically sound findings. Optimised and reproducible analysis conditions are also essential for further identification and validation steps following the initial discovery phase[170,171]. Fine tuning the different laser and detection settings and pre-analytical variability has shown improved consistency indicating reliable biomarker discovery[172,173]. Since there is no consensus on specific profiling conditions and/or settings, and given the differences introduced by the different biological characteristics of each cohort analysed, project tailored optimisation remains the backbone of any MS based study. Here, experiments were conducted to identify optimal analysis conditions for serum samples using the SELDI 4011 Enterprise platform. Standardised sample handling was adhered to in all experiments to minimise pre-analytical variation. Sample type, serum fractionation, laser energy, different binding affinity profiles resulting from different arrays (Q10, CM10, H50 and IMAC Cu+2), the effect of sample reduction by DTT, day of analysis and other conditions were all investigated to reach the optimal environment for differential profiling analysis. In this section, the optimisation steps followed and findings in the Guernsey serum proteomic analysis are described in details.

The presence of thousands of different tissue samples in tumour banks can be an attractive venue for differential proteomics analysis. Not only do these samples save time required for collection, they also carry a wealth of knowledge including a long follow up profile and possible pre pathology prediction biomarkers. Tests were performed to explore the efficacy and reproducibility of such studies in the world of proteomics. Given the different pre-analytical variability issues and the potential *ex vivo* proteolytic effect associated with prolonged and or sub optimal storage[57, 108, 176], this area remained controversial in the world of unbiased proteomics. Although the pre analytical variability within old bio-bank collections may be a limiting factor for such analysis, this cannot be altered. However, limited efforts have been made that reject the hypothesis of archival samples yielding reproducible proteomic markers. Addressing this question is key, as the amount of information that can be extracted from such archival collections is vast, and to miss the opportunity of analysing them is unsatisfactory. Thus, this study aimed to assess the usability of archival urine and serum samples from the Guernsey cohort using MS analysis.

3.1.2 Materials & Methods

A detailed description of the reagents and methods involved in SELDI analysis were described in the methods chapter earlier. Briefly, pooled reference samples (representing all groups in each study) were analysed using SELDI. The buffers and/or concentrations, chromatographic surfaces, energy absorbing molecules, laser settings and other variables were altered to generate a gradient of conditions to act as a road map to characterise the best conditions for current and potential future SELDI analysis (Figure1). This scheme was adapted in all projects. The proteomic profiling of archival samples (urine and serum) was performed to assess the possibility of generating meaningful data from old bio-bank collections.

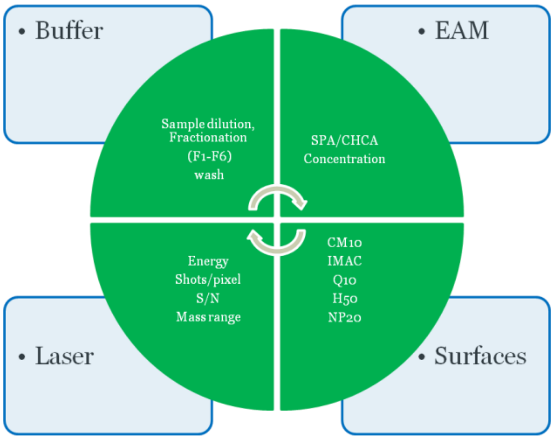


Figure 1. MS condition optimisation workflow

A diagram showing the interactive stages for optimising MS profiling. This is a critical phase prior to pilot and generic biomarker discovery studies. Several steps are involved starting with sample preparation / enrichment / dilution assessment. Energy absorbing molecule (EAM) choice and concentration is also an important variable along with the characterisation of surfaces of best/distinct profiling yields. Finally, laser and analytical settings optimisation per study is the last step in this complex approach. Regular quality control checks during the analysis are essential, to ensure biologically meaningful discovery.

3.1.3 Results

3.1.3.1 Archival urine proteomic profiling of the Guernsey cohort

In a pilot study, 20 urine samples from the Guernsey cohort were analysed. Firstly, WCX protein arrays were used to analyse the samples as described earlier. Assessment of pooled urine samples using SELDI profiling revealed scarce protein ion peaks and a noisy background. Despite efforts to improve peak yields by altering the laser settings, only minimal improvement was achieved (Figure 2). Shifting the energy absorbing molecule between different acids (SPA and CHCA) in an attempt to reduce the LMW region noise (the urinary proteome is predicted to fall within this region) did not improve the resulting profiles. Overall, a generally less noisy condition was shown to be generated at a higher laser energy setting using sinapinic acid (SPA) (Figure 2.c).

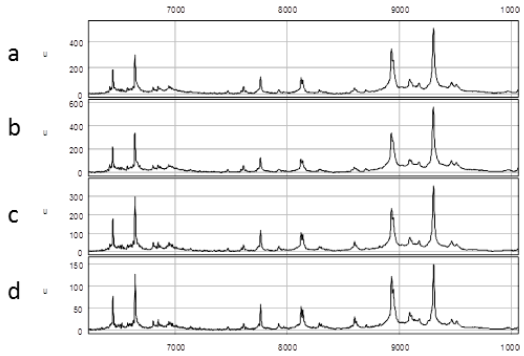


Figure 2. Archival urine sample analysis using WCX protein arrays
For this and subsequent MS figures; the X and Y axes represent the m/*z* (corresponding to protein mass) and peak intensities in μA respectively.

80μl urine was mixed with 30 μl U9 and either 50% SPA (a,c) or 100% CHCA(b,d) were used as matrices. Profiling was conducted on (WCX) CM10 arrays using the following settings:
a,b. LE 1,500nj, shots 530, FM 1.5kDa, MR 0-50kDa, MA 1kDa.
c,d. LE 2,500nj, shots 530, FM 1.5kDa, MR 0-50kDa, MA 1kDa.

Overall, the spectra from the archival urine samples appeared to be very noisy, and although more peaks were detected with higher LE using SPA (c), only very few peaks (15) were detected compromising the statistical power of any further analysis.

The effect of binding affinity on profile improvement was also tested using reverse phase (H50) and copper charged immobilised metal affinity conditions (Figures 3-5). Similarly, a higher laser energy using SPA matrix improved the S/N of the very few peaks detected (Figure 3). However, it was interesting to note the lack of differential profiling using the two chemically distinct chromatographies (Figure 4).

Furthermore, a third copper charged immobilised metal affinity chromatographic array (IMAC Cu^{+2}) was tested in an attempt to elicit more protein peaks and signals. This analysis also revealed very few proteomic peaks, although using SPA and higher laser energies between 2,000-2,500 nj slightly improved the profiling (Figure 5).

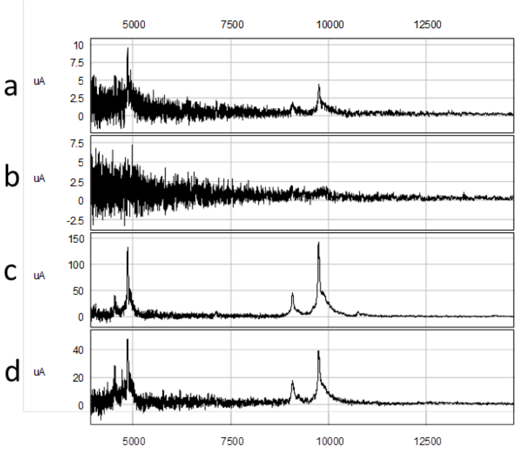


Figure 3. Archival urine proteomic profiling using H50 arrays
MS spectra from the G3 Urine test (80μl urine was mixed with 30μl U9), 50% SPA (a,c), and 100% CHCA (b,d) matrices respectively. Reverse phase (H50) arrays adopting the following settings were used:

a,b: LE 1,500nj, 530 shots, FM 1.5kDa, MR 0-50kDa, MA 1kDa
c,d: LE 2,500nj, 530 shots, FM 1.5kDa, MR 0-50kDa, MA 1kDa

This experiment again showed a very noisy background spectra, although better resolution peaks were detected with higher LE combined with SPA matrix (c). Yet, only very few peaks were detected compromising the statistical power of any further analysis.

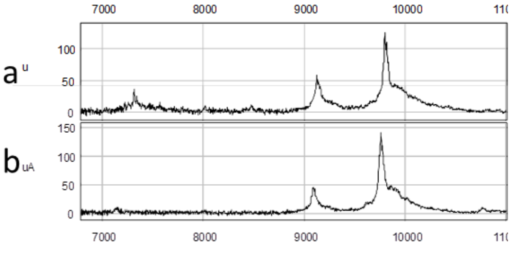


Figure 4. WCX vs. H50 archival urine proteomic profiling
Further G3 Urine analysis (80μl urine was mixed with 30μl U9), using 50% SPA comparing two different arrays (a: CM10, and b: H50 arrays). The following settings were used:
(a,b) LE 2,500nj, shots 530, FM 1.5kDa, MR 0-50kDa, MA 1kDa.
Noisy spectra and very few peaks detected at high laser energies were noted. The scarce number of peaks detected was believed to compromise the statistical power of any further analysis. Interestingly, no obvious spectral differences were detected between the two different chromatographic conditions (a,b). This might be related to presence of very few proteins/peptides in the sample.

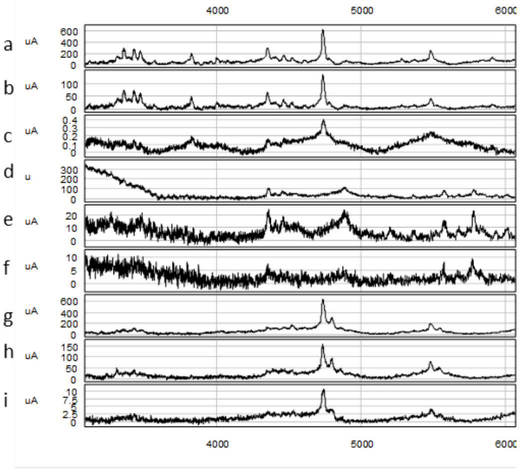


Figure 5. Archival urine profiling array optimisation spectra

Representative MS spectra from the G3 Urine test (samples: 80μl urine mixed U9 30μl), 50% SPA was used in combination with three different arrays (a-c: CM10, d-f: H50 and g-i: IMAC Cu^{++} arrays). Gradient of 11 different LE powers between 500-6000nj were tested.

(a ,d , g) LE 2,000 nj, shots 530, FM 1.5kDa, MR 0-50kDa, MA 1kDa.

(b ,e , h) LE: 2,500 nj, shots 530, FM 3kDa, MR 0-50kDa, MA 1kDa.

(c ,f , i) LE: 6,000 nj, shots 530, FM 3kDa, MR 0-50kDa, MA 1kDa.

The reverse phase (H50) surface revealed the worst profiles in terms of peak detection and signal (d-f)

Using LE 2,000nj and 2,500 nj shows better detection in CM10 and IMAC Cu profiles, whereas the use of LE 6,000 nj yielded higher noise background probably related to detector saturation. Overall, only very few peaks were detected compromising the statistical power of any further analysis. Noted is the reproducibility of the profiles noted in WCX, IMAC (a-c, g-i) spectra. Although more peaks were detected with higher laser energy (LE), only very few additional peaks were detected with noisy spectra background. It was concluded that further analysis using these conditions would compromise the statistical power of future work. Interestingly, very few spectral differences were detected between different chromatographic conditions.

The urine MS analysis revealed that using higher laser energy lead to noisy backgrounds across the different arrays (Figure 5 c, f, i). Altering the matrix between SPA and CHCA in all arrays combined with a gradient of 11 different laser settings revealed better profiles generated by WCX profiling and a poorer spectra using the reverse phase arrays. No advantage was noted in using either matrix (Figure 6).

Over all, the scarce peak number detected in all profiles limited any meaningful analysis. It was concluded that archival urine sample profiling by SELDI is less likely to reveal meaningful profiles, and was therefore not continued.

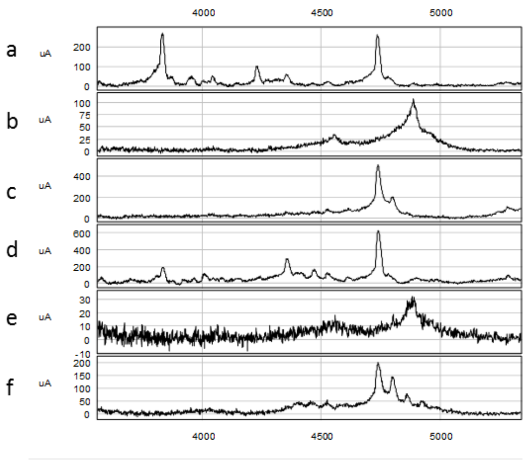


Figure 6. Archival urine proteomic profiling laser optimisation
G3 Urine SELDI test (80µl urine was mixed with 30µl U9), 50% CHCA was used in a-c, and 50% SPA used in (d-f). Three array conditions (a,d: CM10, b,e: H50 and c,f: IMAC Cu^{++} arrays) were used along with a gradient of 11 different LE between 500-6000 nj. Best profiles were generated using WCX arrays; no advantage was noted in using either matrix. CM10 profiles showed more peaks compared to the other profiles, no advantage however was noted with either of the SPA or CHCA matrices. IMAC profiles showed slight increase in the peak numbers using SPA matrix. H50 remained a poor profiling surface for urine analysis. The scarce peak number detected limited meaningful analysis.

3.1.3.2 Archival serum MS analysis optimisation

Serum is a highly complex, variable and dynamic proteomic environment which renders differential profiling a challenge. Thus, efforts focusing on reducing the complexity, variability and abundance bias are vital to achieve novel biomarker discovery[9, 69, 202, 203]. To overcome such hurdles, low abundance and low molecular weight protein detection must be improved to enable in depth biomarker mining. Moreover, clinically matched samples and standardised protocols should be used in any profiling comparison to minimise non-specific variability. Here, work was done to establish the optimal settings for archival serum sample analysis. Matched archival serum samples were fractionated using strong anion exchange (SAX) and sequential pH gradient[203]. Serum fractionation combined with a LMW SELDI protocol was then tested in an attempt to overcome the challenges of complexity, dynamic range and non-pathologic variability.

First, WCX arrays (CM10) were compared to immobilised metal affinity arrays (IMAC Cu^{+2}) arrays. WCX arrays with their carboxylated surface affinity provide a higher binding attraction to positively charged proteins. The metal coordination of the two oxygen groups in the phosphoproteome is believed to lead to stronger binding of phosphorylated molecules on IMAC arrays. IMAC Cu^{+2} arrays therefore provide a good platform for phosphoproteins and signaling pathway discovery.

MS profiling of pooled serum samples using these arrays revealed more complex spectra (greater peak numbers) with the WCX compared to the copper charged IMAC Cu^{+2} profiles. Different laser settings were tested to identify the best profiling conditions for low and higher mass protein peaks. The optimisation also confirmed WCX profiles to provide the greater peak number and better signal detection (Figure 7).

Different serum fractions showed distinct ion peak profiles from fractions 1, 5 and 6 generating the largest number of peaks; hence enhancing their biomarker discovery probabilities. Fractions 2, 3 and 4 had very low ion peak yield (Figure 8). Repeated analysis of these fractions showed similar results (data not shown). Other reports have shown a similarly poor peak signal in F2 (pH 7.0)[204]. This poor yield was not investigated further here as this would require repeated fractionation/analysis, a time and resource consuming exercise which was beyond the scope of this work.

Similar findings were repeatedly illustrated in three different tests, and reproducibility of the analysis platform was confirmed; thus, further analysis of archival serum samples was initiated using these WCX analysis conditions.

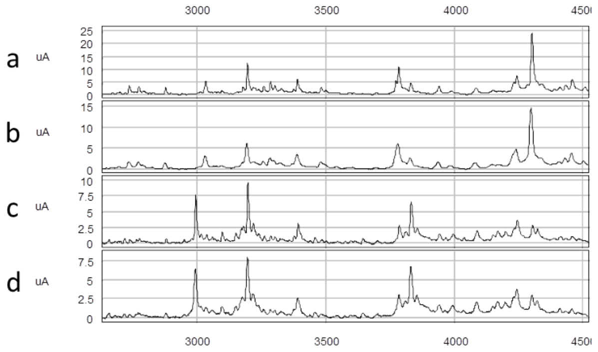


Figure 7. Archival serum proteomic profiling optimisation using WCX and IMAC
Representative spectra from the Guernsey archival serum assessment study. Pooled Guernsey (G4) serum sample analysed using CM10 (a,b) and IMAC Cu^{+2} (c,d) arrays used for comparison of generic profiles. CM10 pH 4.0 buffer was used for WCX arrays (a,b) and IMAC Cu^{+2} buffer for IMAC profiling (c,d). The following settings were used: LE 2,800 (a,c) vs 3,600nj (b,d) used, shots 530, FM 4kDa, MR 0-100kDa, MA 1kDa. 50% SPA matrix used. Reproducibility and inter array profile differences were clearly demonstrated with more complex spectra generated in the WCX profiles.

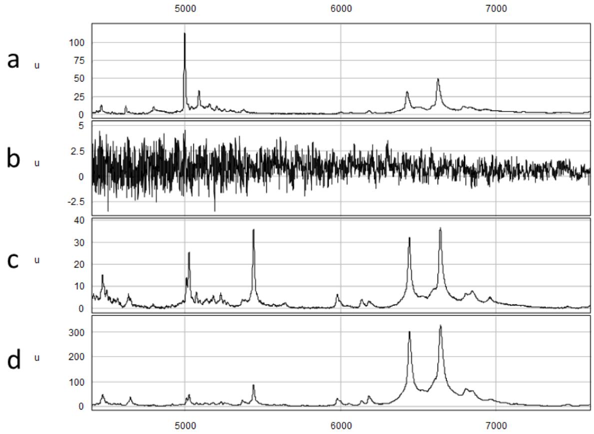


Figure 8. Representative spectra from the archival serum pilot study.
Pooled Guernsey (G4) serum sample were analysed using CM10 arrays. SAX pH gradient fractionation was used prior to SELDI analysis. The effect of fractionation was illustrated (F1,F2,F5,F6) using CM10 arrays and pH 4.0 CM buffer with 50% SPA. The F3 and F4 showed a scarce number of peaks which could be explained by the rich albumin content in these fractions masking many low abundance peaks and yielding spectra with few informative peaks (Table 2). The following settings were used: LE 2,800, 484 shots, FM 6kDa, MR 0-100kDa, MA 2.5kDa.

3.1.3.3 Archival serum SELDI analysis

3.1.3.3.1 Introduction

Breast cancer is the most commonly diagnosed malignancy in women[205]. Whilst in the United Kingdom 5 year survival for good prognostic groups is the same as in the general population[15], the outlook is worse for the poorer prognostic groups regardless of advances in treatment[17, 18]. Early detection and intervention are critical factors affecting survival and overall outcome[23, 206]. Despite improvements in screening and diagnosis, less than two thirds of breast cancers are localised at the time of diagnosis[123]. The developments of novel pre-symptomatic screening approaches therefore play a significant role in breast cancer screening and diagnosis, and have the potential to reduce disease related mortality. Proteomic profiling is a rapidly evolving strategy for biomarker discovery where several advanced proteomic profiling techniques have shown promise in biomarker discovery for the diagnosis of several tumour types including breast cancer[52, 63, 65, 87-92].

Variations in sample collection, storage, and preparation conditions significantly affect the results obtained from clinical proteomic studies[108]. This is a major concern over the use of SELDI and other profiling techniques for the identification of cancer biomarkers and more so in archival serum collections. Effects of sample handling, storage temperature, collection tubes, coagulation temperature, as well as freeze-thaw cycles on MS analysis have been reported previously[108, 109]. Yet none have investigated the effectiveness of analysing samples stored for decades. It is not known if biomarkers reported in recent standardised studies can be reliably detected, and retain predictive power in older tissue and serum stored in tissue banks. Assessing archival sample profiles is the first step towards such application.

3.1.3.3.2 Methods

The detailed technical SELDI methodology was described in detail in the previous chapter. Briefly, serum samples from the Guernsey cohort were defrosted and mixed with U9 buffer incubated on ice for 30 minutes. Each sample (100 μl) was robotically loaded onto CM10 arrays in a randomised manner. Samples were analysed in duplicate, and inter/intra assay reproducibility was assessed by a reference sample made from a pool of all samples. The analysis was performed in two centres using identical protocols. Here, the same samples were reanalysed using identical conditions 3 months later as part of a validation study at the Bio-Rad Laboratories in Copenhagen, Denmark.

3.1.3.3.3 Archival serum profiling results

The practicality and effectiveness of SELDI profiling of prospective archival (30 years old) serum samples (n=14). This work profiled serum from patients who, at the time of collection, had non-metastatic breast cancer or were matched healthy controls. This study thus addressed the possibility that archival collections of serum can potentially be used to identify breast cancer from healthy patients in samples which have been stored for up to three decades.

Having established reliable profiling conditions (Table 4), further analysis of a representative cohort from the archival serum samples was conducted. Weak cation exchange arrays (CM10, pH 4.0) generated the highest quality peak-rich spectra of any array type and were consequently employed for the analysis. SELDI analysis generated 128 peak clusters across the spectra from all analysed fractions (Figure 9). The results of a Bland and Altman analysis showed that the vast majority of peaks were within two standard deviations of the mean with a concordance coefficient of 1 (Figure 10). This concurrence was especially evident within the low molecular weight mass region, which is the main focus of SELDI profiling. CVs between the distinct MS runs were 18%-22% (Figures 11 and 12), which lies within the acceptable range for this methodology[207, 208].

Array	Fraction	No. Peaks	CV Range	Optimal Buffer
WXC (CM10)	**F1**	42	19-26%	**CM pH 4.0-4.5**
	F2	6	22-28%	CM pH 4.5
	F3	12	18-31%	CM pH 4.0
	F4	15	21-27%	CM pH 4.5
	F5	24	16-20%	**CM pH 4.0**
	F6	29	18-26%	**CM pH 4.0-4.5**
	Serum/U9	112	16-24%	**CM pH 4.0**
IMAC	F1	18	22-37%	IMAC Cu^{+2}
	F2	7	24-33%	IMAC Cu^{+2}
	F3	14	20-25%	IMAC Cu^{+2}
	F4	14	18-30%	IMAC Cu^{+2}
	F5	18	21-29%	IMAC Cu^{+2}
	F6	22	25-28%	IMAC Cu^{+2}
	Serum/U9	46	19-27%	IMAC Cu^{+2}

Table 4. Characterisation of optimisation conditions for the Guernsey (G4) cohort.
The different array, and SAX fractionation (F1-F6) spectral yield from the archival serum optimisation analysis. The analysis showed reproducible results with an inter-assay coefficient of variation between 16-37%. In bold are conditions revealing distinct complex reproducible profiles appropriate for future analysis.

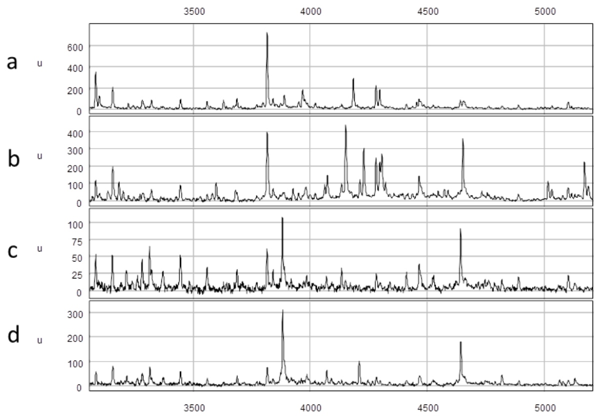


Figure 9. Archival serum WCX SELDI analysis

Representative archival serum profiles following optimisation of CM 10 analysis conditions using CM10 buffer pH 4.5, LE 2,000nj, 848 shots, MA 2.5kDa, MR 0-100kDa and FM 10kDa. Cancer (a,b) versus control (c,d) spectra were generated. All samples were analysed in duplicates. Cancer and control samples were repeatedly compared and showed effective generation of reproducible differential profiles.

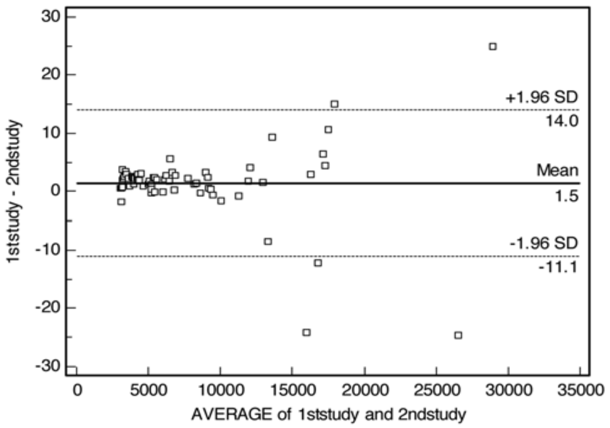


Figure 10. Bland Altman plot of the two comparative archival serum profiling studies

A Bland and Altman plot of the two comparative archival serum profiling studies. Independent analyses at the University of Southampton and Bio-Rad laboratories (Copenhagen) to test the reproducibility of SELDI analysis of archival serum samples were performed. The data generated in both centres illustrated concordance, with a correlation coefficient of 1, indicating a high degree of agreement between the two results. This result indicated the suitability of future archival proteomic analysis on a larger cohort.

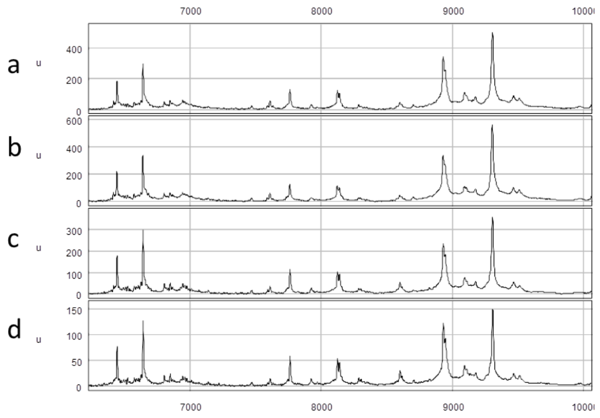


Figure 11. Archival serum proteomic analysis reproducibility
Guernsey cohort (G4) pooled sample reproducibility testing. Optimised conditions using CM 10 (WCX) arrays, CM10 buffer pH 4.5, LE 1,750nj, 848 shots, FM 6kDa, MA 2.5kDa, MR 0-200kDa were tested. Samples were analysed on days 0 (a,c) and 7 (b,d). The average intra array CV on day 0 was 17.6% compared to an average inter assay CV of 19%. All profiles were calibrated and normalised to total ion current (TIC) prior to CV calculation. The average intra array CV on day 7 was 22.1% compared to an average inter assay CV of 20.8%. All profiles were calibrated and normalised to TIC prior to CV calculation.

A batch effect analysis was also performed comparing the spectra obtained over a period of 7 days to compare the CVs of the historic samples with others recently collected (Wessex serum samples). The CVs in both analyses were comparable with no significant difference found between the CVs in both groups. The Wessex reference sample CVs compared to the 30 year old Guernsey cohort samples were 16.4% *vs.* 18.2%, respectively. Analysing the same samples seven days later showed comparable CVs of 18.2% and 20.6% (Figure 12).

Although beyond the aim of the study, the data analysis demonstrated differentially expressed peaks between the cancer and non-cancer serum samples. 16 statistically significant differentially expressed peaks with *P*-values <0.05, differentiated the two groups in this study (Table 5). Seven overexpressed and nine underexpressed peaks were detected in serum samples from cancer patients of this cohort (Figures 13.1-13.6).

The most plausible differentiating protein ion peaks ($P \leq 0.01$) between the two groups were the 3.4, 6.3, 8.6, 8.7, 10.9, 13.9 and 17.4kDa peaks (Figure 13). These peaks showed reproducible differentially expressed group specific behaviour in both studies. Specifically, the 3.4, 8.6 and 10.9kDa peaks were overexpressed in sera from the breast cancer group, whereas the 6.3, 8.7, 13.9 and 17.4kDa signals were underexpressed in all sera from women in the breast cancer group.

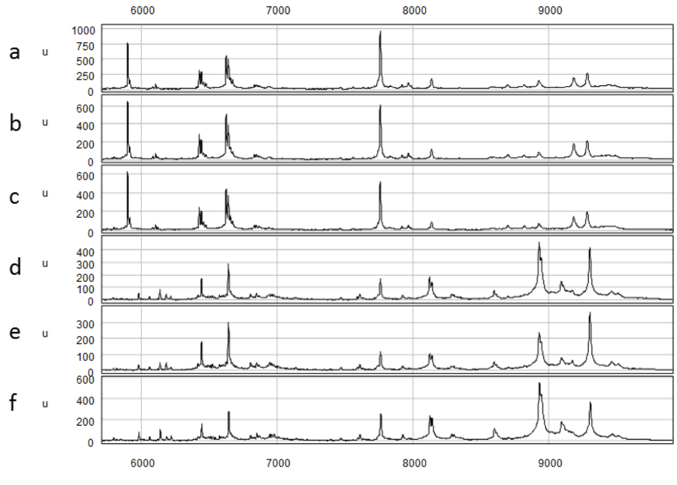


Figure 12. The Guernsey vs. Wessex serum proteomic analysis reproducibility

Comparing the CVs from both the Guernsey cohort serum (a-c) and Wessex serum pools (d-f). Optimised conditions using CM 10 (WCX) arrays, CM10 buffer pH 4.5, LE 1,750nj, 848 shots, FM 6kDa, MA 2.5kDa, MR 0-200kDa were adopted. All profiles were calibrated and normalised to TIC prior to CV calculation. Acceptable reproducibility and CV for the Guernsey serum profiles ranged between 18.2-20.6%, compared to 16.4-18.2% for the Wessex profiles.

m/*z* (kDa)	***P* value**	**Mean intensity (μA) Cancer**	**Mean Intensity (μA) Non-cancer**
<u>3.0</u>	0.04	135.1	56.3
<u>3.4</u>	0.01	33.9	8.37
3.8	0.02	43.5	87.6
6.3	0.01	4.03	5.34
6.8	0.025	6.78	3.62
8.3	0.013	11.16	2.62
<u>8.6</u>	0.01	42.7	8.6
8.7	0.01	7.67	12.54
10.0	0.035	37.01	48.44
10.9	0.01	1.154	0.286
12.7	0.03	15.3	22.61
13.9	0.01	3.22	4.84
17.4	0.01	1.36	2.79
28.3	0.03	0.441	1.493
28.9	0.02	0.81	0.37
56.0	0.025	0.025	0.055

Table 5. Candidate biomarker peaks in the Guernsey cohorts

SELDI protein peaks in the archival Guernsey serum discriminating cancer from non-cancer groups ($P<0.05$). Overall, 16 potential biomarkers between 3.0-56.0kDa were detected (bold). Underlined are peaks with m/*z* similar to previously reported SELDI-derived biomarkers peaks (3.0, 3.4 and 8.6kDa) in breast cancer.

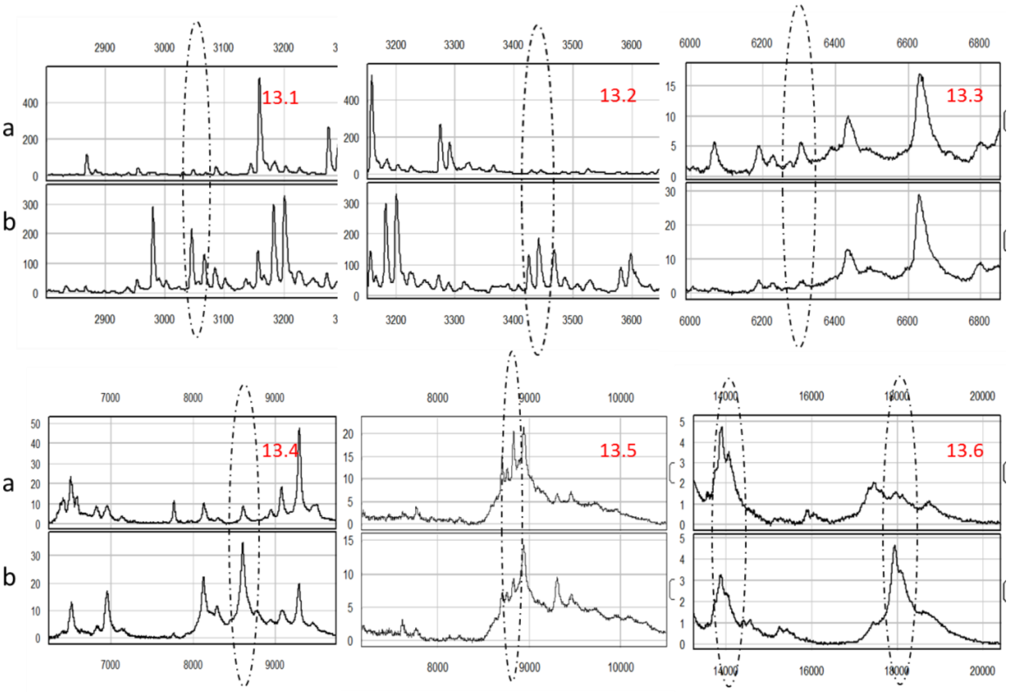


Figure 13 (1-6). Guernsey analysis candidate proteomic biomarkers

Representative SELDI spectra from the Guernsey (G4) cohort comparing serum samples from healthy volunteers (a) and volunteers who had breast cancer at the time of serum collection (b). The highlighted regions represent differentially expressed protein peaks using WCX (CM10) arrays. Protein ions of 3.0, 3.4 and 8.6kDa (Figures 13.1, 13.2 and 13.5 respectively) were overexpressed in cancer patients whereas the peaks at 6.3, 8.7, 13.9 and 17.4kDa (Figures 13.3, 13.4 and 13.6) were under expressed in cancer patients.

3.1.4 Discussion of the Guernsey breast cancer cohort analysis

Disease specific peptides and proteolytic fragments resulting from tumour related proteolytic activity which could be present in the urine were the main target of our archival urinary sample MS analysis[209]. Here, archival urine samples were shown to be of little value for proteomic SELDI analysis as poor quality proteomic profiles were generated despite using a wide range of analytical conditions and settings. In the case of the Guernsey breast cancer cohort, it was evident that larger scale MS analyses using the archival urine samples would not be successful as the MS proteomic signals were generally deprived, limiting meaningful interpretation. Such a finding was unsurprising and could be attributed to several technical and biological factors. Firstly, sample decay might have occurred as the Guernsey samples were stored at -20°C for over 30 years. Secondly, urine samples have relatively low urinary protein content, particularly in individuals with no renal or urological pathologies. Also, the lack of protein enrichment/fractionation steps prior to MS analysis may have limited in depth analysis of samples. Finally, the low resolution and limited sensitivity of the MS platform used in this analysis may have limited the study[210-212]. Despite these factors, our findings could indicate the lack of reliable top down proteomic urinary markers within the Guernsey breast cancer cohort. Although the urinary proteome was previously thought to be more stable than that of blood subsets, our findings did not support this. Despite the poor urinary proteomic profiles generated in this work, it should be noted that over 1,000 urinary proteins/peptides have been previously described, many of which were identified in healthy subjects[213-215]. Moreover, recent urinary exosomes were reported as potential clinical markers[216, 217]. Thus, further urinary proteomic analysis could yield interesting breast cancer biomarkers.

SELDI analysis using WCX profiling of the Guernsey archival serum samples following thorough optimisation of analytical conditions, was shown to be better than IMAC $Cu2^+$ in terms of spectral complexity. Fewer proteins were reported in the inherently albumin rich fractions (Albumin *p*I≈4.7) F3 and F4. This abundant protein has a strong ion suppression effect and co-binding affinity which might explain the poorer signal in these fractions. The low proteomic signal shown in F2 (pH 7.0) is, however, unexpected. This may be related to preparation or analytical errors. It was thought that the protein array preparation protocol which involved a final dH_2O (pH 7.0) washing step, might provide a chemical explanation. The F2 fraction consisted of a proteomic environment rich with neutral proteins (*p*I 7), washing these arrays with dH_2O could have led to the reduction of protein binding affinity. Further investigation involving a comparison between washing with dH2O versus no washing, revealed no difference between the two preparations. Thus, it was concluded that the F2 poor proteomic signal was probably related to a preparatory fault and this fraction was, therefore, excluded from further analysis.

Serum sample profiles were shown to provide reproducible spectra in a multi-centre analysis. Sixteen proteomic peaks were validated by SELDI analysis and differentiated the breast cancer from healthy volunteer sera. Although this is encouraging evidence that supports the analysis of the Guernsey archival samples, these results should be interpreted with caution. A generalised conclusion that all archival samples can be used for proteomic profiling based on the findings described here would be misleading. Instead, these results illustrate the need for further efforts to explore the usability of different archival samples worldwide through similar pilot studies. This would add new evidence regarding the potential use of valuable archival bio bank serum samples in MS profiling.

3.1.5 The Guernsey breast cancer cohort project summary

In this section, work was undertaken to identify the optimal analysis conditions for the Guernsey cohort. MS analysis for urine samples from this cohort was not successful as the proteomic signals were generally poor. However, serum analysis using the WCX array revealed richer profiles compared to other conditions, and was thus used in this pilot study following optimisation of other conditions. Subsequently, two independent analyses of a small subset of the Guernsey serum samples were conducted in two different centres and revealed consistent profiling results. This has added new evidence on the potential use of valuable archival bio bank serum samples in MS profiling.

3.2 Serum proteomic profiling of the Wessex breast cancer cohort

3.2.1 Introduction

Building on the experience gained from the Guernsey cohort analysis, a larger cohort of serum samples (the Wessex cohort) was used to conduct a diagnostic breast cancer biomarker discovery study. Potential diagnostic serum markers differentiating healthy controls, benign breast disease (BBD) and invasive breast cancer (IDC) patients were targeted in this project. Although a similar MS based profiling approach was adopted, biomarker purification, identification and validation were also performed as stronger evidence was provided by the larger cohort size.

3.2.2 Materials and methods

Serum samples were collected from breast disease and healthy volunteers from the Wessex region (UK) prior to any intervention. This collection was approved by the Southampton General Hospital NHS Trust Ethics Committee (Ethical Approval 05/Q1702/13, R&D reference no. RHMCAN0392) and informed consent was obtained from all participants in the study. Samples were collected from volunteers who were recalled for further investigations following routine 3 yearly mammographic screening. Serum samples were prepared after 45 minutes coagulation time, by centrifugation at 4000 rpm for 15 minutes and aliquots were stored at -80°C within 15 minutes of collection.

The detailed protocols used in each stage have previously been explained in the main methods section. Briefly, array/laser/buffer and matrix conditions were tested to reach the most complex and reproducible profiling conditions. This was followed by automated preparation and MS analysis using the SELDI. Samples were analysed in duplicate and a second blinded validation analysis was performed. A further independent validation study was also conducted using a different MS method (MALDI) at the Medical Biomics Centre (St. George's University of London). The candidate markers passing all three analyses were subject to biomarker purification, identification, verification and immunovalidation.

3.2.3 Results

3.2.3.1 Weak cation exchange (WCX) profiling assessment

SAX fractions from the Wessex cohort serum samples (F1-F6) were used for the characterisation and optimisation of MS analysis. Initially, the profiling abilities of WCX arrays were explored using a low stringency CM10 wash/binding buffer (pH 4.5). This was based on previous optimisation results obtained from the Guernsey cohort analysis, revealing consistent profiles under such conditions. Applying 50% SPA, a laser energy of 2,000nj and an average of 220 shots, only very few peaks were detected and spectra with poor signal and noisy background were generated (Figure 14).

To investigate the effect of laser settings on improving the profiles, different laser energy/number of shots were tested, and more favourable spectra were created using higher laser settings (LE 2,500-3600nj, 636 shots) (Figure 15). More intense proteomic peaks, higher S/N and improved resolution typified the clearer spectra. This work demonstrated the effect of serum sample ion exchange fractionation in generating distinct proteomic patterns (Figure 15).

Although fractionation profiles indicated a reduction of complexity in terms of distinct patterns generated by different fractions, the attenuation of the serum dynamicity was not fully established. The presence of albumin in fractions 3 and 4 was not surprising[203], however its presence in fraction 5 at high concentrations was unexpected (Figure 16). This was thought to partly explain the fewer proteomic signals in this fraction. Fraction 2 profiles were shown to be very noisy with almost no signal. Fractions 2 sub-optimal profiles can be explained by either preparation error or, to a lesser extent, the sample nature. Overall, serum profiling by WCX arrays appeared to be successful which was consistent with previous findings in the pilot study.

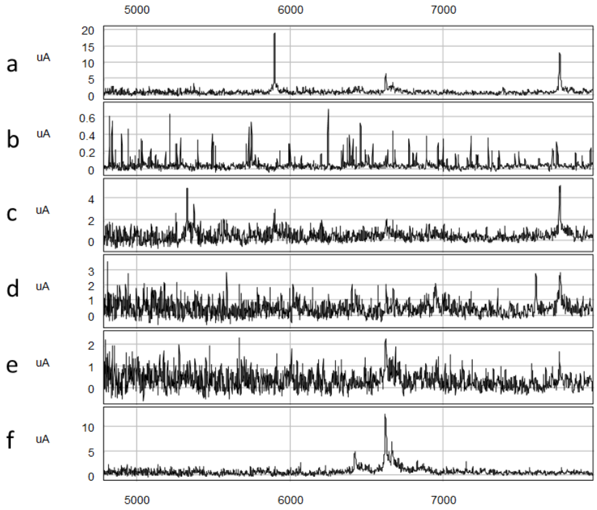


Figure 14. Pooled Wessex serum samples SAX fractionation

Pooled Wessex serum samples fractionated using SAX ion exchange and pH gradient elution ranging between pH 9-organic acid (F1-F6), and analysed by SELDI TOF MS on WCX arrays. Fractions 1-6 (a-f respectively) are shown. The settings used in this analysis were: SPA 50%, LE 2,000nj, average shots 220, MR 0-100kDa, MA 3kDa, FM 5kDa. Very few peaks detected, poor signal and noisy spectra. Ion suppression effect is a possible cause of such profiles. Sub optimal laser energy could also contribute to the results shown. Potential S/N improvement using higher LE settings was tested in subsequent steps.

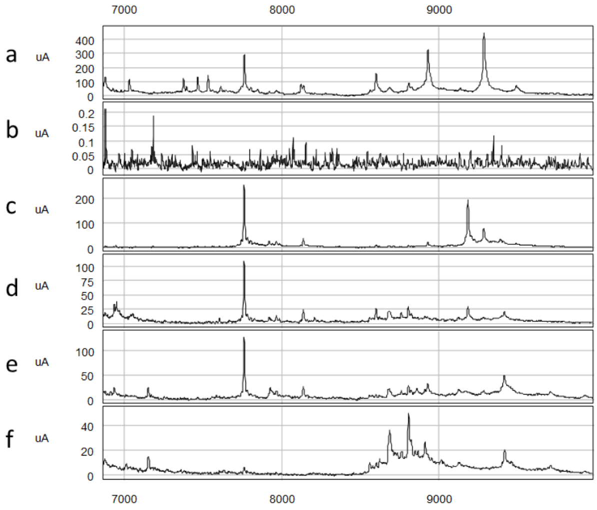


Figure 15. WCX (CM10) SELDI profiling of fractionated Wessex samples

WCX (CM10) SELDI profiling of fractionated (a-f) pooled Wessex reference sample (F1-F6) illustrating the improvement of peak detection and resolution at higher laser settings. Increasing the LE to 3,600nj and shots to 636 in the previous experiment (Figure 15) improved S/N. Fraction 2 (b) however showed very poor signal in repeated analyses.

The fractionation effect is demonstrated by the differential profiling in each fraction originating from the same pooled samples.

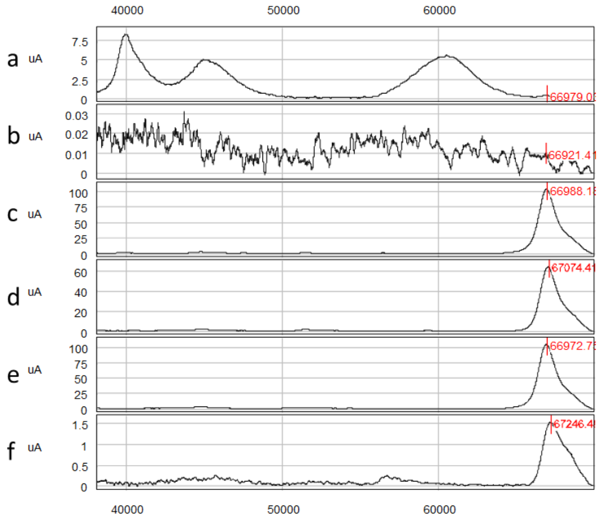


Figure 16. Albumin representation in fractionated Wessex sera

The six (a-f) fractions from the Wessex cohort using SAX resin solid phase ion exchange chromatography and pH gradient sequential elution analysed on WCX (CM10) arrays. The presence of ion peaks at 67kDa demonstrated the presence of Albumin in fractions 3, 4 and 5 corresponding to pH elutions 5, 4 and 3 respectively. This was shown to be present in a relatively high concentration indicated by total ion current (TIC) normalised peak intensities.

3.2.3.2 Strong anion exchange (SAX) profiling assessment

In an attempt to explore the quality of serum profiling using different chromatographic arrays in order to establish the utility of different array analysis in future work, experiments were conducted applying the same pooled serum fractions (F1-F6) from the Wessex cohort to SAX (Q10) chips. Tris HCl pH 8.0 buffer was used in combination with 50% SPA for sample preparation. Although very few proteomic signals were detected, laser energy/shots gradient was tested revealing acceptable peak S/N and resolution at laser energies around 1,700- 2,800 nj combined with an average of 636 shots (Figures 17 and 18).

To exclude instrumental and preparation errors and to confirm the suitability of the previously established settings, a repeat laser setting gradient test was conducted. The best profiles were generated at energy levels around 2,000 nj using 636 shots; using energies around 1,500 nj was also meaningful, but of less quality (Figure 19 e and f). Overall, the profiles generated using SAX arrays showed noisy background and had limited peak numbers so were deemed unsuitable for future analysis.

To improve the quality of SAX generated spectra, on spot preparation (as per manufacturer's recommendations) was performed. This has showed no improvement in the spectra in terms of noise or complexity (i.e. peak number). Laser settings here were identical to the automated preparation parameters (Figure 20).

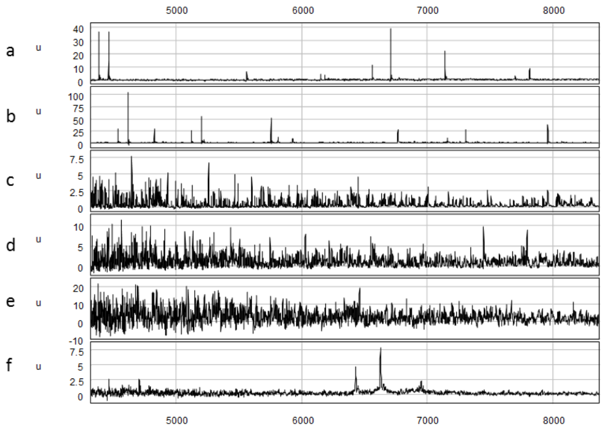


Figure 17. Q10 analysis of SAX fractionated sera

Representative Wessex serum fractions (F1-F6) (a-f) analysis using SAX (Q10) arrays. 50% SPA and Tris HCL pH 8.0 buffer were used with different laser energies (LE).

LE used for optimisation was as follows: a: LE 800nj, 1 shot, b: LE 800nj, 20 shots, c: LE 1,200nj, 20 shots, d: LE 1,500nj, 20 shots, e: LE 2,000nj, 636 shots, f: LE 2,800nj, 636 shots. Poor spectra were noticed in all runs. Spectra (f) shows slightly improved protein ion signals. This indicated that higher laser energy improved the S/N.

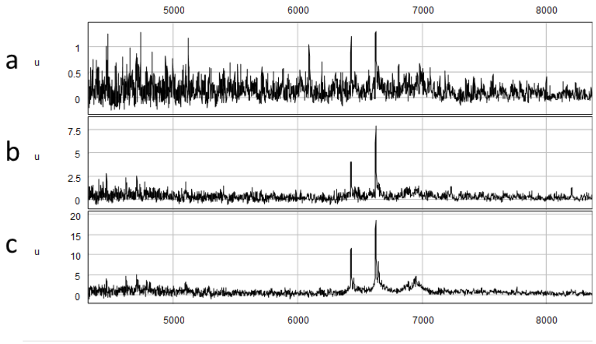


Figure 18. Laser modefication of Q10 array analysis

SAX (Q10) array analysis using pooled serum sample from F6 fraction (a-c) of the Wessex cohort comparing the LE effect on the spectra S/N. a: LE 1,500nj, shots 371. b: LE 1,700nj, 371 shots, c: LE 1,700nj 530 shots. A slight improvement of the signal was noted by higher laser energy (a,b) and laser shot (b,c) settings. Nonetheless, the spectra were generally of poor in quality. The same observation was noted in different fractions indicating a possible suboptimal fractionation, analysis conditions and/or array chemistry.

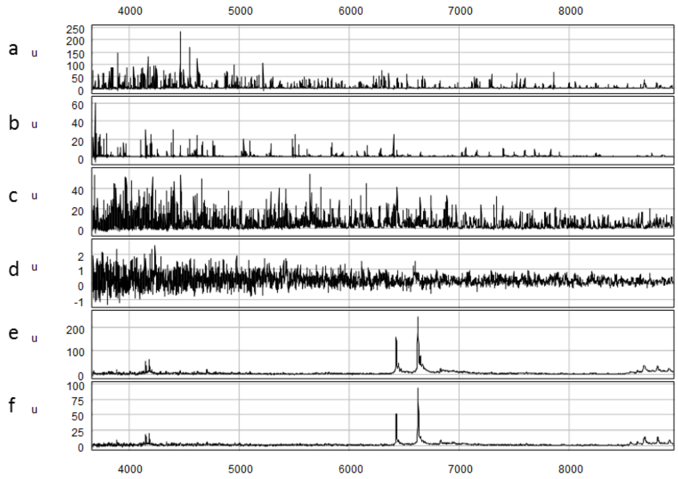


Figure 19. On spot preparation of fraction 2 using Q10 arrays

SAX (Q10) optimisation analysis using Wessex pooled serum sample from F2 (a-c) and U9 treated serum (d-f) using on spot preparation. Poor spectra were noticed in all runs. Spectra in (e,f) showed improved protein ion signals (e,f) which supports improved profiling by increased laser energy/shots. Settings used were as follows: a: LE 800 nj, 1 shot, b: LE 800 nj, 20 shots, c: LE 1,200 nj, 20 shots, d: LE 1,500 nj, 20 shots, e: LE 2,000nj, 636 shots, f: LE 1,500 nj, 636 shots.

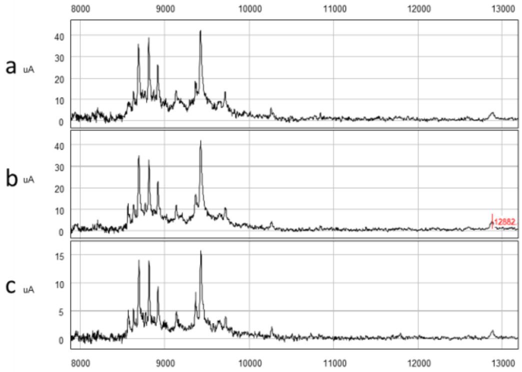


Figure 20. On spot preparation of serum samples using Q10 arrays

SAX (Q10) on spot preparation of serum sample in triplicate, and SELDI profiling. The following settings were used: (a) LE 1,500 nj, 318 shots. (b) LE 1,700 nj, 530 shots. (c) LE 1500 nj, 636 shots. No major differences in the peak/profile complexity were noted. This could be related to poor sample quality/preparation and/or sub optimal array conditions.

Further optimisation tests were implemented to assess the effect of sample dilution on S/N and peak detection. Pooled fractions were diluted with a dH_2O (1:1, 1:10, 1:50 and 1:100 dilutions) gradient. Interestingly, this appeared to improve the S/N by reducing background noise (Figure 21). This can be explained by the reduction of the salt concentration in the sample which suppresses ion profiling and generates less baseline “grass” effect. Yet, no effect was shown on peak number/detection, and the number of peaks remained scarce. It was concluded that SAX profiling provides a less favourable profiling yield and thus less expression profiling data compared to WCX profiles.

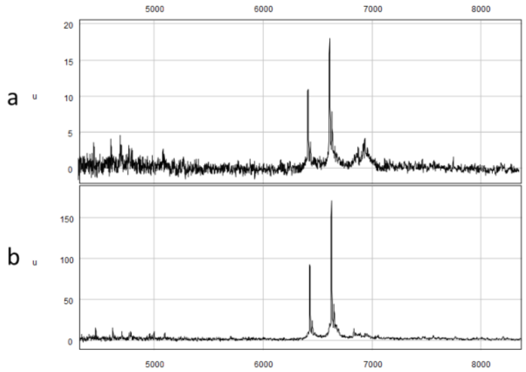


Figure 21. dH20 dilutional effect on Q10 array analysis
SAX (Q10) chromatography MS analysis using pooled serum sample from U9 treated serum (a,b) comparing the dilution effect on the spectra S/N. (a) Neat serum/U9 sample, (b) 1:50 dilution (dH2O) of the same sample improved signal detection on Q10 arrays.

3.2.3.3 Reverse phase (H50) profiling assessment

As a continuation of the preliminary analysis of the Wessex cohort, profiling using hydrphobic affinity arrays was performed. Samples were treated with H50 buffer and 50% SPA used based on previous reverse phase array analysis. Reverse phase hydrophobic affinity analysis was tested on all six fractions used earlier (F1-F6) using different laser energy settings. Laser energy ranged between 1,500-5,000 nj and 120-630 shots were tested. This work indicated that laser energies of approximately 2,000 nj using 630 shots had better peak yields with less saturation (Figure 22). However, there were very few peaks detected (Table 4), and aside from a prominent signal at 17.5kDa, the rest of the spectra appeared to be noisy. The poor quality profiling using this surface was confirmed in repeated nalyses eliminating this condition from further optimisation (Figure 23). The limited number of peaks generated by serum reverse phase chromatography is consistent with the limited number of membrane and hydrophobic proteins in the highly hydrophilic serum. Regardless of any other contributing factors, it appeared that any further optimisation would be time consuming and could only achieve very minor improvement of the analysis.

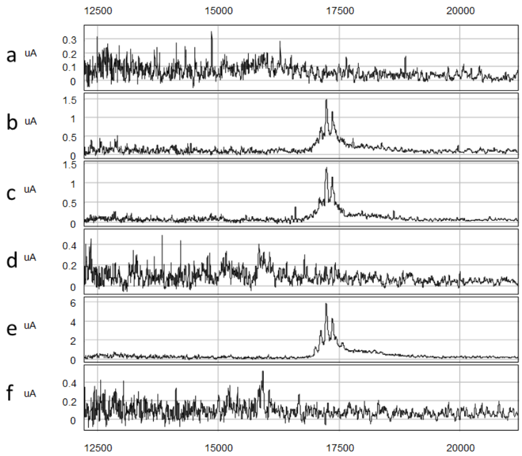


Figure 22. Reverse phase (H50) array analysis of the Wessex samples

Reverse phase array (H50) assessment of the Wessex samples. 50% SPA and H50 buffer were used before profiling of F1-F6 (a-f). The following settings were used: LE 2,000nj, 630 shots, FM 5kDa, MR 0-100,000, MA 3kDa. Poor spectra in all fractions were reported. This was tested over a range of LE and shots (LE between 1500-5000nj, shots between 120-630) and confirmed the poor quality of profiles generated by this surface. Of note was a prominent peak at the 17kDa region.

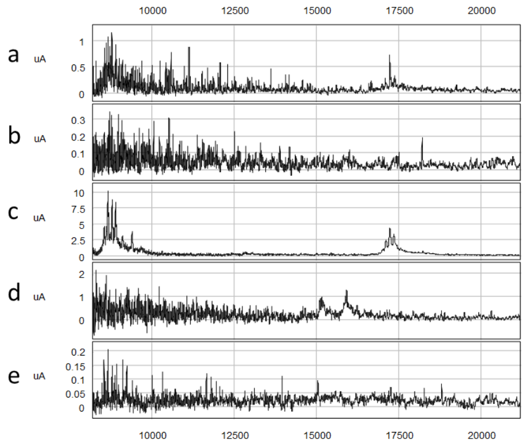


Figure 23. Laser modification effect in H50 array analysis

From previous experiments (Figure 21), using the same conditions (H50) arrays and buffer, 50% SPA profiling F1,3,4,5 and F6) from the Wessex serum pools, poor spectra in all fractions was still noted. This was tested over a range of LE and shots (LE between 1500-5000 nj, shots between 120-630) FM 5kDa, MR 0-100,000, MA 3kDa. Fraction 5 LE optimisation, LE: a: 1,500nj / 630shots, b: 1,800nj / 630 shots, c: 2,000nj / 630 shots, d: 2,000nj /100 shots, e: 2,000nj / 200 shots. Optimal laser settings were illustrated at LE 2,000nj/ 630 shots. Yet, very few peaks were detected using this array condition.

3.2.3.4 Immobilised metal affinity capture (IMAC Cu^{2+}) profiling assessment

Proteomic profiles generated by covalently bound proteins on copper loaded immobilised metal affinity (IMAC Cu^{2+}) arrays were studied to complement the chromatographic display of the Wessex cohort serum profiling. Laser settings were characterised and tested resulting in optimal laser settings of 2,800 nj and 630 shots (data not shown). The differential profiling resulting from fractionation is demonstrated in Figure 24. Albumin was concentrated in fractions 3 and 4 (pH 5.0 and 4.0 respectively). Fraction 2 (pH 7.0) had noisy background, and no peaks. This raised the possibility of a technical error involving this fraction. A comparison of the number from IMAC to those from CM10 profiles indicated larger peak numbers with CM10 arrays (Table 6). Based on these findings, further work focused on WCX SELDI optimisation and setting fine-tuning to achieve a reproducible platform and reliable analysis.

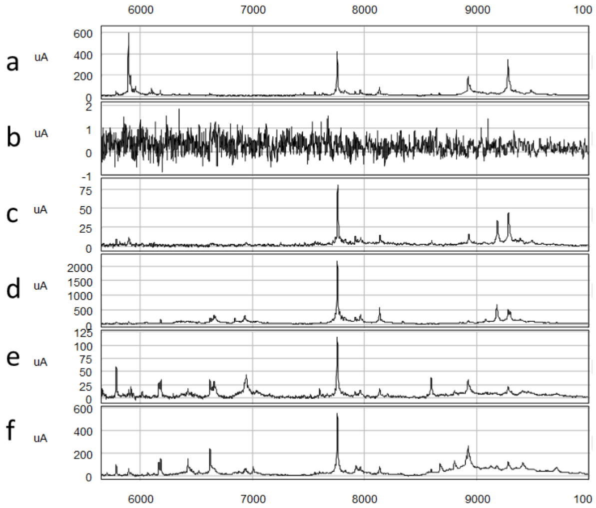


Figure 24. IMAC Cu2^{+} profiling of the Wessex cohort

IMAC Cu^{2+} profiling using the same F1-F6 (a-f) (SAX) fractions of the Wessex cohort. LE 2,800nj, 630 shots, FM 5kDa, MR 0-100kDa, MA 3kDa and 50% SPA matrix were used. Differential profiling between different fractions was demonstrated. Fraction 2 had poor signals. Overall, IMAC Cu^{2+} profiles revealed less peaks than CM10 profiles. Therefore, CM10 profiling was used for the main study. IMAC Cu^{2+} analysis would be a good option for future work (Table 4).

Array	**Fraction**	**No. Peaks**	**CV Range**	**Optimal Buffer**
WXC (CM10)	**F1**	42	21-27%	**CM pH 3.5-4.5, 7.5**
	F2	5	25-33%	CM pH 3.5-4.5,7.0
	F3	13	17-20%	CM pH 6.0, 7.5
	F4	18	20-27%	CM pH 4.5, 7.5
	F5	20	21-24%	**CM pH 4.5-5.5**
	F6	28	18-28%	**CM pH 3.5-4.5, 6.5**
	Serum/U9	108	20-30%	**CM pH 3.5-6.0**
SAX (Q10)	F1	6	27-32%	Tris HCl pH 7
	F2	3	30-34%	N/A
	F3	5	29-35%	Tris HCl pH 8.0
	F4	9	31-34%	Tris HCl pH 8.5
	F5	11	24-29%	N/A
	F6	10	22-29%	Tris HCL pH 8-10
	Serum/U9	47	21-30%	Tris HCL pH 8.0
IMAC Cu^{+2}	**F1**	38	27-34%	IMAC Cu^{+2}
	F2	9	20-29%	IMAC Cu^{+2}
	F3	13	22-27%	IMAC Cu^{+2}
	F4	6	18-21%	IMAC Cu^{+2}
	F5	17	24-32%	IMAC Cu^{+2}
	F6	22	16-28%	IMAC Cu^{+2}
	Serum/U9	89	23-26%	IMAC Cu^{+2}
Reverse phase (H50)	F1	4	30-39%	10% ACN, 0.1% TFA
	F2	2	33-38%	10% ACN, 0.1% TFA
	F3	3	40-41%	10% ACN, 0.1% TFA
	F4	2	29-42%	10% ACN, 0.1% TFA
	F5	4	33-46%	10% ACN, 0.1% TFA
	F6	3	32-41%	10% ACN, 0.1% TFA
	Serum/U9	5	29-35%	10% ACN, 0.1% TFA

Table 6. Characterisation of optimisation conditions for the Wessex cohort
F1-F6 represent SAX chromatographic fractions 1-6 respectively covering a pH range between 9-2. In bold are conditions revealing distinct reproducible profiles suitable for future analysis.

3.2.3.5 Laser re-optimisation, reproducibility and variability

Work was initiated to establish the best settings and protocols for bulk analysis of the Wessex cohort using the CM10 conditions. Unfortunately, technical problem(s) ensuing from a laser malfunction necessitated a re-check of the optimal settings at an early stage. To examine the new laser settings, laser optimisation tests were conducted and revealed reproducible detection at laser energy levels between 1,750-5,500nj combined with an average of 636 shots (Figure 25 and 26).

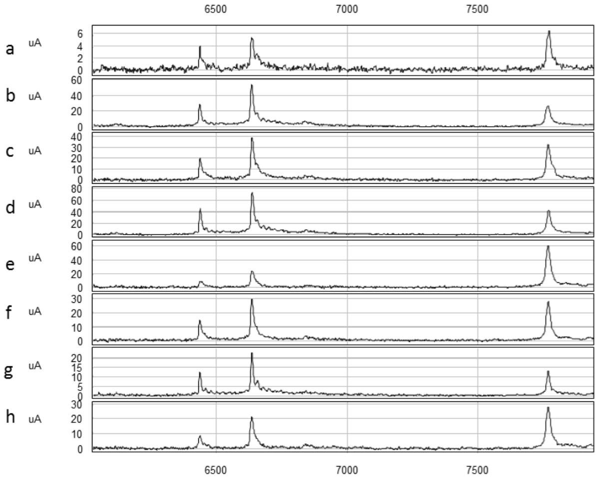


Figure 25. Laser optimisation of the Wessex analysis

Further Wessex serum laser energy (LE) optimisation using a pooled reference sample. WCX (CM 10) arrays, pH 4.5 buffer and 50% SPA were used. Representative spectra for LE gradient were used in the following order: a. LE 1,350nj, 742 shots, b. LE 1,500nj, 795 shots, c. LE 1,650nj, 636 shots, d. LE 1,650nj, 795 shots, e. LE 1,650nj, 848 shots, f. LE 1,500nj, 1060 shots, g. LE 1,500nj, 1060 shots, h. LE 1,500nj, 1060 shots. Spectra resolution improved when LE was increased to > 1,500nj (a *vs.* b-h). Laser shots didn't reveal any significant effect after 795 shots (e.g. d,e). Reproducible peak detection was demonstrated (f-h). This test showed the variable peak intensities between spectra (b-h) which could be related to ion suppression. This is usually adjusted by TIC normalisation prior to analysis.

To assess the batch effect on the profiles, arrays were analysed on the day of preparation and a week later. Reproducible results were observed with peak intensity average CV of 21.6% (Figure 26). Assessing the batch variability 60 days from preparation also showed reproducibility of the peak detection. However, after 60 days higher variability was detected with inter and intra assay CVs of 29 and 36% respectively (Figure 27). This indicated the possibility of re-analysing the arrays up to a week from the day of preparation with reliable results but not after prolonged storage periods.

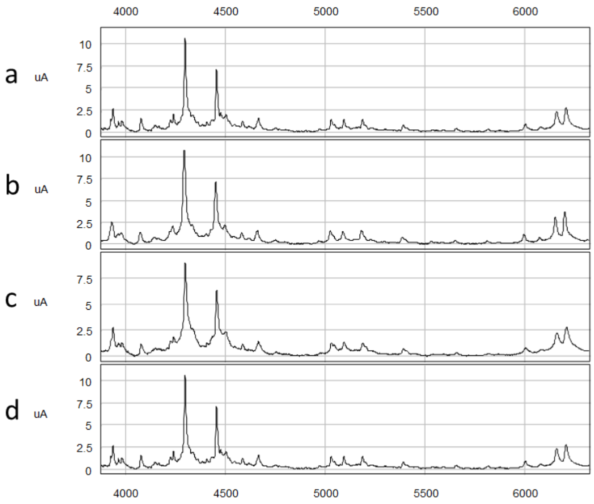


Figure 26. Testing of batch effect in the Wessex proteomic analysis

The pooled serum sample from the Wessex cohort was analysed using: LE 2,800, 3,600, 5,500 and 2,800nj (a-d respectively), FM 5kDa, MR 0-100kDa, MA 3kDa, WCX arrays and SPA 50% matrix. Samples a and b were analysed day 0, samples c and d were analysed on day 7, reproducibility of profiling was illustrated in all runs. (Spectra shown for illustration).

To establish the analytical variability of the analysis, a pooled sample was analysed using 5 simultaneous robotically prepared WCX arrays, using identical protocols. Each array was analysed in quadruplicate. The inter assay CVs ranged between 18-22% and intra assay CVs were 20-23% respectively (Figure 28). Although this work has indicated that analysis up to one week following preparation has an acceptable level of analytical variability[204], all subsequent analysis was performed using identical conditions and was conducted immediately following preparation to avoid any source of analytical bias.

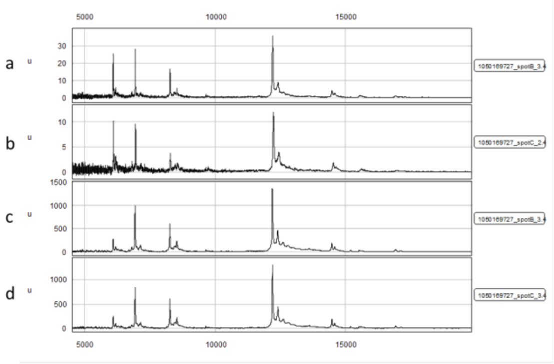


Figure 27. Variability and batch effect demonstration in SELDI test
Variability and batch effect assessment of the SELDI platform using NP20 and All in One protein standards (Bio-Rad), with 50% SPA matrix preparation. Spectra a and b correspond to day 1 and 3 was compared to spectra corresponding to day 60 (c and d) using LE 2,000nj and 538 shots. Reproducibility was clear in terms of peak detection a-d, and mass/intensity accuracy (c and d). Analytical variability showed an inter and intra assay CVs of 29 and 36% respectively.

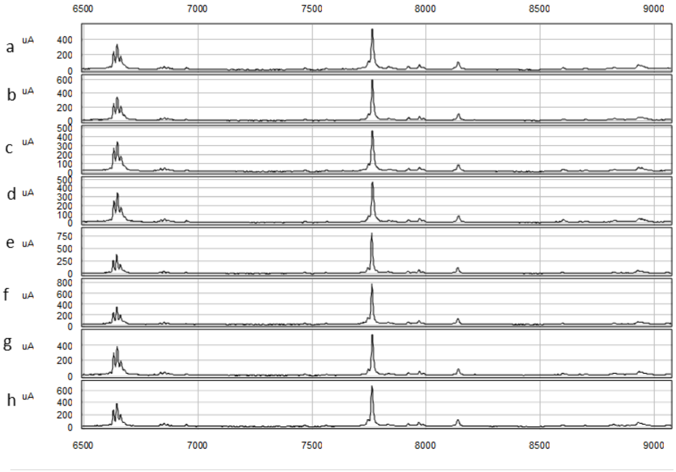


Figure 28. Inter and itra- assay reproducibility and variability in SELDI
A pooled Wessex serum sample was analysed in quadriplicate. The following settings were used: LE 2,500nj, 848 shots, FM 5kDa, MR 0-50kDa, MA 2.5kDa, 50% SPA was applied. Repeated analysis using the same array (a-d) and inter assay analysis using same conditions and different arrays (e-h) is shown. This revealed satisfactory inter assay (18-22%) and intra assay (20-23%) CVs.

3.2.3.6 The effect of dilution, pre-wetting, protein reduction and surfactant

Several steps were tested to explore different sample preparation methods to identify the method giving the highest number of peaks. Firstly, the effect of sample dilution using the WCX analysis was tested. A gradient of sample dilutions were tested (1:1, 1:10, 1:50 and 1:100 with H_2O) for all fractions as well as neat serum (Figure 29). In contrast to the improvement noted by diluting samples profiled on SAX arrays (Figure 21), no major improvement of the signal, resolution or detection was achieved by diluting the samples in WCX profiles. Further, it appeared that diluting serum under these profiling conditions produced worse MS profiles than neat samples (Figure 29 g, h).

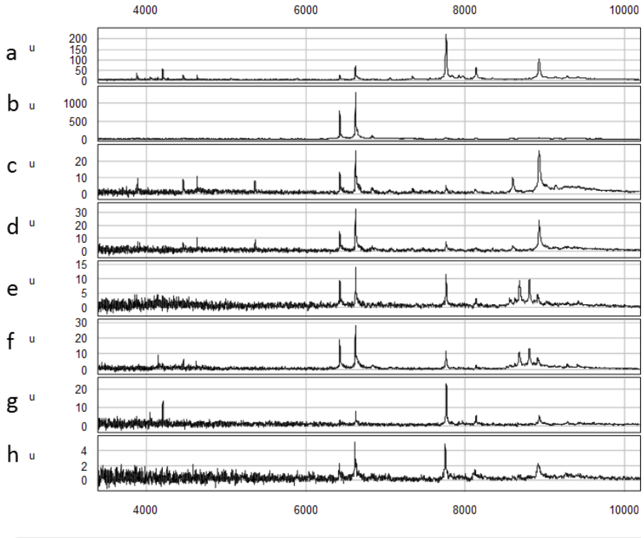


Figure 29. Dilutional effect in WCX array analysis by SELDI

SELDI profiling to test the effect of sample dilution on profiling generated by CM10 analysis. Dilution of samples with dH_2O gradient (1:1, 1:10, 1:50 and 1:100) was tested. Pooled sample from the Wessex cohort and CM10 buffer pH4.5 were used for sample/array preparation and wash. The following settings were adopted: LE 1,350nj, shots 636, FM 5kDa, MR 2-200kDa. a. Crude serum, b. Serum/U9, c. F1, d. F1 (1:100), e. F6, f. F6 (1:100), g. Serum 1:50 dilution (dH2O), h. Serum 1:100 dilution (dH2O). Clearly better profiles generated using this array (CM10) compared to other arrays (Q10, H50 and IMAC). Reproducibility between duplicates (c, d) and (e, f), serum and 1:50 diluted serum (a *vs.* g and h) was demonstrated. No advantage was seen by diluting the serum (a, g, h). U9 seemed to have a denaturing effect and reduced serum complexity (a, b). Spectrum in (h) represents diluted unfractionated serum which had a noisy background. This could be related to a preparatory error.

This work also explored the effect of pre-wetting the arrays with dH_2O prior to preparation and sample addition. Samples were randomised to minimise false interpretation. Surprisingly, array pre wetting moderately improved over all spectra by reducing background noise (figure 30). The most notable improvement was in fraction 2, which had previously revealed poor peaks. Nonetheless, the number of peaks detected in this fraction remained too low to justify further analysis.

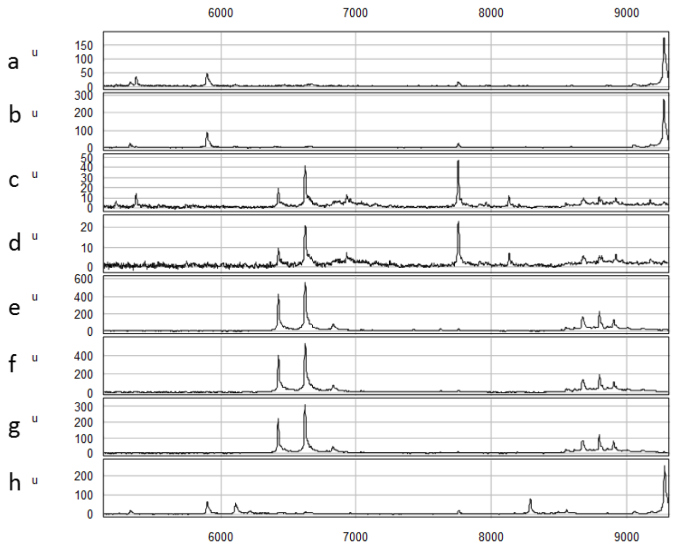


Figure 30. Array pre wetting effect in MS analysis
WCX (CM 10) optimisation testing the effect of pre-witting arrays with 3μl dH2O prior to preparation. Settings used in this experiment were: LE 1,700nj, 848 shots, FM 6kDa, MR 0-85kDa, MA 2.5kDa. a. F2 pooled sample, b. F2 pooled sample (pre-witted), c. F4 pooled sample (pre-witted), d. F4 pooled sample, e. F6 pooled sample, f. F6 pooled sample (pre-witted), g. F6 pooled sample (pre-witted), h. F2 pooled sample (pre-witted). Reproducible spectra were noted (a,b,h), (c,d), (e-g). In addition, there was a slight improvement in the S/N by pre wetting (a *vs.*b and h) and (c,d). Pre wetting changes were less remarkable in the F6 analysis (e-g). These observations were reproducible and led to the incorporation of array pre-wetting in subsequent analysis.

Experiments were also performed to assess Triton X-100 detergent effect on the spectra. This has demonstrated no advancement in profiling quality or complexity by using a non-ionic surfactant (Triton X-100) (Figure 31). This may be explained by the absence of membrane proteins in the sera, the main environment where Triton X-100 exhibits enhanced protein extraction. Although Triton X-100 was previously shown to alter the albumin structure, this did not alter the profiles generated in this work.

The effect of dithiothreitol (DTT) on the spectra quality was also tested. DTT treated serum samples were shown to have higher peak intensities and improved signals (Figure 32). Repeated analysis confirmed the former result and demonstrated changes in the profile signals which may be due to disulfide bond reduction (Figure 33). Although DTT improved the S/N in all fraction profiles, this resulted in saturated peaks, which could compromise LMW signal detection. Thus, DTT was not used in sample preparation during the Wessex cohort MS analysis.

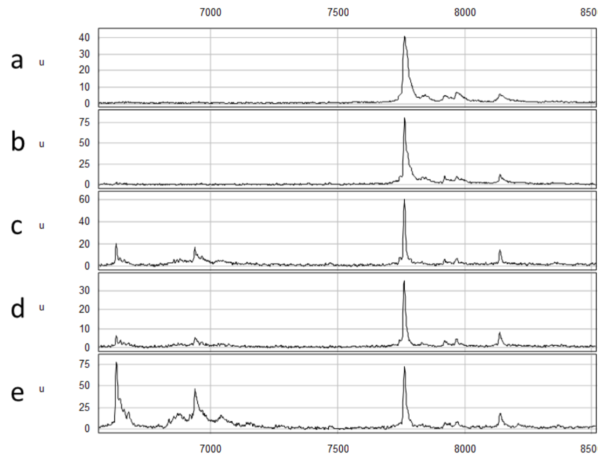


Figure 31. Triton X-100 effect on SELDI preparation

Representative spectra from Triton X-100 tests studying the effect on SELDI profiling. The following samples were used: (a.b) F3 pooled sample, (c,d,e) F4 pooled sample. (a,c) treated with Triton 0.1% X-100, (b,d) treated with 0.01% Triton X-100, (e) not treated with Triton X-100, (a-d) LE 1,750nj, 1060 shots, (e) LE 2,000nj, 848 shots. No clear differences were observed following addition of 0.1% or 0.01% triton X-100 to sample buffer (a-d). Increasing the laser energy to 2,000nj has improved spectra S/N without Triton X-100 treatment (e).

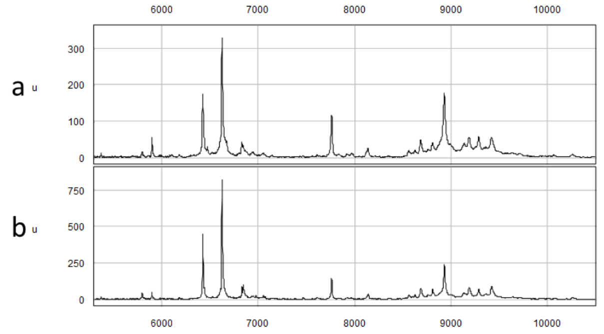


Figure 32. DTT effect on CM10 MS profiling

The effect of DTT on WCX array profiles was tested here using the following settings: LE 2,000 nj, 848 shots, MA 2.5kDa, MR 0-100kDa. Comparison of (1:10) U9 treated serum pre (a) and post (b) DTT addition, showed no significant improvement in terms of peak number. The signal was however markedly increased by 0.1M DTT treatment as illustrated by change in peak intensities in (b). This resulted in saturation of abundant peaks, potentially reducing the detection efficiency of MS profiling.

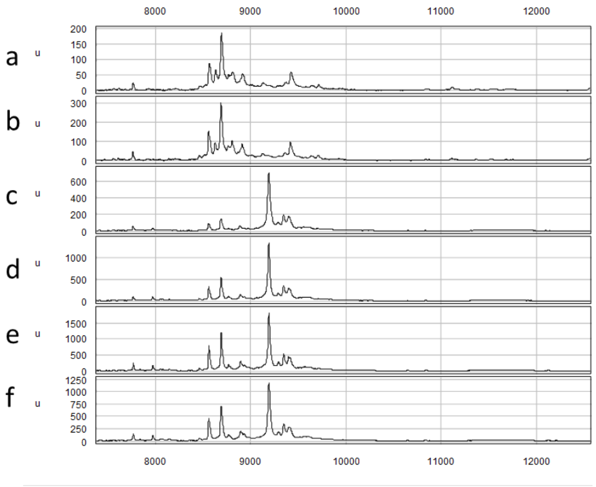


Figure 33. DTT and U9 effect on CM10 profiling
CM10 pH 4.5 buffer was used for sample preparation on WCX arrays to test the effect of DTT addition to the sample. Using the settings: LE 2,000nj, 848 shots, FM 2.5kDa, MR 0-100kDa, MA 2.5kDa SELDI spectra were generated. (a,b) (1:10) U9 treated serum, (c,d) (1:10) U9 treated serum + 0.05M DTT, (e,f) (1:10) U9 treated serum + 0.1M DTT. The reproducibility of spectra is illustrated below (a,b), (c,d), (e,f). The DTT reducing effect was demonstrated by the shift in spectra (c-f) indicated disulfide bond dissociation. Also noted was the increase in the S/N as the DTT concentration increased (c, d vs. e, f). Although this improved S/N is usually a pre-condition for analysis, the S/N achieved without DTT addition was satisfactory and the saturation effect associated with higher peak intensities in DTT treated samples threatened to mask low abundance peaks. DTT was, therefore not used in further analysis.

3.2.3.7 WCX binding/wash buffer assessment

Adjusting the buffer used in array/sample preparation is a crucial element of the SELDI assay optimisation[204]. This was achieved by testing arrays with varying pH buffers and assessing the resultant profiles. The lower the pH of the solution from the p*I* of a protein, the higher its overall positive charge and therefore the stronger the binding to the WCX surface and vice versa. A series of different 50mM buffers, with increasing pH (3.5-7.5) were used to alter the binding pattern of proteins to the WCX arrays. This combined with the original SAX fractionation provided a 2D MS profiling tool. Each SAX fraction was tested using optimized favorable laser settings. This work revealed the best profiles were achieved using CM10 buffer pH 4.5. These profiles possessed larger number of peaks especially at the LMW region, and a high S/N and resolution compared to the other buffers used (Figure 34, 35).

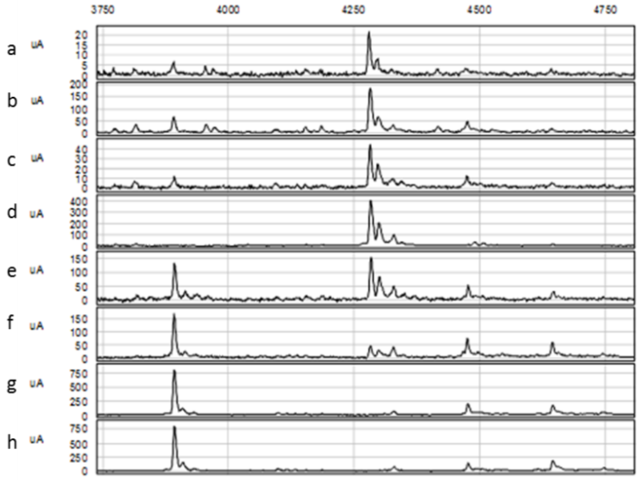


Figure 34. WCX buffer gradient MS optimisation analysis

CM10 ammonium acetate buffer gradient (pH 3.5-7.5) for U9 treated serum (1:10) pooled sample analysis optimisation. The following settings were used: LE 2,000nj, 848 shots, FM 5kDa, MR 0-100kDa, MA 2.5kDa. CM10 buffer with pH 4.5 (b) revealed the most peak numbers with consistent resolution and S/N. Albumin was detected in this fraction. Using other laser energies (LE) between 1,500-2000nj showed similar spectra with good quality (Table 6).

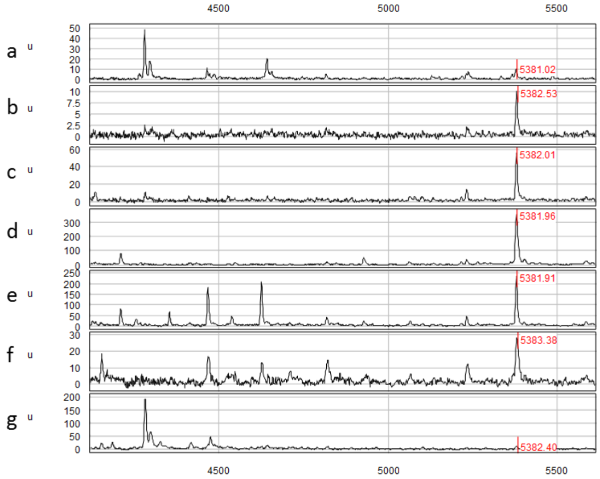


Figure 35. Overall optimisation effect in the Wessex an MS analysis

Illustration of the overall optimised fractionation profiles from pooled Wessex serum. F1-F6 and (1:10) U9 treated serum (a-g) were analysed using SELDI. The different peaks detected in different fractions are illustrated. Overall, LE 2,000nj, 848 shots were optimal for use with all fractions. FM 5kDa, MA 2.5kDa, MR 0-100kDa. CM10 buffer pH 4.5 was successfully applied in all preparations. Noticeably, a peak at 5.4kDa was detected in fractions F2-F6 (b-g) and probably represents an abundant protein peak.

3.2.3.8 The effect of albumin ultrafiltration and combinatorial peptide ligand library (CPLL) enrichment

The poor peak detection in fractions 3 and 4 was thought to be due to a LMW signal masking effect due to the high albumin concentrations in these fractions. Microcon centrifugal filters sieves are after used to concentrate and desalt samples. Here, this ultracentrifugation tool was used to eliminate high abundance proteins from samples prior to analysis. To improve LMW proteomic profiles from SAX fractions 3 and 4 using SELDI, albumin was eliminated using MW exclusion ultrafiltration columns with a size cut off of 50kDa (YM50, Millipore). The exclusion of albumin from the filtrates from both fractions 3 and 4 was successfully achieved (Figure 36). Using 50kDa cut off filters, the retentate had a high albumin signal compared to poor trace in the filtrates from fractions 3 and 4 respectively.

To assess the effect of albumin depletion on profiling, YM50 filtrates were analysed on WCX arrays. This showed no improvement in fractions 3 and 4 or on the LMW peak detection. However, there was a slight improvement in peak resolution in other fractions (Figure 37). The absence of LMW peaks in fractions 3 and 4 after albumin filtration may be due to co-binding of the LMW proteome to albumin, and possible preparatory sample loss.

As proteopeptide co-binding resulting in signal loss and variation due to additional pre-processing cannot be eliminated during subsequent profiling, this step was not used in the profiling analysis. Yet, this approach was considered as a potential aid for candidate biomarker purification.

As the dynamic range in serum proteomics extends over several orders of magnitude, a non-depleting protein enrichment approach was tested using the CPLL (Proteominer, Bio-Rad). This method applied multiple washing/eluting steps through a hexpeptide ligand library, resulting in a large number of combined peptides that specifically binds to ligands based on their distinct chemical properties. This aimed to fractionate biological samples without the loss of proteins. Here, work was done to illustrate the effect of CPLL enrichment on the profiles generated from WCX analysis. Although the number of peaks detected markedly increased in the CPLL treated samples, the resolution of these peaks remained sub-optimal (Figure 38).

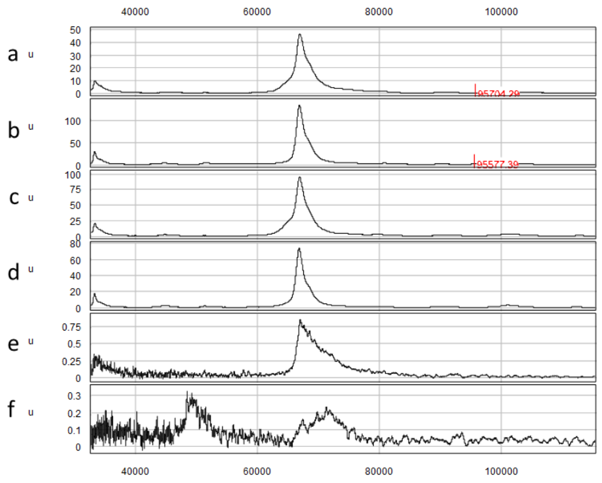


Figure 36. 50kDa (Microcon) size cut off filtration n effect on albumin profiles
The profiles represent duplicates of albumin rich F3, F4 (a, b respectively), their retentate (c, d respectively) and the filtrates (e, f respectively) from pooled Wessex serum samples. The following settings were used: LE 2,000nj, MA 5kDa, MR 0-200kDa, FM 12kDa. CM 10 arrays used with CM pH 4.5 buffer. The albumin depleting effect of the filtration columns was demonstrated in the filtrates showing trace albumin peak signals (e, f) which indicated effective albumin exclusion from samples. The risk of co binding and protein loss was however considered and this approach was not used in sample preparation preceding MS analysis.

The CPLL preparation however did not show an increased saturation effect, a phenomenon previously reported when using DTT (Figure 38). However, it was evident that using a lower CPLL enrichment step affected the equilibration and profiling of the LMW proteome and of some high abundance proteins. Therefore, this method was thus excluded from the main profiling study. Overall, this tool was thought to also be a useful adjunct in biomarker purification given the enhanced low abundance signal in a signal non saturating fashion.

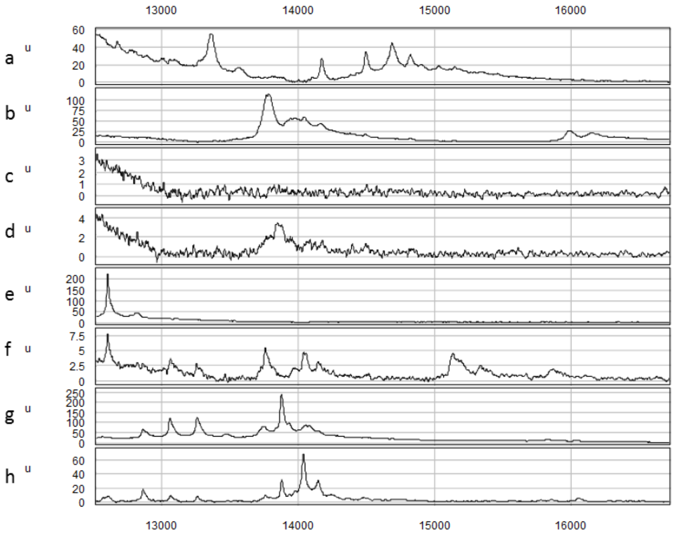


Figure 37. Ultrafiltration effect on serum fraction profiles

CM10 profiling of the Wessex cohort serum using pH 4.5 CM 10 buffer. LE 2,000 nj, FM 12kDa, MA 10kDa. Spectra from the following samples: a. F1 filtrate, b. F2 filtrate, c. F3 filtrate, d. F4 filtrate, e. F5 filtrate, f. F6 filtrate, g. Serum/U9 filtrate, h. CPLL (Proteominer, Bio Rad) filtrate. Despite effective albumin depletion illustrated by the improved signal in fractions (a,b) and (e-h), Fractions 3 and 4 (c, d) showed no improvement in the MS peak detection ability. This was tested applying different LE between 1,500-2,000 nj which also showed similar results. Although this was reproducibly shown to be an effective method for albumin depletion, the possibility of LMW protein co-binding and loss of differentiation signals, in addition to the freeze thaw cycle effect / limitation lead to the exclusion of filtration in the initial analysis.

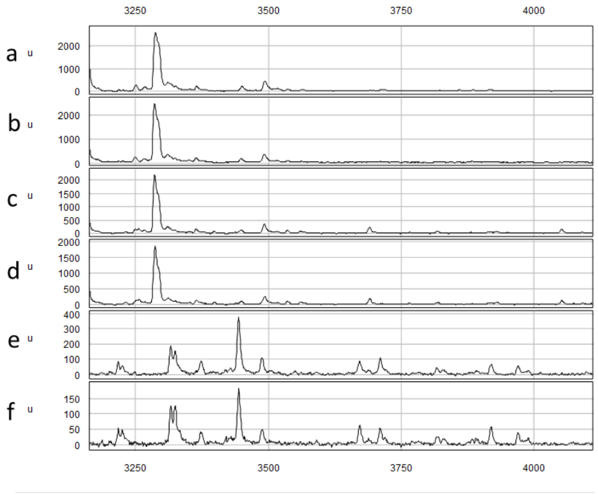


Figure 38. Combinatorial peptide ligand library (CPLL) enrichment effect on SELDI profiling

Improving the low abundance protein detection by combinatorial peptide ligand library (CPLL) enrichment (Proteominer, Bio Rad) was tested. Using pooled serum samples from the Wessex cohort, the following settings were adopted: LE 2,000nj, 848 shots, MA 2.5kDa, MR 0-100kDa, FM 5kDa. (a, b) Serum/U9, (c, d) Serum/U9 + 0.1M DTT, (e, f) Serum + Proteominer. The spectra showed high reproducibility. A signal enhancing effect of DTT was noticed with few extra peaks detected in these profiles (c,d). Enriching the samples with the CPLL method illustrated more peaks in the LMW region (e,f), nonetheless, the peak signal was less optimal than the original profiles (a,b).

3.2.3.9 The Wessex cohort SELDI profiling

Based on the optimal conditions revealed by the extensive optimisation process, samples were fractionated using (SAX) ion exchange chromatography. In parallel, paired samples were denatured with U9 buffer (9M Urea, 2% CHAPS, 50mM Tris HCl, Bio-Rad). Experiments were designed to identify the chromatographic arrays, fractions and conditions with the most informative mass spectra (Table 6), to be used in this analysis. Serum protein concentrations ranged between 60 – 520ng/μl. This was consistent across all fractions and was measured using a Qubit™ fluorometer and the Quant-IT™ Protein assay kit, and confirmed by Nanodrop and a modified Bradford assay. Generally, Fraction 2 (F2) showed very sparse peaks on all chromatographic conditions. F3 and F4 samples showed high albumin concentration demonstrated by very high peak intensity in these fractions. F5 fractions had a small number of peaks, probably due to the masking effect of relatively high albumin contents. Such high concentrations of HMW molecules would potentially mask low abundance and LMW signals of interest in biomarker discovery.

In this analysis, standardised and highly optimised preparation protocols were used to profile fraction F1, master fraction F5/6 (Figure 39) and U9 treated samples from 347 serum samples (210 healthy, 73 breast cancer and 64 benign breast disease) by SELDI (Figures 40-45). The clinicopathological characterestics of the patients are shown below (Table 7). This large cohort was essential to ensure adequate study power and to alleviate outlier effects.

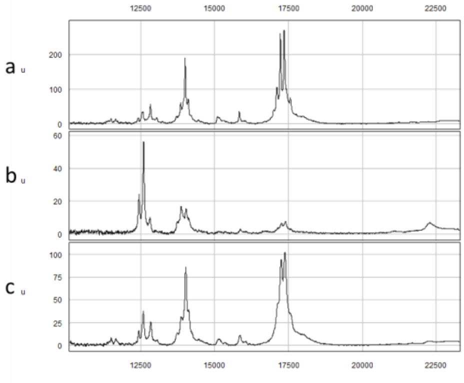


Figure 39. The Wessex cohort fractions 5 and 6 pooling effect on SELDI profiling

Figure illustrating the effect of pooling the two fractions F5 and F6 into one master fraction. (a) F5, (b) F6, (c) F5/F6. CM10 buffer pH 4.5 was used, combined with the following settings: LE 2,000nj, 848 shots, FM 5kDa, MR 0-100kDa, MA 2.5kDa. Pooling effect on fractions 5 and 6 (F5/6) was tested as the Wessex serum F5 yield was comparatively low. Fractions F5 and F6 revealed moderate number of peaks using CM10 arrays (Table 4). Pooling both fractions into F5/F6 master fraction for analysis, more proteomic signals were generated, prior to further analysis of individual fractions. Although this can mask potentially relevant signals, it was used as an initial profiling step which can be followed by individual F5 and/or F6 analysis.

Clinical and Pathologic Features of the Wessex cohort serum samples

Characteristics	**Percentage**
Diagnostic category	
Control	210 (60.5%)
BBD	64 (18.4%)
IDC	73 (21%)
Age median [range] (SD)	63.6 [50.8-75.1] (7.33)
Menopausal status	
Pre-menopause	62 (18%)
Peri-menopause	134 (38.6 %)
Post-menopause	139 (40.1%)
Unknown	12 (3.5%)
Tumour type	
Microcalcifications	7 (2%)
Duct adenosis	4 (1.2%)
Fibroadenoma	14 (4.0%)
Fibrocystic changes	11 (3.2%)
Typical hyperplasia	3 (0.9%)
Sclerosing adenosis	2 (0.6%)
Cyst	7 (2%)
ADH	0 (0%)
Others	8 (2.3%)
DCIS element	17 (4.9%)
ER negative	17 (4.9%)
ER Positive	36 (10.4%)
IDC	
IDC	44 (12.7%)
ILC	0 (0%)
Mixed	29 (8.4%)
Grade*	
Low	102 (29.4%)
Intermediate	35 (10.1%)
High	33 (9.5%)
ER	
Positive	59 (17.0%)
Negative	4 (1.2%)
Unknown	10 (2.9%)
PR	
Positive	24 (6.9%)
Negative	4 (1.2%)
HER2	
Positive	32 (9.2%)
Negative	23 (6.6%)
Lymph node involvement	
Negative	26 (7.5%)
Positive	16 (4.6%)

Table 7. The clinicopathological characteristics of the Wessex breast cancer cohort

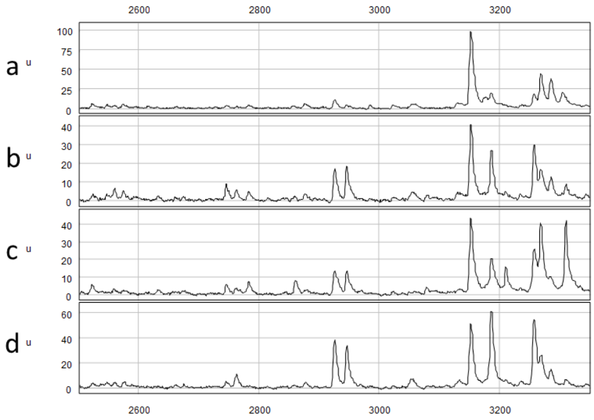


Figure 40. Illustration of the MS spectra following optimised WCX profiling
Representative Wessex F1 profiles following optimisation of CM 10 analysis conditions using CM10 buffer pH 4.5, and low mass range detection settings: LE 3,000nj, 742 shots, MA 1kDa, MR 0-50kDa and FM 3.5kDa. All samples were analysed in duplicate. LMW detection power achieved by MS profiling demonstrated by protein signals below 3kDa. (a and b) Cancer,
(c and d) Control.

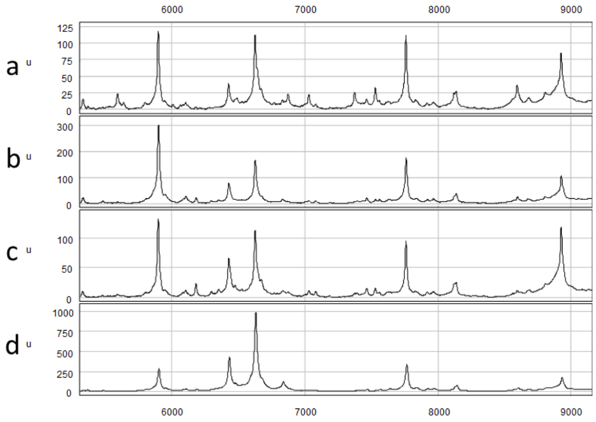


Figure 41. Representative Wessex F1 profiles following WCX array optimisation
Using CM10 buffer pH 4.5, and high mass range detection settings: LE 6,000 nj, 759 shots, MA 5kDa, MR 6-200kDa and FM 25kDa. All samples were analysed in duplicate. (a and b) Cancer, (c and d) Control.

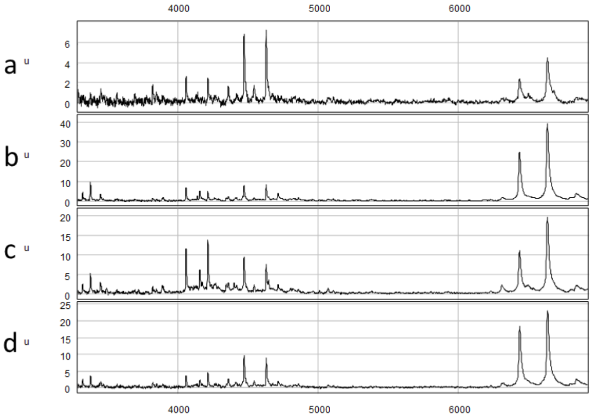


Figure 42. Representative Wessex F5/F6 profiles following optimisation

Representative Wessex F5/F6 profiles following optimisation of CM 10 analysis conditions using CM10 buffer pH 4.5, and low mass range detection settings: LE 3,000nj, 742 shots, MA 1kDa, MR 0-50kDa and FM 3.5kDa. All samples were analysed in duplicate. Low detection power achieved by MS profiling demonstrated by protein masses below 4kDa detection.

(a and b) Cancer, (c and d) Control.

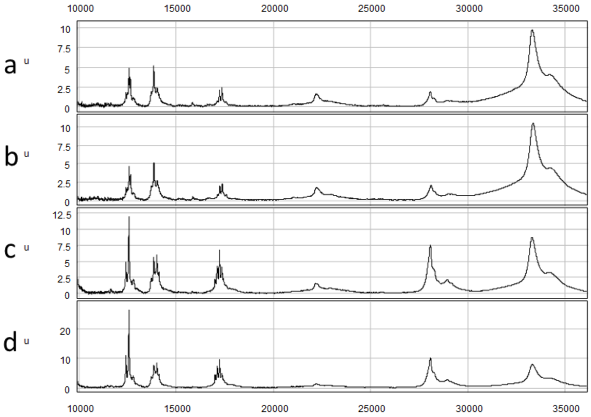


Figure 43. Representative Wessex F5/F6 profiles

Following optimisation of CM 10 analysis conditions using CM10 buffer pH 4.5, and high mass range detection settings: LE 6,000nj, 759 shots, MA 5kDa, MR 6-200kDa and FM 25kDa. All F5/F6 samples were analysed in duplicate. (a and b) Cancer, (c and d) Control.

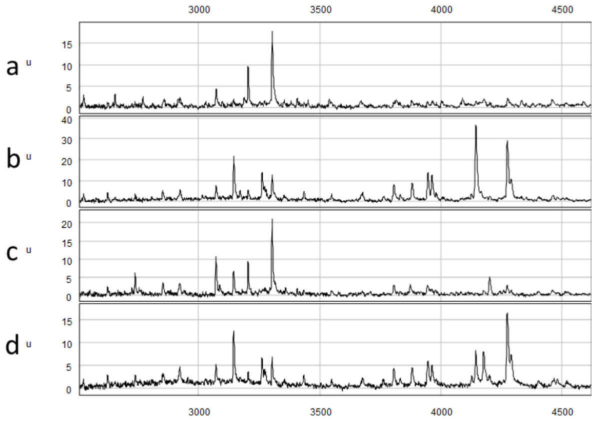


Figure 44. Representative Wessex Serum/U9 profiles during pre-analytical optimisation
CM 10 analysis of serum/U9 samples from the Wessex cohort. Optimal conditions were obtained using CM10 buffer pH 4.5, and the low mass range detection settings: LE 3,000nj, 742 shots, MA 1kDa, MR 0-50kDa and FM 3.5kDa. All samples were analysed in duplicate. Low detection power achieved by MS profiling demonstrated by protein masses below 3kDa detection. (a and b) Cancer, (c and d) Control.

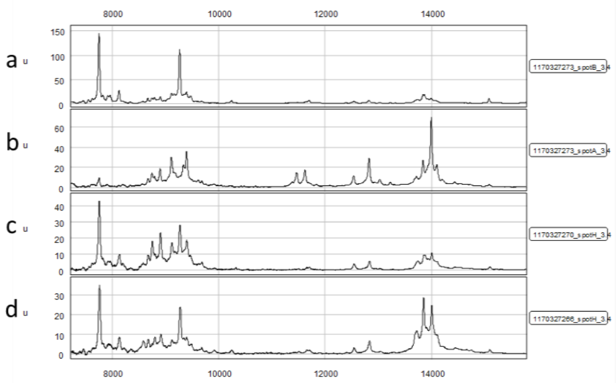


Figure 45. Wessex Serum/U9 profiles following optimisation of CM 10 conditions
Analysis conditions using CM10 buffer pH 4.5, and the high mass range detection settings: LE 6,000nj, 636 shots, MA 5kDa, MR 6-200kDa and FM 25kDa. All samples were analysed in duplicates. (a and b) Cancer, (c and d) Control.

The weak cation exchange arrays (CM10) analysis gave reproducible and complex spectra of all chip surfaces, as determined by number of peaks detected and range of peak intensity CVs (Table 6). To determine the reproducibility and batch effect, sera were run looking for statistical differences in peak intensities on days 0 and 7 of analysis (Figure 46).

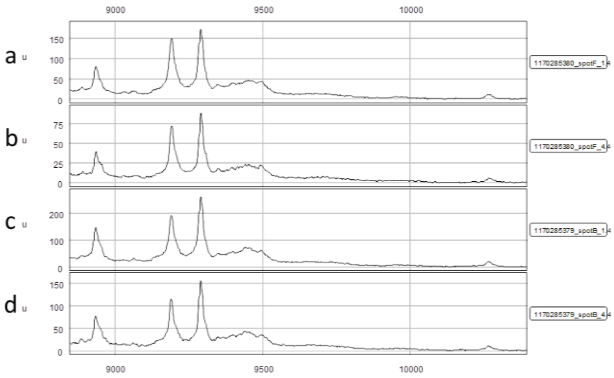


Figure 46. Wessex cohort pooled serum sample reproducibility assessment
Optimised conditions using CM 10 (WCX) arrays, CM10 buffer pH 4.5, LE 1,750nj, 848 shots, FM 6kDa, MA 2.5kDa, MR 0-200kDa.. The average intra array CV on day 0 was 19.6% compared to an average inter assay CV of 22.7%. The average intra array CV on day 7 was 22.6% compared to an average inter assay CV of 25.3%. All profiles were calibrated and normalised to TIC prior to CV calculation. a. Sample analysed on day 0 (array 1), b. Sample prepared on day 7 (array 1), c. Sample prepared on day 0 (array 2), e. Sample prepared on day 7 (array 2).

A total of 27 markers were identified and validated in house using an independent set and identical protocols. Nine markers were differentially expressed between the cancer and healthy group, eleven between the cancer and benign disease group and seven discriminated the benign disease group from healthy controls (Table 8).

Candidate biomarker (kDa) Wessex cohort **Cancer *vs.* Healthy**	**Candidate biomarker (kDa) Wessex cohort** **Cancer *vs.* Benign**	**Candidate biomarker (kDa) Wessex cohort** **Benign *vs.* Healthy**	**Candidate biomarker (kDa) Guernsey cohort** **Cancer *vs.* Healthy**
🟩 6.4 ↑	🟨 2.7 ↓	🟦 6.6 ↑	🟩 3.0 ↑
🟩 6.9 ↓	🟦 2.9 ↓	🟨 6.8 ↑	🟨 3.4 ↑
🟨 9.5 ↑	🟨 4.2 ↓	🟨 8.2 ↑	🟨 3.8 ↓
🟦 12.8 ↑	🟩 6.6 ↑	🟦 12.6 ↓	🟩 6.3 ↓
🟦 13.0 ↑	🟦 6.8 ↑	🟦 13.3 ↑	🟩 6.8 ↑
🟦 66.9 ↑	🟨 8.6 ↓	🟦 15.9 ↑	🟦 8.3 ↑
🟨 90.2 ↑	🟨 9.3 ↑	🟩 149.3 ↑	🟦 8.6 ↑
🟩 101.2 ↑	🟩 12.8 ↑		🟩 8.7 ↓
🟦 134.3 ↑	🟨 13.1 ↑		🟨 10.0 ↓
	🟦 13.3 ↑		🟨 10.9 ↑
	🟩 168.3 ↑		🟦 12.7 ↑
			🟦 13.9 ↓
			🟦 17.4 ↓
			🟩 28.3 ↓
			🟨 28.9 ↑
			🟩 56.0 ↓

Table 8. Candidate proteomic breast cancer serum markers detected by MS analysis. In the Wessex and Guernsey cohort analyses.

PCDM bioinformatic software and non- parametric statistical analysis were used. In the Wessex cohort; nine markers were differentially expressed between the cancer and healthy group, eleven between the cancer and benign disease group and seven discriminated the benign disease group from healthy controls. The SAX fractions from which each marker was identified are indicated by colour coding as follows: Serum/U9 in green, F1 in yellow and F5/F6 in blue. Expression of the differentially expressed candidate biomarkers between groups is designated ↑ and↓ respectively indicating over and under expression ($P<0.05$) in the first comparison group.

3.2.3.10 MALDI-TOF MS biomarker validation

In a confirmatory step to focus on the most reliable and potential “real world” biomarkers, I conducted a validation step in the laboratories of our collaborators in the Medical Biomics Centre at St. George’s University of London (SGUL). This was undertaken using MALDI TOF MS as an alternative top down profiling platform to minimise false discovery rates, select the most stable biological alteration which would facilitates a more clinically valid biomarker, and to stratify markers in terms of a higher success rate for future validation.

Here, 12 months after the original analysis, identical robotic preparation of WCX protein arrays using aliquots from the same samples was performed. This was followed by MS analysis using a different MS platform; MALDI-TOF MS, (Bruker Daltonics, Bremen, Germany; detailed methods described earlier). This platform has inherently a higher resolution and mass accuracy than the SELDI study platform. The statistical analysis was performed by an independent researcher using the Lucid System Software (Bio-Rad, Hercules, USA).

Among the identified biomarkers previously described (Table 8), three ion peak markers qualified as discriminatory markers (*P*<0.05) between the breast disease groups in the MALSI analysis. These were the ion peaks 6.4, 9.5kDa (showing higher levels in the cancer group compared to controls), and the peak at 15.9kDa which was over expressed in the benign disease group compared to controls (Figure 47).

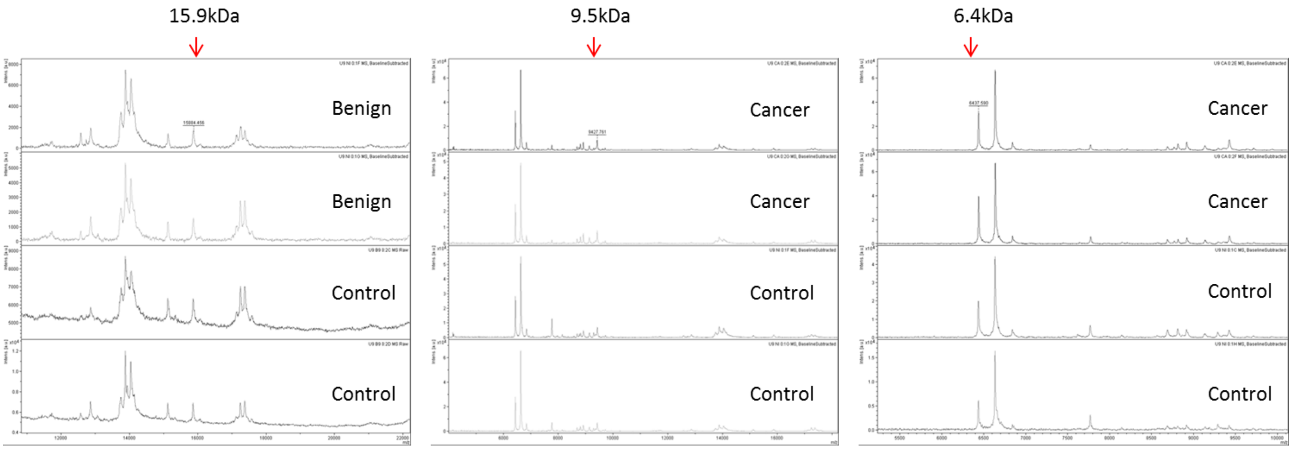


Figure 47. MALDI validated markers from the Wessex cohort

Illustrative MALDI spectra representing the three validated serum markers confirmed by an independent analysis of the Wessex samples in SGUL. These breast cancer serum markers qualifying for further immunovalidation included: A 15.9kDa marker which was over-expressed in benign breast disease compared to controls. On the other hand, a 9.5kDa marker was over-expressed in the cancer group compared to healthy controls. Finally a 6.4kDa was over-expressed in the cancer compared to the control group.

3.2.3.11 WCX spin column preparation and gel electrophoresis purification

The methodology adapted here was described earlier. Briefly, experiments were performed to mimic the original sample preparation condition using WCX columns (Bio-Rad). Following this; the eluents (in Tris HCl pH 10.5) were electrophoresed using hand cast SDS PAGE (4% stacking and 10% resolving gel). Samples were pre filtered using YM50 microcon columns (Millipore) to eliminate albumin. Following this, samples were vacuum dried and then re-suspended in Laemmli sample buffer prior to gel runs. Gel running buffer was used and electrophoresis performed at 120V. Repeatedly, this was found to be of low resolution; inadequate for proteomic bands below 16kDa (Figure 48).

To increase the gel resolution at the LMW range, a precast gradient gel (4-15% Tris-Glycine gels) (Bio-Rad) was used to analyse the same samples. This showed slight improvement in the number of bands detected; yet, few bands were seen below 16kDa (figure 49). This work was repeated several times and showed similar results indicating a possible sample quality and/or technical problem.

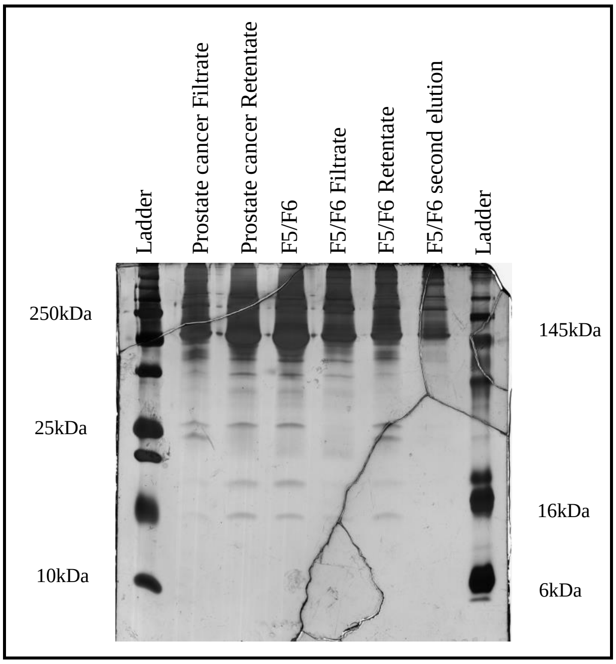


Figure 48. SDS PAGE purification of the Wessex serum samples

An SDS gel (4% stacking-10% resolving) was used; samples were prepared on WCX columns (WCX). A size filtration step using 50kDa cut off ultrafiltration was used. Filtrate and retentate samples were then loaded. 50μl sample and 5 μl ladder was loaded in each corresponding well. Staining was performed using Biosafe Coomassie (Bio-Rad). No bands were evident below 16kDa.

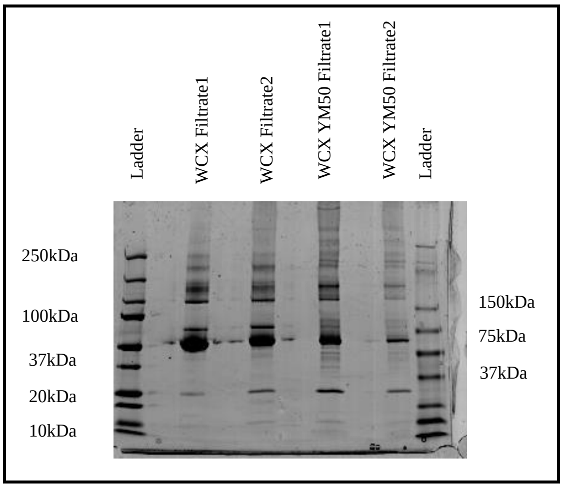


Figure 49. Gradient Tris-Glycine (TG) GE of the Wessex serum samples
A 4-15% gradient TG gel was used. Samples were prepared on WCX columns (WCX) and eluted twice (Lanes 3 and 5) with 50mM, pH 10.5 tris HCL buffer. An extra size filtration step using 50kDa cut off was also used and samples were then eluted twice (lanes 7 and 9). 15μl sample and 5 μl ladder was loaded in each corresponding well. Staining was performed using Biosafe coomassie (Bio-Rad). Reproducible but weaker bands were demonstrated in the second elution samples (Filtrate2). Noted also was the weak signal below 16kDa.

To investigate this further, Tris-Tricine gels were used comparing different LMW standards and staining methods (Coomassie *vs.* silver staining) (Figure 50). LMW resolution was achieved using the Tris-Tricine gels and silver staining, where protein standards were detected at 1.7kDa. These gels and standards were utilised to analyse pooled clinical samples from the Wessex cohort. Nonetheless, this showed very poor resolution of the clinical serum samples compared to the standard ladder (Figure 51). This was attributed to sample decay (samples were stored in -20°C for one week) or protein loss during cut off filtration. To exclude poor sample quality, freshly prepared samples using the same preparation conditions were tested. This however has continued to show poor detection at the LMW range (Figure 52). The presence of large number of proteins below the 50kDa cut off in the retentate samples was observed so tests were performed using only the WCX column sample preparation step. In an attempt to increase the sample concentration, samples were compared between vacuum dried and none dried (to a minimum volume of 20μl). In addition, filtrate from the WCX column preparations were tested for biomarker signals (Figure 53). As expected, drying the samples was shown to improve the band density; nonetheless, the band resolution remained poor at the LMW region. The WCX column filtrate was also shown to have signals at around 10kDa level. The WCX filtrate proteins should theoretically contain no proteins of interest as they favourably bind the resins during the preparation process and therefore would not have been present in the original SELDI experiment. This could indicate sub optimal fractionation using this method.

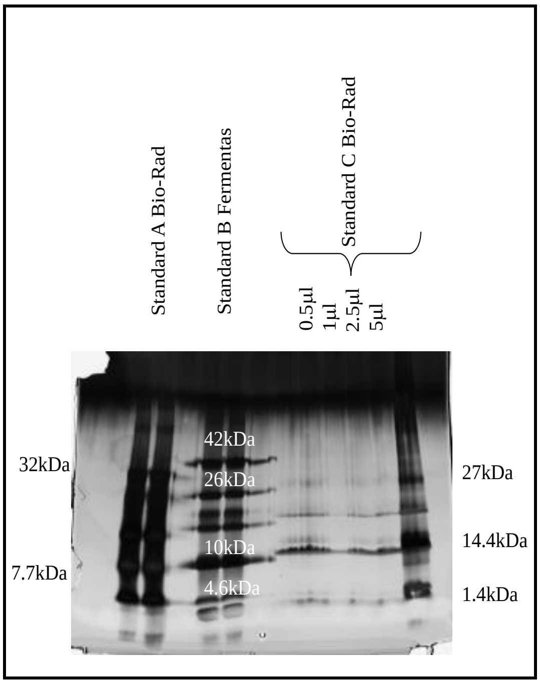


Figure 50. Tris trycine gel biomarker purification test

Tris tricine (10-20%) (Bio-Rad) standard gradient gels were used and then silver stained and developed for 6 minutes. Duplicates of two different protein standards (A and B) and a concentration gradient of Standard C were resolved. Unless stated otherwise, 10µl was loaded in each lane. Effective staining using silver staining was shown. Poor resolution was noted when using standard A.

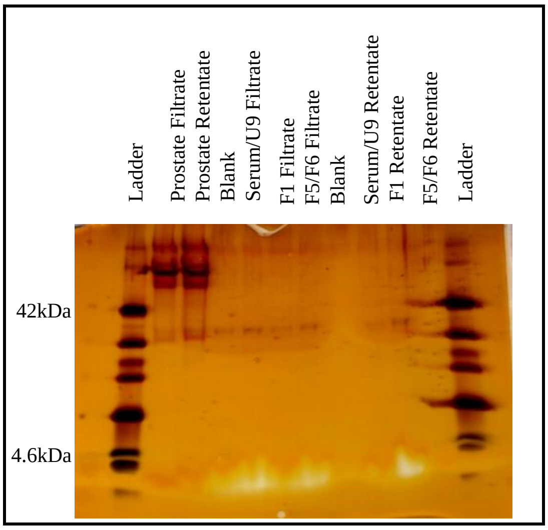


Figure 51. Tris Tricine gel biomarker purification optimisation

WCX column prepared Wessex serum samples were subjected to size filtration (YM50 Millipore) and then analysed on a 10-20% Tris-Tricine gel (Bio-Rad). The gel was silver stained (development for 6 minutes). Serum, F1, F5 /F6 correspond to the Wessex cohort samples. Very poor resolution of the clinical samples was observed compared to the high resolution seen with the LMW standards. Sample protein concentrations were measured and ranged between 100-420ng/µl.

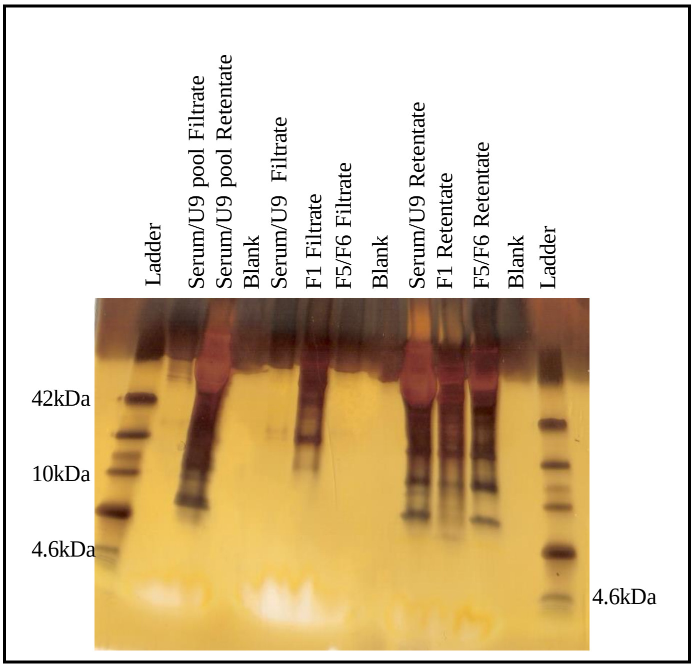


Figure 52. Effect of sample concentration and size exclusion on biomarker purification
Serum samples prepared with WCX columns (Bio-Rad), YM 50 columns (Millipore) and vacuum dried on room temperature. The gel was silver stained and developed for 6 minutes. Despite using fresh serum samples, LMW detection was still very poor, where no bands below 10kDa were reported.

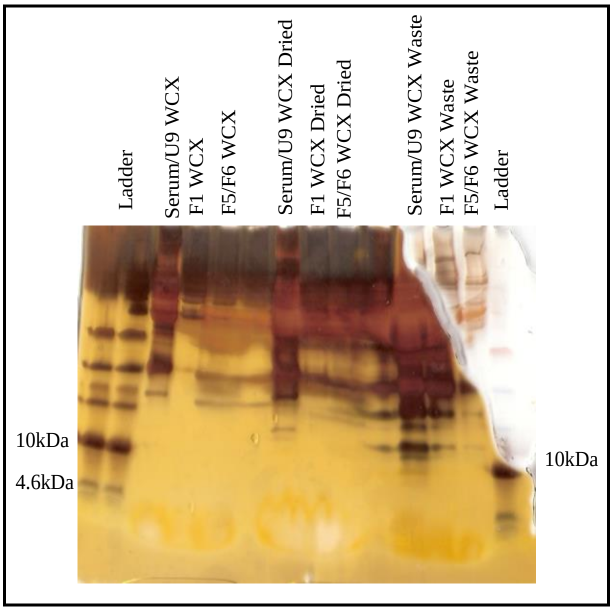


Figure 53. Drying/concentration effect on gel based marker purification
Wessex serum samples were treated with WCX columns (Bio-Rad) with and without vacuum drying at room temperature. Waste from the WCX preparation was also assessed for presence of potential LMW proteins. The gel was silver stained and developed for 6 minutes. The LMW detection was still very poor, where no bands below 10kDa were detected. Drying the samples enhanced band intensities, and there were few signals around 10kDa from the waste samples indicating possible sample loss during sample preparation.

To assess the sample quality as a confounding factor in repeatedly inadequate visualisation of LMW bands, a recombinant standard peptide sample (2.3kDa) (a kind gift from Professor G Packham's laboratories), sonicated cell lysate, and BSA concentration gradient samples were tested under identical analysis preparation conditions. The detection of ladder bands down to 1.7kDa, but a failure to see the recombinant protein at 2.3kDa and cell lysate proteins below 10kDa remained a problem (Figure 54). Silver staining with up to 30 minutes development was undertaken with no further band detection (data not shown).

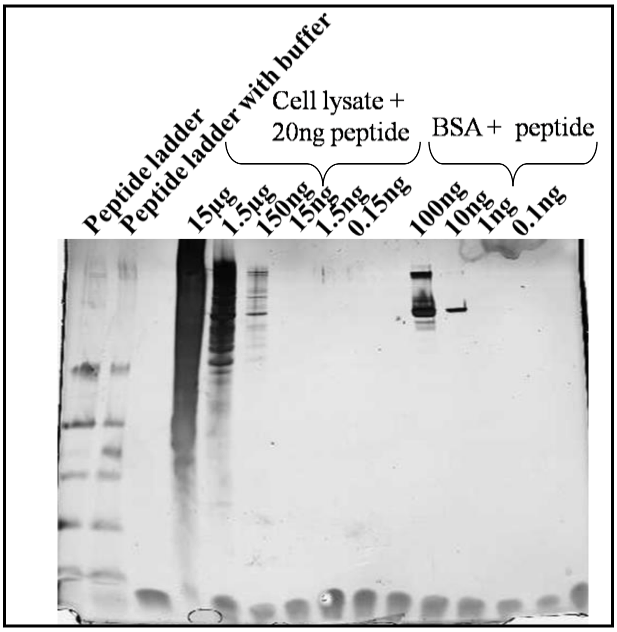


Figure 54. Illustration of poor resolution of gel based biomarker purification at the LMW range
Silver stained Tris-Tricine (10-20%) (Bio-Rad) GE assessing the LMW detection of the Wessex samples compared to standard proteins and cell lystaes. Same ladder in figure 52 used. This showed very good detection of the LMW of the standard peptide ladder with and without sample buffer. BSA was detected in addition to cell lysate proteins. Yet, LMW detection of a recombinant protein (2.3kDa) and cell lysates remained unachievable. This has confirmed failure of gel based methods in detecting proteins of interest below 10kDa.

The absence of LMW protein/peptides from biological samples, contrary to standard ladder peptides, suggested possible preparation error. To exclude loss of LMW proteins from the gel during the wash steps, an extra fixation step (using 30% methanol and 10% acetic acid solution) was performed. Despite 24 hour fixation and silver staining developing steps for 30 minutes, no improvement of the detection was noted (figure 55). Finally, the possibility of sample run off at the end of the gel was explored. Electrophoresis was performed at very low voltage and was stopped when the dye front reached the middle of the gel. However, this was not the cause of the LMW band misdetection (figure 56).

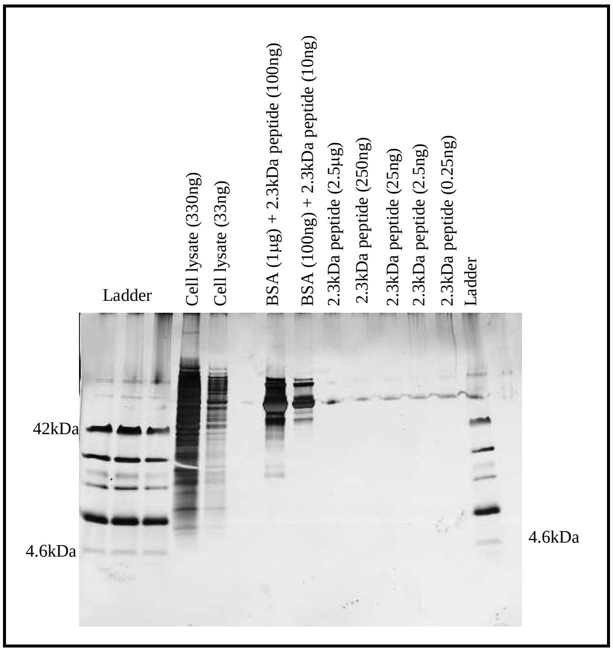


Figure 55. Testing LMW gel band loss in biomarker purification
The same samples used in Figure 54 were used here to exclude LMW protein permeation during gel washes. Tris-Tricine gel (10-20%) (Bio-Rad) was fixed overnight with 30% methanol and 10% acetic acid solution. The gel was then silver stained and developed for 8 minutes. No improvement in band detection below 10kDa was observed.

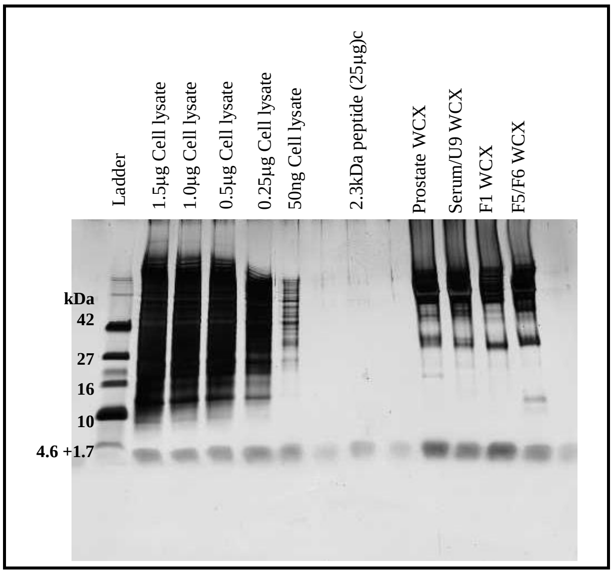


Figure 56. Assessment of experimental errors in gel based purification
The same samples used in figure 55, in addition to the Wessex and a parallel prostate study samples were used to exclude LMW protein run off at the gel end. The Tris tricine (10-20%) (Bio-Rad) was also fixed overnight with 30% methanol and 10% acetic acid solution. Gels were run at 80V current and then silver stained and developed for 15 minutes. No bands were detected in the LMW range.

3.2.3.12 Gel free biomarker purification and identification

Free Flow Electrophoresis (FFE) fractionation

Following unsuccessful gel based biomarker purification, a gel free based purification and identification method was investigated. This work adopted pre-fractionation using IEF-FFE, which previously has previously been shown to provide high-resolution fractionation in accordance with minimal sample loss[218, 219]. In addition, solid phase 2D chromatography was tested.

Enrichment and purification of candidate biomarkers is an essential step prior to identification. Although reported as an efficient purification adjunct, common depletion approaches frequently lack specificity and may remove potentially important proteins of interest[69]. In contrast, FFE has the potential to preserve all sample constituents. Here, samples were subjected to carrier-free IEF on an FFE instrument (BD). This technique combines IEF with the benefits of gel free electrophoresis by showing improved sample recovery due to absence of carrier material and enhanced sample loading capacity through continuous sample flow. SAX fractionated serum samples from the pooled stock were equilibrated with equal amounts of separation media. Samples were loaded with a continuous flow of 0.8 ml/h on a linear pH-gradient covering pH 3-10 and 96 fractions were collected. To make further analysis more feasible, four successive fractions at a time were combined generating twenty four master fractions (Appendix 1). MS showed that these had displayed well resolved protein signals across the entire pH-range. In comparison to the native serum, fractionation of serum by IEF-FFE allowed the detection of an increased number of LMW protein peaks (Figure 57 and 58). Albumin, the most prominent protein, normally masking proteins of lower abundance, was primarily present in fractions SAX F5/F6 - FFE 13 and SAX F5/F6 - FFE 14. Consequently, the remaining fractions were almost entirely free of albumin, which helped to overcome signal suppression by albumin and enabled visualization of lower abundance candidate ion peaks of interest. These were shown to be concentrated in FFE master fractions SAX F5/F6 - FFE 5, SAX F1 - FFE 8, SAX F5/F6 - FFE 8, SAX F5/F6 - FFE 9, SAX F1 - FFE 11 and SAX F1 - FFE 22 (Table 9).

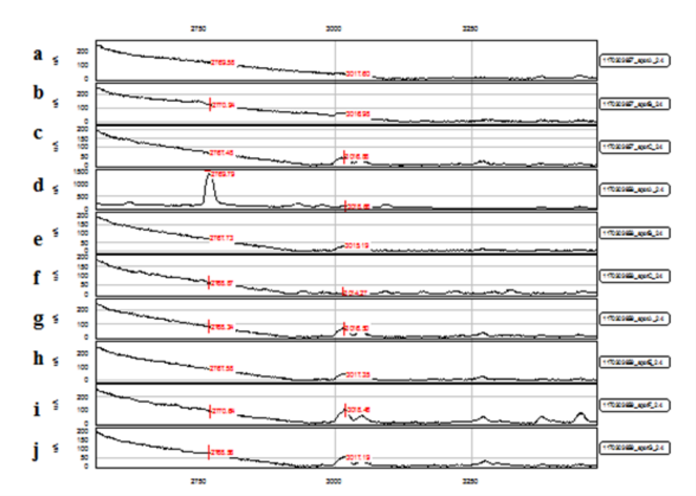


Figure 57. FFE based biomarker purification (2.7kDa) illustration
Representative spectra from the Wessex cohort pooled serum fractionated using IEF- FFE and prepared with WCX (columns or on spot retentate harvesting). Master fractions 1-3 (a-c), 9-11 (d-f), 17 (g), 21-23 (h-j) shown respectively. A 2.7kDa protein peak was purified in master fraction 9 (d) which represents an optimal identification medium for this marker.

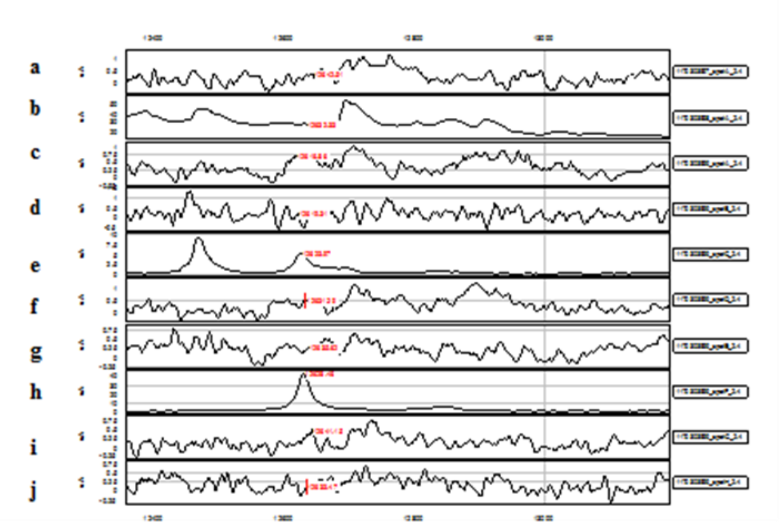


Figure 58. FFE based biomarker purification (12.6kDa) illustration
Spectra from pooled Wessex cohort serum fractionated using IEF- FFE and prepared with WCX (columns or on spot retentate harvesting). Shown are master fraction pools 1 (a), 9 (b), 17-24 (c-j) respectively. Proteomic peak at 12.6kDa representing a candidate biomarker was isolated in master fraction 22 (h).

Following IEF-FFE characterisation of samples, the fractions containing the peaks of interest were divided into three aliquots. One aliquot was trypsin digested in solution. The second aliquot was applied to a WCX arrays and retentates harvested and trypsin digested in solution. Both digested samples were then analysed by tandem mass using MALDI-TOF/TOF and/or ESI LC MS/MS. This decoded several potential candidate proteins pending validation (Table 9). The third FFE aliquot was analysed by SDS PAGE (4-15%) and Tris-Tricine (10-20%) (Bio-Rad), which failed to detect most corresponding bands (data not shown). This could be explained by the low detection power of gel based purification at the LMW range exacerbated by a dilutional effect from further FFE fractionation.

SELDI m/*z*	**SAX-FFE fraction**	**Protein/Peptide ID**	**Protein function**
8.6kDa	SAX F1 - FFE 8	Beta-defensin (D108B)	Poorly explored. Possible Immune/tumour suppressor function.
8.6kDa	SAX F1 - FFE 11	Ca/calmodulin-dependentprotein kinase II Inhibitor CK2N1	Cell cycle progression.
8.6kDa	SAX F1 - FFE 22	Apolipoprotein C3	Plasma lipid transport.
13.3kDa	SAX F5/F6 - FFE 5	Transthyretin (TTR)	Transport protein for retinol-binding protein and thyroxin.
15.9kDa	SAX F5/F6- FFE 8	Annexin A3 (ANX A3)	Vesicle trafficking, calcium signalling, apoptosis, cell growth and division.
15.9kDa	SAX F5/F6 - FFE 9	Osteopontin (OPN)	Adhesion, migration, invasion, chemotaxis and cell survival.
9.5kDa	SAX F1 - FFE 11	Apolipoprotein C1	Plasma lipid transport.

Table 9. potentially identified markers from the Wessex cohort
A list of potentially identified markers from the Wessex cohort. The SAX and FFE fractions from which each marker was purified are illustrated. The confirmation of the identity of these markers was subject to immunoprecipitation (IP) verification using the original SELDI platform. IP verified markers would then be re-validated by immuno assays (ELISA/WB) on an independent cohort of serum samples and potential breast cancer pathways involved would be subsequently targeted.

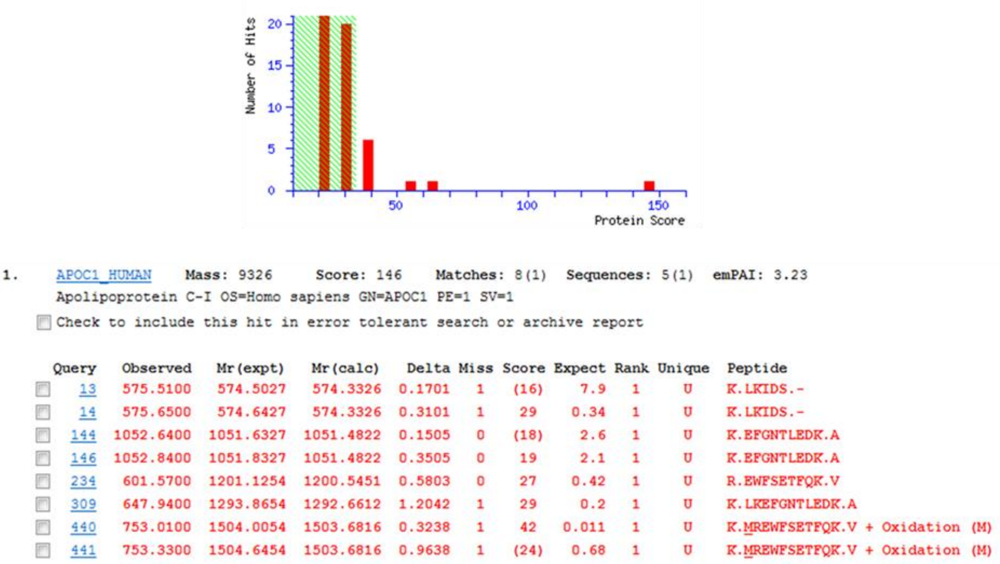
3.2.3.13 Biomarker verification by immunoprecipitation-MS (IP- MS)

Following the MALDI-TOF MS validation results, markers identified at 15.9, 9.5 and 6.4kDa were subjected to further evaluation to confirm their identities and correlation to the original MS spectra. The identities of two of the three ion peaks of interest were assigned as apolipoprotein C1 (Apo C1) at 9.5kDa and either annexin A3 (ANX A3) or osteopontin (OPN) at 15.9kDa. The last marker at 6.4kDa is still under investigation pending its *bona fide* identification. Briefly, IP-MS tests involved the re-analysis of fractions of the samples containing the marker of interest, MS analysis of the same sample after marker immuno-precipitation (IP), and finally MS analysis of the IP enriched eluate to confirm the presence of the same size marker.

Firstly, the 9.5kDa marker identified by MS/MS (Figure 59) was subjected to IP-MS verification. The depletion was performed using μMACs columns (Miltenyi Biotech) and a mouse monoclonal antibody (ab54800, Abcam) as described in the methods section. The MS spectra confirmed the identity of this marker as ApoC1 as the original MS signal was significantly depressed and strongly expressed in the depleted and enriched samples respectively (Figure 60). Interestingly, the 9.5kDa signal detected here represents the molecular weight of Apo C1 precursor. Apo C1 is normally truncated following secretion to a 6.6kDa molecule [220]. This could indicate that Apo C1 differential expression in this occasion is related to cell death rather than over secretion.

Similarly, OPN verification by IP-MS was performed using a monoclonal anti OPN antibody (Abcam, ab69498). Here, the IP failed to eliminate the 15.9kDa band from the original sample, and no enhancement was seen in the theoretically OPN IP enriched eluate (Figure 61). These findings indicated that there was either gel band/FFE identification SELDI peak mismatch; which would exclude OPN as a potential marker in this analysis set, or an antibody failure. An antibody peptide motif mismatch was also a possible explanation of this result, but investigation of this was not possible at this stage.

To exclude antibody related failure, western blotting of recombinant OPN protein was performed, which reproducibly revealed good bands at the expected detection size (Figure 62). This suggests the antibody was functional. Although other OPN antibodies covering different epitope(s) could have been tested to confirm this negative IP-MS result, this was not the primary aim of this work and was decided to be an alternative option should the other potential marker at 15.9kDa fail the validation step. Thus, OPN was excluded from further validation work.



1. APOC1_HUMAN Mass: 9326 Score: 146 Matches: 8(1) Sequences: 5(1) emPAI: 3.23
Apolipoprotein C-I OS=Homo sapiens GN=APOC1 PE=1 SV=1
☐ Check to include this hit in error tolerant search or archive report

	Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
☐	13	575.5100	574.5027	574.3326	0.1701	1	(16)	7.9	1	U	K.LKIDS.-
☐	14	575.6500	574.6427	574.3326	0.3101	1	29	0.34	1	U	K.LKIDS.-
☐	144	1052.6400	1051.6327	1051.4822	0.1505	0	(18)	2.6	1	U	K.EFGNTLEDK.A
☐	146	1052.8400	1051.8327	1051.4822	0.3505	0	19	2.1	1	U	K.EFGNTLEDK.A
☐	234	601.5700	1201.1254	1200.5451	0.5803	0	27	0.42	1	U	R.EWFSETFQK.V
☐	309	647.9400	1293.8654	1292.6612	1.2042	1	29	0.2	1	U	K.LKEFGNTLEDK.A
☐	440	753.0100	1504.0054	1503.6816	0.3238	1	42	0.011	1	U	K.MREWFSETFQK.V + Oxidation (M)
☐	441	753.3300	1504.6454	1503.6816	0.9638	1	(24)	0.68	1	U	K.MREWFSETFQK.V + Oxidation (M)

Figure 59. MS/MS identification of Apo C1

Tryptic peptides identified by product ion MALDI-TOF TOF analysis of the 9.5kDa marker. Tandem mass sequence coverage of Apo C1 with a high identification score (>100) indicated very high identification confidence.

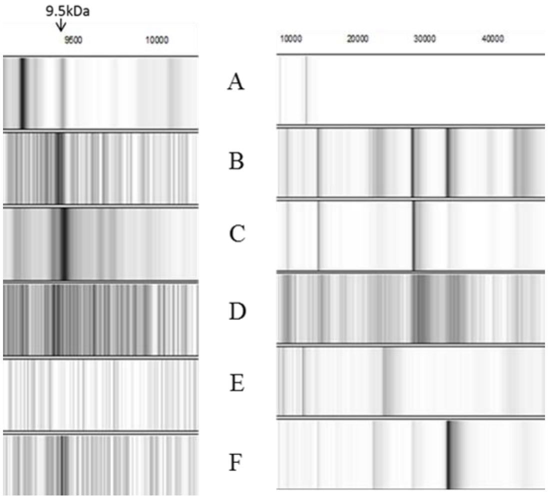


Figure 60. IP-MS verification of Apo C1 as a potential breast cancer biomarker

IP-MS verification of the 9.5kDa (red arrow) marker identified as ApoC1. Antibody capture of ApoC1 was performed using monoclonal antibody (ab54800, Abcam). The antibody was linked to μMACS protein A/G columns (Miltenyi Biotecs) and incubated with pooled serum fractions containing ApoC1. Shown are original SELDI pseudo-gel views of ApoC1 serum fraction (*A*) , along with mass spectra of the captured protein from depleted specimen (*B*) and the enriched ApoC1 IP eluate (*C*), respectively. Negative controls illustrated as follows: ApoC1 abtibody + beads (*D*), Glycine and PBST (*E*), anti ApoC1 antibody (*F*).

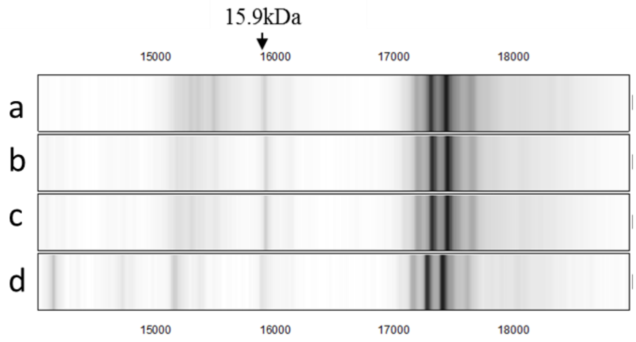


Figure 61. Exclusion of OPN as a breast cancer marker
Osteopontin IP-MS verification test as the 15.9kDa marker (black arrow). The serum fraction containing the 15.9kDa marker (*A*) was re-analysed using SELDI along with the depleted fractions (*B,C*) and the enriched IP eluate (*D*). The 15.9kDa signal was preserved in the depleted samples (*B,C*) and not specifically enriched in the IP eluate spectra (*D*). These findings indicated that either an anti-body failure or failed verification of OPN as the candidate 15.9kDa marker.

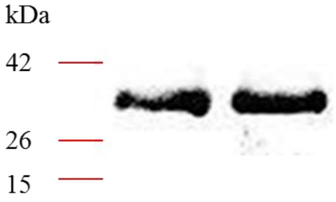
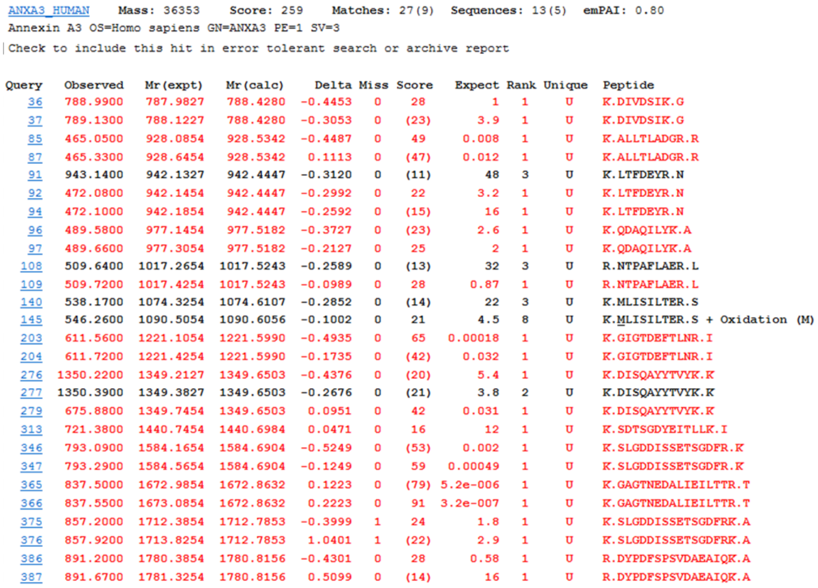


Figure 62. WB confirming effective OPN antibodies used in IP-MS
Western blot (WB) analysis testing anti OPN antibody (ab69498, Abcam) binding. 10µg recombinant OPN protein (ab92964, Abcam) was loaded in each well of 12% SDS PAGE. WB was carried out as described earlier. Reproducible detection of OPN at the expected size (35kDa) region was confirmed, ruling out antibody binding problem.

Finally, IP-MS verification of ANX A3 as the marker representing the 15.9kDa peak (Figure 63) was performed using a monoclonal anti ANX A3 antibody (Santa Cruz, sc-134260). Here, the ANX A3 depleted samples showed fading 15.9kDa signals and an enriched signal was evident in the ANX A3 rich eluate MS (Figure 64). This step was repeated using a different monoclonal anti ANX A3 antibody (AbD serotec, MCA2930Z) confirming the correlation between ANX A3 and the 15.9kDa marker (Figure 65).



```
ANXA3_HUMAN      Mass: 36353     Score: 259     Matches: 27(9)   Sequences: 13(5)   emPAI: 0.80
Annexin A3 OS=Homo sapiens GN=ANXA3 PE=1 SV=3
Check to include this hit in error tolerant search or archive report

Query   Observed    Mr(expt)    Mr(calc)     Delta Miss Score   Expect Rank Unique  Peptide
   36   788.9900    787.9827    788.4280   -0.4453    0    28        1    1      U    K.DIVDSIK.G
   37   789.1300    788.1227    788.4280   -0.3053    0   (23)     3.9    1      U    K.DIVDSIK.G
   85   465.0500    928.0854    928.5342   -0.4487    0    49    0.008    1      U    K.ALLTLADGR.R
   87   465.3300    928.6454    928.5342    0.1113    0   (47)   0.012    1      U    K.ALLTLADGR.R
   91   943.1400    942.1327    942.4447   -0.3120    0   (11)      48    3      U    K.LTFDEYR.N
   92   472.0800    942.1454    942.4447   -0.2992    0    22      3.2    1      U    K.LTFDEYR.N
   94   472.1000    942.1854    942.4447   -0.2592    0   (15)      16    1      U    K.LTFDEYR.N
   96   489.5800    977.1454    977.5182   -0.3727    0   (23)     2.6    1      U    K.QDAQILYK.A
   97   489.6600    977.3054    977.5182   -0.2127    0    25        2    1      U    K.QDAQILYK.A
  108   509.6400   1017.2654   1017.5243   -0.2589    0   (13)      32    3      U    R.NTPAFLAER.L
  109   509.7200   1017.4254   1017.5243   -0.0989    0    28     0.87    1      U    R.NTPAFLAER.L
  140   538.1700   1074.3254   1074.6107   -0.2852    0   (14)      22    3      U    K.MLISILTER.S
  145   546.2600   1090.5054   1090.6056   -0.1002    0    21      4.5    8      U    K.MLISILTER.S + Oxidation (M)
  203   611.5600   1221.1054   1221.5990   -0.4935    0    65  0.00018    1      U    K.GIGTDEFTLNR.I
  204   611.7200   1221.4254   1221.5990   -0.1735    0   (42)   0.032    1      U    K.GIGTDEFTLNR.I
  276  1350.2200   1349.2127   1349.6503   -0.4376    0   (20)     5.4    1      U    K.DISQAYYTVYK.K
  277  1350.3900   1349.3827   1349.6503   -0.2676    0   (21)     3.8    2      U    K.DISQAYYTVYK.K
  279   675.8800   1349.7454   1349.6503    0.0951    0    42    0.031    1      U    K.DISQAYYTVYK.K
  313   721.3800   1440.7454   1440.6984    0.0471    0    16       12    1      U    K.SDTSGDYEITLLK.I
  346   793.0900   1584.1654   1584.6904   -0.5249    0   (53)   0.002    1      U    K.SLGDDISSETSGDFR.K
  347   793.2900   1584.5654   1584.6904   -0.1249    0    59  0.00049    1      U    K.SLGDDISSETSGDFR.K
  365   837.5000   1672.9854   1672.8632    0.1223    0   (79) 5.2e-006    1      U    K.GAGTNEDALIEILTTR.T
  366   837.5500   1673.0854   1672.8632    0.2223    0    91  3.2e-007    1      U    K.GAGTNEDALIEILTTR.T
  375   857.2000   1712.3854   1712.7853   -0.3999    1    24      1.8    1      U    K.SLGDDISSETSGDFRK.A
  376   857.9200   1713.8254   1712.7853    1.0401    1   (22)     2.9    1      U    K.SLGDDISSETSGDFRK.A
  386   891.2000   1780.3854   1780.8156   -0.4301    0    28     0.58    1      U    R.DYPDFSPSVDAEAIQK.A
  387   891.6700   1781.3254   1780.8156    0.5099    0   (14)      16    1      U    R.DYPDFSPSVDAEAIQK.A
```

Figure 63. MS/MS identification of ANX A3

Sequence coverage of ANX A3 following biomarker purification. Tryptic peptides determined by product ion LC-ESI-MS^2 analysis of ANX A3 was confirmed by IP MS verification. ANX A3 molecular weight is 36kDa, the marker identified here could represent a truncated ANX A3 product, and/or a doubly charged MS signal.

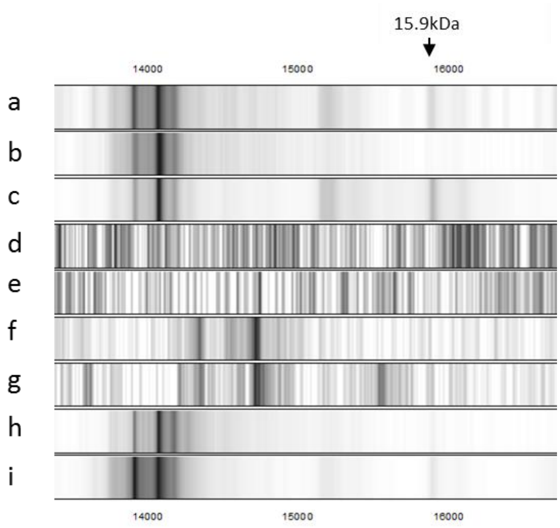


Figure 64. IP-MS verification of ANX A3 as a breast cancer marker

IP-MS verification of the 15.9kDa (black arrow) marker identified as ANX A3. Antibody capture of ANX A3 was performed using monoclonal antibody (sc-134260, Santa Cruz). The antibody was linked to μMACS protein A/G columns (Miltenyi Biotecs) and incubated with pooled serum fractions containing ANX A3. Shown are original mass spectra of ANX A3 serum fraction (a) , along with mass spectra of the captured protein from depleted specimen (b) and the enriched ANX A3 IP eluate (c), respectively. Negative controls illustrated as follows: ANX A3 abtibody + beads (d), Glycine and PBST (e,f), anti ANX A3 antibody (g), anti IgG antibody + beads (h), mouse serum (i).

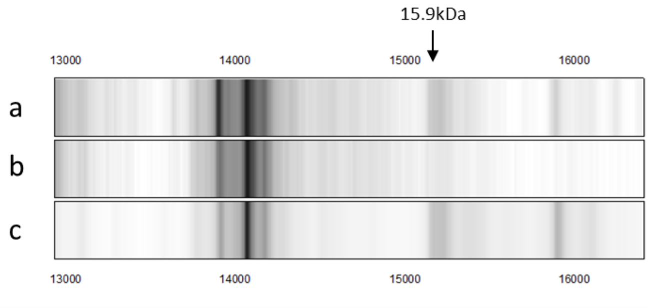
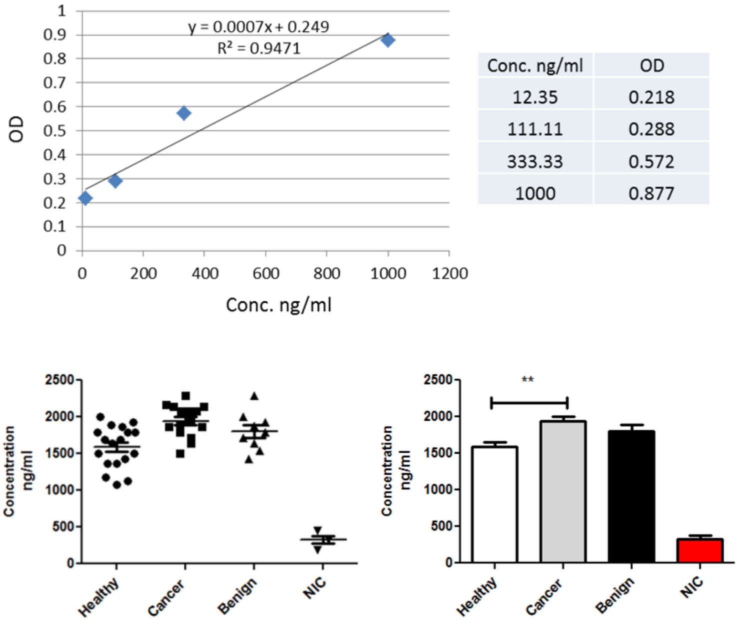


Figure 65. IP-MS verification of ANX A3 as a breast cancer marker using an alternative antibody
Further confirmation of the 15.9kDa biomarker as ANX A3 using a different monoclonal anti ANX A3 antibody (AbD serotec, MCA2930Z) following the same IP-MS protocol described previously. Shown are the original mass spectra of ANX A3 serum fraction (a) , along with mass spectra of the captured protein from depleted specimen (b) and the enriched ANX A3 IP eluate (*c*), respectively. This further confirmed that the 15.9kDa marker was ANX A3.

Based on the IP-MS verification, further evaluation of Apo C1 and ANX A3 as potential clinically valid/reliable markers was carried out.

3.2.3.14 Apo C1 ELISA validation

A commercial ELISA kit (E90252Hu, Uscn Life Science Inc.) was used to validate the weight of Apo C1 as a breast cancer serum marker. The manufacturers recommended protocol was followed (detailed in the methods section). Forty two samples representing the three comparison groups were analysed. Here, the ELISA analysis revealed significant over-expression of Apo C1 in the cancer compared to the control group (*P* = 0.002). The small size of the ELISA validation cohort however compromised further ROC analysis. However, although the sample size was small, this finding combined with previous MS findings strongly indicate a potential role for ApoC1 as a diagnostic breast cancer marker (Figure 66). Further multi-centre validation and translational efforts are key to confirm and elucidate the ApoC1 role in breast tumourigenesis pathways.



Conc. ng/ml	OD
12.35	0.218
111.11	0.288
333.33	0.572
1000	0.877

Figure 66. Serum Apo C1 ELISA validation

Serum apolipoprotein C1 ELISA results comparing breast cancer, benign breast disease and healthy control groups. A total of 42 samples (18 control, 15 cancer and 9 benign) were analysed. A standard curve used in protein concentration calculation is shown. The analysis indicated significant over expression of ApoC1 levels in sera from cancer patients compared to the control group (*P* =0.002). No significant differential expression was detected between the other groups. Apo D protein, secondary antibody and sample deficient reactions were used as negative controls (NIC).

3.2.3.15 ANX A3 ELISA validation

To confirm the performance of ANX A3 as a breast disease diagnostic marker, ELISA validation of SAX fractionated as well as unfractionated serum samples was carried out. Sandwich ELISA assays were developed in house using a mouse monoclonal anti ANX A3 antibody (AbD serotic, MCA2930Z) for plate coating, rabbit polyclonal anti ANX A3 antibody (Abcam, ab33068) as an antigen binding antibody (primary antibody) and an anti-rabbit HRP tagged antibody as a detection, or secondary antibody (GE Healthcare, NA9340-1ML). Assay optimisation was tested using recombinant ANX A3 protein serial dilutions to establish a titration curve (recombinant protein concentration 1μg - 1pg).

The coating antibody was assessed using a concentration gradient (1:10) covering concentrations between 1μg/ml-1pg/ml. This was tested in combination with both primary antibody concentration gradients of 1μg/ml-0.1ng/ml and detection antibody concentration gradients of 0.2μg/ml-0.1μg/ml. These parameters were used based on manufactures' instructions and in collaboration with Professor Mark Cragg and Dr. Oliver Poetz. The signal levels from all combinations were very low and had a high level background preventing a standard curve establishment (Figure 67). The mean control OD was 0.08 and OD of all conditions ranged between 0.076 -0.156 (mean 0.09). The same conditions were then tested using two different ELISA plates to exclude plate related binding issues. Titration curves were not achieved and the signal showed low and irreproducible signal patterns (Figures 68, 69). Wider concentration ranges were also tested to exclude dilutional/saturation effects. Here the antibodies were diluted 1:10 to cover the following concentration ranges (coating antibody: 10μg/ml-0.1pg/ml, primary antibody: 10μg/ml-1pg/ml, secondary antibody: (1μg/ml-1ng/ml). Nonetheless, irreproducibility, variably, a low signals were still features of this assay (Figure 70).

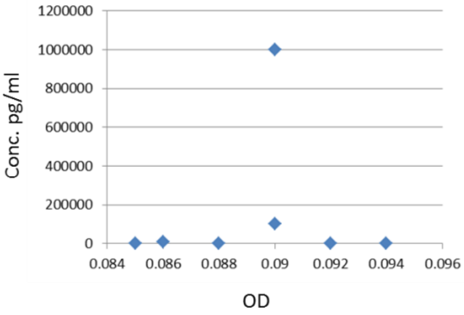


Figure 67. Initial ANX A3 ELISA testing

A figure representing part of the ANX A3 ELISA optimisation experiments using BD falcon plates (BD Biosciences, Cat. No. 353279). A mouse monoclonal anti ANX A3 antibody (AbD serotic, MCA2930Z) was used for plate coating (1μg/ml), rabbit polyclonal anti ANX A3 antibody (Abcam, ab33068) was used as a primary antibody (1μg/ml) and an anti-rabbit HRP tagged antibody (0.1μg/ml)(GE Healthcare, NA9340-1ML). A titration curve using recombinant ANX A3 protein concentration gradients (1μg/ml -1pg/ml) was not achieved. OD was measured at a wave length of 450nm. Low non-linear OD signals prevented the generation of a usable curve.

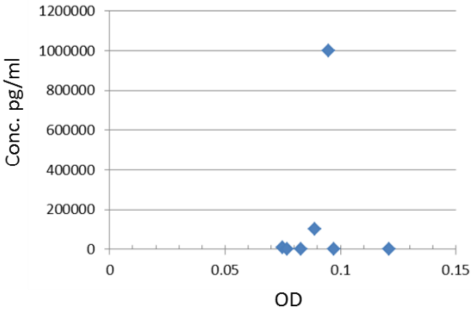


Figure 68. ANX A3 ELISA platform optimisation

A figure representing part of the ANX A3 ELISA optimisation experiments using BD falcon plate (BD Biosciences, Cat. No. 353279). Mouse monoclonal anti ANX A3 antibody (AbD serotic, MCA2930Z) was used for plate coating (1μg/ml), rabbit polyclonal anti ANX A3 antibody (Abcam, ab33068) was used as a (primary) antigen binding antibody (1μg/ml) and an anti-rabbit HRP tagged (secondary) detection antibody (0.1μg/ml) (GE Healthcare, NA9340-1ML). Titration curve using recombinant ANX A3 protein concentration gradients (1μg/ml -1pg/ml) was not achieved. OD was measured at a wave length of 450nm.

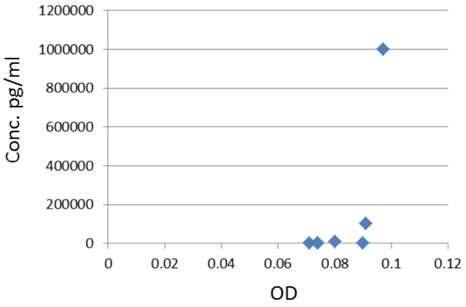


Figure 69. ANX A3 ELISA optimisation, testing plate fault

Figure representing part of the ANX A3 ELISA optimisation experiments. Using plate Pierce 96-well plates (Cat. 15041) Mouse monoclonal anti ANX A3 antibody (AbD serotic, MCA2930Z) was used for plate coating (1μg/ml), rabbit polyclonal anti ANX A3 antibody (Abcam, ab33068) was used as a (primary) antigen binding antibody (1μg/ml) and an anti-rabbit HRP tagged (secondary) detection antibody (0.1μg/ml)(GE Healthcare, NA9340-1ML). Failure to establish a titration curve was illustrated. OD was measured at a wave length of 450nm.

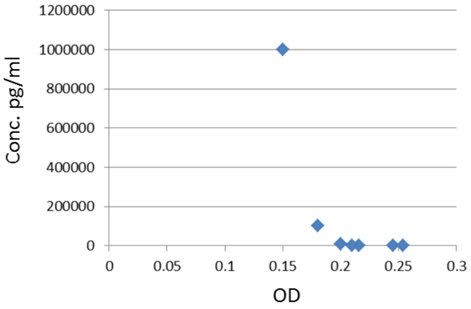


Figure 70. Final ANX A3 ELISA design testing

Figure representing part of the ANX A3 ELISA optimisation experiments. Using BD falcon plate (BD Biosciences, Cat. No. 353279). Mouse monoclonal anti ANX A3 antibody (AbD serotic, MCA2930Z) was used for plate coating (10μg/ml), rabbit polyclonal anti ANX A3 antibody (Abcam, ab33068) was used as a (primary) antigen binding antibody (10μg/ml) and an anti-rabbit HRP tagged (secondary) detection antibody (1μg/ml)(GE Healthcare, NA9340-1ML). Titration curve using recombinant ANX A3 protein concentration gradients (1μg/ml -1pg/ml) was not achieved. OD was measured at a wave length of 450nm.

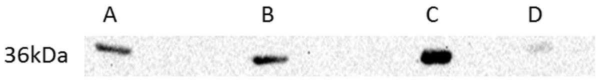


Figure 71. WB excluding faulty ANX A3 antibodies

Western blotting of recombinant ANX A3 protein (Abcam, ab92929). 0.1μg ANX A3 was loaded in each well, and different antibody concentrations were used (according to the manufacturers' recommendations) for detection. Four antibodies previously used in the ANX A3 ELISA customisation were tested. Mouse monoclonal anti ANX A3 antibody (AbD serotic, MCA2930Z) (*A*), rabbit polyclonal anti ANX A3 antibody (Abcam, ab33068) (*B*), (Santa Cruz, sc-134260) (*C*), rabbit polyclonal antibody (Kind gift from Dr Egle Solito, Imperial College- London) (*D*) were successful in detecting the recombinant protein excluding antibody and/or protein related ELISA errors.

Alternative antibodies/designs were used in various combinations to establish a more sensitive and reproducible ELISA test. First, we tested an ELISA whereby the coating utilised the polyclonal antibody combined with monoclonal antibodies were used as primary antibodies. Several other antibodies were also tested including monoclonal anti ANX A3 antibody (sc-134260, Santa Cruz), an in-house rabbit polyclonal antibody (A kind gift from Dr Egle Solito, Imperial College- London) and alternative HRP tagged antibodies matching the primary antibody species (mouse polyclonal Abcam, ab2110 and rabbit polyclonal , ab2115). However, these failed to achieve a reliable titration curve. In addition, other possible sources of error including reading wave lengths, freshly prepared buffers, extra washing steps, and antibody dilution with PBS alone without BSA were tested revealing no improvement (data not shown). To exclude faulty recombinant protein (Abcam, ab92929) or antibodies, western blotting of 0.1μg ANX A3 recombinant protein was performed. A reproducible signal was detected at the expected size using the different antibodies previously tested for ANX A3 ELISA design (Figure 71). It was concluded that further optimisation was beyond the scope of this project (due to time limitation), and a newly available commercial ANX A3 sandwich ELISA kit (Uscn Life Sciences, Cat. No. E94786Hu) was tested. Here, titration curves were successfully established using the ANX A3 concentration gradients between 0-10ng/ml as per manufacturer's instructions (Figures 72 A-C).

Following the successful ANX A3 ELISA titration tests, pooled serum fractions representing the 3 disease groups (IDC, BBD and controls) were used to test assay performance on serum samples from the Wessex cohort. The pooling approach was adopted due to limited fractioned sample volume. Here, 10μl from 10 samples was used to constitute each pool. Overall, 5 IDC, 5 controls and 4 BBD pools were tested (Figure 73). OD signals from all samples were satisfactory and enabled meaningful interpretation of the ANX A3 concentrations. Interestingly, this revealed significantly higher ANX A3 level (*P*= 0.016) in BBD compared to controls which was consistent with the MS findings. Surprisingly, the levels of ANX A3 in BBD were also significantly higher than the cancer group (*P*= 0.016) which was not found in the MS analysis. Finally, the ANX A3 expression levels in the cancer and control groups were not significantly different (*P*= 0.81), which was consistent with the original MS findings.

The differential expression revealed here was consistent with the original MS findings where, although no significant over expression was detected between BBD and IDC by MS, the ANX A3 levels were shown to be higher in BBD than IDC. This could be explained by higher sensitivity and favourable quantification abilities of ELISA compared to top-down MS analysis. However, these results were interpreted with caution as pooling the samples could contribute to skewed results.

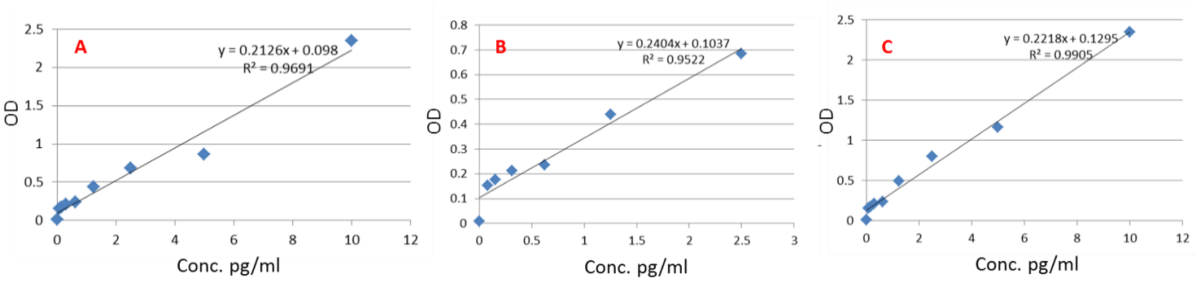


Figure 72. (A-C) Optimal ANX A3 ELISA standard curves

Titration curves (73 A-C) established using the ANX A3 ELISA kit (Uscn Life Sciences, Cat. No. E94786Hu). Reproducible linear standard curves were produced (R^2 >0.95). This allowed confident accurate measurement of serum ANX A3 levels in further analysis. OD was measured at a wave length of 450nm.

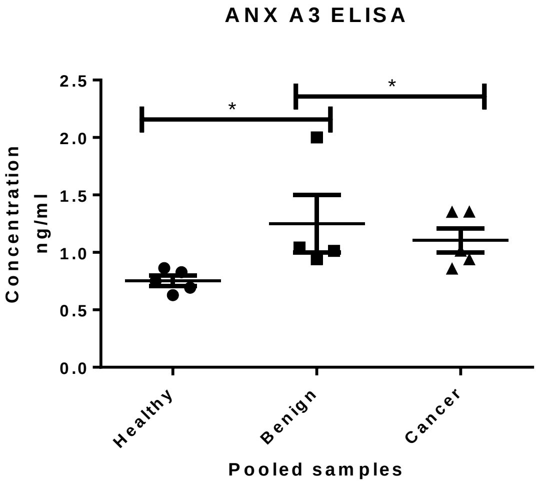


Figure 73. Validation of ANX A3 as a BBD biomarker using pooled samples

ANX A3 immuno-validation using pooled serum fractions from the cancer (IDC), BBD and controls. Each pool was constituted from 10μl aliquot from 10 SAX F5/F6 fractionated samples totaling to 100μl which was then used as a representative sample (n=1). Normal (n=5), BBD (n=4) and cancer (n=5) pools were tested and BBD levels were significantly higher than controls and IDC groups. No significant difference in ANX A3 levels was detected between the IDC group and the controls.

*(*P*< 0.05)

To validate the previous MS and ELISA findings, further ANX A3 ELISA analysis was carried out using SAX fractionated serum representing the three groups (14 controls, 21 BBD and 16 IDC). In addition, recombinant ANX V (Kind gift from Professor Mark Cragg-Cancer Science Unit, University of Southampton) was used as a negative control to test the specificity of the ELISA assays. The ELISA confirmed consistent differential expressions with significantly higher ANX A3 levels in the BBD compared to IDC and controls ($P<0.0005$) with no differential expression between the IDC and control groups. ANX V signal was very minimal indicating high ANX A3 detection specificity (Figures 74).

The sensitivity and specificity of ANX A3 as a predictor of BBD compared to IDC was 81%, whereas ANX A3 differentiated BBD from controls with a sensitivity of 95% and a specificity of 86% (Figure 75). As such, serum ANX A3 levels could represent a promising adjunct diagnostic marker pending further multi-centre evaluation.

To date, there are no reports linking ANX A3 to breast disease. Further multi-centre validation efforts are essential to confirm the role of ANX A3 as a BBD marker, elucidating the molecular pathways linking ANX A3 to key breast cancer hormonal and apoptotic events, which would enable better understanding and validation of this discovery. To investigate such links, pathways analysis using IPA software (Ingenuity Systems, USA) was conducted including ANX A3 protein, ER, PR and HER-2 receptors in a generic literature search. Here, ANX A3 was shown to be directly linked to two key players in mammary development, prolactin and beta-oestradiol (Figure 76). Such findings support the potential value of ANX A3 as a breast disease marker. Further evaluation of prolactin and estradiol levels in these samples could be the starting step towards further investigation in this direction.

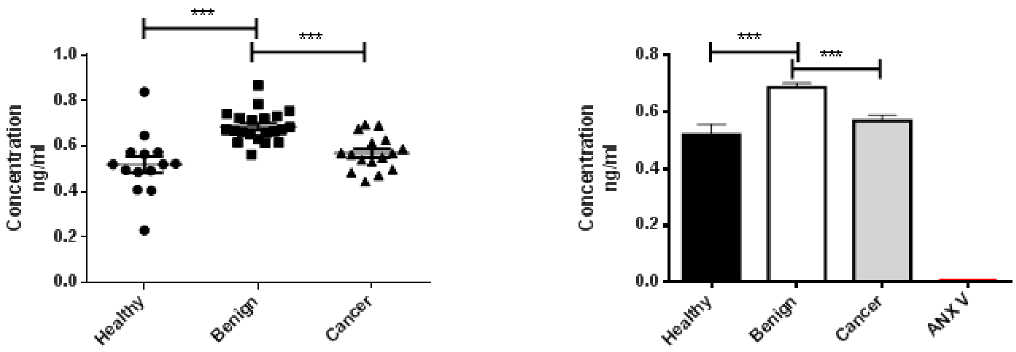


Figure 74. ANX A3 ELISA validation as a BBD biomarker

ANX A3 immuno-assay validation by sandwich ELISA (Uscn Life Sciences, Cat. No. E94786Hu). Differential expression at the protein level. ANX A3 levels in BBD were significantly higher than the IDC and control groups. The specificity of the assay to ANX A3 detection was illustrated by poor signal from ANX V protein (an ANX A3 homolog). ***($P<0.0005$)

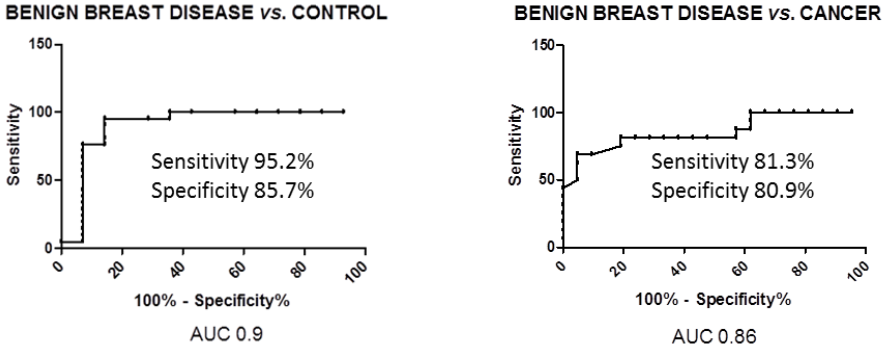


Figure 75. ROC analysis of serum ANX A3 in breast disease
Serum ANX A3 levels - women with benign breast disease were compared to healthy controls and women with invasive breast cancer. The ANX A3 differentiating accuracy is shown in the receiver operating characteristic (ROC) analysis and the area under curve (AUC) where a promising group prediction performance (>80% sensitivity and specificity) is illustrated in the differentiation of BBD from IDC and controls.

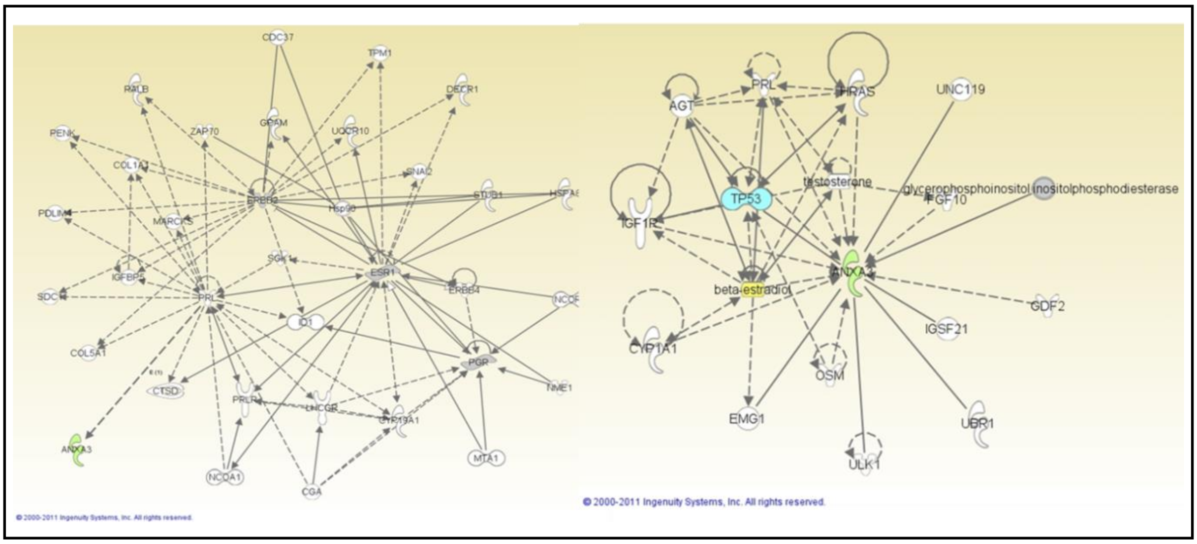


Figure 76. Pathway analysis of the potential role of ANX A3 in breast cancer pathways
Biochemical pathway analysis (Ingenuity systems) indicating the regulation of ANX A3 (in green) by beta-oestrodiol (in yellow) and testosterone. Moreover prolactin (PRL) has been reported to have a regulatory effect on ANX A3. Prolactin levels were shown to be of significance in rat prostate cancer models. In cells expressing human Oestrogen Receptor protein(s) from human breast tissue, beta-estradiol is involved in the expression of human ANXA3 mRNA. Thus, a potential link between ANX A3 and major hormonal receptors justifies future ANX A3 functional experiments.

3.2.4 Discussion of the Wessex breast cancer serum proteomic analysis

Current serum and hormonal markers for breast cancer are far from perfect[42]. Other markers including carcinoembryonic antigen (CEA) and carbohydrate antigen (CA) 15-3 are occasionally used in metastatic disease to monitor response to treatment, but there is an urgent need to discover effective and reliable serum biomarkers for the prediction, diagnosis and stratification of breast tumours, and proteomic studies have considerable potential in this arena[44].

In the previous section, detailed breast cancer serum SELDI optimisation was reported. Several biochemical conditions were tested and WCX arrays were shown to provide the most complex profiles compared to SAX, reverse phase and IMAC arrays. It was also evident that fractions 2, 3 and 4 of the SAX fractionated sera were unfavourable due to high albumin content (Fractions 3 and 4) and poor MS signal (Fraction 2). These analyses were similar to the serum optimisation methods utilized for the Guernsey serum analysis. A successful final re-optimisation step was conducted to confirm the reproducibility of the WCX conditions prior to the actual analysis. Acceptable inter and intra-analysis coefficient of variations (< 30%) were obtained in all checks with up to a one week gap from original analysis (batch effect analysis). Array pre-wetting in combination with 50 mM ammonium acetate (pH 4.5) buffer were shown to provide the best WCX profiling results. However, CPLL and size exclusion columns were found to be of importance in the purification rather than the initial analysis stage.

This study was designed to address a key cancer biomarker challenge; the identification of diagnostic breast cancer serum markers[221]. 347 samples representing controls, benign breast disease (BBD) and invasive breast cancer (IDC) groups initially revealed 27 potential biomarkers differentiating the groups. These were challenged by a final validation analysis using a different platform (MALDI-TOF MS) in a partner centre (St. George's University of London), where only 3 of the 27 markers were validated. Two of these candidate markers were over-expressed in the cancer group compared to controls (6.4 and 9.5kDa), whereas the 15.9kDa marker was over-expressed in the BBD compared to the control group.

Based on these findings, purification, identification and verification steps were conducted to elucidate the nature of the three validated proteomic peaks. Following a comprehensive gel based and gel free purification steps, Apo C1 (9.5kDa) and ANX A3 (15.9kDa) were identified and verified by IP-MS as breast disease biomarkers in this study. The identity of the third marker at 6.4kDa is under continuing investigation.

Literature profiling for the potential role of these markers in breast cancer, both candidates (Apo C1, ANX A3) were of biological interest. Firstly, the 9.5kDa marker identified as apolipoprotein C1 (Apo C1) has been reported to be a diagnostic and prognostic breast cancer marker[222, 223]. Interestingly, the second biomarker corresponding to the MS peak at 15.9kDa was the calcium and phospholipid binding protein annexin A3 (ANX A3), reported as a major cancer biomarker[224-226]; which has never been previously reported in relation to breast disease. Based on the positive confirmatory MS validation, immunovalidation of these markers was conducted using sandwich ELISA on independent serum samples representing the former three groups. Immunovalidation by sandwich ELISA was chosen as the preferred validation method for several reasons. Firstly, the MS signal intensity indicated low serum ANX A3 levels hence; ELISA was chosen as a more sensitive tool than WB. In addition, the use of sandwich ELISA was anticipated to overcome non-specific background resulting from the highly abundant protein in serum WB. Finally, as ELISA is a clinically applicable platform, this would expedite biomarker translation to a bedside test.

Here, in agreement with the MS analysis findings, Apo C1 serum levels were shown to be over-expressed in breast cancer sera, whereas ANX A3 was over-expressed in the benign disease group compared to controls. Interestingly, ANX A3 ELISA also indicated a significant over-expression in the benign group compared to cancer, which is a similar but more significant expression trend revealed by the MS analysis.

Our findings indicate Apo C1 as a potential diagnostic breast cancer biomarker are in agreement with previous findings describing it as an early diagnostic marker and a marker of early breast cancer relapse[222, 227, 228]. Although the exact molecular role of Apo C1 in neoplastic breast disease remains unclear, an Apo C1 anti apoptotic role was previously illustrated in pancreatic cancer cells and a similar mechanism could thus explain over expression in breast cancer patient sera[229]. Further blinded validation studies involving larger cohorts representing different molecular sub-groups and *in situ* disease remain essential. More importantly, the molecular role(s) of Apo C1 in mammary neoplasia warrants further exploration.

This study also revealed a 15.9kDa breast disease serum biomarker that was identified as the phospholipid and calcium binding protein human annexin A3 (ANX A3). Until now, differential expression of ANX A3 in breast pathology has not been reported. We validated our MS findings by ELISA which showed that high expression levels of serum ANX A3 correlated with BBD. The levels of this or other annexins were previously found to be differentially expressed in intestinal, pancreatic, hepatocellular, prostatic, renal and lung malignancies, as well as hairy cell leukaemia[230-232]. Differential ANX A3 expression also correlated to thyroid tumour progression as well as ovarian cancer chemo resistance[224, 226].

In this study, serum ANX A3 was significantly elevated in benign breast disease (BBD) compared to both invasive breast cancer (IDC) and controls. Although this did not reach statistical significance in the MS analysis, validation by quantitative ELISA has confirmed the significance of ANX A3 as a biological BBD discriminator. This could be related to the limited semi-quantitative abilities of top down MS analysis compared to the more highly sensitive and quantitative ELISA assay. More importantly, both analyses have revealed ANX A3 as a potential BBD diagnostic marker. The sensitivities and specificities of ANX A3 levels as a BBD differentiator in the ELISA tests ranged from 81-95%. These results suggest that combined with other markers and/or diagnostic tests, ANX A3 could be a reliable biomarker for BBD diagnosis. However, these remain preliminary findings that should be interpreted with caution and require further multi centre testing.

Proliferative lesions of the breast such as atypical ductal hyperplasia (ADH) are associated with increased risk of breast cancer[233-235]. Further evaluation of ANX A3 expression and function among different benign breast disease groups such as hormonally driven (eg. atypical ductal hyperplasia and columnar cells changes) ER-negative lesions (such as apocrine metaplasia) is thus an essential part of future ANX A3 validation efforts. Equally, the expression behaviour of ANX A3 in DCIS and different IDC molecular sub groups (basal, ERBB-2 and luminal) needs larger sub-entity focused studies involving biological fluids, tissue expression and translational functional studies.

The potential functional role of ANX A3 in malignant transformation and/or progression is intriguing as ANX A3 was defined previously as an independent prognostic factor related to neoplastic progression[236]. In addition, in breast neoplasms, patients diagnosed with probable BBD who presented later with IDC are more likely to be older, postmenopausal women and patients with a strong family history of breast cancer; implying both familial and hormonal components of pathogenesis[237]. ANX A3 could thus play a role in follow up risk assessment in these patients.

Oestrogen is known to synergise with progesterone to promote breast proliferation and both play key roles in breast tumourigenesis[238, 239]. Interestingly, a potential link between ANX A3 and major hormonal receptors as well as prolactin can be mapped using pathway analysis (IPA Ingenuity Systems, USA)[240]. This in addition to recent reports highlighting the role of ANX A3 as a marker in other hormone dependent cancers raises the question of whether ANX A3 is functionally associated with major cancer hormonal pathways[226, 241, 242].

Our data showing ANX A3 elevation in the BBD group is encouraging as other studies showed down regulation of ANX A3 in prostate cancer and up-regulation in advanced and metastatic prostate cancer (US patent 2003/0108963). In addition, ANX A3-driven therapeutic evaluation for urological and intestinal tumours is on-going (US Patent 7732148, Jun 2010).

Our findings have several limitations, most notably the lack of multi-centre validation and the absence of DCIS samples. With the use of a large sample size and an independent multi-platform discovery validation followed by immunovalidation, we aimed to minimise potential false biomarker discovery. In fact, the growing evidence of the importance of Apo C1 and ANX A3 in cancer suggests that our findings are plausible[224, 226, 236, 241-244]. However, an expansion of this effort to include multiple centres to explore wider histological categories could identify subtype-specific Apo C1 and ANX A3 profiles and elucidate their potential biological role(s) in breast cancer.

In summary, our Wessex analysis results demonstrated that Apo C1 and ANX A3 are probable discriminatory breast cancer markers that have considerable potential as diagnostic biomarkers. Prospective multi-centre studies are essential to establish the role of these markers inbreast cancer diagnosis. Our current efforts are focusing on the molecular role of these markers in the breast cancer disease process and their potential as therapeutic targets.

3.2.5 The Wessex breast cancer serum proteomic analysis summary

In the previous section, detailed MS analysis optimisation methodologies were covered. Several biochemical conditions were tested and WCX arrays were shown to provide the most complex MS profiles. A standardised and independently validated MS analysis validated three potential breast cancer serum biomarkers. Two of these markers were over-expressed in the cancer group compared to the controls (6.4 and 9.5kDa), whereas the 15.9kDa marker was over-expressed in the benign disease compared to the control group. Apo C1 (9.5kDa) and ANX A3 (15.9kDa) were identified and verified by IP-MS as breast cancer biomarkers. The identity of the third marker at 6.4kDa is currently under investigation. Our current efforts are focusing on the molecular role of these markers in the disease process and their potential as therapeutic targets.

3.3 Early onset breast cancer prognosticators, the "POSH" serum analysis

3.3.1 Introduction

Breast cancer is the most common cancer among women worldwide, affecting up to 12% of all women in Europe and North America[245]. Approximately 2-5% of all breast cancer diagnosis are patients under the age of 40[16, 189]. Despite its low prevalence among breast cancer populations, early onset breast cancer is reported as a major malignancy and the most common cause of cancer related deaths in this age group[246, 247]. The natural history of early onset breast cancer tends to be of a more aggressive and recurrent phenotype compared to later onset disease. The disease in this group tends to present with high grade, high stage and less responsive forms[248-251]. Overall, young women with breast cancer have a decreased overall disease- free survival rate and a higher ratio of tumours with poor pathologic features[252, 253].

Additionally, recent reports have shown that the risk of death rose by 5% for every 1-year reduction in age at diagnosis[254]. Even without nodal involvement early onset breast cancer has a 25% recurrence rate which could occur up to 12 years post-surgery[252]. Generally, the risk of local, regional and distant relapses after surgical treatment is higher in this group of patients[252]. Therefore, it is vital to find biomarkers to identify young women who have an increased risk of breast cancer recurrence and/or relapse, and would therefore benefit from intensified surveillance and/or adjuvant therapy. Moreover, markers modulating early onset disease outcome could provide insight into treatments altering novel molecular targets/pathways and overall prognosis. This work was conducted in order to achieve these targets and to stratify patients with early-onset breast cancer by elucidating the proteomic molecular targets that underlie treatment response.

3.3.2 Methods

In an effort to unveil new prognosticators of early onset breast cancer, analysis of matched serum samples from the "POSH" cohort was performed using a novel 3D MudPIT application[11]. POSH is a large prospective cohort study recruiting 3,000 women aged 40 years or younger at breast cancer diagnosis[194]. An additional recruitment category included women who were BRCA1 or BRCA2 gene carriers aged 41 to 50 years and were diagnosed with their first breast cancer during the study recruiting period. Most of the recruited patients received an anthracycline based adjuvant chemotherapy regimen (Table 10).

Samples used here were from patients treated with the same chemotherapeutic agents (anthracycline based), who underwent surgery after the completion of neoadjuvant chemotherapy at different centres. Briefly, the POSH proteomics profiling project involved patients with none metastatic early stage (stage1-2) invasive ductal carcinoma (IDC). A total of 399 serum samples (203 good outcome *vs.* 196 poor outcome) were collected under strict SOPs; to explore prognostic markers key to patient outcomes. Only 29 patients (19 good and 10 poor outcome samples) had confirmed BRCA1/2 mutations.

The good outcome group was defined as patients who underwent treatment for breast cancer and had a disease free survival (DFS) of at least 5 years following treatment. Patients who showed less favourable outcomes (uni/bilateral recurrence, metastasis or death) within 2 years of treatment were denoted as the poor outcome group. This novel quantitative proteomic analysis (3D MudPIT) was adopted building on evidence from previous breast cancer tissue[201] and benign prostate hyperplasia serum studies[11] which indicated a robust and more advanced biomarker discovery method compared to SELDI TOF MS. Here, pooled samples from each group were analysed. Triplicate analysis of all samples involved high-performance size-exclusion chromatography (SEC) for the pre-fractionation of serum proteins followed by dialysis exchange and solution phase trypsin proteolysis. The tryptic peptides were then offline fractionated with hydrophilic interaction chromatography (HILIC). Each fraction was then analysed with online reverse phase nano ultra-performance chromatography connected to nanoelectrospray ionisation - tandem mass spectrometry (RP nUPLC - nESI MS/MS) using an ion trap mass analyser. For the spectral processing, the SpectrumMill and InsPecT software programs were used. Label free analysis was performed at the BRFAA (see methods for full details) for the qualitative and quantitative assessment of the proteins and phosphoproteins. The DAVID gene ontology tool and the Ingenuity Pathway Analysis program (IPA Ingenuity Systems, USA) were used for the functional annotation of the quantitative proteome and phosphoproteome analysis and associated biochemical and molecular biology pathways and networks.

Candidate biomarkers revealed at this stage were then subject to further ELISA validation based on their relevance in the disease pathway analysis and/or literature. This was performed using commercial ELISA kits, where samples were analysed in duplicate using recombinant standard proteins to generate linear titration curves. In addition, reactions with no secondary/detection antibodies/antigen were used as negative controls.

3.3.3 Results

3.3.3.1 3D MudPIT MS/MS analysis

This study resulted in the quantitative profiling of over 6,500 serum proteins and over 1200 phosphoproteins at very high confidence ($P < 0.05$; FDR<5%). As such, this modified platform enabled a 3-fold increase in the identified proteins compared to the original proof of concept study[11]. Among these, 156 prognosticators were reproducibly differentially expressed ($P< 0.05$) between the good and poor outcome groups (Table 11). These included established diagnosis, disease progression and drug efficacy biomarkers of breast cancer and its pharmacotherapy.

Pathway analysis of the identified biomarkers indicated their role in key molecular pathways including DNA damage response, hereditary breast cancer, molecular mechanisms of cancer and prolactin signalling (Figure 77, 78).

Further analysis of breast cancer biological pathways (using protein markers which have been identified in both groups) by IPA software (Ingenuity Systems, USA) revealed that 7 of these markers qualify as strong potential biomarkers according to existing molecular pathways. These candidate markers were: nitric oxide synthase 2 (NOS2), protein patched homolog 1 (PTCH1), DNA polymerase subunit gamma-1 (DPOG1), apolipoprotein C-I (ApoC1), 40S ribosomal protein S4 X isoform (RPS4X), polycomb complex protein (BMI-1), and DNA (cytosine-5)-methyltransferase 1 (DNMT1).

Although these markers were favourable candidates for immunovalidation and functional studies, other markers were also identified as “interesting” based on supportive literature and/or biological evidence, and will be studied in future work. In addition, technical aspects related to the peptide sequences involved in quantification were carefully considered prior to relatively long and costly validation processes. As a general guide, biomarkers identified with identical peptide hits showing significant differential expression were considered more appropriate candidates for further evaluation. Such stringent limits should reduce false positives and facilitate a clinically reliable conclusion. Based on the formerly described factors, and with time and resources limitations, further immunovalidation was conducted on four of the qualifying markers. These were ANX A2, ApoC1, NOS2 and DNMT1.

Characteristic	Good outcome (n=203)	Poor Outcome (n=196)	Study cohort (n=399)
Age, Years			
Median	37	36	36
Range	25-40	18-41	18-41
Follow up, years			
Median	4.71	3	4.12
Range	5.0-5.7	0.4-6.5	0.4-6.5
Histology			
Invasive Ductal Carcinoma	203 (100)	191(97)	394 (99)
Invasive Lobular Carcinoma	0	0	0
Unknown	0	5 (3)	5 (1)
Grade 1	10 (5)	5 (3)	15 (4)
Grade 2	76 (37)	47 (24)	123 (31)
Grade 3	113 (56)	139 (70)	252 (63)
Unknown	4 (2)	5 (3)	9 (2)
Lymph node status			
Negative	104 (51)	60 (31)	164 (41)
Positive	96 (47)	128 (65)	224 (56)
Undetermined	3 (2)	8 (4)	11 (3)
Oestrogen receptor status			
Positive	139 (68)	108 (55)	247 (62)
Negative	62 (21)	88 (45)	150 (38)
Unknown	2 (1)	0	2 (0)
Progesterone receptor status			
Positive	88 (43)	75 (38)	163 (41)
Negative	80 (39)	86 (44)	166 (42)
Unknown	35 (18)	35 (18)	70 (17)
HER2 receptor status			
Positive	52 (26)	83 (42)	135 (34)
Negative	59 (29)	92 (47)	151(38)
Unknown	92 (45)	21(11)	113 (28)
BRCA1/2 mutation status			
Positive	19 (9)	10 (5)	29 (7)
Negative	182 (90)	182 (93)	364 (91)
Unknown	2 (1)	4 (2)	6 (2)
Resection margin			
R0 resection	141(70)	141(72)	282 (71)
R1 resection	25 (12)	21 (11)	46 (12)
Unknown	37 (18)	34 (17)	71 (17)
Chemotherapy			
FEC	68 (34)	71 (36)	139 (35)
ECMF	29 (14)	32 (16)	61 (15)
FEC + Decotaxel	23 (11)	13 (7)	36 (9)
AC	17 (8)	16 (8)	33 (8)
EC + Paclitaxel	14 (7)	12 (6)	26 (7)
EC + Paclitaxel + Gemcitabine	9 (4)	8 (4)	17 (4)
EC	9 (4)	6 (3)	15 (4)
Nul	22 (11)	7 (4)	29 (7)
Other	12 (7)	31 (16)	43 (11)

Table 10. The clinicopathological and treatment details of the POSH serum samples used in the MudPIT analysis.

A: Adriamycin; C: Cyclophosphamide; E: Epirubicin; F: 5 FU; M: Methotrexate. Percentages are in parentheses.

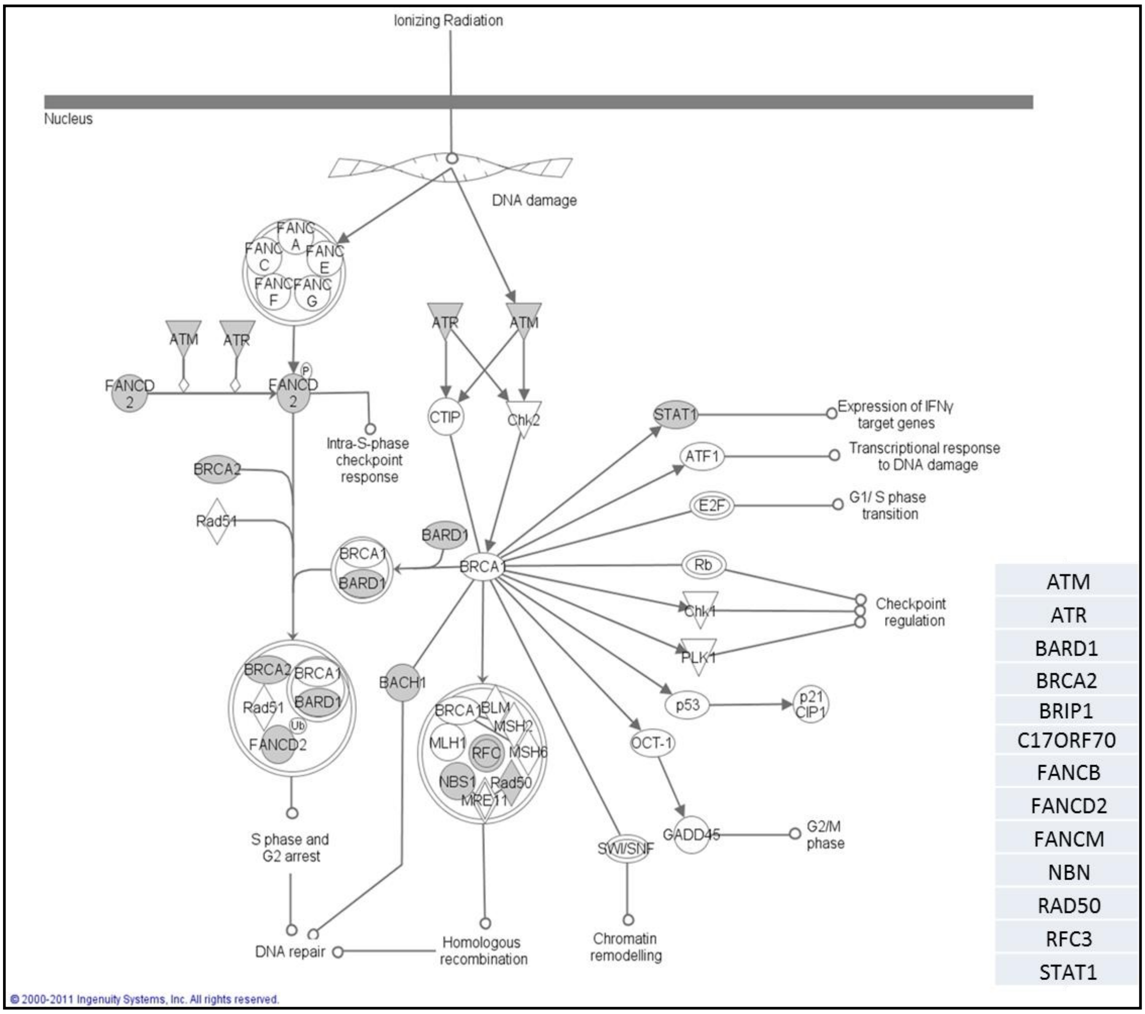


Figure 77. Pathway analysis of EBC prognosticator from the POSH analysis

Ingenuity pathway analysis (IPA) involving 13 candidate prognosticators of early onset breast cancer response (highlighted in table 11). These were found to be involved in several DNA repair pathways illustrated in this graph. Hereditary breast cancer proteins such as BRCA2 were among these markers as illustrated above. Markers identified from the POSH analysis are highlighted in grey.

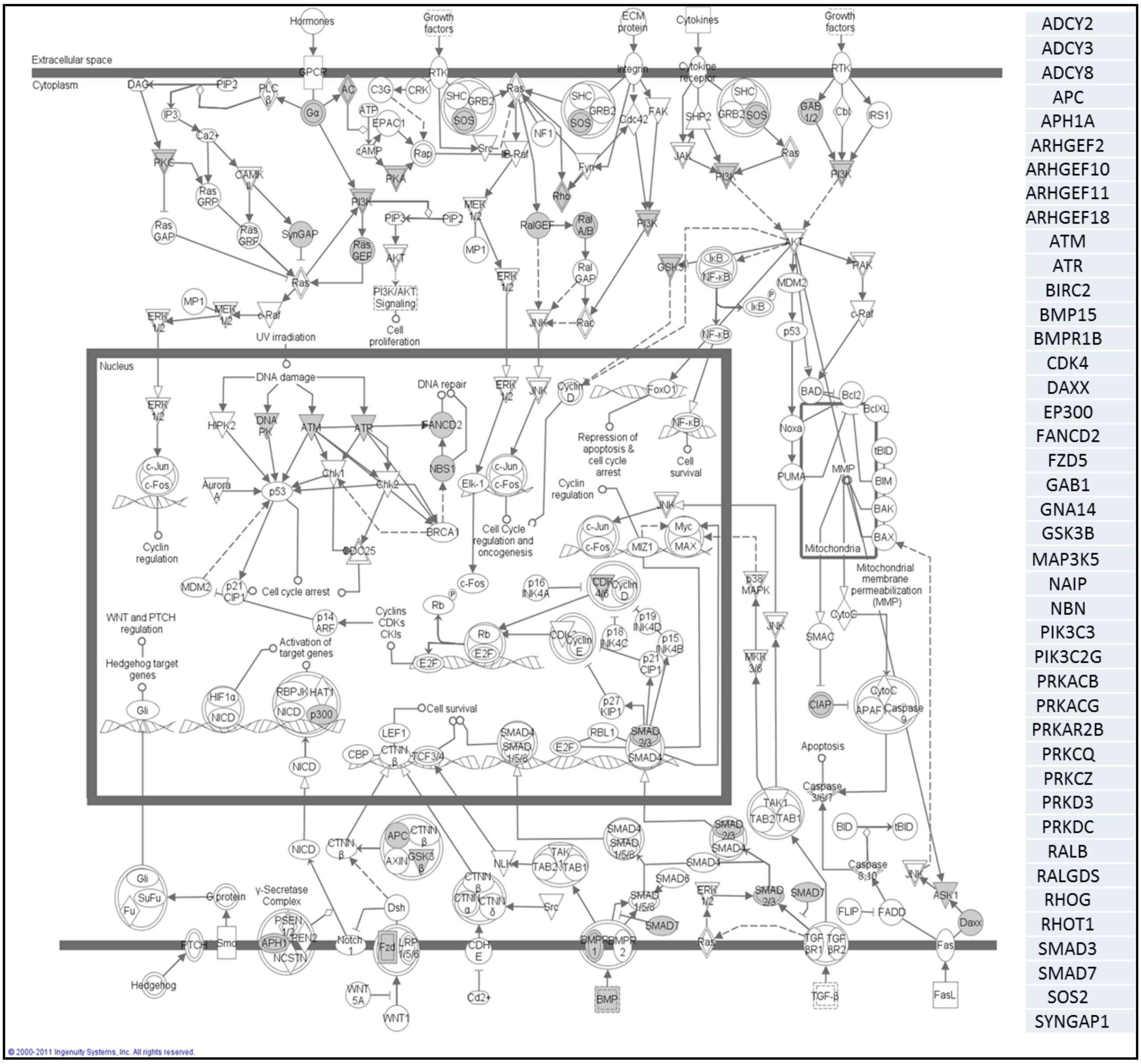
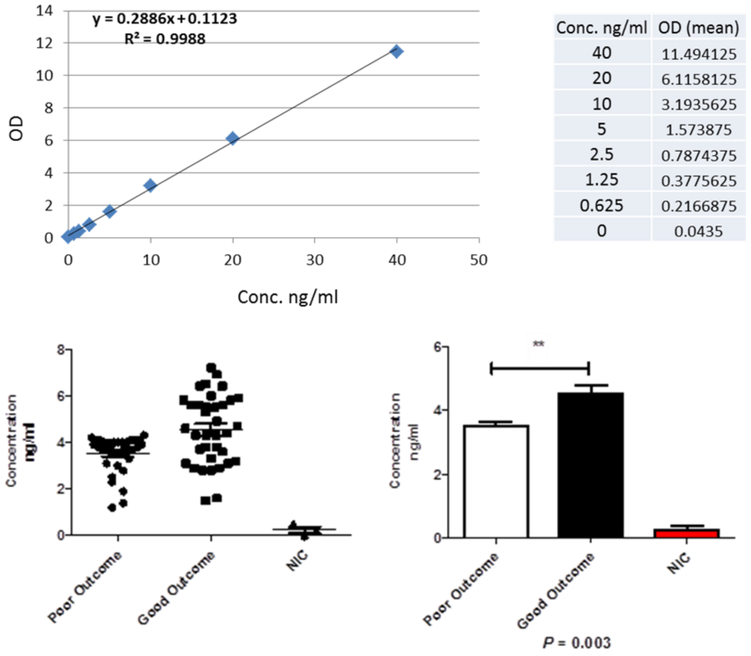


Figure 78. Pathway analysis of 42 prognosticators revealed in the POSH analysis

Ingenuity pathway analysis (IPA) illustrating the key involvement of 42 candidate prognostic biomarkers identified from the POSH analysis (highlighted in table 11) in different steps of carcinogenesis. Such correlation indicated the reliability of the initial analysis results and appropriateness of further validation. Markers identified from the POSH analysis are highlighted in grey.

3.3.3.2 ANX A2 ELISA validation

The Wessex analysis revealed ANX A3 as a promising biomarker in breast disease and therefore prompted further validation of ANX A2 here. ANX A2 levels were significantly higher in the serum of the good outcome compared to the poor outcome group (Figure 79). The mean ANX A2 concentration was 3.52 compared to 4.5 ng/ml in the poor and good outcome groups, respectively. This was consistent with the results revealed by the original MudPIT analysis confirming the significance of ANX A2 as a potential prognosticator.



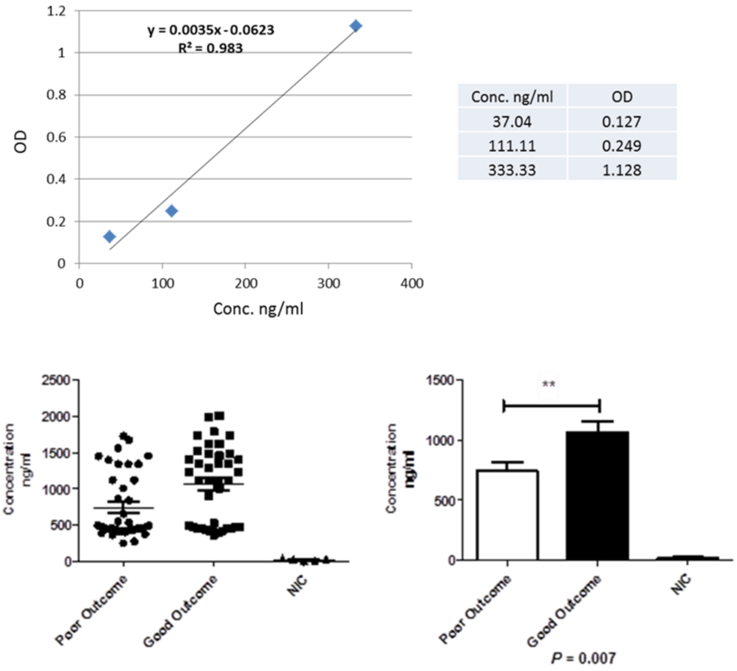
Conc. ng/ml	OD (mean)
40	11.494125
20	6.1158125
10	3.1935625
5	1.573875
2.5	0.7874375
1.25	0.3775625
0.625	0.2166875
0	0.0435

Figure 79. ANX A2 ELISA validation as a prognostic marker in the POSH analysis

ANX A2 levels in serum from the POSH cohort using a commercially available assay (E91944Hu, Uscn Life Science Inc.). Analysis of 36 poor vs. 36 good outcome samples revealed significantly higher ANX A2 levels in the good outcome patients (*P* = 0.003). Illustrated also are the standard curve and negative controls used in this analysis.

3.3.3.3 ApoC1 ELISA validation

Apo C1 was found to be a promising breast cancer marker in previous work as well as in other reports. Moreover, IPA pathway analysis indicated its strong connection to mammary tumourogenesis. Hence, immunovalidation analysis was performed to support the evidence revealed by our MudPIT analysis. Here, a higher serum Apo C1 level was shown to be associated with better outcome in patients recruited in the POSH study (Figure 80). The mean serum Apo C1concentration was 1μg/ml in the good outcome group, whereas the mean concentration in the poor outcome group was 0.74 μg/ml. These findings were based on ELISA assays involving 40 *vs.* 39 poor and good outcome serum samples respectively. These results added further evidence to the role of Apo C1 as a prognostic breast cancer biomarker and the successfully validate the MudPIT study.



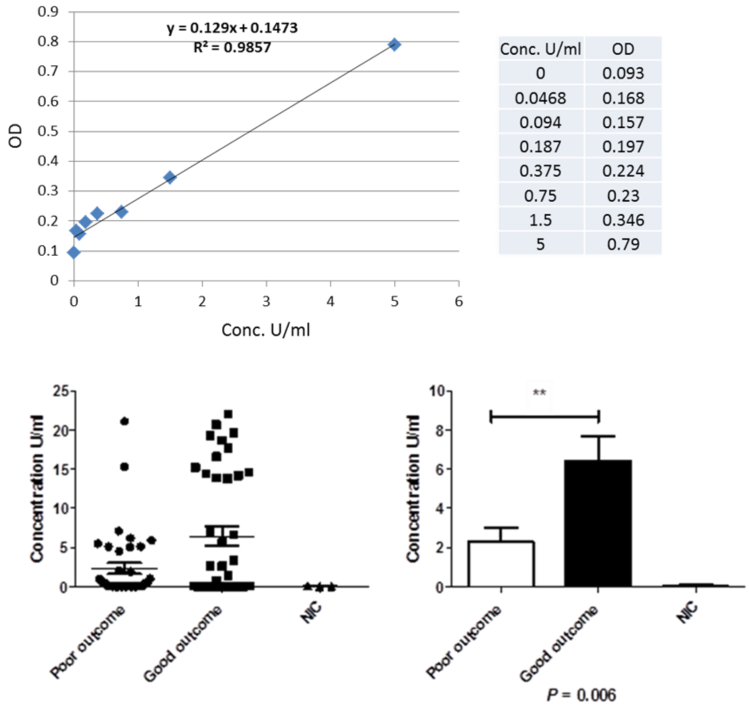
Conc. ng/ml	OD
37.04	0.127
111.11	0.249
333.33	1.128

Figure 80. ELISA validation of Apo C1 as a prognosticator of EBC

Immunovalidation of serum ApoC1 levels in the POSH cohort using a commercial ELISA assay (E90252Hu, Uscn Life Science Inc.). 40 poor vs. 39 good outcome samples were analysed. Significantly higher ANX A2 levels were shown in the good prognostic outcome patients (*P* = 0.007). Illustrated also are the standard curve and negative controls used in this analysis.

3.3.3.4 NOS2 ELISA validation

Based on the pathway analysis of the candidate prognosticators reported in the POSH MudPIT analysis, NOS2 was subjected to ELISA analysis in an attempt to assess its performance as a serum breast cancer prognosticator. A significantly higher NOS2 level was confirmed in the good outcome group than the poor outcome group (Figure 81). The mean concentration of NOS2 in the good outcome group was 6.4 U/ml. On the other hand, the poor outcome sera had a mean concentration of 2.3 U/ml. The findings illustrated here were consistent with the original study results and indicated 2.78 fold higher levels of NOS2 in the good outcome samples.



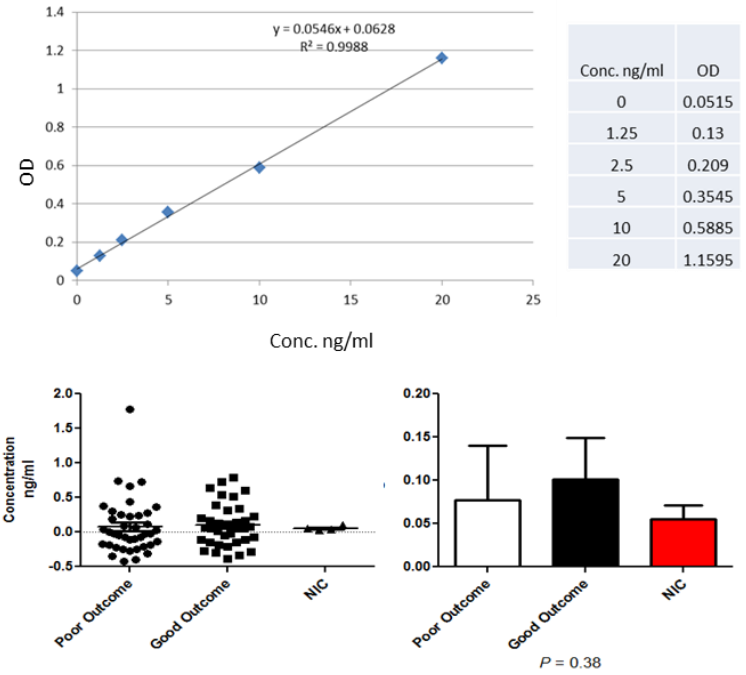
Conc. U/ml	OD
0	0.093
0.0468	0.168
0.094	0.157
0.187	0.197
0.375	0.224
0.75	0.23
1.5	0.346
5	0.79

Figure 81. ELISA test confirming NOS2 as a candidate prognosticator in EBC

Immunovalidation of serum NOS2 levels in the POSH cohort using a commercial ELISA assay (E90837Hu, Uscn Life Science Inc.). 39 poor vs. 39 good outcome samples were analysed. Significantly higher NOS2 levels were shown in the good prognostic outcome patients (*P* = 0.006). Illustrated also are the standard curve and negative controls used in this analysis.

3.3.3.5 DNMT1 ELISA validation

DNMT1 is a recognised cancer marker. A total of 78 samples were analysed by ELISA and showed no significant difference between the two groups (Figure 82). The mean DNMT1 serum concentration was 0.8 *vs.* 1 ng/ml in the poor and good outcome groups correspondingly; contrasting with the data from the MudPIT analysis.



Conc. ng/ml	OD
0	0.0515
1.25	0.13
2.5	0.209
5	0.3545
10	0.5885
20	1.1595

Figure 82. Failed DNMT1 ELISA validation as an EBC biomarker in the POSH analysis

Immunovalidation of serum DNMT1 levels in the POSH cohort using a commercial ELISA assay (E98244Hu, Uscn Life Science Inc). 40 poor vs. 38 good outcome samples were analysed. Although lower DNMT1 levels were initially shown in the good outcome group, inconsistently with the MudPIT analysis DNMT1 levels were higher in the good outcome group. This was not statistically significant (*P* = 0.38) and DNMT1 validation was deemed unsuccessful. Illustrated also are the standard curve and negative controls used in this analysis.

Protein Name	**Poor/Good Ratio**	***P*-value**
Ecto-NOX disulfide-thiol exchanger 1	0.012225795	0.00035843
Keratin, type I cytoskeletal 9	0.012269939	0.00036076
Glutathione peroxidase 3	0.021980977	0.0010329
Protein Jade-1	0.026269702	0.00142616
G2/mitotic-specific cyclin-B3	0.02901538	0.00170763
Nitric oxide synthase2	0.039160839	0.00294205
Golgin subfamily B member 1	0.041245791	0.00323265
Apolipoprotein L6	0.045075125	0.00379836
Forkhead box protein J1	0.045383412	0.00384569
Calcium-dependent secretion activator 1	0.04622051	0.00397553
Gem-associated protein 6	0.053046001	0.00510613
L_PI3H4 Large structural protein	0.057580779	0.00592698
Corepressor interacting with RBPJ 1	0.061356932	0.00665218
Protein FAM184A	0.064782908	0.00734242
Apolipoprotein F	0.065023512	0.00739204
Peroxisome biogenesis factor 10	0.065136012	0.00741529
Ig kappa chain V-I region HK101 (Fragment)	0.069497	0.00834173
Solute carrier family 12 member 3	0.0775	0.01016724
Phosphatidate phosphatase LPIN1	0.079335793	0.01060852
Protein diaphanous homolog 2	0.080952381	0.01100398
Zinc finger CCCH domain-containing protein 3	0.083129584	0.01154669
Thrombospondin type-1 domain-containing protein 7B	0.084745763	0.01195702
Transmembrane protein 63B	0.08585034	0.01224109
Coiled-coil domain-containing protein 123, mitochondrial	0.086019874	0.01228496
Tripartite motif-containing protein 59	0.088202867	0.01285591
Cell division protein kinase 9	0.091094891	0.01362985
Ig kappa chain V-II region GM607 (Fragment)	0.092514833	0.01401711
Rho GTPase-activating protein SYDE2	0.095849057	0.01494512
Ribosome biogenesis protein BMS1 homolog	0.097826087	0.01550765
Keratin, type II cytoskeletal 2 oral	0.098382749	0.01566768
Apoptosis-stimulating of p53 protein 2	0.098591549	0.01572789
Serine/threonine-protein kinase ICK	0.098901099	0.01581734
NCK-interacting protein with SH3 domain	0.101068999	0.01644998
Atrial natriuretic peptide-converting enzyme	0.101123596	0.01646605
Lumican	0.101338432	0.01652936
Outer capsid glycoprotein VP7 OS=Rotavirus A	0.101827676	0.01667393
Transcriptional regulator ATRX	0.105680736	0.01783157
Condensin-2 complex subunit G2	0.111891069	0.01976791
Shugoshin-like 1	0.112586605	0.01999013
ATP-binding cassette sub-family A member 12	0.113611615	0.02031956
Ig heavy chain V-I region EU	0.11808209	0.02178328
DPOG1_HUMAN DNA polymerase subunit gamma-1	0.12102405	0.02277025
LCA5_HUMAN Lebercilin	0.129812981	0.02582883

Protein Name	**Poor/Good Ratio**	***P*-value**
SWI/SNF complex subunit SMARCC2	0.129860325	0.02584575
Nebulin	0.136919315	0.02841993
Nitric oxide synthase, inducible	0.140684411	0.02983457
Ig kappa chain V-I region Lay	0.144201434	0.03118175
Zinc fingers and homeoboxes protein 1	0.148877691	0.03301092
T-lymphoma invasion and metastasis-inducing protein 2	0.152489627	0.03445301
Apolipoprotein C-I	0.152745802	0.03455625
Ephrin type-A receptor 10	0.155178268	0.03554281
Amphoterin-induced protein 3	0.156415695	0.03604901
Leucine-rich repeats and immunoglobulin-like domains prote	0.156648452	0.03614455
Envoplakin	0.160150376	0.03759437
ATP-dependent DNA helicase 2 subunit 2	0.160955348	0.03793089
Laminin subunit beta-4	0.162314749	0.03850193
Plexin-A2	0.163278272	0.03890875
POLG_HE701 Genome polyprotein enterovirus 70	0.164658635	0.03949457
Paired amphipathic helix protein Sin3b	0.16773557	0.04081298
Annexin-2 receptor	0.168262855	0.04104065
Zinc finger CCCH domain-containing protein 6	0.168845935	0.041293
ENV_HV1JR Envelope glycoprotein gp160	0.170068027	0.04182389
Myotubularin-related protein 8	0.171014493	0.0422369
Bloom syndrome protein	0.175615627	0.04426749
Sarcosine dehydrogenase, mitochondrial	0.17761807	0.04516292
Phosphatidylethanolamine-binding protein 4	0.178226514	0.04543639
Epidermis-specific serine protease-like protein	0.178352941	0.04549329
R1AB_CVHN1 Replicase polyprotein 1ab coronavirus HKU1	0.181642512	0.04698368
CAP-Gly domain-containing linker protein 1	0.182644628	0.04744142
Putative tyrosine-protein phosphatase TPTE	0.185210544	0.04862128
VE2_HPV05 Regulatory protein E2	0.18627451	0.0491138
ATP-dependent RNA helicase DDX24	4.327300151	0.04970212
Inter-alpha-trypsin inhibitor heavy chain H3	4.333333333	0.04957651
Myomegalin	4.339130435	0.04945627
Interferon-induced very large GTPase 1	4.428571429	0.04765492
Lysine-specific demethylase 4C	4.433333333	0.04756178
Potassium voltage-gated channel subfamily A member 1	4.440944882	0.04741346
DNA (cytosine-5)-methyltransferase 1	4.54893617	0.04538147
Glycine dehydrogenase [decarboxylating], mitochondrial	4.554376658	0.04528257
Arginase-1	4.596287703	0.04453133
VP3_ROTHT Protein VP3 OS=Rotavirus A	4.603082852	0.04441128
Prickle-like protein 3	4.632780083	0.04389228
Histone H2B type 1-A	4.684044234	0.04301748
Ribosome biogenesis regulatory protein homolog	4.696428571	0.04281005
Ig kappa chain V-I region Wes	4.739936102	0.04209309
Protein patched homolog 1	4.7578125	0.0418037

Protein Name	**Poor/Good Ratio**	***P*-value**
Sickle tail protein homolog	4.763636364	0.04171006
Ig heavy chain V-II region NEWM	4.814814815	0.04090054
Dynein heavy chain 9, axonemal	4.865979381	0.04011452
Phosphatidylinositol-4-phosphate 3-kinase C2	4.905325444	0.03952534
DNA-dependent protein kinase catalytic subunit	4.929032258	0.03917658
Neuronal pentraxin-1	4.967948718	0.03861398
Ig kappa chain V-III region SIE	4.968264406	0.03860947
Cullin-associated NEDD8-dissociated protein 1	4.989473684	0.03830798
N-terminal acetyltransferase complex ARD1 subunit	5.058219178	0.03735466
PLSI_HUMAN Plastin-1	5.112403101	0.03662809
CAN3_HUMAN Calpain-3	5.128205128	0.03642018
Zinc finger CCHC domain-containing protein 10	5.204225352	0.03544428
THO complex subunit 7 homolog	5.397683398	0.03313063
RNA-binding protein 28	5.473015873	0.03229061
Zinc finger protein 429	5.520746888	0.0317748
Parkin coregulated gene protein	5.52866242	0.03169046
Interferon-inducible GTPase 5	5.532894737	0.0316455
PHD finger protein 6	5.559550562	0.03136452
USP6 N-terminal-like protein	5.582278481	0.03112789
EF-hand calcium-binding domain-containing protein 3	5.641255605	0.03052622
Ferredoxin-fold anticodon-binding domain-containing protein	5.68013468	0.03013908
Protein tyrosine phosphatase domain-containing protein 1	5.744186047	0.02951717
R3H and coiled-coil domain-containing protein 1	5.894060995	0.02813514
DENN domain-containing protein 4C	6.070422535	0.02662984
Uncharacterized protein C12orf52	6.136263736	0.02609875
Outer capsid protein VP4	6.215775159	0.02547834
CD2-associated protein	6.220588235	0.0254415
KDEL motif-containing protein 2	6.23556582	0.02532736
Ig gamma-3 chain C region	6.268076573	0.02508225
E1A-binding protein p400	6.373417722	0.02431208
Dystrophia myotonica WD repeat-containing protein	6.846153846	0.0212584
Endophilin-A3	6.953900709	0.02064332
Nuclear pore membrane glycoprotein 210	6.988188976	0.02045319
Metal transporter CNNM2	7.268468468	0.01899323
40S ribosomal protein S4, X isoform	7.284023669	0.01891685
Nucleotide-binding oligomerization domain-containing pro.	7.285714286	0.01890857
VP4_ROTHP Outer capsid protein VP4	7.298319328	0.01884705
Polycomb complex protein BMI-1	7.335839599	0.0186657
A-kinase anchor protein 9	7.476190476	0.01801011
Ig alpha-2 chain C region	7.481398975	0.01798645
Netrin-G1	7.505882353	0.01787586
Ubiquitin-like domain-containing CTD phosphatase 1	7.600732601	0.01745697
Grainyhead-like protein 1 homolog	7.844611529	0.01644538

Protein Name	**Poor/Good Ratio**	***P*-value**
VP4_ROTHN Outer capsid protein VP4	7.994974874	0.01586506
Tropomyosin alpha-1 chain	8.064377682	0.01560756
Cyclic nucleotide-binding domain-containing protein 1	8.26929982	0.01488294
Peroxisomal Lon protease homolog 2	8.373333333	0.01453431
Brain-specific angiogenesis inhibitor 2	8.915712799	0.01290156
Sodium/glucose cotransporter 1	8.958115183	0.01278572
Ig lambda chain V-III region SH	9.149748744	0.01228126
Formin-binding protein 4	9.214285714	0.0121181
Putative uncharacterized protein C1orf197	9.331010453	0.0118312
Synaptotagmin-2	9.337837838	0.01181474
Mixed lineage kinase domain-like protein	9.726973684	0.01093061
Large proline-rich protein BAT3	9.808121019	0.01075883
Uncharacterized protein C9orf104	10.03926702	0.01029125
Coagulation factor XIII B chain	10.23125	0.0099258
Anoctamin-6	10.46581197	0.00950525
SWI/SNF complex subunit SMARCC1	10.64308682	0.00920499
Ig heavy chain V-I region HG3	10.86479592	0.00884917
Fibrinogen beta chain	11.43093923	0.00802936
Latent-transforming growth factor beta-binding protein 4 O	13.21787709	0.00607648
Mediator of RNA polymerase II transcription subunit 26	13.65811966	0.00570551
NS2A_CVHOC Non-structural protein 2a	13.6977887	0.00567378
Zinc finger protein 680	13.90666667	0.00551097
Myosin-IXa	14.89928058	0.00482587
Ig kappa chain V-III region GOL	15.68559838	0.00437036
ATP-binding cassette sub-family A member 2	19.44224924	0.00288617
Ig lambda chain V-VI region WLT	28.26984127	0.00139531
Ig kappa chain V-II region Cum	43.15396825	0.00061107

Table 11. The potential prognostic biomarkers differentiating good and poor outcome patients from the POSH cohort proteomic analysis.

A total of 156 biomarkers were identified with at least 4 fold difference between the groups (*P*<0.05, FDR<%5). Candidate biomarkers qualifying as potentially clinically relevant (as indicated by IPA pathway analysis) are highlighted in grey. These markers are currently under further investigation for their clinical use and potential prognostic and/or therapeutic targets.

3.3.4 Discussion of the POSH proteomic analysis

Despite the increased use of neoadjuvant chemotherapy, and an overall decrease in breast cancer mortality in general, early onset breast cancer (EBC) is still assosciated with poor prognosis [255]. Although the exact causes of this phenomenon are still unclear, EBC is thought to be a unique disease entity with a more aggressive and less responsive trend. This can be partly attributed to higher grade and less hormonally responsive tumours in many EBC patients[255, 256].

Although several reports showed survival gain in patients treated with anthracycline based adjuvant therapy in early breast cancer, such benefit was modest and was accompanied by a risk of increased chemotoxicity. In particular, outcomes in the control arms of these trials show that 70% of patients derive adequate disease control from non-anthracycline-containing therapies, challenging the need to treat 'all comers' with anthracyclines[257, 258]. A risk benefit balance necessitates attempts to identify patient subgroups for which anthracycline and/or other chemotherapy based treatments offer an advantage. Selection of patients that are more likely to benefit from chemotherapy would also allow cost effective use of chemotherapy and reduce the risk of patients developing chemotoxic side effects with little or no efficacy. Finally, identifying EBC prognostic biomarkers could highlight novel resistance and/or predisposition pathways that can then be targeted to increase treatment efficacy.

From a treatment efficacy perspective, administering chemotherapy to patients who are likely to respond and not to those who are unlikely to respond is key. To this end, predictive factors including biomarkers of response to treatment need to be explored. So far, evidence suggests an association between aggressive tumour phenotypes, (including estrogen receptor ER negativity, higher grade and Ki-67 expression), and favourable response to anthracycline-based regimes[259]. Nonetheless, ER positive sub sets of HER-2 positive tumours were shown to have better response. Overall, molecular players involved in such observations remain scarce. A promising prognostic marker; topoisomerase II alpha (TOP2A) was reported as a potential prognosticator in EBC patients[260].

Functionally, anthracyclines bind TOP2A, and stabilise the DNA double-strand breaks, resulting in cell cycle arrest and apoptosis. Moreover, they induce the formation of formaldehyde and its associated oxygen radicles leading to cell death. Both mechanisms are believed to induce cell death. Despite conflicting evidence on the value of TOP2A as a marker of anthracycline adjuvant treatment response, It was reported that in the ER receptor positive subgroup of HER-2 positive EBC patients TOP2A over expression had better anthracycline treatment response[261]. TOP2A expression was shown to be a surrogate marker for tumour proliferation, but not as strongly associated with outcome as Ki-67[262]. Co expression of both HER-2 and TOP2A can be explained by their gene co-amplification due to their chromosomal location (chromosome 17)[260, 263]. The role of TOP2A in HER-2 negative tumours remains unclear.

The POSH study presented here targeted a critical area in breast cancer research. This involved analysing matched samples from patients who developed breast cancer and received neoadjuvant treatment before the age of 40. We aimed to elucidate novel EBC proteomic prognosticators that could lead to improved treatment stratification and personalised medical intervention. Critical biological pathways revealed by this effort could also be exploited for potential new pharmaceutical agents to alter overall outcome in both early and late onset breast cancer patients. The study design was chosen to alleviate individual non-specific variations using a large cohort, minimise bias by randomised sample collection, facilitate identification of controls that could be matched for all potential confounding factors and ensure accurate standardised and high quality data collection.

Applying a modified 3D MudPIT MS/MS analysis[264]; 6,500 serum proteins were identified of which over 1200 were phosphoproteins. 156 prognostic biomarkers were found to differentiate good from poor outcome sera with high confidence ($P< 0.05$, FDR < 5%). Further molecular pathway analysis demonstrated that these markers were involved in relevant cancer pathways including DNA damage response, hereditary breast cancer, molecular mechanisms of cancer and prolactin signalling. In addition, several markers emphasised in the pathway analysis were the focus of further immunovalidation. Three of these markers (ANX A2, Apo C1 and NOS2) were successfully validated by ELISA assays using an independent validation cohort, illustrating consistent results with the original MudPIT analysis. Here, it was confirmed that their over-expression was associated with a more favourable post treatment outcome in early onset breast cancer. A fourth candidate marker (DNMT1) however, failed the ELISA validation and showed no differential expression between the two groups. This can be explained by several factors. Firstly, a potential mismatch may have occurred between the conformational and linear epitope in the ELISA as opposed to the amino acid sequence based MS/MS detection. A lower detection limit threshold advantage in MS/MS analysis is another potential source for such finding. Moreover, interfering molecules and inter assay variability are reported limitations in ELISA validation[265, 266]. It is possible that the ELISA was sub optimal, suggested by an OD of below the detection threshold. This could be further investigated by repeated/additional ELISA analysis. However, this was deemed to be beyond the scope of this project. Despite this, three prognostic markers (ANX A2, Apo C1 and NOS2) were validated by ELISA as potential clinically applicable fit for purpose prognosticators of EBC.

The role of ANX A2 in breast cancer

Calcium-binding proteins, including annexins, are key molecules in signaling pathways. ANX A2 specifically plays a significant role in angiogenesis by activating the plasminogen/plasmin system, aggressiveness and metastasis[267, 268]. In addition, ANX A2 has been shown to be over expressed in several malignancies including prostate, hepatocellular and pancreatic carcinomas[269-271]. Enhanced expression of annexin A2 in immunohistochemistry (IHC) and the concurrent decreased expression of annexin A1 in breast cancers biopsies were reported to be associated with a poor response to neoadjuvant chemotherapy in large and locally advanced tumours[272]. Although an anti-apoptotic role of ANX A2 was thought to be the cause of chemoresistance in lung and breast cancer, ANX A2's role in early onset breast cancer has not been fully explored[273]. Recent evidence suggested an association between ANX A2/ tPA binding on breast cell surfaces and plasmin activation leading to invasive breast cell line migration and enhanced angiogenic activity[267, 268]. Moreover, multi drug resistant MCF7 cell lines (MCF7/ADR) were shown to have a higher invasive/metastatic propensity associated with ANX A2 over expression, a trend which was reversed following ANX A2 inhibition[274]. However, ANX A2 was not directly involved in the acquisition of multi-drug resistance.

In agreement with previous discoveries that indicated a role for ANX A2 in breast cancer, our results indicated that annexin A2 over expression was associated with good outcome in EBC patients following anthracycline based neoadjuvant chemotherapy. Discrepancies of ANX A2 expression behaviour between different studies need further investigation. We propose that conflicting evidence could be explained by the different *in vitro* cellular behaviour, *in vivo* pathway(s) in early onset breast tumours, or a unique pathway activation following intervention in these patients. ANX A2 in breast cancer patient sera was not studied in any of the previous studies, which could also be a source of discrepancy. Finally, false positive discovery is also a potential source for such discrepant finding; necessitating further validation and functional studies.

Apo C1 as a breast cancer marker

Bibliographic mining involving the candidate serum markers identified by pathway analysis was carried out. Here few of these markers were reported as diagnosis and/or prognosis breast cancer biomarkers. There are few reports that describe Apo C1 as a marker associated with breast cancer, though one study showed it to be an early predictor of early breast cancer recurrence[228]. Goncalves and colleagues illustrated that lower Apo C1 levels were associated with higher metastatic recurrence risk[223]. This is consistent with our finding were ApoC1 levels in the good prognosis group were significantly higher than the poor outcome group. Other work has reported lower Apo C1 levels associated with breast cancer compared to healthy controls[222]. Interestingly, Apo C1 was identified as a diagnostic breast cancer marker in the previous Wessex analysis and a prognostic marker in the POSH cohort. ApoC1 over expression was demonstrated to be diagnostic in IDC patients from the Wessex cohort and associated with good outcome in post treatment EBC. This could be explained by several factors including pre analytical variability, such as sample collection, storage and handling, different cohorts/populations as well as analysis platforms. Also, the molecular action of Apo C1 could differ from a diagnostic and prognostic point of view depending on its biological role in each process. This could also indicate that EBC patients possessing higher Apo C1 levels represent a sub group mimicking late onset sporadic disease, which generally poses a more favourable response. Overall, the frequent presence of Apo C1 as a breast cancer marker strongly suggests a possible biological role in breast disease tumourogenesis.

NOS2 expression in breast cancer

The POSH analysis revealed that higher NOS2 levels correlated with good prognosis. Increased NOS2 levels (indicating an inflammatory tumour environment) have been shown to be associated with Akt phosphorylation and its down-stream pro-apoptotic effectors caspase 9 and BAD (inhibiting their pro-apoptotic function). Other factors such as ER receptor status, HER2/neu status, p53 mutation, nodal involvement, menopausal status, BMI, smoking, age and ethnicity, were not associated with Akt pathway activation[275]. Clinical studies have found an association of iNOS expression and poor prognosis in breast cancer patients[276, 277], and the use of non-steroidal anti-inflammatory drugs (NSAIDs) (potentially affecting its pathway) with a reduction in the risk of breast cancer[278, 279]. Glynn *et al.* previously showed that ER negative patients and patients with basal tumours had a poorer prognosis as well as higher NOS2 levels[280, 281]. The evidence from the POSH cohort is in contrast to these findings and warrants further validation in multi-centre studies to confirm the exact role of NOS2 in breast pathology. The contrasting expression behaviour could be explained by different biological behaviour of early onset breast cancer patients. Clearly, this observation warrants further validation and functional evaluation.

DNMT1 in breast cancer

The POSH MudPIT analysis showed that DNMT 1 levels were five-fold higher in the poor outcome group. DNMT 1 plays a key role DNA hypermethylation and transcriptional silencing of tumour suppressor genes[282]. DNMT 1 has been shown to be involved in the development of many malignancies[283, 284]. DNMT 1 antagonist 5-AZA was shown to reverse ER receptor negative status and restore tamoxifen sensitivity in ER negative breast cancer cells[285, 286]. An increase of DNMT1 activation plays a role in the methylation of the PTEN promoter contributing to persistent Akt activation and tamoxifen resistance[287] . As such, DNMT 1 could be a potential therapeutic target for reversing tamoxifen resistance in breast cancer. The effect of DNMT1 on breast cancer and a BRCA1 association has been reported in different ethnic populations[288-290]. In addition, it has been shown to play a role in docetaxel chemoresistance[291]. Although DNMT1 ELISA validation in the POSH study was not achieved; the former reports are consistent with the POSH MudPIT analysis evidence indicating poorer prognosis was associated with higher serum DNMT 1 levels.

3.3.5 The POSH cohort proteomic analysis summary

The study presented here targeted a critical area in breast cancer research. Analysing matched samples from patients who have developed breast cancer before the age of 40 aimed to elucidate novel proteomic prognosticators. It was anticipated that differentially expressed markers from this study could lead to improved treatment stratification and personalised medical intervention. Critical biological pathways revealed by this effort could be exploited for potential new pharmaceutical agents to alter overall outcome. The study design was chosen to alleviate individual variations, minimise ascertainment bias, facilitate identification of controls that could be matched for all potential confounding factors and ensure accurate standardised and high quality data collection.

The application of 3D MudPIT MS/MS approach to the POSH samples generated 6,500 serum proteins and over 1200 phosphoproteins, of which 156 prognostic biomarkers were found. Several markers emphasised in the pathway analysis were the focus of further immunovalidation.

Three of these markers (ANX A2, Apo C1 and NOS2) were successfully validated by ELISA assays, illustrating consistent results with the original MudPIT analysis. Here, it was confirmed that their over-expression was associated with a more favourable post treatment outcome in early onset breast cancer. A fourth candidate marker (DNMT1), failed the ELISA validation and showed no differential expression between the two groups.

This work adds new evidence to the field of prognosticators in early onset breast cancer and opens new venues for novel biomarkers and pathway discovery where critical biological pathways could be exploited for new therapeutics. These findings should however be interpreted with caution and are subject to further confirmation by independent studies.

Chapter 4

Final Discussion

4. Final Discussion

Breast cancer is believed to result from the accumulation of oncogene mutations or rearrangements and silencing of tumour suppressor genes. During tumour development and/or growth, specific tumour and host secreted proteins, and plasma protein digested by tumour/host secreted proteases would be released into the blood. Generally, expression profiling biomarker discovery aims to provide new means for tumour diagnosis, classification and prognostication by targeting such molecules. This could lead to the creation of early detection modules, and personalised effective cancer therapies which would improve outcomes. To date, genomic expression profiling has successfully contributed to the classification and outcome prediction of breast cancer [38]. Genomic profiling has led to the use of several markers including ER, PR and HER-2 in directing breast cancer treatment as well as estimating prognosis [292, 293]. Generally, patients with hormone receptor (ER and PR) amplification have better prognosis as opposed to patients with HER-2 overexpression [293]. Consequently, this has led to the introduction of hormonal therapy involving either Tamoxifen, or an Aromatase Inhibitors in hormone receptor positive patients, and HER-2 monoclonal antibody therapy (Herceptin) for patients with HER-2 amplifying tumours [294]. Yet, this simple classification does not completely reflect the heterogeneity of breast disease as more markers and molecular subtypes have been illustrated over the last decade [37]. Such new molecular subtyping is expected to alter patients' stratification. Thus, more personalised and targeted treatments are expected to influence current management regimes.

Whilst huge strides have been made by the application of genomic techniques in this area, many crucial changes at the protein level including global patterns and PTMs may prove to be key players in carcinogenesis, and may still be missed by studies at the genomic level. It is, therefore, essential to correlate and study biological proteomic patterns and correlate them to tumour classification and clinical outcomes. The consistency of this approach was illustrated by recent proteomic studies, leading to sub grouping of breast cancer; parallel to the genomic profiling classification[47, 145]. Although proteomic research is challenged by the absence of real time protein monitoring, de novo sequencing and global profiling, biomarker discovery is a key field which has proven effective despite these factors[54, 68, 87, 88]. The dynamic range of the proteome coupled with varying isoforms are also tough challenges facing these efforts[221]. In addition, heterogeneity of the breast and its pathological status will continue to play part in this challenge. Recent evidence demonstrated heterogeneity within breast tumour cell lines [295], this in addition to reports demonstrating similarities of the molecular subtypes in both DCIS and IDC raise the question of how mutually exclusive are breast pathologies [296] ? Similarly, normal breast tissue is subject to a degree of heterogeneity under the influence of local environment and hormonal and/or cyclical tissue changes, which could lead to more frequent encountering of incidental benign breast changes in “normal” breast tissue.

Despite these challenges, breast cancer serum profiling in particular, remains the focus of many of these studies applying MS technology[63, 67, 123, 131, 136, 144, 157, 297, 298].

The detection of breast and other cancers at an early stage would improve the likelihood of cure, life expectancy and quality of life. Current breast cancer screening and diagnostic tools are limited. So it is important to identify alternative approaches/markers to detect breast cancer earlier. But why target serum proteomics despite many hurdles involving serum complexity, dynamic range and sample heterogeneity?

Firstly; proteins and/or their alterations represent the sum total of biologically active molecules responsible for most cellular function. The proteome is, therefore, an attractive area for novel biomarker discovery[91, 299]. Secondly, the use of blood samples for proteomic analysis is a simple, minimally invasive, reproducible and cost effective method which can be used as a simple adjunct in addition to current screening and diagnostic tests, and may possibly be used as an alternative to current screening tests in the future.

In our work, MS analysis was used in diagnostic and prognostic breast cancer biomarker discovery. Firstly, unbiased top down MS profiling was applied to explore the usability of archival samples in proteomic biomarker discovery; the Guernsey cohort analysis. Secondly, a similar approach was adapted to interrogate sera from patients with breast neoplasms and healthy controls for diagnostic proteomic marker discovery; the Wessex cohort analysis. SELDI with its high throughput capabilities and its ability to simplify crude biological samples using different chromatographic surface, represented a promising opportunity for semi-global breast cancer proteomic profiling in both these studies. Both studies were designed to add more evidence in the field of breast cancer diagnostics and possible monitoring of therapies.

In a parallel project, a shot gun proteomic approach applying a novel 3D MudPIT analysis platform was embraced to mine the sera of post treatment early onset breast cancer (EBC) patients for novel prognosticators[264]. Post treatment early relapse within the first two years was a striking phenomenon in many of the patients in the POSH cohort[194]; necessitating further evaluation and molecular elucidation; the POSH cohort analysis. This work was anticipated to reveal markers which could be applied in outcome prediction as well as targeting new aggressive disease pathways.

Archival samples MS proteomic analysis -The Guernsey Cohort analysis-

Orthodox proteomic practice advocates would debate the use of archival samples due to concerns over sample variability and decay. Clinical proteomics is, however, challenged by a paradox; where on one hand standardised sample handling is critical for relevant biomarker discovery, yet, clinically relevant biomarkers must survive the "real world" sampling and handling protocols[111]. A balanced study design with multicentre validations of candidate "archival" biomarkers is essential and could provide answers to such debates.

In an attempt to address this area, MS analysis of archival urine and serum samples was conducted in a pilot study. Optimisation of the analysis conditions was a key step in this work. It was evident that archival urine samples from the Guernsey cohort generated poor proteomic profiles and thus compromised the statistical significance of further analysis. Despite optimization of analysis conditions, the deficiency of protein peaks in archival urine profiles persisted. In the absence of renal and/or urological pathologies poor protein signals were expected. Additionally, added to a sample ageing/degradation where thought to affect these archival urine samples. We concluded that archival urine proteomic profiling from the Guernsey cohort was impractical.

In contrast to the urine analysis, profiling archival serum from the same cohort revealed promising results[300]. It was clear that WCX profiling was the optimal condition for higher peak yield. This was shown consistently and high peak detection correlation and reproducibility were achieved in two different centres[300]. In addition, analytical variability seen in this work was within the expected values for proteomic analysis. Interestingly, this was shown to be comparable to recent serum cohort profiling reproducibility, even when analysis was deferred for up to a week from preparation. The reproducibility of these results, across two independent centres, provided a new insight in proteomic research[300], where thousands of samples in tissue banks worldwide could potentially be utilised. The evidence provided here indicated that historic biobank serum collections may be used for the identification of disease biomarkers. This is a promising avenue; where large cohorts of predictive biomarker characteristics could be analysed.

We reported 16 potential biomarker peaks detected in this two cohort study. This unique finding was recently supported by reports translating MS based tumour bank serum discovery into immune based quantitative assays[110]. In conclusion, this work demonstrated the potential for proteomic biomarker discovery using archival serum samples. In agreement to our findings, more reports described archival cancer biomarker discovery[301, 302]. Further validation studies involving larger cohorts in a multicentre fashion are nonetheless needed to confirm the future of archival sample proteomics.

Diagnostic breast cancer serum biomarker discovery – The Wessex Cohort analysis-
In the second phase of this project, top down MS proteomic profiling of a prospective collection (The Wessex cohort[300]) was used to probe potential diagnostic serum breast cancer biomarkers. Many MS based diagnostic studies rely upon the comparison between serum protein profiles obtained from healthy individuals and from diseased patients in the search for clinical markers[88, 130, 224, 300]. Similarly, we hypothesised that there would be detectable differences within a complex serum protein background in breast cancer groups. Considering the extreme dynamic range and variable concentration of proteins present in serum, it was unlikely that a simple single dimension fractionation coupled with a limited dynamic range detection MS in global profiling mode could achieve novel biomarker discovery. Therefore, it seemed unlikely that low abundant proteins would be reliably detected by the simple SELDI platform. In addition to the broad affinity characteristics provided with its chromatographic arrays, further enrichment tools were needed to uncover low abundance biomarkers in serum. In the last decade, many promising biomarkers have been invalidated prior to clinical use. This may be due to many pre, peri and post analytical parameters complicating biomarker discovery, and validation risk resulting in “false discovery”[303]. Hence, serum samples from the Wessex cohort were matched for age, menopausal status, and stage. They were also enriched by (SAX) chromato-focusing fractionation. Optimisation of this cohort followed stringent protocols testing different analytical conditions to control for pre and peri analytical variables. It was shown that the most informative profiling settings would involve analysing denatured serum, SAX fraction 1 (F1) and pooled SAX fractions 5 and 6 (F5/F6) using WCX and pH 4.5 buffer conditions. Fraction 5 (F5) had a relatively low peak yield, thus samples from this fraction were pooled with corresponding samples from fraction 6 (F6) prior to analysis. Although this may have resulted in data loss from both fractions, it was proposed as an efficient technique to avoid poor profiling, and compromising stringent discovery from the peak poor fraction 5 (Personal communication with Bio-Rad and collaborators). The success of this pooling step was examined and revealed improved reproducible profiles. Following the assessment and optimisation of several preparations and profiling conditions, robotic array preparation was conducted to minimise peri analytical variability, and MS profiles generated using WCX arrays. In an effort to minimise false discovery, all samples were analysed blindly in duplicate and two overlapping analyses were performed for all samples using low and high mass range detection settings. Moreover, a pooled serum (QC) sample made from equal volumes of each fraction was included on one spot of each array to control analytical peri/post variability.

This standardised and highly optimised preparation protocol was then used to profile fractions from 347 serum samples (210 healthy, 73 breast cancer and 64 benign breast disease) by SELDI. This large cohort was essential to ensure adequate study power. Putative proteomic peaks were validated using an independent set and identical protocols. Initially, nine differentially expressed peaks were detected between the cancer and healthy control groups. In addition, eleven peaks differentiated the cancer and benign disease group. Finally, seven candidate biomarkers showed differential patterns between the healthy and benign breast disease group. Reproducibility is a prerequisite for any comparison analysis, thus we tested within-run and between-run reproducibility in order to evaluate the precision of our protein profiling. Variations between spectra were determined by comparing the relative peak intensities of selected protein mass signals across a wide mass and intensity range. In line with previous reports[81, 123, 131], our findings demonstrated successful detection of potential breast cancer biomarkers based on MS analysis. Evidently, it was difficult to compare results of different studies, as there are huge variations in the analytical conditions used which would lead to discovery of different subsets of biomarkers. In addition the bioinformatics involved in data analysis, may well be another source of discrepancy in the profiling efforts.

To further validate our findings and to minimise the risk of false biomarker discovery, an independent, but identical MALDI analysis of the same cohort was conducted in collaboration with St. George's University of London. Here, three of the 27 ion peaks (6.4, 9.5 and 15.9kDa) were validated as potential breast cancer biomarkers. The failure of several markers originally identified by SELDI profiling may be related to several factors including platform sensitivity/resolution, sample decay, and analysis methods. More importantly is the validation of these three markers across centres, a promising sign of their performance as "real world" markers[303].

Biomarker purification experiments were performed to isolate the proteins of interest prior to their identification. This was proved to be a lengthy and challenging stage, typical of top down MS biomarker discovery[175]. Purification of candidate markers is usually dependent on gel electrophoresis (GE) and many limitations can prevent successful biomarker identification. The tendency towards LMW bias in MS profiling leads to difficulty when using LMW band deficient GE for protein purification. In addition, the mass detection may differ between the two platforms, which cannot be ignored. Finally, proteome behaviour within the two distinct analysis environments can add to purification problems. An additional purification approach involving the use of gel free purification methods was tested. For this, FFE was used to fractionate pooled samples from different fractions. This pre-fractionation using IEF-FFE, had previously been shown to provide high-resolution fractionation with minimal sample loss[218, 219].

This was followed by a second dimension GE providing multilevel purification and concentration of low-mass serum proteins. This approach ensured increased signal intensities of the desired markers, aiding downstream analysis as shown in other profiling studies[304]. This extensive process led to the identification of two biomarkers, Apo C1 (9.5kDa) and ANX A3 (15.9kDa). The validity of both markers was confirmed by ELISA using independent serum sample sets. On-going work is investigating the identity of the third marker reported at 6.4kDa.

The identification of over expressed Apo C1 as a marker of IDC is an interesting observation. Apolipoproteins are synthesised in the liver and mainly act within lipid carrier lipoprotein metabolism. In addition to their recognised role in lipid transport, apolipoproteins have been reported to regulate many cellular functions. Apo C1 over expression was previously described as an anti-apoptotic factor associated with poor prognosis in pancreatic cancer[229]. Although the molecular role of Apo C1 in breast cancer is unclear, it was previously described as a potential predictor of metastatic relapse in early onset breast cancer[223]. Moreover, Apo C1 over expression was shown to be a potential early breast cancer prediction marker[227]. However, Apo C1 is an acute phase reactant and this lack of disease specificity could limit its application as a clinical biomarker[223, 229, 243]. Discrepancies in serum Apo C1 over/under expression in this and other cancer studies is confusing. This may be explained by differences in sample population, collection, processing and storage. Moreover, trends of Apo C1 differential expression in pre-diagnostic stages of breast cancer compared to symptomatic or detectable stages could reflect a unique biological role in different disease groups[223, 227]. A potential element of false discovery is also possible. Nonetheless, the molecular role of Apo C1 in breast cancer could be exciting given its recurrent identification as a breast cancer biomarker and the scarce reports on its function in mammary pathology.

In our study, serum ANX A3 was validated as an over-expressed marker in benign breast disease. Annexins are unique membrane binding proteins with various molecular functions[305]. Each annexin is composed of two principal domains: a variable NH2 terminal “head” (16 or less amino acids) and a conserved COOH-terminal protein core harbouring the calcium and membrane binding sites outlining an annexin core comprising four segments of internal homology[305] (Figure 83).

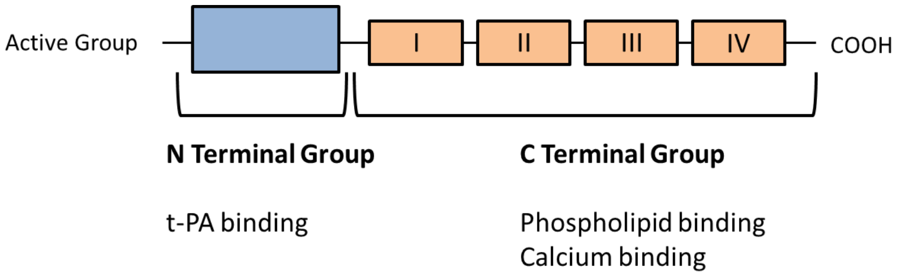


Figure 83. A schematic diagram of annexin A3 structure
The N-terminal domain consists of tissue plasminogen activator (t-PA) binding sites. The C-terminal domain consists of four repeating domains, each contains ANX A3 consensus sequences and the biding sites for calcium and phospholipids.

Although MS analysis showed ANX A3 over expression in BBD compared to IDC, this trend was not statistically significant was significant in the ELISA validation. This could be explained by the higher sensitivity and favourable quantification abilities of targeted ELISA compared to global top-down MS analysis.

More importantly, we described a large scale, multi-site, top down MS proteomic biomarker discovery study. A 15.9kDa serum biomarker was confirmed as a BBD marker in different centres using different top down analysis platforms. This peak was identified as the phospholipid and calcium binding protein human annexin A3 (ANX A3). Differential expression of ANX A3 in breast pathology has not been previously reported. The novel MS findings were validated by ELISA which showed that high levels of serum ANX A3 correlated with BBD. The levels of annexins were previously found to be differentially expressed in intestinal, pancreatic, hepatocellular, prostatic, renal and lung malignancies, as well as hairy cell leukaemia, thyroid tumour progression and ovarian cancer chemoresistance[224, 226, 230-232].

The sensitivities and specificities of ANX A3 levels as a BBD differentiator ranged between 81-95%. These results suggest that combined with other markers and/or diagnostic tests, ANX A3 could be a reliable biomarker for BBD diagnosis. However, these remain preliminary findings that should be interpreted with caution. Mammographic detection of small, non-palpable breast abnormalities has led not only to improved early-stage breast cancer detection, but also a raising undetermined breast lesions incidence. Consequently, more surgical biopsies are needed involving surgery related risks and representing a significant element of screening costs. In this respect, ANX A3 could be a beneficial marker in BI-RADS category 3 (probably benign) and 4 (suspicious) patients or impalpable breast lesions. Compared to the general population risk of 5%, the risk of breast cancer increases by 1% with the diagnosis of non-proliferative breast tumours[234]. Yet, patients with lower risk non proliferative BBD undergo invasive intervention with associated morbidity[235, 306], to help establishing definite diagnosis. Although reports suggest that the evaluation of mass margins and morphologic of features of calcifications help distinguish a malignant lesion from a probably benign finding, this remains a subjective evaluation. Surrogate markers such as serum ANX A3 could play a key supporting role in such scenarios[237]. In these patients, ANX A3 levels along with other markers could be used as part of a risk stratification system influencing the need for invasive biopsies and avoiding unnecessary risk and cost. More efforts and ANX A3 validation studies should aim at developing inexpensive and cost-effective tests, allowing specific identification of high-risk BBD patients at diagnosis and managing them more aggressively.

Hyperproliferative lesions of the breast such as ductal hyperplasia are associated with increased risk of breast cancer[233-235]. Further evaluation of ANX A3 expression and function among different benign breast disease groups such as hormonally driven (eg. atypical ductal hyperplasia (ADH) and columnar cells changes) ER-negative lesions (such as apocrine metaplasia) is thus an essential part of future ANX A3 validation efforts. Equally, the expression behaviour of ANX A3 in DCIS and different IDC molecular sub groups (basal, ERBB-2 and luminal) needs larger sub-entity focused studies involving biological fluids, tissue expression and translational functional studies.

The role of ANX A3 in malignant transformation and/or progression should be explored. ANX A3 was previously defined as an independent prognostic factor related to neoplastic progression[236]. In breast neoplasms, the subsequent diagnosis of IBC following BBD diagnosis was reported to be more likely in older, postmenopausal women, and patients with a strong family history of breast cancer[237]. Clearly, biomarkers such as ANX A3 could provide new insights into the molecular and cellular basis of mammary tumourigenesis. This could hold potential relevance in earlier detection, stratification, intervention and new treatment options.

As the role of ANX A3 in breast cancer remains unclear, elucidating its potential connection to hormonal pathways is of interest. Oestrogen synergises with progesterone to promote breast proliferation and both play a key role in breast tumourigenesis[238, 239]. A potential link between ANX A3 and major hormonal receptors as well as prolactin was identified using pathway analysis, justifying future ANX A3 functional experiments. Moreover, ANX A3 has been previously described as a phospholipase A2 inhibitor and is reported to play anti-apoptotic, angiogenic and growth regulatory roles[307-309]. Although the actual role of ANX A3 in carcinogenesis is less defined, based on the established role of ANX A3 and other homologus annexins, a theoretical ANX A3 pathway involving plasminogen activation, ECM degradation and angiogenesis promotion can be postulated (Figure 84).

In the present study, ANX A3 was down regulated in cancer and control groups compared to the benign group. BBD ANXA3 was up-regulated by an average of 1.25 fold (between 1.2 and 1.3-fold with 95% confidence; $p< 0.05$) suggesting that at least in certain types of benign breast disease ANX A3 abundance may be involved in non-cancerous disease phenotypes. A paradoxical down regulation of ANX A3 in prostate cancer as opposed to an up-regulation in advanced and metastatic prostate cancer has been shown (US patent 2003/0108963). In addition, ANX A3 driven therapeutic evaluation for urological and intestinal tumours is on-going (US Patent 7732148, Jun 2010). A functional role of annexin A3 in breast cancer as suggested by our findings is therefore worthwhile exploring.

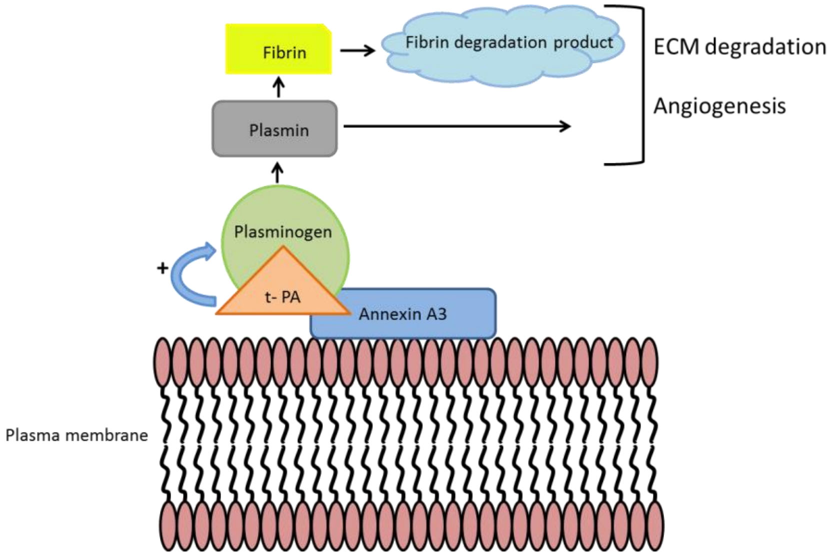


Figure 84. Proposed mechanism of annexin A3 role in carcinogenesis and metastasis
Proposed mechanism of annexin A3 promoting carcinogenesis and metastasis through plasminogen activation system. Cell surface ANX A3 binds to t-PA and activates plasminogen conversion to plasmin. Plasmin results in activation of metalloproteases and leads to extracellular matrix degradation. Increased annexin A3 expression results in increased plasmin generation and enhances angiogenesis, cancer invasion and metastasis.

Could ANX A3 under expression negatively affect the lactogenic environment, and hence lead to the loss of a tumour protective microenvironment? Does the differential expression of ANX A3 modulate the phenotypic behaviour of morphologically benign epithelial cells and hence may be an important determinant in initiating, or promoting human mammary cancers? To date, there are no reports on the association of ANX A3 with breast cancer and the biological functions of ANX A3 and its significance in breast tumourigenesis is unclear. Whether this ANX A3 up-regulation persists over years representing a stable protective phenomenon is for future follow up studies to clarify.

In conclusion, using MS-based profiling approach on a breast cancer screening population prior to intervention, we have identified a serum proteomic profile that may differentiate breast cancer, benign breast disease and healthy control sera. Two markers (Apo C1 and ANX A3) were identified and validated using an independent cohort and ELISA. Validation on a larger and independent population as well as identification of the third marker (6.4kDa) reported in this study is currently in progress. These data suggest a diagnostic value for serum in breast cancer patients and supports the prospective collection of serum samples in translational studies to diagnose and predict clinical outcome and/or therapeutic response.

Early onset breast cancer prognostic biomarker discovery -The POSH cohort analysis-
Although neoadjuvant therapies are widely used in the younger breast cancer age groups, and despite an overall decrease in mortality from breast cancer, early onset breast cancer mortality rate has not decreased over the years[185, 186]. Early onset breast cancer (EBC) is commonly associated with poor prognosis[187]. This paradoxical behaviour between early and later onset disease can be explained by early diagnosis in the older age group who are targeted in screening programs, and more importantly, a distinct biological behaviour between the two groups. This was demonstrated in a large study involving over 5,000 patients with EBC who had higher grade, larger tumours, which were more frequently lymph node positive over a median 11 year follow up period[185]. In our study, most of the 45% of deaths occurred within the first five years. Such poor prognosis has been previously attributed to higher grade tumours in this group of patients[188]. In agreement with these findings, over 40% of grade 3 tumours and 29% of grade 2 tumours died whereas no mortalities were recorded in grade 1 disease[185]. Other reports indicated that HER-2 positive and p53 over expression was associated with younger breast cancer patients and that HER-2 over expression correlated with lymph node metastasis[189]. Tissue microarray studies illustrated a link between basal type tumours and the proteins Fhit, Wwox, AP2α[48]. Yet, few studies have been directed towards early onset tumours which are a biologically distinct entity. The biological basis for the variable responses to neo-adjuvant chemotherapy and surgical treatment in EBC remains an unresolved issue. Although pre-treatment tumour grade and nuclear proliferative activity in the general population were shown to have a prognostic effect on response to anthracycline based chemotherapy[190], inconsistent findings in other studies cast doubt on this[191-193] .

Clearly, more evidence is needed from large cohorts representing EBC. Such evidence should focus not only on chemo-responsiveness, but also overall outcome, to enable effective personalised treatment plans. The distinct aggressive behaviour of early onset breast cancer necessitates extensive investigations of the molecular pathways involved in the disease in order to predict and improve outcome.

Potential prognostic serum markers in EBC were explored in this project using the POSH cohort and applying a 3D MudPIT MS/MS approach[264]. Here, post treatment serum samples from 399 early stage (stage 1 and 2), non-metastatic invasive ductal carcinoma patients were analysed. Three candidate prognosticators were identified and validated (ANX A2, Apo C1, and NOS2) using an independent cohort. The biological relevance of these markers in breast and other malignancies were discussed earlier. The identification of ANX A2 and Apo C1 as prognostic early onset breast cancer markers in the POSH cohort is an exciting finding as ANX A3 and Apo C1 were identified as diagnostic breast cancer biomarkers in the Wessex cohort. Although the molecular pathways of these markers in both disease entities may be distinct, this analogous observation in mammary neoplasm sera supports further exploration of their role as markers and therapeutic targets in breast tumourigenesis.

To date, several studies have investigated prognosticators of EBC[258, 258, 261]. Previous reports suggest that HER2/neu and p53 are markers of aggressiveness in EBC. These studies found that most of the patients presented with IDC and had lymph node involvement which correlated to HER-2 overexpression. In our cohort, we focused on early stage non-metastatic IDC, and found no prognostic role for HER-2 or p53. In agreement with our dismissed prognostic HER-2 findings, anthracycline based chemotherapy response was previously shown to be favoured by high grade and proliferative activity (mitotic activity and mitosin staining) and not by stage, HER-2 or lymph node status[190]. The small cohort size remains a major limiting factor towards any definite conclusions in this and similar studies[189].

The importance of novel prognostic biomarkers in EBC is related to the variable overall response to treatment. Reports suggest that 15% of EBC patients achieve a complete response and a similar proportion of patients display minimal change or progressive disease following treatment[190]. Currently, the underlying mechanism for the different responses is unknown. Causative factors may include the genetic and hormonal settings of the tumour. So far, studies have generated conflicting results regarding the positive or negative value of different markers in predicting breast cancer outcome. Therefore, the prognostic EBC biomarkers presented in this and similar work could play role in improved EBC treatment by aiding a personalised treatment stratification model, or elucidating new therapeutic targets. A major obstacle facing such efforts lie in conflicting results reported in different studies. This could be attributed to the various treatment protocols, different chemotherapy regimens, patient selection criteria, ethnic and genetic variation, and different biomarker assessment methods. More importantly, the vast molecular heterogeneity of breast cancer is likely to play a role. Eliminating non biological factors is essential to establish key disease pathways and help providing personalised treatment options optimising outcome and reducing sided effects.

In summary, the POSH analysis illustrated ANX A2, Apo C1 and NOS2 as prognostic serum markers in EBC patients. Further multi centre validation of the clinical utility and the functional role of these markers in EBC is a vital step towards future bedside translation of these findings.

In conclusion, this thesis added more evidence in the field of breast cancer biomarker discovery. Firstly; the potential role of archival serum sample in proteomic profiling was illustrated. Secondly; Apo C1 and ANX A3 were shown as diagnostic serum breast cancer biomarkers. Finally, ANX A2, Apo C1 and NOS2 were confirmed as early breast cancer prognosticators.

Future directions

- Independent multi-centre blinded validation of the diagnostic/prognostic performance of the previously reported markers is a long term object of this work. It is anticipated that a combined panel of markers is more likely to achieve clinically reliable diagnostic/prognostic value. To further understand the performance of each marker in different breast disease entities, larger cohorts covering different breast cancer sub groups as well as DCIS should also be involved in such efforts.

- Assessment of the candidate breast cancer biomarkers identified in this work and their expression in breast tissue from the same cohort is a key step towards evaluating the source of such differential expression in serum samples. Moreover, this may indicate the cellular localisation patterns and possible role as tissue tumour marker once confirmed in independent cohorts.

- Proteins secreted by cells into conditioned media *in vitro* should be studied to establish possible similar behaviour *in vivo*. Although *in vitro* conditions only partially represent *in vivo* settings, it important to map the potential biological pathways involved in tumourigenesis[310].

- Finally, functional studies on breast cell lines and animal models to dissect the specific mechanism(s) of action involved in mammary neoplasia echoed by these markers differential behaviour is an interesting venue for future projects.

References

1. Sawyers CL. The cancer biomarker problem. Nature 2008;452(7187):548-552.

2. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5(11):845-856.

3. Curtis C, Shah SP, Chin SF et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012.

4. Zeidan BA, Townsend PA. SELDI-TOF proteomic profiling of breast carcinomas identifies clinicopathologically relevant groups of patients similar to previously defined clusters from cDNA expression. Breast Cancer Res 2008;10(3):107.

5. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10(8):789-799.

6. Saraswat-Ohri S, Vetvicka V. New insights into procathepsin D in pathological and physiological conditions. N Am J Med Sci 2011;3(5):222-226.

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

8. Collins BC, Sposny A, McCarthy D et al. Use of SELDI MS to discover and identify potential biomarkers of toxicity in InnoMed PredTox: a multi-site, multi-compound study. Proteomics 2010;10(8):1592-1608.

9. Apweiler R, Aslanidis C, Deufel T et al. Approaching clinical proteomics: current state and future fields of application in fluid proteomics. Clin Chem Lab Med 2009;47(6):724-744.

10. Washburn MP, Wolters D, Yates JR, III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 2001;19(3):242-247.

11. Garbis SD, Roumeliotis TI, Tyritzis SI, Zorpas KM, Pavlakis K, Constantinides CA. A Novel Multidimensional Protein Identification Technology Approach Combining Protein Size Exclusion Prefractionation, Peptide Zwitterion-Ion Hydrophilic Interaction Chromatography, and Nano-Ultraperformance RP Chromatography/nESI-MS(2) for the in-Depth Analysis of the Serum Proteome and Phosphoproteome: Application to Clinical Sera Derived from Humans with Benign Prostate Hyperplasia. Anal Chem 2010.

12. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 2009;11:49-79.

13. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62(1):10-29.

14. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60(5):277-300.

15. Lawrence G. "An audit of screen detected breast cancers for the year of screening April 2006", NHSBSP, ABS at BASO. 11-6-2008. West Midlands Cancer Intelligence Unit, NHS Cancer Screening Programmes.

16. Varga D, Koenig J, Kuhr K et al. Comparison of early onset breast cancer patients to older premenopausal breast cancer patients. Arch Gynecol Obstet 2010;282(4):427-432.

17. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55(2):74-108.

18. Boyle P, Ferlay J. Cancer incidence and mortality in Europe, 2004. Ann Oncol 2005;16(3):481-488.

19. Wolfe JN. Breast parenchymal patterns and their changes with age. Radiology 1976;121(3 Pt. 1):545-552.

20. El-Bastawissi AY, White E, Mandelson MT, Taplin SH. Reproductive and hormonal factors associated with mammographic breast density by age (United States). Cancer Causes Control 2000;11(10):955-963.

21. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996;276(1):33-38.

22. Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. JAMA 2005;293(10):1245-1256.

23. Tabar L, Yen MF, Vitak B, Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet 2003;361(9367):1405-1410.

24. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137(5 Part 1):347-360.

25. Carney PA, Miglioretti DL, Yankaskas BC et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138(3):168-175.

26. Hansson SF, Simonsen AH, Zetterberg H et al. Cystatin C in cerebrospinal fluid and multiple sclerosis. Ann Neurol 2007;62(2):193-196.

27. Adam BL, Qu Y, Davis JW et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62(13):3609-3614.

28. Petricoin EF, III, Ornstein DK, Paweletz CP et al. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst 2002;94(20):1576-1578.

29. Mischak H, Apweiler R, Banks RE et al. Clinical proteomics: A need to define the field and to begin to set adequate standards. Proteomics Clin Appl 2007;1(2):148-156.

30. Poon TC. Opportunities and limitations of SELDI-TOF-MS in biomedical research: practical advices. Expert Rev Proteomics 2007;4(1):51-65.

31. Watson CJ, Khaled WT. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development 2008;135(6):995-1003.

32. LaMarca HL, Rosen JM. Minireview: hormones and mammary cell fate--what will I become when I grow up? Endocrinology 2008;149(9):4317-4321.

33. McNally S, Martin F. Molecular regulators of pubertal mammary gland development. Ann Med 2011;43(3):212-234.

34. Oakes SR, Hilton HN, Ormandy CJ. The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res 2006;8(2):207.

35. Ginestier C, Hur MH, Charafe-Jauffret E et al. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell 2007;1(5):555-567.

36. Eroles P, Bosch A, Perez-Fidalgo JA, Lluch A. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. Cancer Treat Rev 2012;38(6):698-707.

37. Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-10874.

38. Sorlie T, Wang Y, Xiao C et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics 2006;7:127.

39. Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-10874.

40. Parker JS, Mullins M, Cheang MC et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27(8):1160-1167.

41. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009;59(4):225-249.

42. Piccart-Gebhart MJ, Procter M, Leyland-Jones B et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 2005;353(16):1659-1672.

43. Fang R, Elias DA, Monroe ME et al. Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach. Mol Cell Proteomics 2006;5(4):714-725.

44. Diamandis EP. Cancer biomarkers: can we turn recent failures into success? J Natl Cancer Inst 2010;102(19):1462-1467.

45. van 't Veer LJ, Dai H, van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415(6871):530-536.

46. Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-10874.

47. Brozkova K, Budinska E, Bouchal P et al. SELDI-TOF proteomic profiling of breast carcinomas identifies clinicopathologically relevant groups of patients similar to previously defined clusters from cDNA expression. Breast Cancer Res 2008;10(3):R48.

48. Guler G, Huebner K, Himmetoglu C et al. Fragile histidine triad protein, WW domain-containing oxidoreductase protein Wwox, and activator protein

2gamma expression levels correlate with basal phenotype in breast cancer. Cancer 2009;115(4):899-908.

49. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422(6928):198-207.

50. Rodland KD. Mass spectrometry and biomarker development. Dis Markers 2004;20(3):129-130.

51. Rodland KD. Proteomics and cancer diagnosis: the potential of mass spectrometry. Clin Biochem 2004;37(7):579-583.

52. Petricoin EF, Ardekani AM, Hitt BA et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359(9306):572-577.

53. Zeidan BA, Cutress RI, Hastie C, Mirnezami AH, Packham G, Townsend PA. SELDI-TOF MS proteomics in breast cancer. Clinical Proteomics 2009;Online first(3).

54. Vlahou A, Schellhammer PF, Mendrinos S et al. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. Am J Pathol 2001;158(4):1491-1502.

55. Ranish JA, Yi EC, Leslie DM et al. The study of macromolecular complexes by quantitative proteomics. Nat Genet 2003;33(3):349-355.

56. Han DK, Eng J, Zhou H, Aebersold R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat Biotechnol 2001;19(10):946-951.

57. Villanueva J, Shaffer DR, Philip J et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 2006;116(1):271-284.

58. Liotta LA, Ferrari M, Petricoin E. Clinical proteomics: written in blood. Nature 2003;425(6961):905.

59. Muller CA, Markovic-Lipkovski J, Klatt T et al. Human alpha-defensins HNPs-1, -2, and -3 in renal cell carcinoma: influences on tumor cell proliferation. Am J Pathol 2002;160(4):1311-1324.

60. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the norwegian screening program. Ann Intern Med 2012;156(7):491-499.

61. Hegar V, Oliveira K, Kakarala B, Mangram A, Dunn E. Annual mammography screening: is it necessary? Am Surg 2012;78(1):104-106.

62. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Ann Intern Med 2011;155(1):10-20.

63. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin Chem 2002;48(8):1296-1304.

64. Li J, Orlandi R, White CN et al. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. Clin Chem 2005;51(12):2229-2235.

65. Laronga C, Becker S, Watson P et al. SELDI-TOF serum profiling for prognostic and diagnostic classification of breast cancers. Dis Markers 2003;19(4-5):229-238.

66. Becker S, Cazares LH, Watson P et al. Surfaced-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer. Ann Surg Oncol 2004;11(10):907-914.

67. Pawlik TM, Fritsche H, Coombes KR et al. Significant differences in nipple aspirate fluid protein expression between healthy women and those with breast cancer demonstrated by time-of-flight mass spectrometry. Breast Cancer Res Treat 2005;89(2):149-157.

68. Pusztai L, Gregory BW, Baggerly KA et al. Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. Cancer 2004;100(9):1814-1822.

69. Zolotarjova N, Martosella J, Nicol G, Bailey J, Boyes BE, Barrett WC. Differences among techniques for high-abundant protein depletion. Proteomics 2005;5(13):3304-3313.

70. Righetti PG, Boschetti E, Lomas L, Citterio A. Protein Equalizer Technology : the quest for a "democratic proteome". Proteomics 2006;6(14):3980-3992.

71. Huber LA, Pfaller K, Vietor I. Organelle proteomics: implications for subcellular fractionation in proteomics. Circ Res 2003;92(9):962-968.

72. Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. Proteomics 2005;5(13):3292-3303.

73. Hirabayashi J. Lectin-based structural glycomics: glycoproteomics and glycan profiling. Glycoconj J 2004;21(1-2):35-40.

74. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jorgensen TJ. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics 2005;4(7):873-886.

75. Rush J, Moritz A, Lee KA et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol 2005;23(1):94-101.

76. Boschetti E, Righetti PG. The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species. J Proteomics 2008;71(3):255-264.

77. Sidransky D. Emerging molecular markers of cancer. Nat Rev Cancer 2002;2(3):210-219.

78. Barry MJ. Screening for prostate cancer--the controversy that refuses to die. N Engl J Med 2009;360(13):1351-1354.

79. Pietrowska M, Marczak L, Polanska J et al. Mass spectrometry-based serum proteome pattern analysis in molecular diagnostics of early stage breast cancer. J Transl Med 2009;7:60.

80. Lei L, Wang XJ, Zheng ZG et al. Identification of serum protein markers for breast cancer relapse with SELDI-TOF MS. Anat Rec (Hoboken) 2011;294(6):941-944.

81. Mathelin C, Cromer A, Wendling C, Tomasetto C, Rio MC. Serum biomarkers for detection of breast cancers: A prospective study. Breast Cancer Res Treat 2006;96(1):83-90.

82. Callesen AK, Christensen R, Madsen JS et al. Reproducibility of serum protein profiling by systematic assessment using solid-phase extraction and matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom 2008;22(3):291-300.

83. Druker BJ, Talpaz M, Resta DJ et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 2001;344(14):1031-1037.

84. Baselga J. Targeting tyrosine kinases in cancer: the second wave. Science 2006;312(5777):1175-1178.

85. Bouchal P, Roumeliotis T, Hrstka R, Nenutil R, Vojtesek B, Garbis SD. Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. J Proteome Res 2009;8(1):362-373.

86. Iwamoto T, Bianchini G, Booser D et al. Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. J Natl Cancer Inst 2011;103(3):264-272.

87. Petricoin EF, III, Ornstein DK, Paweletz CP et al. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst 2002;94(20):1576-1578.

88. Albrethsen J, Bogebo R, Gammeltoft S, Olsen J, Winther B, Raskov H. Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1-3) in colon cancer serum and tumours: a biomarker study. BMC Cancer 2005;5:8.

89. Vlahou A, Laronga C, Wilson L et al. A novel approach toward development of a rapid blood test for breast cancer. Clin Breast Cancer 2003;4(3):203-209.

90. Chuthapisith S, Layfield R, Kerr ID, Eremin O. Principles of proteomics and its applications in cancer. Surgeon 2007;5(1):14-22.

91. Zhao J, Qiu W, Simeone DM, Lubman DM. N-linked glycosylation profiling of pancreatic cancer serum using capillary liquid phase separation coupled with mass spectrometric analysis. J Proteome Res 2007;6(3):1126-1138.

92. Chuthapisith S, Layfield R, Kerr ID, Hughes C, Eremin O. Proteomic profiling of MCF-7 breast cancer cells with chemoresistance to different types of anti-cancer drugs. Int J Oncol 2007;30(6):1545-1551.

93. Alcolea MP, Casado P, Rodriguez-Prados JC, Vanhaesebroeck B, Cutillas PR. Phosphoproteomic analysis of leukemia cells under basal and drug-treated conditions identifies markers of kinase pathway activation and mechanisms of resistance. Mol Cell Proteomics 2012.

94. Yuan Y, Tan CW, Shen H et al. Identification of the Biomarkers for the Prediction of Efficacy in First-Line Chemotherapy of Metastatic Colorectal Cancer Patients

using SELDI-TOF-MS and Artificial Neural Networks. Hepatogastroenterology 2012;59(120).

95. Adam BL, Qu Y, Davis JW et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62(13):3609-3614.

96. Yasui Y, Pepe M, Thompson ML et al. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. Biostatistics 2003;4(3):449-463.

97. Qu Y, Adam BL, Yasui Y et al. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. Clin Chem 2002;48(10):1835-1843.

98. McLafferty FW, Fridriksson EK, Horn DM, Lewis MA, Zubarev RA. Techview: biochemistry. Biomolecule mass spectrometry. Science 1999;284(5418):1289-1290.

99. Chait BT. Chemistry. Mass spectrometry: bottom-up or top-down? Science 2006;314(5796):65-66.

100. Reid GE, McLuckey SA. 'Top down' protein characterization via tandem mass spectrometry. J Mass Spectrom 2002;37(7):663-675.

101. Siuti N, Kelleher NL. Decoding protein modifications using top-down mass spectrometry. Nat Methods 2007;4(10):817-821.

102. Uttenweiler-Joseph S, Claverol S, Sylvius L, Bousquet-Dubouch MP, Burlet-Schiltz O, Monsarrat B. Toward a full characterization of the human 20S proteasome subunits and their isoforms by a combination of proteomic approaches. Methods Mol Biol 2008;484:111-130.

103. Zhang Q, Faca V, Hanash S. Mining the plasma proteome for disease applications across seven logs of protein abundance. J Proteome Res 2011;10(1):46-50.

104. Wolters DA, Washburn MP, Yates JR, III. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 2001;73(23):5683-5690.

105. Timms JF, Cramer R, Camuzeaux S et al. Peptides generated ex vivo from serum proteins by tumor-specific exopeptidases are not useful biomarkers in ovarian cancer. Clin Chem 2010;56(2):262-271.

106. Ceciliani F, Ceron JJ, Eckersall PD, Sauerwein H. Acute phase proteins in ruminants. J Proteomics 2012.

107. Song J, Patel M, Rosenzweig CN et al. Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. Clin Chem 2006;52(6):1045-1053.

108. Banks RE, Stanley AJ, Cairns DA et al. Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. Clin Chem 2005;51(9):1637-1649.

109. Traum AZ, Wells MP, Aivado M, Libermann TA, Ramoni MF, Schachter AD. SELDI-TOF MS of quadruplicate urine and serum samples to evaluate changes related to storage conditions. Proteomics 2006;6(5):1676-1680.

110. Kim BK, Lee JW, Park PJ et al. The multiplex bead array approach to identifying serum biomarkers associated with breast cancer. Breast Cancer Res 2009;11(2):R22.

111. Zeidan BA, Cutress RI, Murray N et al. Proteomic analysis of archival breast cancer serum. Cancer Genomics Proteomics 2009;6(3):141-147.

112. Troester MA, Hoadley KA, Sorlie T et al. Cell-type-specific responses to chemotherapeutics in breast cancer. Cancer Res 2004;64(12):4218-4226.

113. Bertucci F, Finetti P, Rougemont J et al. Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. Cancer Res 2005;65(6):2170-2178.

114. Sotiriou C, Neo SY, McShane LM et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci U S A 2003;100(18):10393-10398.

115. Baines CJ. Are there downsides to mammography screening? Breast J 2005;11 Suppl 1:S7-10.

116. Kaplan RM, Porzsolt F. The natural history of breast cancer. Arch Intern Med 2008;168(21):2302-2303.

117. van BS, V, Duijm LE, Voogd AC, Jansen FH, Louwman MW. Mammographic changes resulting from benign breast surgery impair breast cancer detection at screening mammography. Eur J Cancer 2012.

118. Ciocca DR, Elledge R. Molecular markers for predicting response to tamoxifen in breast cancer patients. Endocrine 2000;13(1):1-10.

119. Colomer R, Shamon LA, Tsai MS, Lupu R. Herceptin: from the bench to the clinic. Cancer Invest 2001;19(1):49-56.

120. Bourhis XL, Toillon RA, Boilly B, Hondermarck H. Autocrine and paracrine growth inhibitors of breast cancer cells. Breast Cancer Res Treat 2000;60(3):251-258.

121. Zou J, Hong G, Guo X et al. Reproducible cancer biomarker discovery in SELDI-TOF MS using different pre-processing algorithms. PLoS One 2011;6(10):e26294.

122. Zhang K, Yuan K, Wu H et al. Identification of potential markers related to neoadjuvant chemotherapy sensitivity of breast cancer by SELDI-TOF MS. Appl Biochem Biotechnol 2012;166(3):753-763.

123. Li J, Zhao J, Yu X et al. Identification of biomarkers for breast cancer in nipple aspiration and ductal lavage fluid. Clin Cancer Res 2005;11(23):8312-8320.

124. van Winden AWJ, Gast M, Beijnen JH et al. Validation of previously identified serum biomarkers for breast cancer with SELDI-TOF MS: a case control study. BMC Medical Genomics 2009;2(4).

125. Issaq HJ, Veenstra TD, Conrads TP, Felschow D. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. Biochem Biophys Res Commun 2002;292(3):587-592.

126. Tang N, Tornatore P, Weinberger SR. Current developments in SELDI affinity technology. Mass Spectrom Rev 2004;23(1):34-44.

127. Li P, Zhang D, Guo C. Serum biomarker screening for the diagnosis of early gastric cancer using SELDI-TOF-MS. Mol Med Report 2012;5(6):1531-1535.

128. Jemal A, Tiwari RC, Murray T et al. Cancer statistics, 2004. CA Cancer J Clin 2004;54(1):8-29.

129. Silverstein MJ, Barth A, Poller DN et al. Ten-year results comparing mastectomy to excision and radiation therapy for ductal carcinoma in situ of the breast. Eur J Cancer 1995;31A(9):1425-1427.

130. Atahan K, Kupeli H, Gur S et al. The value of serum biomarkers (Bc1, Bc2, Bc3) in the diagnosis of early breast cancer. Int J Med Sci 2011;8(2):148-155.

131. Belluco C, Petricoin EF, Mammano E et al. Serum proteomic analysis identifies a highly sensitive and specific discriminatory pattern in stage 1 breast cancer. Ann Surg Oncol 2007;14(9):2470-2476.

132. Reynolds T. Study clarifies risk of breast, ovarian cancer among mutation carriers. J Natl Cancer Inst 2003;95(24):1816-1818.

133. Pavlou MP, Kulasingam V, Sauter ER, Kliethermes B, Diamandis EP. Nipple aspirate fluid proteome of healthy females and patients with breast cancer. Clin Chem 2010;56(5):848-855.

134. Alexander H, Stegner AL, Wagner-Mann C, Du Bois GC, Alexander S, Sauter ER. Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. Clin Cancer Res 2004;10(22):7500-7510.

135. Varnum SM, Covington CC, Woodbury RL et al. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. Breast Cancer Res Treat 2003;80(1):87-97.

136. Sauter ER, Shan S, Hewett JE, Speckman P, Du Bois GC. Proteomic analysis of nipple aspirate fluid using SELDI-TOF-MS. Int J Cancer 2005;114(5):791-796.

137. Sauter ER, Zhu W, Fan XJ, Wassell RP, Chervoneva I, Du Bois GC. Proteomic analysis of nipple aspirate fluid to detect biologic markers of breast cancer. Br J Cancer 2002;86(9):1440-1443.

138. Bateman A, Singh A, Jothy S, Fraser R, Esch F, Solomon S. The levels and biologic action of the human neutrophil granule peptide HP-1 in lung tumors. Peptides 1992;13(1):133-139.

139. Mizukawa N, Sugiyama K, Kamio M et al. Immunohistochemical staining of human alpha-defensin-1 (HNP-1), in the submandibular glands of patients with oral carcinomas. Anticancer Res 2000;20(2B):1125-1127.

140. Zhou J, Trock B, Tsangaris TN et al. A unique proteolytic fragment of alpha1-antitrypsin is elevated in ductal fluid of breast cancer patient. Breast Cancer Res Treat 2010;123(1):73-86.

141. Noble JL, Dua RS, Coulton GR, Isacke CM, Gui GP. A comparative proteinomic analysis of nipple aspiration fluid from healthy women and women with breast cancer. Eur J Cancer 2007;43(16):2315-2320.

142. Pawlik TM, Hawke DH, Liu Y et al. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. BMC Cancer 2006;16(6):68.

143. Kuerer HM, Goldknopf IL, Fritsche H, Krishnamurthy S, Sheta EA, Hunt KK. Identification of distinct protein expression patterns in bilateral matched pair breast ductal fluid spicemens from women with unilateral invasive breast carcinoma. High-throuput biomarker discovery. Cancer 2002;95(11):2276-2282.

144. Ricolleau G, Charbonnel C, Lode L et al. Surface-enhanced laser desorption/ionization time of flight mass spectrometry protein profiling identifies ubiquitin and ferritin light chain as prognostic biomarkers in node-negative breast cancer tumors. Proteomics 2006;6(6):1963-1975.

145. Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol 2005;23(29):7350-7360.

146. Nakagawa T, Huang SK, Martinez SR et al. Proteomic profiling of primary breast cancer predicts axillary lymph node metastasis. Cancer Res 2006;66(24):11825-11830.

147. Goncalves A, Charafe-Jauffret E, Bertucci F et al. Protein profiling of human breast tumor cells identifies novel biomarkers associated with molecular subtypes. Mol Cell Proteomics 2008;7(8):1420-1433.

148. Leong S, Christopherson RI, Baxter RC. Profiling of apoptotic changes in human breast cancer cells using SELDI-TOF mass spectrometry. Cell Physiol Biochem 2007;20(5):579-590.

149. Chen DX, Schwartz PE, Li XG, Yang Z. Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses. Obstet Gynecol 1988;72(1):23-27.

150. Thesleff I, Viinikka L, Saxen L, Lehtonen E, Perheentupa J. The parotid gland is the main source of human salivary epidermal growth factor. Life Sci 1988;43(1):13-18.

151. Navarro MA, Mesia R, ez-Gibert O, Rueda A, Ojeda B, Alonso MC. Epidermal growth factor in plasma and saliva of patients with active breast cancer and breast cancer patients in follow-up compared with healthy women. Breast Cancer Res Treat 1997;42(1):83-86.

152. Streckfus CF, Bigler L, Dellinger T, Kuhn M, Chouinard N, Dai X. The expression of the c-erbB-2 receptor protein in glandular salivary secretions. J Oral Pathol Med 2004;33(10):595-600.

153. Streckfus CF, Mayorga-Wark O, Arreola D, Edwards C, Bigler L, Dubinsky WP. Breast cancer related proteins are present in saliva and are modulated secondary to ductal carcinoma in situ of the breast. Cancer Invest 2008;26(2):159-167.

154. Lawrence HP. Salivary markers of systemic disease: noninvasive diagnosis of disease and monitoring of general health. J Can Dent Assoc 2002;68(3):170-174.

155. Kaufman E, Lamster IB. The diagnostic applications of saliva--a review. Crit Rev Oral Biol Med 2002;13(2):197-212.

156. Tabak LA. A revolution in biomedical assessment: the development of salivary diagnostics. J Dent Educ 2001;65(12):1335-1339.

157. Streckfus CF, Bigler LR, Zwick M. The use of surface-enhanced laser desorption/ionization time-of-flight mass spectrometry to detect putative breast cancer markers in saliva: a feasibility study. J Oral Pathol Med 2006;35(5):292-300.

158. Lebrecht A, Boehm D, Schmidt M, Koelbl H, Grus FH. Surface-enhanced Laser Desorption/Ionisation Time-of-flight Mass Spectrometry to Detect Breast Cancer Markers in Tears and Serum. Cancer Genomics Proteomics 2009;6(2):75-83.

159. Simonson MS, Tiktin M, Debanne SM et al. The renal transcriptome of db/db mice identifies putative urinary biomarker proteins in patients with type 2 diabetes: a pilot study. Am J Physiol Renal Physiol 2012;302(7):F820-F829.

160. Rouse R, Siwy J, Mullen W, Mischak H, Metzger J, Hanig J. Proteomic candidate biomarkers of drug-induced nephrotoxicity in the rat. PLoS One 2012;7(4):e34606.

161. Konvalinka A, Scholey JW, Diamandis EP. Searching for new biomarkers of renal diseases through proteomics. Clin Chem 2012;58(2):353-365.

162. Papale M, Pedicillo MC, Thatcher BJ et al. Urine profiling by SELDI-TOF/MS: monitoring of the critical steps in sample collection, handling and analysis. J Chromatogr B Analyt Technol Biomed Life Sci 2007;856(1-2):205-213.

163. Atkinson JM, Pennington CJ, Martin SW et al. Membrane type matrix metalloproteinases (MMPs) show differential expression in non-small cell lung cancer (NSCLC) compared to normal lung: correlation of MMP-14 mRNA expression and proteolytic activity. Eur J Cancer 2007;43(11):1764-1771.

164. Munro NP, Cairns DA, Clarke P et al. Urinary biomarker profiling in transitional cell carcinoma. Int J Cancer 2006;119(11):2642-2650.

165. Rogers MA, Clarke P, Noble J et al. Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. Cancer Res 2003;63(20):6971-6983.

166. M'Koma AE, Blum DL, Norris JL et al. Detection of pre-neoplastic and neoplastic prostate disease by MALDI profiling of urine. Biochem Biophys Res Commun 2007;353(3):829-834.

167. Zhang D, Yang H, Kong X et al. Proteomics analysis reveals diabetic kidney as a ketogenic organ in type 2 diabetes. Am J Physiol Endocrinol Metab 2011;300(2):E287-E295.

168. Wu J, Chen YD, Yu JK, Shi XL, Gu W. Analysis of urinary proteomic patterns for type 2 diabetic nephropathy by ProteinChip. Diabetes Res Clin Pract 2011;91(2):213-219.

169. Slupsky CM, Steed H, Wells TH et al. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. Clin Cancer Res 2010;16(23):5835-5841.

170. Hou MF, Lin SB, Yuan SS et al. Diagnostic value of urine deoxypyridinoline for detecting bone metastases in breast cancer patients. Ann Clin Lab Sci 2003;33(1):55-61.

171. Guney N, Soydine HO, Derin D et al. Serum and urine survivin levels in breast cancer. Med Oncol 2006;23(3):335-339.

172. Adams SV, Newcomb PA, Shafer MM et al. Urinary cadmium and mammographic density in premenopausal women. Breast Cancer Res Treat 2011.

173. Ward DG, Nyangoma S, Joy H et al. Proteomic profiling of urine for the detection of colon cancer. Proteome Sci 2008;6:19.

174. Ye B, Skates S, Mok SC et al. Proteomic-based discovery and characterization of glycosylated eosinophil-derived neurotoxin and COOH-terminal osteopontin fragments for ovarian cancer in urine. Clin Cancer Res 2006;12(2):432-441.

175. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics 2004;3(4):367-378.

176. Villanueva J, Philip J, Chaparro CA et al. Correcting common errors in identifying cancer-specific serum peptide signatures. J Proteome Res 2005;4(4):1060-1072.

177. Findeisen P, Sismanidis D, Riedl M, Costina V, Neumaier M. Preanalytical impact of sample handling on proteome profiling experiments with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Clin Chem 2005;51(12):2409-2411.

178. Timms JF, rslan-Low E, Gentry-Maharaj A et al. Preanalytic influence of sample handling on SELDI-TOF serum protein profiles. Clin Chem 2007;53(4):645-656.

179. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. J Natl Cancer Inst 2005;97(4):307-309.

180. McLerran D, Grizzle WE, Feng Z et al. Analytical validation of serum proteomic profiling for diagnosis of prostate cancer: sources of sample bias. Clin Chem 2008;54(1):44-52.

181. Engwegen JY, Mehra N, Haanen JB et al. Validation of SELDI-TOF MS serum protein profiles for renal cell carcinoma in new populations. Lab Invest 2007;87(2):161-172.

182. Galvao ER, Martins LM, Ibiapina JO, Andrade HM, Monte SJ. Breast cancer proteomics: a review for clinicians. J Cancer Res Clin Oncol 2011.

183. Antoniou A, Pharoah PD, Narod S et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 2003;72(5):1117-1130.

184. Rennert G, Bisland-Naggan S, Barnett-Griness O et al. Clinical outcomes of breast cancer in carriers of BRCA1 and BRCA2 mutations. N Engl J Med 2007;357(2):115-123.

185. Sundquist M, Thorstenson S, Brudin L, Wingren S, Nordenskjold B. Incidence and prognosis in early onset breast cancer. Breast 2002;11(1):30-35.

186. Duffy SW, Tabar L, Olsen AH et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. J Med Screen 2010;17(1):25-30.

187. Shavers VL, Harlan LC, Stevens JL. Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. Cancer 2003;97(1):134-147.

188. Kollias J, Elston CW, Ellis IO, Robertson JF, Blamey RW. Early-onset breast cancer--histopathological and prognostic considerations. Br J Cancer 1997;75(9):1318-1323.

189. Maru D, Middleton LP, Wang S, Valero V, Sahin A. HER-2/neu and p53 overexpression as biomarkers of breast carcinoma in women age 30 years and younger. Cancer 2005;103(5):900-905.

190. Wang J, Buchholz TA, Middleton LP et al. Assessment of histologic features and expression of biomarkers in predicting pathologic response to anthracycline-based neoadjuvant chemotherapy in patients with breast carcinoma. Cancer 2002;94(12):3107-3114.

191. Owainati AA, Robins RA, Hinton C et al. Tumour aneuploidy, prognostic parameters and survival in primary breast cancer. Br J Cancer 1987;55(4):449-454.

192. Sjostrom J, Krajewski S, Franssila K et al. A multivariate analysis of tumour biological factors predicting response to cytotoxic treatment in advanced breast cancer. Br J Cancer 1998;78(6):812-815.

193. Aas T, Geisler S, Paulsen T et al. Primary systemic treatment with weekly doxorubicin monotherapy in women with locally advanced breast cancer; clinical experience and parameters predicting outcome. Acta Oncol 1996;35 Suppl 5:5-8.

194. Eccles D, Gerty S, Simmonds P, Hammond V, Ennis S, Altman DG. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. BMC Cancer 2007;7:160.

195. Fentiman IS, Wang DY, Allen DS et al. Bone density of normal women in relation to endogenous and exogenous oestrogens. Br J Rheumatol 1994;33(9):808-815.

196. Fentiman IS, Hanby A, Allen DS, Key T, Meilahn EN. Hormone dependency of breast tumours developing in the Guernsey Cohort study. Breast Cancer Res Treat 2006;97(2):205-208.

197. Pepe MS, Cai T, Longton G. Combining predictors for classification using the area under the receiver operating characteristic curve. Biometrics 2006;62(1):221-229.

198. Moritz RL, Clippingdale AB, Kapp EA et al. Application of 2-D free-flow electrophoresis/RP-HPLC for proteomic analysis of human plasma depleted of multi high-abundance proteins. Proteomics 2005;5(13):3402-3413.

199. Moritz RL, Skandarajah AR, Ji H, Simpson RJ. Proteomic Analysis of Colorectal Cancer: Prefractionation Strategies Using two-Dimensional Free-Flow Electrophoresis. Comp Funct Genomics 2005;6(4):236-243.

200. Finehout EJ, Lee KH. Comparison of automated in-gel digest methods for femtomole level samples. Electrophoresis 2003;24(19-20):3508-3516.

201. Bouchal P, Roumeliotis T, Hrstka R, Nenutil R, Vojtesek B, Garbis SD. Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. J Proteome Res 2009;8(1):362-373.

202. Boschetti E, Righetti PG. The art of observing rare protein species in proteomes with peptide ligand libraries. Proteomics 2009;9(6):1492-1510.

203. Sahab ZJ, Iczkowski KA, Sang QX. Anion exchange fractionation of serum proteins versus albumin elimination. Anal Biochem 2007;368(1):24-32.

204. Rollin D, Whistler T, Vernon SD. Laboratory methods to improve SELDI peak detection and quantitation. Proteome Sci 2007;5:9.

205. Tom J. Cancer Research UK CancerStats monograph2004. 2004.
Ref Type: Generic

206. Suzuki T, Toi M, Saji S et al. Early breast cancer. Int J Clin Oncol 2006;11(2):108-119.

207. Vanhoutte KJ, Laarakkers C, Marchiori E et al. Biomarker discovery with SELDI-TOF MS in human urine associated with early renal injury: evaluation with computational analytical tools. Nephrol Dial Transplant 2007;22(10):2932-2943.

208. Whistler T, Rollin D, Vernon SD. A method for improving SELDI-TOF mass spectrometry data quality. Proteome Sci 2007;5:14.

209. Atkinson JM, Pennington CJ, Martin SW et al. Membrane type matrix metalloproteinases (MMPs) show differential expression in non-small cell lung cancer (NSCLC) compared to normal lung: correlation of MMP-14 mRNA expression and proteolytic activity. Eur J Cancer 2007;43(11):1764-1771.

210. Theodorescu D, Wittke S, Ross MM et al. Discovery and validation of new protein biomarkers for urothelial cancer: a prospective analysis. Lancet Oncol 2006;7(3):230-240.

211. Schaub S, Wilkins J, Weiler T, Sangster K, Rush D, Nickerson P. Urine protein profiling with surface-enhanced laser-desorption/ionization time-of-flight mass spectrometry. Kidney Int 2004;65(1):323-332.

212. Kiyokawa I, Sogawa K, Ise K et al. Adsorption of urinary proteins on the conventionally used urine collection tubes: possible effects on urinary proteome analysis and prevention of the adsorption by polymer coating. Int J Proteomics 2011;2011:502845.

213. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol 2006;7(9):R80.

214. Pieper R, Gatlin CL, McGrath AM et al. Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. Proteomics 2004;4(4):1159-1174.

215. Sun W, Li F, Wu S et al. Human urine proteome analysis by three separation approaches. Proteomics 2005;5(18):4994-5001.

216. Zhou H, Pisitkun T, Aponte A et al. Exosomal Fetuin-A identified by proteomics: a novel urinary biomarker for detecting acute kidney injury. Kidney Int 2006;70(10):1847-1857.

217. Wang Z, Hill S, Luther JM, Hachey DL, Schey KL. Proteomic analysis of urine exosomes by multidimensional protein identification technology (MudPIT). Proteomics 2012;12(2):329-338.

218. Weber G, Bocek P. Recent developments in preparative free flow isoelectric focusing. Electrophoresis 1998;19(10):1649-1653.

219. Weber G, Bocek P. Stability of continuous flow electrophoresis. Electrophoresis 1998;19(18):3094-3095.

220. Mahley RW, Innerarity TL, Rall SC, Jr., Weisgraber KH. Plasma lipoproteins: apolipoprotein structure and function. J Lipid Res 1984;25(12):1277-1294.

221. Hanash S, Taguchi A. The grand challenge to decipher the cancer proteome. Nat Rev Cancer 2010;10(9):652-660.

222. Fan Y, Wang J, Yang Y et al. Detection and identification of potential biomarkers of breast cancer. J Cancer Res Clin Oncol 2010;136(8):1243-1254.

223. Goncalves A, Esterni B, Bertucci F et al. Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. Oncogene 2006;25(7):981-989.

224. Jung EJ, Moon HG, Park ST et al. Decreased annexin A3 expression correlates with tumor progression in papillary thyroid cancer. Proteomics Clin Appl 2010;4(5):528-537.

225. Kollermann J, Schlomm T, Bang H et al. Expression and prognostic relevance of annexin A3 in prostate cancer. Eur Urol 2008;54(6):1314-1323.

226. Yan X, Yin J, Yao H, Mao N, Yang Y, Pan L. Increased expression of annexin A3 is a mechanism of platinum resistance in ovarian cancer. Cancer Res 2010;70(4):1616-1624.

227. Opstal-van Winden AW, Krop EJ, Karedal MH et al. Searching for early breast cancer biomarkers by serum protein profiling of pre-diagnostic serum; a nested case-control study. BMC Cancer 2011;11:381.

228. Goncalves A, Esterni B, Bertucci F et al. Postoperative serum proteomic profiles may predict metastatic relapse in high-risk primary breast cancer patients receiving adjuvant chemotherapy. Oncogene 2006;25(7):981-989.

229. Takano S, Yoshitomi H, Togawa A et al. Apolipoprotein C-1 maintains cell survival by preventing from apoptosis in pancreatic cancer cells. Oncogene 2008;27(20):2810-2822.

230. Mussunoor S, Murray GI. The role of annexins in tumour development and progression. J Pathol 2008;216(2):131-140.

231. Wozny W, Schroer K, Schwall GP et al. Differential radioactive quantification of protein abundance ratios between benign and malignant prostate tissues: cancer association of annexin A3. Proteomics 2007;7(2):313-322.

232. Lam F, Jankova L, Dent OF et al. Identification of distinctive protein expression patterns in colorectal adenoma. Proteomics Clin Appl 2010;4(1):60-70.

233. Page DL, Dupont WD. Anatomic markers of human premalignancy and risk of breast cancer. Cancer 1990;66(6 Suppl):1326-1335.

234. Hartmann LC, Sellers TA, Frost MH et al. Benign breast disease and the risk of breast cancer. N Engl J Med 2005;353(3):229-237.

235. Armstrong K, Eisen A, Weber B. Assessing the risk of breast cancer. N Engl J Med 2000;342(8):564-571.

236. Liu YF, Xiao ZQ, Li MX et al. Quantitative proteome analysis reveals annexin A3 as a novel biomarker in lung adenocarcinoma. J Pathol 2009;217(1):54-64.

237. Lehman CD, Rutter CM, Eby PR, White E, Buist DS, Taplin SH. Lesion and patient characteristics associated with malignancy after a probably benign finding on community practice mammography. AJR Am J Roentgenol 2008;190(2):511-515.

238. Russo J, Hu YF, Yang X, Russo IH. Developmental, cellular, and molecular basis of human breast cancer. J Natl Cancer Inst Monogr 2000;(27):17-37.

239. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15(1):17-35.

240. Horseman ND. A prolactin-inducible gene product which is a member of the calpactin/lipocortin family. Mol Endocrinol 1989;3(5):773-779.

241. Yin J, Yan X, Yao X et al. Secretion of annexin A3 from ovarian cancer cells and its association with platinum resistance in ovarian cancer patients. J Cell Mol Med 2011.

242. Kollermann J, Schlomm T, Bang H et al. Expression and prognostic relevance of annexin A3 in prostate cancer. Eur Urol 2008;54(6):1314-1323.

243. Cohen M, Yossef R, Erez T et al. Serum apolipoproteins C-I and C-III are reduced in stomach cancer patients: results from MALDI-based peptidome and immuno-based clinical assays. PLoS One 2011;6(1):e14540.

244. Xue A, Scarlett CJ, Chung L et al. Discovery of serum biomarkers for pancreatic adenocarcinoma using proteomic analysis. Br J Cancer 2010;103(3):391-400.

245. Pollan M. Epidemiology of breast cancer in young women. Breast Cancer Res Treat 2010;123 Suppl 1:3-6.

246. Marrett LD, Frood J, Nishri D, Ugnat AM. Cancer incidence in young adults in Canada: preliminary results of a cancer surveillance project. Chronic Dis Can 2002;23(2):58-64.

247. Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. CA Cancer J Clin 2003;53(1):5-26.

248. Gajdos C, Tartter PI, Bleiweiss IJ, Bodian C, Brower ST. Stage 0 to stage III breast cancer in young women. J Am Coll Surg 2000;190(5):523-529.

249. Figueiredo JC, Ennis M, Knight JA et al. Influence of young age at diagnosis and family history of breast or ovarian cancer on breast cancer outcomes in a population-based cohort study. Breast Cancer Res Treat 2007;105(1):69-80.

250. Bowen RL, Duffy SW, Ryan DA, Hart IR, Jones JL. Early onset of breast cancer in a group of British black women. Br J Cancer 2008;98(2):277-281.

251. Anders CK, Hsu DS, Broadwater G et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol 2008;26(20):3324-3330.

252. Haffty BG, Harrold E, Khan AJ et al. Outcome of conservatively managed early-onset breast cancer by BRCA1/2 status. Lancet 2002;359(9316):1471-1477.

253. Zhou P, Recht A. Young age and outcome for women with early-stage invasive breast carcinoma. Cancer 2004;101(6):1264-1274.

254. Han W, Kang SY. Relationship between age at diagnosis and outcome of premenopausal breast cancer: age less than 35 years is a reasonable cut-off for defining young age-onset breast cancer. Breast Cancer Res Treat 2010;119(1):193-200.

255. Sundquist M, Thorstenson S, Brudin L, Wingren S, Nordenskjold B. Incidence and prognosis in early onset breast cancer. Breast 2002;11(1):30-35.

256. Marcus JN, Watson P, Page DL et al. Hereditary breast cancer: pathobiology, prognosis, and BRCA1 and BRCA2 gene linkage. Cancer 1996;77(4):697-709.

257. Levine MN, Pritchard KI, Bramwell VH, Shepherd LE, Tu D, Paul N. Randomized trial comparing cyclophosphamide, epirubicin, and fluorouracil with cyclophosphamide, methotrexate, and fluorouracil in premenopausal women with node-positive breast cancer: update of National Cancer Institute of Canada Clinical Trials Group Trial MA5. J Clin Oncol 2005;23(22):5166-5170.

258. Munro AF, Cameron DA, Bartlett JM. Targeting anthracyclines in early breast cancer: new candidate predictive biomarkers emerge. Oncogene 2010;29(38):5231-5240.

259. Vincent-Salomon A, Rousseau A, Jouve M et al. Proliferation markers predictive of the pathological response and disease outcome of patients with breast carcinomas treated by anthracycline-based preoperative chemotherapy. Eur J Cancer 2004;40(10):1502-1508.

260. Miyoshi Y, Kurosumi M, Kurebayashi J et al. Predictive factors for anthracycline-based chemotherapy for human breast cancer. Breast Cancer 2010;17(2):103-109.

261. Arriola E, Rodriguez-Pinilla SM, Lambros MB et al. Topoisomerase II alpha amplification may predict benefit from adjuvant anthracyclines in HER2 positive early breast cancer. Breast Cancer Res Treat 2007;106(2):181-189.

262. MacGrogan G, Rudolph P, Mascarel I, I et al. DNA topoisomerase IIalpha expression and the response toprimary chemotherapy in breast cancer. Br J Cancer 2003;89(4):666-671.

263. Masuda H, Masuda N, Kodama Y et al. Predictive factors for the effectiveness of neoadjuvant chemotherapy and prognosis in triple-negative breast cancer patients. Cancer Chemother Pharmacol 2011;67(4):911-917.

264. Garbis SD, Roumeliotis TI, Tyritzis SI, Zorpas KM, Pavlakis K, Constantinides CA. A novel multidimensional protein identification technology approach combining protein size exclusion prefractionation, peptide zwitterion-ion hydrophilic interaction chromatography, and nano-ultraperformance RP chromatography/nESI-MS2 for the in-depth analysis of the serum proteome and phosphoproteome: application to clinical sera derived from humans with benign prostate hyperplasia. Anal Chem 2011;83(3):708-718.

265. Cummings J, Ward TH, Greystoke A, Ranson M, Dive C. Biomarker method validation in anticancer drug development. Br J Pharmacol 2008;153(4):646-656.

266. Chowdhury F, Williams A, Johnson P. Validation and comparison of two multiplex technologies, Luminex and Mesoscale Discovery, for human cytokine profiling. J Immunol Methods 2009;340(1):55-64.

267. Sharma M, Blackman MR, Sharma MC. Antibody-directed neutralization of annexin II (ANX II) inhibits neoangiogenesis and human breast tumor growth in a xenograft model. Exp Mol Pathol 2011;92(1):175-184.

268. Sharma M, Ownbey RT, Sharma MC. Breast cancer cell surface annexin II induces cell migration and neoangiogenesis via tPA dependent plasmin generation. Exp Mol Pathol 2010;88(2):278-286.

269. Inokuchi J, Narula N, Yee DS et al. Annexin A2 positively contributes to the malignant phenotype and secretion of IL-6 in DU145 prostate cancer cells. Int J Cancer 2009;124(1):68-74.

270. Mohammad HS, Kurokohchi K, Yoneyama H et al. Annexin A2 expression and phosphorylation are up-regulated in hepatocellular carcinoma. Int J Oncol 2008;33(6):1157-1163.

271. Vishwanatha JK, Chiang Y, Kumble KD, Hollingsworth MA, Pour PM. Enhanced expression of annexin II in human pancreatic carcinoma cells and primary pancreatic cancers. Carcinogenesis 1993;14(12):2575-2579.

272. Chuthapisith S, Bean BE, Cowley G et al. Annexins in human breast cancer: Possible predictors of pathological response to neoadjuvant chemotherapy. Eur J Cancer 2009;45(7):1274-1281.

273. Huang Y, Jin Y, Yan CH et al. Involvement of Annexin A2 in p53 induced apoptosis in lung cancer. Mol Cell Biochem 2008;309(1-2):117-123.

274. Zhang F, Zhang L, Zhang B et al. Anxa2 plays a critical role in enhanced invasiveness of the multidrug resistant human breast cancer cells. J Proteome Res 2009;8(11):5041-5047.

275. Prueitt RL, Boersma BJ, Howe TM et al. Inflammation and IGF-I activate the Akt pathway in breast cancer. Int J Cancer 2007;120(4):796-805.

276. Loibl S, Buck A, Strank C et al. The role of early expression of inducible nitric oxide synthase in human breast cancer. Eur J Cancer 2005;41(2):265-271.

277. Loibl S, Strank C, von MG et al. Immunohistochemical evaluation of endothelial nitric oxide synthase expression in primary breast cancer. Breast 2005;14(3):230-235.

278. Harris RE, Beebe-Donk J, Doss H, Burr DD. Aspirin, ibuprofen, and other non-steroidal anti-inflammatory drugs in cancer prevention: a critical review of non-selective COX-2 blockade (review). Oncol Rep 2005;13(4):559-583.

279. Khuder SA, Mutgi AB. Breast cancer and NSAID use: a meta-analysis. Br J Cancer 2001;84(9):1188-1192.

280. Glynn SA, Boersma BJ, Dorsey TH et al. Increased NOS2 predicts poor survival in estrogen receptor-negative breast cancer patients. J Clin Invest 2010;120(11):3843-3854.

281. Glynn SA, Prueitt RL, Ridnour LA et al. COX-2 activation is associated with Akt phosphorylation and poor survival in ER-negative, HER2-positive breast cancer. BMC Cancer 2010;10:626.

282. Rhee I, Bachman KE, Park BH et al. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 2002;416(6880):552-556.

283. Gaudet F, Hodgson JG, Eden A et al. Induction of tumors in mice by genomic hypomethylation. Science 2003;300(5618):489-492.

284. Gribnau J, Hochedlinger K, Hata K, Li E, Jaenisch R. Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. Genes Dev 2003;17(6):759-773.

285. Yang X, Phillips DL, Ferguson AT, Nelson WG, Herman JG, Davidson NE. Synergistic activation of functional estrogen receptor (ER)-alpha by DNA methyltransferase and histone deacetylase inhibition in human ER-alpha-negative breast cancer cells. Cancer Res 2001;61(19):7025-7029.

286. Sharma D, Saxena NK, Davidson NE, Vertino PM. Restoration of tamoxifen sensitivity in estrogen receptor-negative breast cancer cells: tamoxifen-bound reactivated ER recruits distinctive corepressor complexes. Cancer Res 2006;66(12):6370-6378.

287. Phuong NT, Kim SK, Lim SC et al. Role of PTEN promoter methylation in tamoxifen-resistant breast cancer cells. Breast Cancer Res Treat 2011;130(1):73-83.

288. Phuong NT, Kim SK, Lim SC et al. Role of PTEN promoter methylation in tamoxifen-resistant breast cancer cells. Breast Cancer Res Treat 2010.

289. Shukla V, Coumoul X, Lahusen T et al. BRCA1 affects global DNA methylation through regulation of DNMT1. Cell Res 2010;20(11):1201-1215.

290. Xiang G, Zhenkun F, Shuang C et al. Association of DNMT1 gene polymorphisms in exons with sporadic infiltrating ductal breast carcinoma among Chinese Han women in the Heilongjiang Province. Clin Breast Cancer 2010;10(5):373-377.

291. Kastl L, Brown I, Schofield AC. Altered DNA methylation is associated with docetaxel resistance in human breast cancer cells. Int J Oncol 2010;36(5):1235-1241.

292. Di Fiore PP, Pierce JH, Kraus MH, Segatto O, King CR, Aaronson SA. erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells. Science 1987;237(4811):178-182.

293. Esteva FJ, Hortobagyi GN. Prognostic molecular markers in early breast cancer. Breast Cancer Res 2004;6(3):109-118.

294. Goddard KA, Weinmann S, Richert-Boe K, Chen C, Bulkley J, Wax C. HER2 evaluation and its impact on breast cancer treatment decisions. Public Health Genomics 2012;15(1):1-10.

295. Mosoyan G, Nagi C, Marukian S et al. Multiple Breast Cancer Cell-Lines Derived from a Single Tumor Differ in Their Molecular Characteristics and Tumorigenic Potential. PLoS One 2013;8(1):e55145.

296. Tamimi RM, Baer HJ, Marotti J et al. Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. Breast Cancer Res 2008;10(4):R67.

297. Imielinski M, Cha S, Rejtar T, Richardson EA, Karger BL, Sgroi DC. Integrated proteomic, transcriptomic, and biological network analysis of breast carcinoma reveals molecular features of tumorigenesis and clinical relapse. Mol Cell Proteomics 2012.

298. Gormley M, Minkeu A, Meng R, Zhong Z, Quong AA. Proteomic profiling of infiltrating ductal carcinoma reveals increased cellular interactions with tissue microenvironment. J Proteome Res 2012.

299. Devine PL, McKenzie IF. Mucins: structure, function, and associations with malignancy. Bioessays 1992;14(9):619-625.

300. Zeidan BA, Cutress RI, Murray N et al. Proteomic analysis of archival breast cancer serum. Cancer Genomics Proteomics 2009;6(3):141-147.

301. Gamez-Pozo A, Sanchez-Navarro I, Calvo E et al. Protein phosphorylation analysis in archival clinical cancer samples by shotgun and targeted proteomics approaches. Mol Biosyst 2011;7(8):2368-2374.

302. Fowler CB, Waybright TJ, Veenstra TD, O'Leary TJ, Mason JT. Pressure-assisted protein extraction: a novel method for recovering proteins from archival tissue for proteomic analysis. J Proteome Res 2012.

303. Diamandis EP. Cancer biomarkers: can we turn recent failures into success? J Natl Cancer Inst 2010;102(19):1462-1467.

304. Gagnon A, Shi Q, Ye B. Surface-enhanced laser desorption/ionization mass spectrometry for protein and Peptide profiling of body fluids. Methods Mol Biol 2008;441:41-56.

305. Gerke V, Moss SE. Annexins: from structure to function. Physiol Rev 2002;82(2):331-371.

306. Fletcher SW, Elmore JG. Clinical practice. Mammographic screening for breast cancer. N Engl J Med 2003;348(17):1672-1680.

307. Harashima M, Niimi S, Koyanagi H et al. Change in annexin A3 expression by regulatory factors of hepatocyte growth in primary cultured rat hepatocytes. Biol Pharm Bull 2006;29(7):1339-1343.

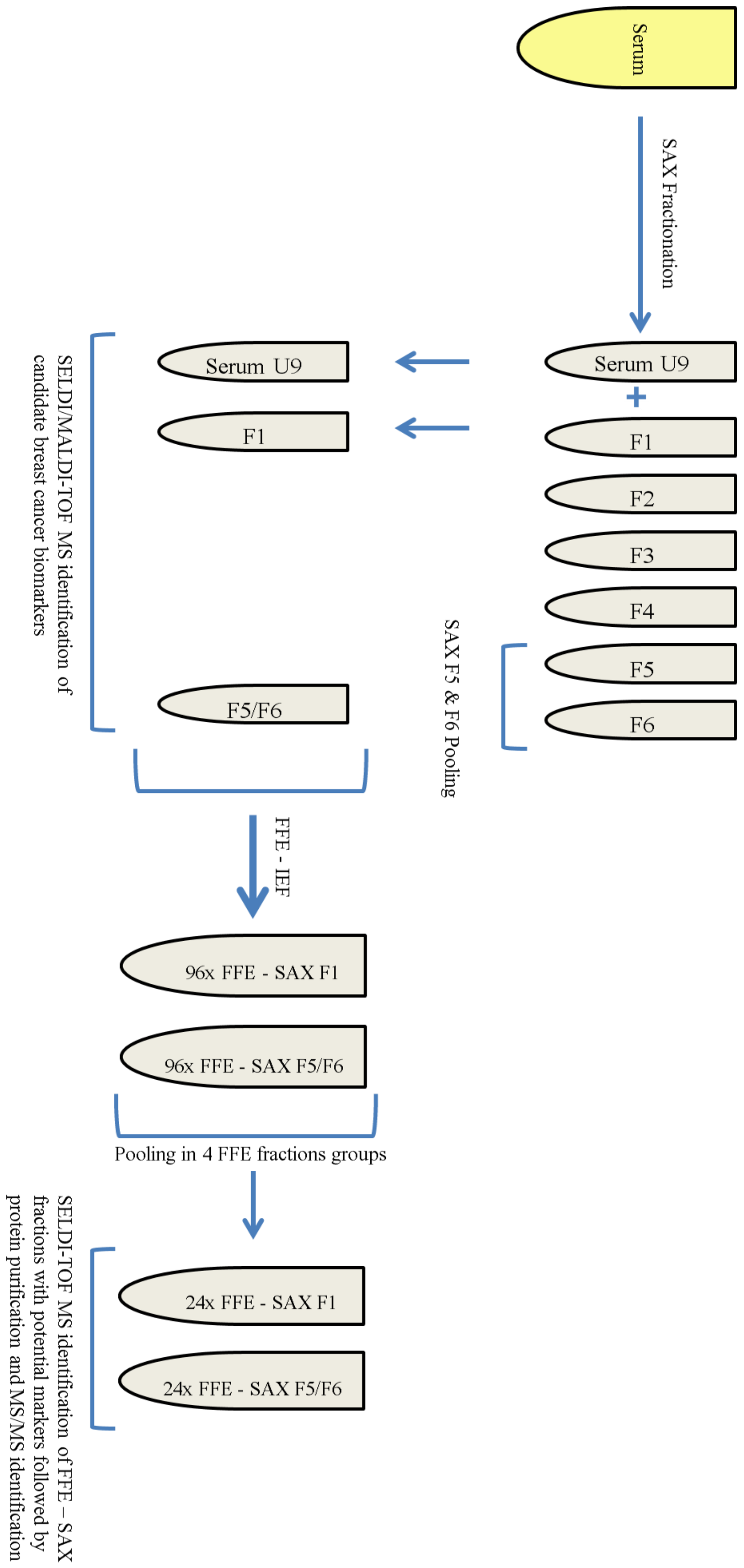
308. Park JE, Lee DH, Lee JA et al. Annexin A3 is a potential angiogenic mediator. Biochem Biophys Res Commun 2005;337(4):1283-1287.

309. Kim S, Ko J, Kim JH, Choi EC, Na DS. Differential effects of annexins I, II, III, and V on cytosolic phospholipase A2 activity: specific interaction model. FEBS Lett 2001;489(2-3):243-248.

310. Stastna M, Van Eyk JE. Secreted proteins as a fundamental source for biomarker discovery. Proteomics 2012.

Appendix 1

Fractionation and pooling steps in the Wessex serum breast cancer biomarker discovery.



Schematic diagram illustrating the different fractionation & pooling steps involved in the Wessex serum breast cancer biomarker discovery proteomic analysis.